SIAM J. Applied Dynamical Systems
Vol. 5, No. 1, pp. 1–11

Simultaneous Zip Bifurcation and Limit Cycles in Three Dimensional Competition Models*

Eduardo Sáez†, Eduardo Stange‡, and Iván Szántó†

Abstract. A three dimensional system is considered, which describes the competition of two predator species for a single prey species. When the predators' functional response is of generalized Holling III type, we prove the simultaneous existence of a hyperbolic attracting limit cycle and a Farkas zip-type bifurcation.

Key words. stability, limit cycles, bifurcations, bioeconomic model

AMS subject classifications. 92D25, 34C, 58F14, 58F21

DOI. 10.1137/040613998

1. Introduction.

We consider two predator species competing for a single prey. The functional response of the predators is assumed to be of a generalized Holling III type. The Holling III functional response is of the form $S^2/(S^2 + a^2)$, where S is the prey density; it has the property that up to a certain value the growth of S increases the predator efficiency and, as a consequence, the predator has a stabilizing effect on the population. Above this value, predator efficiency decreases (see May [5]). In this work the generalized Holling III-type functional response is $S^n/(S^n+a^n)$, where $n > 1$ is an arbitrary integer. We assume that one of the predators could be identified as a K-strategist, and the other as an r-strategist. In simple terms, an r-strategist is a species that tries to ensure its survival by having a relatively high growth rate, and a K-strategist is a species that consumes less, has a lower growth rate, and is able to raise its offspring on a scarce food supply. For a large class of three dimensional models, Farkas [1, 2, 3, 4] introduced in 1984 the phenomenon of *zip bifurcation*. It was assumed that the threshold densities of prey at which the predator growth rate became positive were equal. We prove in this paper that the system has a one dimensional continuum of equilibria for an open set in the parameter space. Moreover, when the system bifurcates, a hyperbolic attracting limit cycle (in the two dimensional sense) and a zip bifurcation (contained in the continuum of equilibria) appear simultaneously.

In the remainder of this section we introduce the model, in section 2 we establish the conditions for which a zip bifurcation occurs, and finally in section 3 we present some computer simulations.

*Received by the editors August 28, 2004; accepted for publication (in revised form) by M. Zeeman September 13, 2005; published electronically January 6, 2006.
http://www.siam.org/journals/siads/5-1/61399.html

†Departamento de Matemática, Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile (eduardo.saez@usm.cl, ivan.szanto@usm.cl). The work of these authors was supported by FONDECYT grants 1030264 and 7030005.

‡Instituto de Matemáticas y Física, Universidad de Valparaíso, Casilla 5030, Valparaíso, Chile (eduardo.stange@uv.cl). The work of this author was supported by DIPUV grant 14/2004.

Let us consider the following system that describes the competition of the two predator species for the prey:

$$\begin{cases} \dot{S} = rS\left(1-\frac{S}{K}\right) - \sum_{i=1}^{2} m_i x_i \frac{S^n}{a_i^n + S^n}, \\ \dot{x_i} = m_i x_i \frac{S^n}{a_i^n + S^n} - d_i x_i, \qquad i = 1, 2, \end{cases} \tag{1}$$

where the densities of prey and the two predators at time t will be denoted by $S(t)$, $x_1(t)$, and $x_2(t)$, respectively. The intrinsic growth rate of the prey and the carrying capacity will be denoted by r and K, respectively, and the maximal birth rate, the half saturation constant, and the death rate of predators i will be m_i, a_i, and d_i, respectively. All these constants are supposed to be positive, and $n \geq 2$ is an integer.

As the shape of the predator functional response is a sigmoidal one, then as S tends to infinity, the per capita predator birth rate tends to m_i. Furthermore, it is clear that for the survival of predator i it is necessary that the maximal birth rate be larger than the death rate:

$$b_i := \frac{m_i}{d_i} > 1, \qquad i = 1, 2,$$

and this will be assumed in what follows. Therefore, we shall consider a predator with a low half saturation constant a *K-strategist*. We assume that

$$a_1 < a_2.$$

The population of predator i begins to grow only if the right-hand side of the second equation of (1) is positive. This is the case if $S > S_i := a_i \sqrt[n]{d_i/(m_i - d_i)} = a_i \sqrt[n]{1/(b_i - 1)}$. We shall assume that

$$S_T := S_1 = S_2 \Rightarrow \frac{a_1^n d_1}{m_1 - d_1} = \frac{a_2^n d_2}{m_2 - d_2}. \tag{2}$$

This means that S_T is the minimum density of prey, which assures the growth of both predators.

The above conditions imply

$$b_1 < b_2.$$

We shall consider a species with a high ratio of maximal birth rate to death rate as an *r-strategist*. Thus, in our system predator 1 may be considered a K-strategist, and predator 2 an r-strategist.

2. Main results. In order to simplify the study of the dynamic of (1) in the space of the realistic variables $\bar{\Omega} = \{(S, x_1, x_2) \mid S, x_i \geq 0,\ i = 1, 2\}$, let the change of time be $t \to \frac{K}{r}(a_1^n + S^n)(a_2^n + S^n)t$.

The model (1) is reduced to the polynomial vector field $\tilde{X}_\eta$ given by

$$\begin{cases} \dot{S} = S(K-S)(a_1^n + S^n)(a_2^n + S^n) - \frac{m_1 K}{r} S^n (a_2^n + S^n) x_1 \\ \qquad - \frac{m_2 K}{r} S^n (a_1^n + S^n) x_2, \\ \dot{x}_1 = \frac{m_1 K}{r} S^n (a_2^n + S^n) x_1 - \frac{d_1 K}{r} (a_1^n + S^n)(a_2^n + S^n) x_1, \\ \dot{x}_2 = \frac{m_2 K}{r} S^n (a_1^n + S^n) x_2 - \frac{d_2 K}{r} (a_1^n + S^n)(a_2^n + S^n) x_2. \end{cases} \tag{3}$$

If in (3) the rescaling of the parameters $\{m_i \to \frac{r}{K}m_i,\ d_i \to \frac{r}{K}d_i\}$, $i = 1, 2$, and the linear transformation of coordinates $\{x_i \to \frac{x_i}{d_i}\}$ are considered, then systems (1) and (3) are equivalent to the polynomial system

$$\begin{cases} \dot{S} = S(K-S)(a_1^n + S^n)(a_2^n + S^n) - \frac{m_1}{d_1}S^n(a_2^n + S^n)x_1 \\ \qquad\quad - \frac{m_2}{d_2}S^n(a_1^n + S^n)x_2, \\ \dot{x}_1 = m_1 S^n(a_2^n + S^n)x_1 - d_1(a_1^n + S^n)(a_2^n + S^n)x_1, \\ \dot{x}_2 = m_2 S^n(a_1^n + S^n)x_2 - d_2(a_1^n + S^n)(a_2^n + S^n)x_2. \end{cases} \tag{4}$$

Let us consider a new rescaling of the parameters $\{m_i \to m_i d_i\}$, $i = 1, 2$. Then the systems (1), (3), and (4) are equivalent to the polynomial system

$$\begin{cases} \dot{S} = S(K-S)(a_1^n + S^n)(a_2^n + S^n) - x_1 m_1 S^n(a_2^n + S^n) \\ \qquad\quad - x_2 m_2 S^n(a_1^n + S^n), \\ \dot{x}_1 = x_1 d_1(a_2^n + S^n)[(m_1 - 1)S^n - a_1^n], \\ \dot{x}_2 = x_2 d_2(a_1^n + S^n)[(m_2 - 1)S^n - a_2^n]. \end{cases} \tag{5}$$

If $\bar{X}$ denotes the vector field associated with the previous system and

$$\bar{X} = P\frac{\partial}{\partial S} + Q_1\frac{\partial}{\partial x_1} + Q_2\frac{\partial}{\partial x_2},$$

then the set of singularities of $\bar{X}$ in $\bar{\Omega}$ is given by

$$Sing(\bar{X}) = P^{-1}(0)\bigcap_{i=1}^{2} Q_i^{-1}(0)\bigcap \bar{\Omega}.$$

We can see by the previous rescaling that the equality $S_1 = S_2$ of (2) is transformed into $\frac{a_1^n}{m_1 - 1} = \frac{a_2^n}{m_2 - 1}$.

Lemma 1. *Let* $\frac{a_1^n d_1}{m_1 - d_1} = \frac{a_2^n d_2}{m_2 - d_2}$, $m_i > d_i$, $i = 1, 2$. *If* $\lambda = a_i\sqrt[n]{\frac{d_i}{m_i - d_i}}$, *the set of singularities of* (5) *is given by*

$$Sing(\bar{X}) = \{(0,0,0), (K,0,0)\} \cup Sing_\lambda, \quad \textit{where}$$

$$Sing_\lambda = \{(S, x_1, x_2) \mid S = \lambda \le K,\ \lambda(K - \lambda) - x_1 - x_2 = 0,\ x_i \ge 0,\ i = 1, 2\}.$$

Proof. To simplify the proof, we will consider a system topologically equivalent to (1). This system can be obtained from (1) by rescaling parameters, rescaling coordinated axes, and a change of time, as we show in the beginning of section 2. So, the hypotheses of the lemma have the following equivalences:

$$\begin{cases} \frac{a_1^n d_1}{m_1 - d_1} = \frac{a_2^n d_2}{m_2 - d_2} & \to & \frac{a_1^n}{m_1 - 1} = \frac{a_2^n}{m_2 - 1}, \\ m_i > d_i & \to & m_i > 1, \\ \lambda = a_i\sqrt[n]{\frac{d_i}{m_i - d_i}} & \to & \lambda = \frac{a_i}{\sqrt[n]{m_i - 1}}. \end{cases}$$

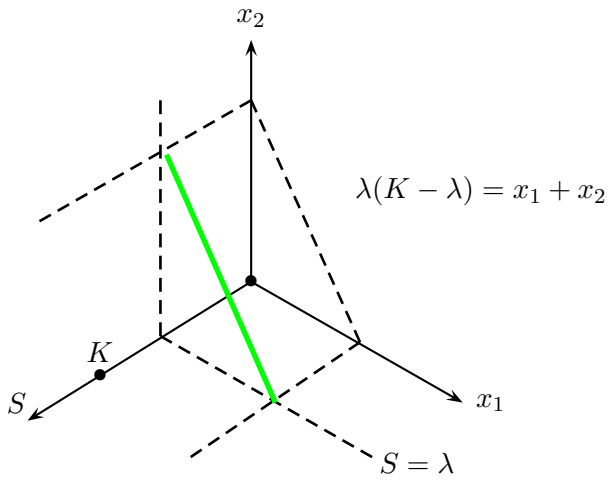


Figure 1.

By direct substitution into (5), it is simple to verify that $(0,0,0)$ and $(K,0,0)$ are singularities of the vector field.

If $x_1, x_2 \geq 0$, from (5) it is easy to see that

$$Q_i|_{S^n=\frac{a_i^n}{m_i-1}} \equiv 0, \qquad i = 1,2.$$

The plane H of equation $S = \lambda$ in $\mathbb{R}^3$ contains singularities of (5) if $P^{-1}(0) \cap H \cap \bar{\Omega} \neq \Phi$. Considering that $(m_i - 1)S^n = a_i^n \Rightarrow x_i m_i S^n = x_i(S^n + a_i^n)$, $i = 1,2$, and replacing them, respectively, in the last two terms in the first equation of system (5), we obtain

$$\dot{S} = (a_1^n + S^n)(a_2^n + S^n)[S(K - S) - x_1 - x_2].$$

If $\lambda \leq K$ (see Figure 1), we have

$$Sing_\lambda = \{(S, x_1, x_2) \mid S = \lambda,\ \lambda(K-\lambda) - x_1 - x_2 = 0,\ x_i \geq 0,\ i = 1,2\}. \qquad \blacksquare$$

Lemma 2. *Let us consider the vector field (5):*

(i) *In the coordinate system Sx_1x_2 of $\mathbb{R}^3$, the coordinated axes and the coordinated planes are invariant manifolds for (5).*

(ii) *The origin of the coordinated system is a hyperbolic saddle, and in a sufficiently small neighborhood of the origin, the axis S is an unstable manifold of dimension one and the two remaining axes x_i, $i = 1,2$, are stable manifolds that generate a stable manifold of dimension two.*

(iii) *For $\lambda \neq K$, the singularity in the point $(K,0,0)$ is a hyperbolic saddle. If $\lambda < K$, in a sufficiently small neighborhood of $(K,0,0)$, then there exists an unstable manifold of dimension two and a stable manifold of dimension one (axis S).*

Proof. (i) From (5), if the vector field is restricted to the coordinated axes and the

coordinated planes, we obtain

$$\begin{cases} \bar{X}(S,0,0) = S(K-S)(a_1^n + S^n)(a_2^n + S^n)\frac{\partial}{\partial S}, \\ \bar{X}(0,x_1,0) = -x_1 d_1 a_1^n a_2^n \frac{\partial}{\partial x_1}, \\ \quad \cdots \quad = \quad \cdots \\ \bar{X}(0,x_1,x_2) = -x_1 d_1 a_1^n a_2^n \frac{\partial}{\partial x_1} - x_2 d_2 a_1^n a_2^n \frac{\partial}{\partial x_2}, \end{cases}$$

and part (i) of the lemma is demonstrated.

(ii) The proof is immediate since the linear part is given by

$$D\bar{X}(0,0,0) = a_1^n a_2^n \begin{pmatrix} K & 0 & 0 \\ 0 & -d_1 & 0 \\ 0 & 0 & -d_2 \end{pmatrix}, \qquad K,\, a_i, d_i > 0, \quad i = 1,2.$$

(iii) At the singularity $(K,0,0)$ of (5), the linear part is a triangular matrix of the form

$$D\bar{X}(K,0,0) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix},$$

where

$$\begin{aligned} a_{11} &= -K(a_1^n + K^n)(a_2^n + K^n) < 0, \\ a_{22} &= d_1(a_2^n + K^n)[-a_1^n + (m_1 - 1)K^n], \\ a_{33} &= d_2(a_1^n + K^n)[-a_2^n + (m_2 - 1)K^n]. \end{aligned}$$

As $\lambda^n = \frac{a_i^n}{m_i - 1}$, $\lambda < K$, we have $-a_i^n + (m_i - 1)K^n > 0$, and then $a_{ii} > 0$, $i = 2,3$. ■

Lemma 3. *The system (1) is bounded in Ω.*

Proof. Let us consider the function $\mathcal{B}(t) = S(t) + x_1(t) + x_2(t)$; then its derivative on the orbits of (1) in Ω is given by

$$\dot{\mathcal{B}} = \dot{S} + \dot{x_1} + \dot{x_2} = rS\left(1 - \frac{S}{K}\right) - d_1 x_1 - d_2 x_2 = \frac{r}{K} S(K - S) - d_1 x_1 - d_2 x_2.$$

If $S \geq K$, we have $\dot{\mathcal{B}}(t) < 0$, and this proves that the trajectories of the system cross the planes $S + x_1 + x_2 = C$ from outside to inside if the constant C is sufficiently large; then (1) is bounded in Ω. ■

Theorem 1. *Let us consider $\frac{a_1^n d_1}{m_1 - d_1} = \frac{a_2^n d_2}{m_2 - d_2}$, $m_i > d_i$, $i = 1,2$, with $a_1 < a_2$ and $\lambda = a_i \sqrt[n]{\frac{d_i}{m_i - d_i}}$:*

(i) *If $(K - 2\lambda)m_1 < n(K - \lambda)(m_1 - d_1)$, the set of singularities $Sing_\lambda$ of (1) in Ω is stable in the Liapunov sense.*

(ii) *If $(K - 2\lambda)m_1 > n(K - \lambda)(m_1 - d_1)$, there exists a point p_z such that $Sing_\lambda^u \cup \{p_z\} \cup Sing_\lambda^s$ is a partition of $Sing_\lambda$, where the singularities of $Sing_\lambda^u$ and $Sing_\lambda^s$ are respectively unstable and stable in the Liapunov sense. The system (1) has a zip bifurcation.*

Proof. As in the proof of Lemma 1, we will consider a system topologically equivalent to (1). This system can be obtained from (1) by rescaling parameters, rescaling coordinated axes, and a change of time, as we show in the beginning of section 2. The hypotheses of this theorem have the following equivalences:

$$\begin{cases} \frac{a_1^n d_1}{m_1-d_1} = \frac{a_2^n d_2}{m_2-d_2} & \to & \frac{a_1^n}{m_1-1} = \frac{a_2^n}{m_2-1}, \\ m_i > d_i & \to & m_i > 1, \\ (K-2\lambda)m_1 \lesseqgtr n(K-\lambda)(m_1-d_1) & \to & (K-2\lambda)m_1 \lesseqgtr n(K-\lambda)(m_1-1), \\ \lambda = a_i \sqrt[n]{\frac{d_i}{m_i-d_i}} & \to & \lambda = \frac{a_i}{\sqrt[n]{m_i-1}}. \end{cases}$$

Then as $a_i^n = (m_i-1)\lambda^n$, $i = 1, 2$, the linear part of the vector field (5), restricted to a any point of $Sing_\lambda$, is

$$D\bar{X}(\lambda, x_1, x_2) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & 0 & 0 \\ a_{31} & 0 & 0 \end{pmatrix},$$

where

$$\begin{aligned} a_{11} &= -\lambda^{2n-1}[\lambda\{(2\lambda-K)m_1m_2 + (\lambda-K)n(m_1+m_2)\} + nm_1(1+m_2)x_1 \\ &\qquad + nm_2(1+m_1)x_2], \\ a_{12} &= -\lambda^{2n}m_1m_2, \\ a_{13} &= -\lambda^{2n}m_1m_2, \\ a_{21} &= \lambda^{2n-1}d_1(m_1-1)m_2nx_1, \\ a_{31} &= \lambda^{2n-1}d_2(m_2-1)m_1nx_2. \end{aligned}$$

Under the hypothesis $\lambda(K-\lambda) - x_1 - x_2 = 0$, the eigenvalues of the singularities of the vector field (5) in $Sing_\lambda$ are the solutions of the characteristic equation

$$\mu[\mu^2 - a_{11}\mu - a_{12}a_{21} - a_{13}a_{31}] = 0.$$

Then in each singularity in $Sing_\lambda$ there exists at least one central manifold of dimension one. If μ_1 and μ_2 denote the roots of the above quadratic equation, we have $\mu_1\mu_2 = -a_{12}a_{21} - a_{13}a_{31}$ and $\mu_1 + \mu_2 = a_{11}$.

As $m_i > 1$, $i = 1, 2$, it is easy to see that $\mu_1\mu_2 = \lambda^{4n-1}m_1m_2n[d_1(m_1-1)m_2x_1 + d_2(m_2-1)m_1x_2] > 0$. Then $\mu_1, \mu_2 \neq 0$ and $SgnRe(\mu_1) = SgnRe(\mu_2)$.

Moreover, it is clear that $a_1 < a_2 \Rightarrow m_1 < m_2$; then if $m_1(K-2\lambda) < n(m_1-1)(K-\lambda)$, we have

$$a_{11} = -\lambda^{2n-1}(m_2\lambda(n(m_1-1)(K-\lambda) - m_1(K-2\lambda)) + nx_2(m_2-m_1)) < 0.$$

So, in each singularity in $Sing_\lambda$, $Re(\mu_1)$ and thus $Re(\mu_2) < 0$, and the set $Sing_\lambda$ is an attractor.

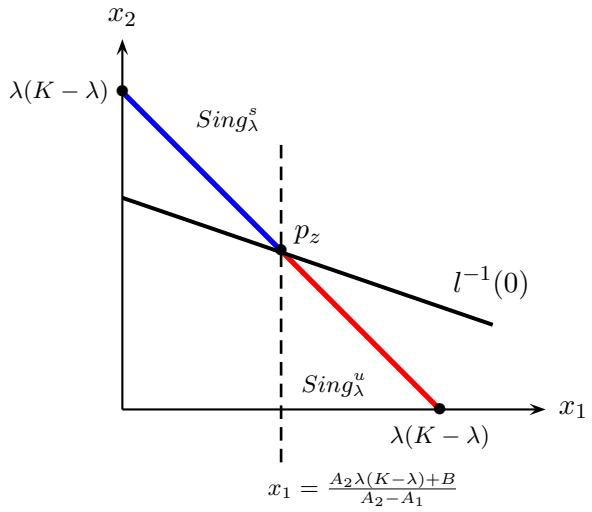


Figure 2.

In the coordinated space x_1x_2 of $\mathbb{R}^2_+$, let us consider the system formed by the two straight lines

$$\begin{cases} x_1 + x_2 = \lambda(K-\lambda), \\ l(x_1,x_2) = A_1x_1 + A_2x_2 + B = 0, \end{cases} \tag{6}$$

where

$$\begin{aligned} A_1 &= nm_1(1+m_2), \\ A_2 &= nm_2(1+m_1), \\ B &= \lambda\{(2\lambda-K)m_1m_2 + (\lambda-K)n(m_1+m_2)\}. \end{aligned}$$

If $m_1(K-2\lambda) > n(m_1-1)(K-\lambda)$, then $\lambda(K-\lambda) > \frac{A_2\lambda(K-\lambda)+B}{A_2-A_1}$, and it follows that $l^{-1}(0) \cap Sing_\lambda = \{p_z\}$ (see Figure 2). Then (6) has a solution in a single point $p_z \in Sing_\lambda$, and the set $Sing_\lambda$ is divided by $Sing_\lambda = Sing^u_\lambda \cup \{p_z\} \cup Sing^s_\lambda$, where the singularities $(\lambda(K-\lambda),0,0) \in Sing^u_\lambda$ and $(0,0,\lambda(K-\lambda)) \in Sing^s_\lambda$. As $a_{11} = -\lambda^{2n-1}l(x_1,x_2)|_{Sing^s_\lambda} < 0$ (resp., $a_{11} = -\lambda^{2n-1}l(x_1,x_2)|_{Sing^u_\lambda} > 0$) and by the fact that $Sing^u_\lambda \subset l^{-1}(-\infty,0)$ and $Sing^s_\lambda \subset l^{-1}(0,\infty)$, we have $Re(\mu_1), Re(\mu_2) < 0$ (resp., > 0). This proves that the sets $Sing^u_\lambda$ and $Sing^s_\lambda$ are respectively unstable and stable in the Liapunov sense. This proves that $Sing_\lambda$ undergoes a zip bifurcation introduced by Farkas in [1]. ■

Theorem 2. *If $\lambda < K$, $m_1 > d_1$, and $\lambda = a_i \sqrt[n]{\frac{d_i}{m_i-d_i}}$, then in $\tilde{\Omega} = \Omega \cap \{(S,x_1,x_2) \mid x_2 = 0\}$ the vector field (1) has the following features:*

(i) *At the singularity $\lambda(1,K-\lambda,0)$, an attracting weak focus of order one if $m_1(2\lambda-K)+n(K-\lambda)(m_1-d_1)=0$.*

(ii) *A hyperbolic limit cycle generated by Hopf bifurcation that encloses the singularity $\lambda(1,K-\lambda,0)$ if $m_1(2\lambda-K)+n(K-\lambda)(m_1-d_1)<0$.*

Proof. As in the proof of Lemma 1, and with the same rescaling and change of coordinates, the hypotheses of the this theorem have the following equivalences:

$$\begin{cases} m_1 > d_1 & \to & m_1 > 1, \\ \lambda = a_i \sqrt[n]{\frac{d_i}{m_i - d_i}} & \to & \lambda = \frac{a_1}{\sqrt[n]{m_1 - 1}}, \\ m_1(2\lambda - K) + n(K-\lambda)(m_1 - d_1) \le 0 & \to & m_1(2\lambda - K) + n(K-\lambda)(m_1 - 1) \le 0. \end{cases}$$

Then as $x_2 = 0$ the system (5) is given by

$$\begin{cases} \dot{S} = S(K-S)(a_1^n + S^n)(a_2^n + S^n) - x_1 m_1 S^n (a_2^n + S^n), \\ \dot{x}_1 = x_1 d_1 (a_2^n + S^n)[(m_1 - 1)S^n - a_1^n]. \end{cases} \tag{7}$$

Let us consider in $\tilde{\Omega}$ the rescaling of time $t \to \frac{t}{a_2^n + S^n}$; then we have

$$\begin{cases} \dot{S} = S(K-S)(a_1^n + S^n) - x_1 m_1 S^n, \\ \dot{x}_1 = x_1 d_1 [(m_1 - 1)S^n - a_1^n]. \end{cases} \tag{8}$$

By Lemma 1, in the plane Sx_1 the set $Sing_\lambda$ is reduced to the only singularity $(\lambda, \lambda(K-\lambda))$. Considering the translation $\{S \to S + \lambda,\ x_1 \to x_1 + \lambda(K-\lambda)\}$ of this singularity to the origin, and by the fact that $a_1^n = (m_1 - 1)\lambda^n$, we obtain

$$Z : \begin{cases} \dot{S} = (K - \lambda - S)(\lambda + S)[\lambda^n(-1 + m_1) + (\lambda + S)^n] \\ \qquad - m_1(\lambda + S)^n[(K-\lambda)\lambda + x_1], \\ \dot{x}_1 = d_1(-1 + m_1)[\lambda^n - (\lambda + S)^n][-K\lambda + \lambda^2 - x_1]. \end{cases} \tag{9}$$

The linear part of the vector field Z at the origin is given by

$$DZ(0,0) = \begin{pmatrix} a10 & a01 \\ b10 & b01 \end{pmatrix},$$

where

$$\begin{aligned} a10 &= -\lambda^n[m_1(2\lambda - K) + n(K-\lambda)(m_1 - 1)], \\ a01 &= -\lambda^n m_1, \\ b10 &= d_1 \lambda^n (K-\lambda)(m_1 - 1)n, \\ b01 &= 0. \end{aligned}$$

As a consequence, it follows that

$$\begin{cases} \det DZ(0,0) = \lambda^{2n} d_1 m_1 n (K-\lambda)(m_1 - 1) > 0, \\ \operatorname{tra} DZ(0,0) = -\lambda^n[m_1(2\lambda - K) + n(K-\lambda)(m_1 - 1)]. \end{cases}$$

(i) The hypothesis $m_1(2\lambda - K) + n(K-\lambda)(m_1 - 1) = 0$ implies that $2\lambda - K < 0$ and $\det DZ(0,0) = \lambda^{2n} d_1 m_1 (K - 2\lambda) > 0$; then the origin is monodromic. Let L_i, $i = 1, 2$, be the

first two Liapunov quantities at the origin of vector field (9). As the origin is monodromic, then $L_1 \equiv 0$ and $m_1 = \frac{n(K-\lambda)}{2\lambda-K+n(K-\lambda)}$. Furthermore, it is known that the second Liapunov quantity depends on the three-jet at the origin of (9). In order to calculate L_2, we introduce a new parameter $A > 0$ such that $d_1 = \frac{A^2}{K-2\lambda}$. Considering the transformation of coordinates $S = x$ and $x_1 = Ay$ and the rescaling of the time $\{t \to \frac{2\lambda-K+n(K-\lambda)}{A\lambda^n(K-\lambda)n}t\}$, the vector field Z is written as

$$\begin{cases} \dot{x} = [(-K+2\lambda+(K-\lambda)n)((K-\lambda-x)(\lambda+x) \\ \qquad (\lambda^n(-1+((K-\lambda)n)/(-K+2\lambda+(K-\lambda)n))+(\lambda+x)^n) \\ \qquad -(K-\lambda)n(\lambda+x)^n((K-\lambda)\lambda+Ay) \\ \qquad /(-K+2\lambda+(K-\lambda)n))]/A(K-\lambda)\lambda^n n, \\ \dot{y} = [(-K+2\lambda+(K-\lambda)n)(-1+(K-\lambda)n/ \\ \qquad (-K+2\lambda+(K-\lambda)n))(\lambda^n-(\lambda+x)^n)(-(K\lambda)+\lambda^2-Ay)] \\ \qquad /(K-2\lambda)(K-\lambda)\lambda^n n, \end{cases}$$

and the three-jet is given by

$$\begin{cases} \dot{x} = -y + \frac{-K^2+3K\lambda-4\lambda^2+K^2n-3K\lambda n+2\lambda^2 n}{2A(K-\lambda)\lambda}x^2 - \frac{n}{\lambda}xy \\ \qquad + \frac{K^2-4\lambda^2-3K^2n+6K\lambda n-6\lambda^2 n+2K^2n^2-6K\lambda n^2+4\lambda^2n^2}{6A(K-\lambda)\lambda^2}x^3 - \frac{n(n-1)}{2\lambda^2}x^2y, \\ \dot{y} = x + \frac{n-1}{2\lambda}x^2 + \frac{A}{(K-\lambda)\lambda}xy + \frac{(n-1)(n-2)}{6\lambda^2}x^3 + \frac{A(n-1)}{2(K-\lambda)\lambda^2}x^2y. \end{cases}$$

Therefore, $L_2 = -\frac{(n-1)(3K-4\lambda)+4\lambda}{16A(K-\lambda)\lambda} < 0$, and since $3K-4\lambda = 3(K-2\lambda)+2\lambda > 0$, the origin is an attracting weak focus of order one.

(ii) Taking a small perturbation of the parameters λ, K, and m_1 such that $m_1(2\lambda-K)+n(K-\lambda)(m_1-1) < 0$, then

$$\operatorname{tra} DZ(0,0) = -\lambda^n[m_1(2\lambda-K)+n(K-\lambda)(m_1-1)] > 0.$$

Then origin of vector field (9) changes its stability to a repelling hyperbolic focus, and therefore an attracting hyperbolic limit cycle is generated by Hopf bifurcation. This completes the proof of the theorem. ■

3. Computer simulations. In this section, using Runge–Kutta application of the Mathematica software [6], a numerical example of system (5) will be presented. The classical Holling III functional response will be considered; i.e., $n = 2$ and the following parameter values are chosen: $d_1 = 1$, $d_2 = 1$, $m_1 = 7/5$, $m_2 = 8/5$, $a_1 = \sqrt{2/5}$, $a_2 = \sqrt{3/5}$, $\lambda = 1$, $K = 4$.

The initial conditions are $S(0) = 1$, $x_1(0) = 3.1$, $x_2(0) = 0.01$.

The next picture (Figure 3) is shown from the viewpoint $(4.9, 2.9, 0.4)$ in the Sx_1x_2-coordinate system. This allows us to give the conjecture that in the realistic three dimensional space there is an attracting invariant manifold that is formed by a continuum of periodic orbits (center in the space), whose border is the attracting limit cycle in the plane $x_2 = 0$ and where the cusp is the singularity p_z.

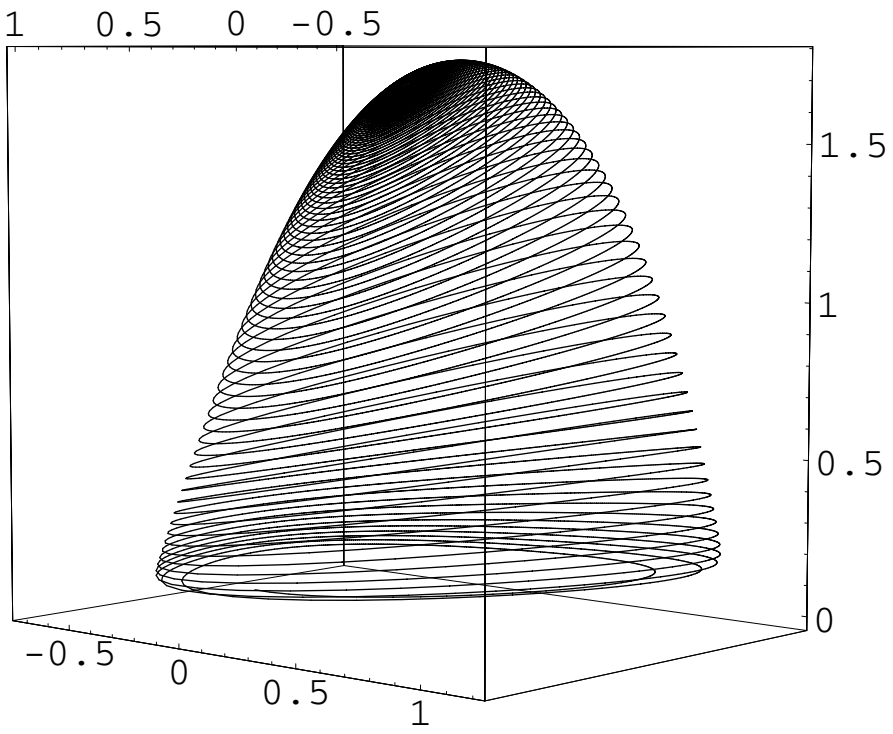


Figure 3.

4. Conclusions. Two predator species competing for a single prey species were studied. In the absence of predation the prey species population follows a logistic dynamics. The predator functional response is of the generalized Holling III type, and both of the two predators begin to grow at the same threshold density of prey. Predator 1 is considered as a K-strategist, and predator 2 an r-strategist.

The inequality in (i) of Theorem 1 means that, given an initial condition in a tubular neighborhood of the continuum of singularities of (5) (which is topologically equivalent to (1)) in the open set Ω, the corresponding w-limit of the orbit is some point of that continuum, and both predators coexist with the prey throughout time.

By the rescaling of the predators $x_i \to \frac{x_i}{d_i}$, $i = 1, 2$, in (3), the zip bifurcation of (5) is C^∞-equivalent to a zip bifurcation of (1) (see Figure 2). In this figure, the set $Sing_\lambda$ is a straight segment formed by a continuum of singularities with a distinguished singularity p_z of greater codimension. The singularity p_z separates the change of stability of those stable and unstable manifolds of the singularities of $Sing_\lambda$ above and below the singularity p_z.

Let us consider $\eta = (H - 2\lambda)m_1 - n(K - \lambda)(m_1 - d_1)$. For $\eta = 0$, p_z reaches the inferior boundary of the segment $Sing_\lambda$, which means that the r-strategist predators disappear. On the other hand, the singularity p_z in the plane $x_2 = 0$ in $\bar{\Omega}$ is a weak focus of order one. For $\eta > 0$, the point p_z moves in $\bar{\Omega}$ along the segment $Sing_\lambda$ from the plane $x_2 = 0$ to the plane $x_1 = 0$. Furthermore, for a given initial condition in a tubular neighborhood of $Sing_\lambda$, the three species coexist in time. Simultaneously, in the plane $x_2 = 0$ a supercritical

Hopf bifurcation takes place, generating an attracting hyperbolic limit cycle whose diameter increases with η.

Acknowledgments. The authors would like to thank Dr. M. Farkas for many stimulating discussions during the preparation of this work. Also we want to express our gratitude to the anonymous referees for their valuable comments and suggestions.

REFERENCES

[1] M. Farkas, *Zip bifurcation in a competition model*, Nonlinear Anal., 8 (1984), pp. 1295–1309.
[2] M. Farkas, *Competitive exclusion by zip bifurcation*, in Dynamical Systems, IIASA Workshop (Sopron, Hungary, 1985), Lecture Notes in Econom. and Math. Systems 287, Springer-Verlag, Berlin, pp. 165–178.
[3] M. Farkas, *Periodic Motions*, Appl. Math. Sci. 104, Springer-Verlag, New York, 1994.
[4] M. Farkas, *Dynamical Models in Biology*, Academic Press, New York, 2001.
[5] R. M. May, ed., *Theoretical Ecology*, 2nd ed., Blackwell, Oxford, UK, 1981.
[6] Wolfram Research, *Mathematica: A System for Doing Mathematics by Computer*, Champaign, IL, 1988.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 1, pp. 12–29

Approximating the Invariant Sets of a Finite Straight Segment near Its Collinear Equilibria*

Jesús F. Palacián†, Patricia Yanguas†, and Susana Gutiérrez–Romero†

Abstract. This paper reviews a method for computing explicitly the asymptotic expressions of some invariant manifolds in the vicinity of equilibrium points corresponding to Hamiltonian systems of m degrees of freedom. The method is based on the use of generalized normal forms. Our goal is to show the power of this technique by applying it to the study of one of the colinear equilibrium points corresponding to the Hamiltonian system defined by the motion of a particle orbiting around a finite straight segment. We make use of three different Lie transformations: (i) we calculate the Hamilton function corresponding to the center manifold of the colinear equilibrium; (ii) we determine the Hamiltonian related to the stable-unstable direction; and (iii) we obtain the Hamiltonian related to one of the two stable-stable directions. By means of (i), we are able to parameterize the center, the stable, and the unstable manifolds of the original system using the direct changes of coordinates of the two transformations. We also compute the normally hyperbolic invariant manifold associated with the equilibrium, together with its stable and unstable manifolds. Using (ii), we compute the invariant 2-tori and quasi-periodic orbits in the neighborhood of the equilibrium chosen. Through the normal form Hamiltonian obtained by (iii) we are able to determine some periodic orbits of the initial Hamiltonian.

Key words. generalized normal forms, invariant structures, straight segment

AMS subject classifications. 34K19, 37G05, 37J15, 70H33

DOI. 10.1137/040614517

1. Introduction and normal forms.

1.1. Scope. The computation of asymptotic (formal) integrals in Hamiltonian systems has received special attention during recent decades due to its utility from a qualitative standpoint. The approach of extending an integral of the principal part of the Hamiltonian to higher orders can be found in [15]. It generalizes the standard theory of normal forms for polynomial Hamiltonians; see, for instance, [13]. The use of generalized normal forms allows one to get a deeper insight into a Hamilton function through its different normal forms. Specifically, one can use this theory for different purposes such as to approximate some invariant manifolds, to decide on the stability of an equilibrium solution, or to analyze the monodromy of a Hamiltonian dynamical system [6]. Applications appear in various fields such as astrodynamics [15] or atomic and molecular physics [24, 20]. Theory on generalized normal forms has been

*Received by the editors September 6, 2004; accepted for publication (in revised form) by S. van Gils September 13, 2005; published electronically January 6, 2006. This work was partially supported by project BFM2002-03157 of Ministerio de Ciencia y Tecnología (Spain), by Resolución 92/2002 of Departamento de Educación y Cultura, Go-bierno de Navarra (Spain), by ACPI2002/04 of Gobierno de La Rioja (Spain), and by API02/20 of Universidad de La Rioja (Spain).
http://www.siam.org/journals/siads/5-1/61451.html
†Departamento de Matemática e Informática, Universidad Pública de Navarra, 31006 Pamplona, Navarra, Spain (palacian@unavarra.es, yanguas@unavarra.es, susana.gutierrez@unavarra.es).

developed in [15]; see also [5, 6]. An algorithm for the case of Hamiltonians with m degrees of freedom (DOF), together with some applications exhibiting the benefit of generalized normal forms, has been detailed in [6].

We focus on Hamiltonians of the form

$$\mathcal{H}(\boldsymbol{x}) = \sum_{n=0}^{\infty} \frac{\varepsilon^n}{n!} \mathcal{H}_n(\boldsymbol{x}), \tag{1.1}$$

where $\boldsymbol{x}$ is a $2m$-dimensional vector in the coordinates $x_1, x_2, \ldots, x_m$ and respective momenta $X_1, X_2, \ldots, X_m$. Each $\mathcal{H}_n$ is a homogeneous polynomial in $\boldsymbol{x}$ of degree $n+2$, and ε stands for a small parameter.

First, we simplify (1.1) by reducing its number of DOF by at least one unit. This goal is achieved through the introduction of a formal integral, that is, an integral up to a certain order of approximation after truncating the tail of the normal form Hamiltonian. Furthermore, when the standard normal form does not imply a decrease of the number of DOF, then our generalized normal forms do it, and they can be used to reduce the number of DOF of the original system.

Both classical and generalized normal forms are managed through normalizing procedures that involve Lie transformations. After the transformation process is performed, the transformed Hamiltonian obtained from (1.1) gives us information on the dynamics associated with the original system from a qualitative point of view. We have taken advantage of this fact to find out some periodic and quasi-periodic orbits of the original system as well as some other invariant objects.

The article has four sections. First, in section 1 we give a summary of the normal form theory. In section 2 we introduce our case study, the Hamiltonian modeling the rotating straight segment. We apply the standard normal form technique, obtaining the stable, unstable, and center manifolds related to one of the two colinear equilibria. Moreover, we determine the normally hyperbolic invariant manifold of the equilibrium we have chosen and its stable and unstable manifolds. In section 3 we use the generalized normal form approach, obtaining some periodic and quasi-periodic orbits and invariant 2-tori. Finally, section 4 is devoted to the conclusions of the work.

Within this paper we extend the results provided in [7], in the sense that here we work out the computations in detail, emphasizing the role played by the different normal forms. Furthermore, we have introduced a new normal form transformation, with the aim of obtaining some new periodic orbits of the original Hamiltonian.

As stated by Guckenheimer and Vladimirsky [4], the task of constructing higher-dimensional invariant manifolds for dynamical systems is computationally expensive. However, in contrast to the numerical approaches followed by various authors towards the approximation of some global invariant manifolds for vector fields (see, for instance, the papers [12, 9, 4]), we proceed analytically, deriving explicit expressions for the (local approximation of) invariant manifolds of the example we have chosen. Global approximations of our invariant manifolds can be determined using either numerical continuation techniques or some methods based on interpolation.

1.2. Normal forms for polynomial Hamiltonians. This section deals with the simplification of Hamiltonian systems through the construction of formal integrals. To achieve this target we use the well-known normal form theorem [13, 21] and a generalization of normal forms explained in [15, 16].

Let $\mathcal{H}$ be an m-DOF Hamiltonian of the type (1.1). It represents an analytic function whose quadratic terms are given by $\mathcal{H}_0(\boldsymbol{x}) = \frac{1}{2}\,\boldsymbol{x}^t\,B\,\boldsymbol{x}$, where B corresponds to a symmetric $2m$-matrix. Let $\mathcal{J}$ be the standard skew-symmetric matrix of dimension $2m$,

$$\mathcal{J} = \begin{pmatrix} 0 & I_m \\ -I_m & 0 \end{pmatrix},$$

where I_m stands for the identity matrix of order m, and let $A = \mathcal{J}\,B$ be the matrix associated with the system defined by $\mathcal{H}_0$.

Through the application of the normal form theorem [13, 21], we are able to decrease the number of DOF of Hamiltonian systems by at least one unit, provided that the matrix A is not nilpotent. More precisely, if $A = S + N$ is decomposed as the sum of its semisimple ($S \neq 0$) and nilpotent (N) components, then the quadratic Hamiltonian $\mathcal{I}_S(\boldsymbol{x}) = -\frac{1}{2}\,\boldsymbol{x}^t\,\mathcal{J}\,S\,\boldsymbol{x}$ becomes a formal integral of the reduced system.

Let $\mathcal{G}$ be a homogeneous polynomial of degree $j \geq 1$ commuting with $\mathcal{H}_0$ (that is, $\mathcal{G}$ is a first integral of $\mathcal{H}_0$). Applying a generalization of the normal form theorem [15], we can construct a normal form Hamiltonian such that the new Hamiltonian commutes with $\mathcal{G}$ up to a given degree. By doing so we can construct a Hamiltonian normal form which has $\mathcal{G}$ as a new first integral, and therefore one can reduce, by at least one, the number of DOF of any Hamiltonian system, in principle even if the associated matrix A is nilpotent, i.e., if $S = 0$. This is usually achieved whenever the Hamiltonian $\mathcal{H}$ fulfills some nonresonant hypotheses. Note that the formal integral does not coincide necessarily with $\mathcal{I}_S(\boldsymbol{x})$, as $\mathcal{G}$ is chosen regardless of the semisimple, nilpotent, or mixed semisimple and nilpotent character of $\mathcal{H}_0$. In all situations, one obtains a symplectic change of variables $\boldsymbol{x} \to \boldsymbol{y}$ that transforms $\mathcal{H}$ into the normalized Hamiltonian $\mathcal{K}$, with

$$\mathcal{K}(\boldsymbol{y};\varepsilon) = \sum_{i=0}^{L} \frac{\varepsilon^n}{n!}\,\mathcal{K}_n(\boldsymbol{y}), \tag{1.2}$$

where $\mathcal{K}_0(\boldsymbol{y}) \equiv \mathcal{H}_0(\boldsymbol{x})$ and each $\mathcal{K}_n$ is a homogeneous polynomial of degree $n+2$ in $\boldsymbol{y}$. Truncating at order L, the error committed in the transformation is of the type $\mathcal{O}(\varepsilon^{L+1})$.

The construction of $\mathcal{K}$ is done order by order (or, equivalently, degree by degree) in an ascendent way from $n = 1$ to $n = L$. The homology equation

$$\{\,\mathcal{W}_n\,,\,\mathcal{H}_0\,\} + \mathcal{K}_n = \widetilde{\mathcal{H}}_n \tag{1.3}$$

is solved with the extra condition $\{\,\mathcal{K}_n\,,\,\mathcal{G}\,\} = 0$ for $n = 1,\ldots,L$ so that each term $\mathcal{K}_n$ commutes with $\mathcal{G}$. Hence, in the end (i.e., when $n = L$), $\mathcal{G}$ becomes an integral of $\mathcal{K}$ after truncation of higher order terms. The operator $\{\ ,\ \}$ denotes the usual Poisson bracket, and the terms $\widetilde{\mathcal{H}}_n$ are those known from the previous orders. The solution of (1.3) is the pair $(\mathcal{W}_n, \mathcal{K}_n)$, where $\mathcal{W}_n$ denotes the generating function determined at order n. The Poisson

bracket $\{\mathcal{W}_n, \mathcal{H}_0\}$ is also written as $\mathcal{L}_{\mathcal{H}_0}(\mathcal{W}_n)$. This is the so-called Lie–Deprit method; see [1] for more details.

The above can be extended somewhat if we assume that $\mathcal{H}_n = 0$, for $n = 1, \ldots, k-1$, and $\mathcal{H}_k \neq 0$ represents a homogeneous polynomial of degree $k+2$. In this situation one still may consider generalized normal forms, though the theory becomes more involved as we cannot split $\mathcal{H}_k$ into semisimple and nilpotent terms.

The main drawback of the generalized method is that $\mathcal{W}_n$ is not necessarily a polynomial function of degree $n+2$, as it occurs with the standard approach of polynomial normal forms. Indeed $\mathcal{W}_n$ can involve rational, logarithmic, or arctangent functions or combinations of those. Thus, we have to exclude the singularities from the domain of validity of the normal form.

The occurrence of polynomial generating functions in a specific normal form computation may be known in advance (i.e., previous to the normal form computations) by analyzing the dimensions of the kernels of various linear spaces of homogeneous polynomials. So, depending on the choice of $\mathcal{G}$, as well as on the form of $\mathcal{H}_0$ and the type of terms in the perturbation, $\mathcal{W}$ could be a polynomial of degree $n+2$. (In this optimal case, each term $\mathcal{W}_n$ belongs to the linear space $\mathcal{R}_n$ corresponding to the spaces of homogeneous polynomials of degree $n+2$ in $\boldsymbol{x} \in \boldsymbol{R}^{2m}$.) These conditions have been settled in [15]. Nevertheless, we do not perform this algebraic analysis previous to the computation of the normal form in the design of the algorithm [6], because we have preferred to try first to solve the homological equation with a homogeneous polynomial. When that is not possible, we resort to finding the solution of the usual partial differential equation (PDE).

1.3. Choices of $\mathcal{G}$. Taking into account the decomposition of A into its semisimple and nilpotent components, we classify the types of reduction into three remarkable cases:

(a) semisimple case, $A = S$;
(b) semisimple plus nilpotent case, $A = S + N$ with $S, N \neq 0$;
(c) nilpotent case, $A = N$.

Next, according to the type of matrix one has at hand, we proceed as follows:

(a) We apply the normal form theorem with $\mathcal{G}(\boldsymbol{x}) = \mathcal{H}_0(\boldsymbol{x})$. Therefore, the procedure yields a polynomial generating function. If $m > 1$ and the reduced Hamiltonian defines a system of zero DOF, we choose another $\mathcal{G}$ and apply the generalized method.
(b) We proceed with the normal form theorem, taking $\mathcal{G}(\boldsymbol{x}) = \mathcal{I}_S(\boldsymbol{x})$.
(c) We perform the transformation, taking $\mathcal{G}(\boldsymbol{x}) = \mathcal{H}_0$. Obviously, we make use of the generalized method laid out above. Besides, if $\mathcal{H}_0 = 0$, then we even could select $\mathcal{G}$ among the integrals of $\mathcal{H}_k$, provided that $\mathcal{H}_n = 0$, for $n = 1, \ldots, k-1$.

Once we have chosen (or determined) $\mathcal{G}$, the next step consists of performing the normal form transformation, calculating $\mathcal{K}$ and $\mathcal{W}$. Now $\mathcal{G}$ corresponds to an integral of $\mathcal{K}$ and

$$\mathcal{I}(\boldsymbol{x}; \varepsilon) = \mathcal{G}(\boldsymbol{x}) + \sum_{n=1}^{L} \frac{\varepsilon^n}{n!} \mathcal{L}^n_{-\mathcal{W}} [\mathcal{G}(\boldsymbol{x})]$$

becomes an asymptotic integral of $\mathcal{H}$ functionally independent of it, and up to an approximation of $\mathcal{O}(\varepsilon^{L+1})$. Here $\mathcal{L}_{-\mathcal{W}}$ refers to the Lie operator $\mathcal{L}_{-\mathcal{W}} : F \longrightarrow \{\mathcal{W}, F\}$, whereas the composition operator $\mathcal{L}^n_{-\mathcal{W}}(\boldsymbol{y})$ denotes the recurrence operator $\mathcal{L}_{-\mathcal{W}}(\mathcal{L}^{n-1}_{-\mathcal{W}}(\boldsymbol{y}))$ for $n \geq 2$. Function $\mathcal{I}(\boldsymbol{x}; \varepsilon)$ is determined through the algorithm of the inverse for Lie transformations

due to Henrard [10]. The method proposed in that paper reduces the number of computations compared with the usual method for the direct and inverse changes of variables proposed by Deprit [1]. For a complete description of the method, see [6, 15] and references therein.

The main features of the approach described above are as follows:

(i) The algorithm is valid for any number of DOF, and it works not only with real coefficients, but also with complex ones.

(ii) The polynomial Hamiltonian may be of any degree. We do not need to start with $\mathcal{H}_0$ being a quadratic polynomial; see an application in [6].

(iii) If $\mathcal{H}_0$ is a quadratic polynomial, its associated matrix does not need to be in diagonal form.

(iv) We can make a qualitative analysis of a certain system, such as the study of monodromy, the nonlinear stability character of equilibria, the computation of periodic solutions and other invariant manifolds, or the determination of versal deformations.

(v) This approach may be used to calculate different normalized Hamiltonians whose flows lie on different reduced phase spaces. Thus, performing several reductions allows us to analyze the original Hamiltonian from different points of view, getting a richer knowledge of the initial system. The number of DOF of the normal form Hamiltonian system depends on the first integral introduced through the transformation. This is accomplished by a careful analysis based on Lie groups and invariant theory; see [14].

2. Application to a finite straight segment.

2.1. Hamiltonian of the problem. The case study we have chosen to apply the theory of section 1 corresponds to a mechanical system dealing with the motion of a point mass under the gravitational field of a massive finite segment. We broach this analysis within a three-dimensional (3D) frame.

The dynamics around an elongated celestial body can be represented approximately by using a massive segment as the representation of this kind of celestial object; see, for instance, [8]. Within this context, the finite straight segment is a model used as an approximation to the gravitational field of irregularly shaped bodies, such as asteroids, comet nuclei, and planetary moons. For this potential, Riaguas [17], Riaguas, Elipe, and Lara [18], and Elipe, Lara, and Riaguas [2] have computed several families of periodic orbits and bifurcations. In addition to that, Riaguas, Elipe, and López–Moratalla [19] and Elipe and Riaguas [3] have analyzed the nonlinear stability of the equilibria in two DOF and three DOF, respectively.

We consider a straight segment of length $2\,\ell$ and mass M that rotates uniformly with constant angular velocity ω about an axis perpendicular to the segment and fixed in space. Then we fix the origin of a reference frame $Ox_1x_2x_3$ at the center of mass O with the segment lying on the axis Ox_1, identifying the axis of rotation with Ox_3. We stress that the reference frame rotates with the straight segment with angular velocity ω.

We follow the approach of [17], [18], and [2] to get the initial Hamiltonian. After making some arrangements and rescaling, the problem is represented in closed form by the Hamilton function given by

$$\mathcal{H}(\boldsymbol{x}) = \frac{1}{2}\left(X_1^2 + X_2^2 + X_3^2\right) - (x_1X_2 - x_2X_1) + U(\boldsymbol{r}), \tag{2.1}$$

where $\boldsymbol{r} = (x_1, x_2, x_3)$ stands for the position of the particle while (X_1, X_2, X_3) represents their velocities or conjugate momenta. The potential U is defined through

$$U(\boldsymbol{r}) = -k \log\left(\frac{r_1 + r_2 + 1}{r_1 + r_2 - 1}\right),$$

where r_1, r_2 are the distances of the particle to the end-points of the segment given by

$$r_1 = \sqrt{(x_1 - 1/2)^2 + x_2^2 + x_3^2}, \qquad r_2 = \sqrt{(x_1 + 1/2)^2 + x_2^2 + x_3^2}$$

and $k = GM/(\omega^2(2\ell)^3) \in (0, \infty)$ stands for a dimensionless parameter that represents the ratio of the gravitational acceleration to centrifugal acceleration, such that $0 < k < 1$ means fast rotation of the segment, whereas $k > 1$ means slow rotation.

Notice that Hamiltonian (2.1) defines a system of three DOF, as no integral other than $\mathcal{H}$ is likely present in the equations of motion.

For our study, we have selected $k = 3$; that is, we are in the slow-rotation regime. In this situation, the system has four equilibrium points, two of them located on the axis Ox_1 at the points $(\pm 3/2, 0, 0)$, the so-called colinear equilibria, and the other two placed on the axis Ox_2, specifically at the points $(0, \pm 3/2, 0)$. The momenta of the equilibria resting on the axis Ox_1 are $X_1 = X_3 = 0$ and $X_2 = \pm 3/2$, whereas the momenta corresponding to the equilibria in the axis Ox_2 are $X_1 = \mp 3/2$ and $X_2 = X_3 = 0$.

Next we choose the equilibrium point whose coordinates are $\boldsymbol{r}_0 = (3/2, 0, 0)$ (or $(3/2, 0, 0, 0, 3/2, 0)$ in the full phase space) and translate it to the origin by means of a linear (and canonical) change of variables, say ψ. After this translation, we keep the same name for the variables and for the Hamiltonian. Thereafter, we Taylor-expand $U(\boldsymbol{r})$ around the equilibrium up to degree eight in x_1, x_2, and x_3, yielding

$$\mathcal{H}(\boldsymbol{x}) = \mathcal{H}_0 + \sum_{i=1}^{6} \mathcal{H}_i(\boldsymbol{r}).$$

We remark that for each i, $\mathcal{H}_i$ is a homogeneous polynomial in $\boldsymbol{x}$ of degree $i+2$, for $i = 0, \ldots, 6$.

The next step consists of diagonalizing the main part of the quadratic term $\mathcal{H}_0$. This is achieved by computing the eigenvalues of the matrix associated with $\mathcal{H}_0$. These are

$$\pm\mu_1 \imath, \qquad \pm\mu_2 \imath, \qquad \pm\mu_3,$$

where

$$\mu_1 = \frac{3\sqrt{2}}{4}, \qquad \mu_2 = \frac{\sqrt{7 + 3\sqrt{17}}}{4}, \qquad \mu_3 = \frac{\sqrt{-7 + 3\sqrt{17}}}{4}.$$

Now we can conclude that our point has a linearization of the type center $\times$ center $\times$ saddle and is, therefore, unstable. The same applies to the equilibrium $(-3/2, 0, 0, 0, -3/2, 0)$. On the other hand, the two equilibria resting on the axis Ox_2 have a linearization of the type center $\times$ center $\times$ center and are linearly stable. Their nonlinear stability is analyzed in [19] and [3]. In Figure 1 we have plotted the contours corresponding to the effective potential $W(\boldsymbol{r}) = U(\boldsymbol{r}) - (x_1^2 + x_2^2)/2$ projected onto the plane Ox_1x_2.

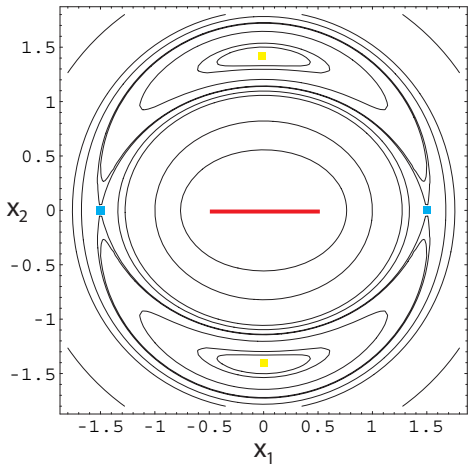


Figure 1. *Contours of the function $W(x_1, x_2, 0)$. The yellow squares correspond to the (linear) stable equilibria, whereas the blue ones correspond to the unstable equilibria. The point under study is the right-most equilibrium on the axis Ox_1. The red line represents the straight segment.*

At this point we need to perform a symplectic linear change of variables that we call φ, such that $\mathcal{H}_0$ is transformed into

$$\mathcal{H}_0(\varphi(\boldsymbol{x})) = \frac{1}{2}\left(X_1^2 + X_2^2 + X_3^2\right) + \frac{1}{2}\left(\mu_1^2 x_1^2 + \mu_2^2 x_2^2 - \mu_3^2 x_3^2\right).$$

By doing this, the quadratic Hamiltonian $\mathcal{H}_0$ is in normal form, and the subsequent computations will be carried out in an easier way. We rename $\mathcal{H}(\varphi) \equiv \mathcal{H}$ in order to avoid cumbersome notation and maintain the same name for the variables.

2.2. The standard normalization. We start by choosing $\mathcal{G} = \mathcal{H}_0$, and our Hamiltonian is ready to apply the Lie transformations explained in section 1. Thus, we may calculate the normal form Hamiltonian up to order six, that is, up to polynomials of degree eight. We call this Hamiltonian $\mathcal{K}$, and it is a function of the new coordinates $\boldsymbol{y} = (y_1, y_2, y_3, Y_1, Y_2, Y_3)$. If we define diagonal complex coordinates $(q_1, q_2, q_3, p_1, p_2, p_3)$ through the linear change of coordinates

$$\begin{aligned} y_1 &= \frac{1}{\sqrt{2}}(q_1 + \imath\, p_1), \quad y_2 = \frac{1}{\sqrt{2}}(q_2 + \imath\, p_2), \quad y_3 = \frac{1}{\sqrt{2}}(q_3 - p_3), \\ Y_1 &= \frac{\mu_1}{\sqrt{2}}(\imath\, q_1 + p_1), \quad Y_2 = \frac{\mu_2}{\sqrt{2}}(\imath\, q_2 + p_2), \quad Y_3 = \frac{\mu_3}{\sqrt{2}}(q_3 + p_3), \end{aligned} \tag{2.2}$$

then $\mathcal{K}$ reads

$$\mathcal{K} = \sum_{\substack{0\le j,k,\ell\le 4 \\ 1\le j+k+\ell\le 4}} a(j,k,\ell)(p_1 q_1)^j (p_2 q_2)^k (p_3 q_3)^\ell, \tag{2.3}$$

where the coefficients $a(j, k, \ell)$ are given in Table 1.

Table 1
Coefficients of the normal form Hamiltonian $\mathcal{K}$.

$a(1,0,0)$	$1.125000000000000000000000000000000000000000000000000000000000\imath$
$a(0,1,0)$	$1.210582304803311353091514347995139442215099854757553831449 74\imath$
$a(0,0,1)$	0.335582304803311353091514347995139442215099854757553831449744
$a(2,0,0)$	0.012030004392096116064681880006045035514583648178933051231676 0
$a(0,2,0)$	$0.066308730845837153713360708735727160836961568549103183918924\imath$
$a(0,0,2)$	$-0.086345894109368605622561330565536170463916665996215216217999$
$a(1,1,0)$	$-0.009126002113334923119912467663874443664951847740191280320 8549$
$a(1,0,1)$	$-0.095227013792664774628311803153716609682486568460503898763322\imath$
$a(0,1,1)$	$-0.297252742477241851459623674892761250708685001551084154338222\imath$
$a(3,0,0)$	$0.000318398384857662229656622037595485756558153744503556688113 57\imath$
$a(0,3,0)$	$0.027145488584395491344331454450057847512821281983895958135504\imath$
$a(0,0,3)$	$-0.043692379682477334385578543306482135635541873260412713783267$
$a(2,1,0)$	$0.056036986263531688900510154521095706235718505605565692123365\imath$
$a(2,0,1)$	$-0.007181048971515578356816427396050399929645195594391462774 4720$
$a(1,2,0)$	$-0.095445019140551020485391679319426103114361856110216448035834\imath$
$a(0,2,1)$	0.065870615091316453833825737336039001085835666557418312095763
$a(1,0,2)$	$-0.006927637253694715777650473729012212898463707168702333369657 7\imath$
$a(0,1,2)$	$-0.106768591640445723910056439151573202462634068271821935291025\imath$
$a(1,1,1)$	0.026918284311774477744419934897706660459522404408789952169848
$a(4,0,0)$	0.000084404313541226691175880660396023467860733144840748899313 48
$a(0,4,0)$	$-0.021531391821351515200894612206090041439329157836071593974346$
$a(0,0,4)$	$-0.037384835495131112197231667394596164922688566168369438454374$
$a(3,1,0)$	0.044014722983883863944044280212617641063119768236126655716244
$a(2,2,0)$	$-0.449792973986013498949950441884140843874921228545777482442 57$
$a(1,3,0)$	0.322239390096914200741027812468130093375354515213973560596 98
$a(3,0,1)$	$0.000677097195404915755713532529007278891929145967028075974023 65\imath$
$a(2,0,2)$	0.001868700475158339364905017413122312686093007170127794480057 6
$a(1,0,3)$	$-0.009596542258008953249177316578688008450271406582342195703274\imath$
$a(0,3,1)$	$0.084259851454671320518081116092249924037833608904071251017 73\imath$
$a(0,2,2)$	0.188070142041829899961152640131552788531372869697402299296 19
$a(0,1,3)$	$-0.149829779314173684571163483469201654267690712898453781711 50\imath$
$a(2,1,1)$	$0.013974751918553899340566685848616490545694007428879981221591\imath$
$a(1,2,1)$	$-0.026153278416863235085351618444028681673631314704894494837680\imath$
$a(1,1,2)$	$-0.022167757568694328712245961256391163513889189566623263026867$

2.3. Invariant manifolds. Once we have carried out the transformation passing from $\mathcal{H}(\boldsymbol{x})$ to $\mathcal{K}(\boldsymbol{y})$, we obtain the explicit expressions for the direct and inverse changes of coordinates. The direct change is given by $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{y})$, whereas the inverse one is $\boldsymbol{y} = \boldsymbol{Y}(\boldsymbol{x})$. Note that $\mathcal{H}(\boldsymbol{x}(\boldsymbol{X}(\boldsymbol{y}))) = \mathcal{K}(\boldsymbol{y})$.

Recalling that ψ and φ have been defined in section 2.1, the expressions of the local invariant manifolds in the initial variables are computed as follows:

- The 1D stable manifold of the equilibrium under study is obtained after calculating $\boldsymbol{x}_S = \psi\,\varphi\,\boldsymbol{X}(0,0,y_3,0,0,-\mu_3 y_3)$.
- The 1D unstable manifold of our equilibrium point is computed through the successive changes: $\boldsymbol{x}_U = \psi\,\varphi\,\boldsymbol{X}(0,0,y_3,0,0,\mu_3 y_3)$.
- The 4D center manifold of the equilibrium is computed as $\boldsymbol{x}_C = \psi\,\varphi\,\boldsymbol{X}(y_1,y_2,0,Y_1,Y_2,0)$.

The expressions $\boldsymbol{x}_S$, $\boldsymbol{x}_U$, and $\boldsymbol{x}_C$ are polynomial approximations of the true stable, unstable, and center manifolds associated with the right-most equilibrium point on the axis Ox_1 corresponding to Hamiltonian (2.1), as we know the true ones exist; see, for example, [13]. As the normal form has been carried out up to degree eight, the expressions are of degree seven. More accurate formulae could be obtained by pushing the computations to higher degrees; however, the convergence is not guaranteed because of the divergent character of the normal form computations.

The intersection of the center manifold with the energy surface $\mathcal{H} = h$ for a fixed value of h gives the normally hyperbolic invariant manifold (NHIM) [23], which is 3D. This invariant set has also its stable and unstable invariant manifolds, which are 4D. The NHIM bounds a 4D surface called "transition state" in chemical reaction dynamics, which partitions the energy surface into volumes characterized as "before" and "after" the transformation. The 4D stable and unstable manifolds associated with the 3D NHIM are impenetrable barriers with the topology of multidimensional spherical cylinders. We have approximated all these invariant objects using the normal form Hamiltonian $\mathcal{K}$; see the details in [24, 20].

More specifically, given an energy value, $h > 0$, the NHIM associated with a Hamilton vector field like (2.1), using (transformed) Cartesian coordinates, is

$$\begin{aligned}\mathcal{M}_h^3(y_1, y_2, Y_1, Y_2) = \big\{&(y_1, y_2, Y_1, Y_2) \;\big|\\ &\tfrac{1}{2}\left(Y_1^2 + Y_2^2\right) + \tfrac{1}{2}\left(\mu_1^2 y_1^2 + \mu_2^2 y_2^2\right) + f_2(y_1, y_2, Y_1, Y_2) = h\big\},\end{aligned} \tag{2.4}$$

where f_2 represents polynomials of degree at least three; i.e., this term is responsible for the nonlinear terms in the Hamiltonian vector field. The NHIM acts as a multidimensional saddle "point." The dynamics occurs in the 5D energy surface given by setting $\mathcal{H}$ in the initial Hamiltonian to be a positive constant h. If we set $y_3 = Y_3 = 0$ in the vector field associated with $\mathcal{H}$, then $\dot{y}_3 = \dot{Y}_3 = 0$. Therefore $y_3 = Y_3 = 0$ is a 4D invariant manifold. Its intersection with the 5D energy surface is the NHIM, given by (2.4). The nonlinear terms are much smaller than the linear terms. Therefore, the NHIM is topologically a deformed sphere for the nonlinear problem and has 4D stable and unstable manifolds in the 5D energy surface since normal hyperbolicity is preserved under perturbations.

The stable and unstable manifolds of $\mathcal{M}_h^3$ are known explicitly. Indeed, they are 4D objects acting as multidimensional separatrices. Their expressions are given by

$$\begin{aligned}W^S\left(\mathcal{M}_h^3\right) = \big\{&(y_1, y_2, y_3, Y_1, Y_2, Y_3) \;\big|\; Y_3 = -\mu_3\, y_3,\\ &\tfrac{1}{2}\left(Y_1^2 + Y_2^2\right) + \tfrac{1}{2}\left(\mu_1^2 y_1^2 + \mu_2^2 y_2^2\right) + f_2(y_1, y_2, Y_1, Y_2) = h\big\},\\ W^U\left(\mathcal{M}_h^3\right) = \big\{&(y_1, y_2, y_3, Y_1, Y_2, Y_3) \;\big|\; Y_3 = \mu_3\, y_3,\\ &\tfrac{1}{2}\left(Y_1^2 + Y_2^2\right) + \tfrac{1}{2}\left(\mu_1^2 y_1^2 + \mu_2^2 y_2^2\right) + f_2(y_1, y_2, Y_1, Y_2) = h\big\}.\end{aligned} \tag{2.5}$$

In particular, using formulae (2.4) and (2.5), we obtain the expressions of the manifolds in the original variables, arriving at the following:

- The 3D NHIM of the equilibrium point under study (e.g., the manifold $\mathcal{M}_h^3$) is obtained as the following composition of the transformations: $\boldsymbol{x}_{\text{NHIM}} = \psi\, \varphi\, \boldsymbol{X}(y_1, y_2, 0, Y_1, Y_2, 0)$, to which we add the constraint $\mathcal{K}(y_1, y_2, 0, Y_1, Y_2, 0) = h$.

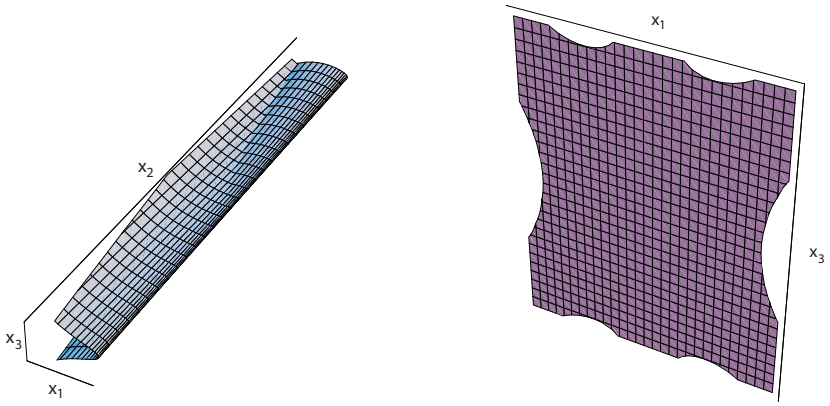


Figure 2. *Surfaces defined by the components x_1, x_2, x_3 of the center manifold; on the left we set $Y_1, Y_2 = 0$, and on the right $y_1, y_2 = 0$.*

- The 4D stable manifold of the NHIM, i.e., the object $W^S(\mathcal{M}_h^3)$, is given by $\boldsymbol{x}_{S\text{NHIM}} = \psi\,\varphi\,\boldsymbol{X}(y_1, y_2, y_3, Y_1, Y_2, -\mu_3 y_3)$, with the constraint $\mathcal{K}(y_1, y_2, y_3, Y_1, Y_2, -\mu_3 y_3) = h$.
- The 4D unstable manifold of the NHIM given through $W^U(\mathcal{M}_h^3)$ is determined as $\boldsymbol{x}_{U\text{NHIM}} = \psi\,\varphi\,\boldsymbol{X}(y_1, y_2, y_3, Y_1, Y_2, \mu_3 y_3)$ together with $\mathcal{K}(y_1, y_2, y_3, Y_1, Y_2, \mu_3 y_3) = h$.

The existence of true invariant sets (e.g., the true NHIM together with its stable and unstable invariant manifolds) of $\mathcal{H}$ close to the ones we have computed can be guaranteed, provided that the global error of the process is kept bounded; see more details on their persistence in [20]. We remark that the expressions of $\boldsymbol{x}_{\text{NHIM}}$, $\boldsymbol{x}_{S\text{NHIM}}$, and $\boldsymbol{x}_{U\text{NHIM}}$ are provided through polynomials of degree seven, since Hamiltonian $\mathcal{K}$ has been calculated up to degree eight.

As our problem at hand has two critical points of the type center × center × saddle, we might determine the NHIM associated with the point whose coordinates are $(-3/2, 0, 0, 0, -3/2, 0)$. Then, using a procedure of Waalkens, Burbanks, and Wiggins [22], we could detect possible heteroclinic connections between the stable manifold of the NHIM corresponding to one of the two unstable equilibria and the unstable manifold of the NHIM of the other equilibrium, as well as some homoclinic connections inside the stable manifold of the NHIM for the two critical points. Note that the NHIMs control the phase space transport across an equilibrium point of center × center × saddle stability type, which is a fundamental mechanism for understanding the physics of the problem.

A projection of the (approximate) center manifold is depicted in Figure 2, whereas in Figure 3 we draw two surfaces corresponding to the projection of the NHIM in the configuration space and in the space of momenta, for a fixed value of $h = 3 \times 10^{-3}$ and the integrals $(Y_1^2 + \mu_1^2 y_1^2)/2 = 10^{-4}$ and $(Y_2^2 + \mu_2^2 y_2^2)/2 = 0.0029003790975507078$.

In order to estimate the error committed after truncating the Lie transformation we have used MATHEMATICA, Version 5.1, with 60 digits of precision. All computations involved in the linear changes of coordinates, in the Taylor expansions, and in the Lie transformations have been performed within this precision. The global error committed when dropping the tail of the transformation has been studied by estimating the difference $E = |\mathcal{H}(\boldsymbol{x}) - \mathcal{K}(\boldsymbol{Y}(\boldsymbol{x}))|$ within $B_{(0,\delta)}$, a ball centered at the colinear point of radius δ. Our study has been developed

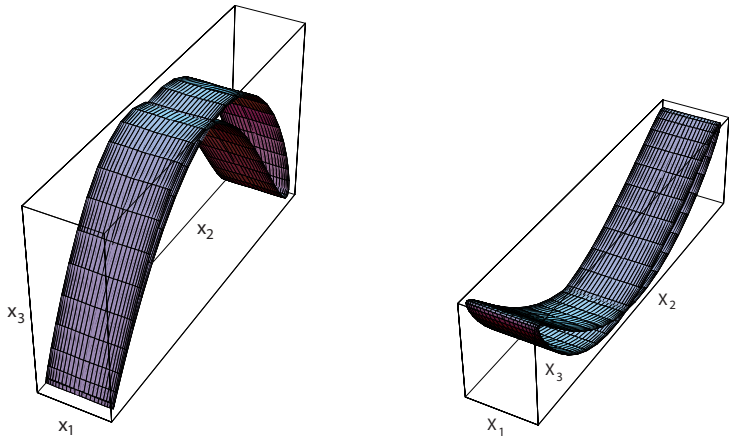


Figure 3. *Left: projection of the NHIM onto the configuration space. Right: projection of the NHIM onto the space defined by the axes* X_1, X_2, X_3.

Table 2

Error committed by the Lie transformation process, after truncating the tail of the normal form at orders one to six, that is, after dropping the polynomials of degrees three and higher to degrees nine and higher.

Order 1	$E = 0.000052784698125$
Order 2	$E = 0.000015958799870$
Order 3	$E = 7.175092835102288 \times 10^{-7}$
Order 4	$E = 1.561351100193047 \times 10^{-7}$
Order 5	$E = 1.1439383299375211 \times 10^{-8}$
Order 6	$E = 7.6559789904728368 \times 10^{-9}$

in a neighborhood of radius $\delta = 10^{-2}$ around the equilibrium. We give the results in Table 2.

Another way of computing the center manifold, the NHIM, and its stable and unstable sets is through the determination of the generalized normal form using the integral $\mathcal{G} = (X_3^2 - \mu_3^2\, x_3^2)/2$. This results in the so-called reduction to the center manifold; see [11]. The strategy is similar to the approach we have followed, and both techniques give equivalent results when the frequencies μ_1 and μ_2 are independent over the rational numbers. In case of commensurability of μ_1 and μ_2 we could not have followed the approach of this paper, as we would have been able to obtain one (or even two) formal integrals out of $\mathcal{H}$, but the invariant manifolds would have been determined very similarly to what we have done here; see, for example, [20].

3. Other invariant sets of the segment.

3.1. Quasi-periodic orbits and invariant 2-tori. We may apply other normal form transformations with the aim of calculating more local invariant manifolds related with the unstable equilibrium point on which we have chosen to perform our analysis. The plan is to approximate (true) periodic orbits and invariant 2-tori.

We choose this time the quadratic integral $\mathcal{G} = \mathcal{G}_1 + \mathcal{G}_2$, with $\mathcal{G}_1 = (X_1^2 + \mu_1^2 x_1^2)/2$ and $\mathcal{G}_2 = (X_2^2 + \mu_2^2 x_2^2)/2$. Thus we calculate the generalized normal form associated with $\mathcal{G}$, say $\mathcal{S}$, and the corresponding direct change of coordinates $\boldsymbol{x} = \boldsymbol{X}^*(\boldsymbol{y}^*)$, up to degree four, that is,

Table 3
Coefficients of the normal form Hamiltonian $\mathcal{S}$.

$b(1,0,0,0)$	$1.125000000000000000000000000000000000000000000000000000000000000\,\imath$
$b(0,1,0,0)$	$1.210582304803311353091514347995139442215099854757553831449 74\,\imath$
$b(0,0,1,1)$	$0.335582304803311353091514347995139442215099854757553831449744$
$b(1,0,1,0)$	$3.634824707673186493250495102357885829267333915512489889 1180\,\imath$
$b(1,0,0,1)$	$-0.080726447867715688281910958496296389887292109746952783700 0\,\imath$
$b(0,1,1,0)$	$-1.90356036537318988890446565329012882956856137657406717358988\,\imath$
$b(0,1,0,1)$	$0.042276499957192756217406856401626700233983520033737888171 13\,\imath$
$b(0,0,3,0)$	$129.040864005748859507016198556538528233553736784577101088946$
$b(0,0,0,3)$	$-0.001413592970556240158304896755127595167352354829610932212$
$b(0,0,2,1)$	$4.215891942174566702354984170999338413113368744656003421 62\,\imath$
$b(0,0,1,2)$	$-0.093631470141442053571191676928864752587088433271232600 02$
$b(2,0,0,0)$	$0.886410098551843434610669406920829413813780008185841007880 21$
$b(0,2,0,0)$	$0.306118364885931497060286299488490305426981219701420771417 97$
$b(1,1,0,0)$	$-0.924952995365646170377586332965110260608061049136296356612 80$
$b(1,0,2,0)$	$-60.261329873355174731798372121243539876539773317904542556589\,\imath$
$b(0,1,2,0)$	$-15.260589952741833305824483579409751879291548890353264223706 9\,\imath$
$b(1,0,1,1)$	$-4.151864997554668685742992182078910134548761524351851838826\,\imath$
$b(0,1,1,1)$	$1.827211544962906181230889568250744289955895507169110007984\,\imath$
$b(1,0,0,2)$	$-0.029723740064180193376835465423773894953343523575710591735\,\imath$
$b(0,1,0,2)$	$-0.007527245248895515811126375175495963013469046812401059291 9\,\imath$
$b(0,0,4,0)$	$-1213.703910843576960464925416370618260065745658108317223064 15$
$b(0,0,0,4)$	$-0.000295285578902026766305639023929043006529343675486903 90$
$b(0,0,3,1)$	$-134.715170011016148143213724031526552598153315876730687414 7$
$b(0,0,2,2)$	$-5.245897609280132845418090838201297781196957119689698490 6$
$b(0,0,1,3)$	$-0.066447897922673947076292371140097546195265559351750805 5$

taking into account second order terms.

In the diagonal complex coordinates defined through (2.2), but using $\boldsymbol{y}^*$ instead of $\boldsymbol{y}$, the Hamilton function $\mathcal{S}$ reads

$$\mathcal{S} = \sum_{\substack{0\le j,k\le 2,\, 0\le \ell,m\le 4\\ 2\le 2j+2k+\ell+m\le 4}} b(j,k,\ell,m)(p_1q_1)^j(p_2q_2)^k p_3^\ell q_3^m, \tag{3.1}$$

where the coefficients $b(j,k,\ell,m)$ are given in Table 3.

This transformation has the following features:

(i) The reduction process associated with the normal form is regular, and therefore the resulting phase space is a plane; see also [16].

(ii) In practice we have already introduced two symmetries in the system, i.e., $\mathcal{G}_1$ and $\mathcal{G}_2$; that is, if we first apply the generalized normal form transformation, picking $\mathcal{G}_1$, and then we apply another generalized normal form, extending the integral $\mathcal{G}_2$, we end up with the same Hamiltonian $\mathcal{S}$. Alternatively, we get the same result if we use first $\mathcal{G}_2$ and then $\mathcal{G}_1$.

(iii) Fixing a ball of radius $\delta = 10^{-2}$ and calling $\boldsymbol{y}^* = \boldsymbol{Y}^*(\boldsymbol{x})$ the inverse change of $\boldsymbol{x} = \boldsymbol{X}^*(\boldsymbol{y}^*)$, the error up to terms of degree four is $E = |\mathcal{H}(\boldsymbol{x}) - \mathcal{S}(\boldsymbol{Y}^*(\boldsymbol{x}))| = 0.000056212888314$.

Next, after fixing the value for the integrals as $\mathcal{G}_1 = j_1 \geq 0$ and $\mathcal{G}_2 = j_2 \geq 0$, we arrive at a Hamiltonian of one DOF. Now, the corresponding reduced Hamiltonian obtained from $\mathcal{S}$ is named $\mathcal{Q}(y_3^*, Y_3^*; j_1, j_2)$. To calculate its relative equilibrium points we need to determine the roots of the algebraic system

$$\left(\frac{\partial \mathcal{Q}}{\partial Y_3^*}, \frac{\partial \mathcal{Q}}{\partial y_3^*}\right) = (0,0). \tag{3.2}$$

The solution of this system of equations yields four nondegenerate equilibrium points:

$$(1,2)\quad \begin{aligned} y_3^{*1,2} &= \left(b_1 + b_2\sqrt{b_3} \pm b_4\sqrt{b_3\, b_5 + b_6\sqrt{b_3}} \pm b_7\sqrt{b_5 + \tfrac{b_6}{\sqrt{b_3}}}\right) \Big/ (b_8\sqrt{b_3}),\\ Y_3^{*1,2} &= c - \tfrac{1}{2}\sqrt{b_3} \pm \tfrac{1}{2}\sqrt{b_5 + \tfrac{b_6}{\sqrt{b_3}}}, \end{aligned}$$

$$(3,4)\quad \begin{aligned} y_3^{*3,4} &= \left(-b_1 + b_2\sqrt{b_3} \pm b_4\sqrt{b_3\, b_5 - b_6\sqrt{b_3}} \mp b_7\sqrt{b_5 - \tfrac{b_6}{\sqrt{b_3}}}\right) \Big/ (b_8\sqrt{b_3}),\\ Y_3^{*3,4} &= c + \tfrac{1}{2}\sqrt{b_3} \pm \tfrac{1}{2}\sqrt{b_5 - \tfrac{b_6}{\sqrt{b_3}}}, \end{aligned}$$

where the b_i's are polynomials of degree one or two in j_1, j_2, for $i = 1, \ldots, 8$, and c is a negative real constant. Now, we assume that $b_3, b_8 \neq 0$. Depending on the signs of b_3, b_5, and b_6, we arrive at these situations:

- Whenever $b_3 > 0$ and $b_5\sqrt{b_3} > |b_6|$, we get four different equilibria (y_3^{*i}, Y_3^{*i}), $i = 1, \ldots, 4$, which correspond to four families of invariant 2-tori of $\mathcal{H}$.
- If $b_3 > 0$ and $-|b_6| < b_5\sqrt{b_3} < |b_6|$, we find two different equilibrium points: either $(y_3^{*1,2}, Y_3^{*1,2})$ when $b_6 > 0$ or $(y_3^{*3,4}, Y_3^{*3,4})$ when $b_6 < 0$. They correspond to two families of invariant 2-tori of the original system.
- If $b_3 > 0$ and $b_5\sqrt{b_3} = |b_6|$ or $b_3 > 0$ and $b_5\sqrt{b_3} = -|b_6|$, there are three equilibria. More specifically, if $b_5\sqrt{b_3} = b_6$, we have that (y_3^{*3}, Y_3^{*3}) coincides with (y_3^{*4}, Y_3^{*4}), whereas if $b_5\sqrt{b_3} = -b_6$, the point (y_3^{*1}, Y_3^{*1}) is the same as (y_3^{*2}, Y_3^{*2}). These equilibria are related with three families of invariant 2-tori of the original system.
- Finally, when $b_3 < 0$ or $b_5\sqrt{b_3} < -|b_6|$, there are no isolated critical points.

Now since the polynomials b_i depend on j_1 and j_2, the conditions $b_5\sqrt{b_3} = |b_6|$ and $b_5\sqrt{b_3} = -|b_6|$ correspond to bifurcation curves in the plane of parameters defined by j_1 and j_2. By going back to the initial variables and undoing the Lie transformation and the other linear changes, these curves correspond to bifurcation curves of invariant tori in the original system.

The approximate invariant tori and the quasi-periodic trajectories can be calculated explicitly as follows. First we compose the different changes of coordinates $\boldsymbol{x} = \psi\varphi\boldsymbol{X}^*(\boldsymbol{y}^*)$. Then we set

$$\mu_1\, y_1^* = \sqrt{j_1}\cos t, \quad Y_1^* = \sqrt{j_1}\sin t, \quad \mu_2\, y_2^* = \sqrt{j_2}\cos u, \quad Y_2^* = \sqrt{j_2}\sin u,$$

with $t, u \in [0, 2\pi)$. Thus, we arrive at an expression of the form $\boldsymbol{y}^*(t, u; j_1, j_2)$. This represents a family of invariant 2-tori (depending on $j_1, j_2 \geq 0$) in the phase space $\mathbf{R}^6$, parameterized by u and t. To obtain a very accurate approximation of the invariant 2-tori we should carry out

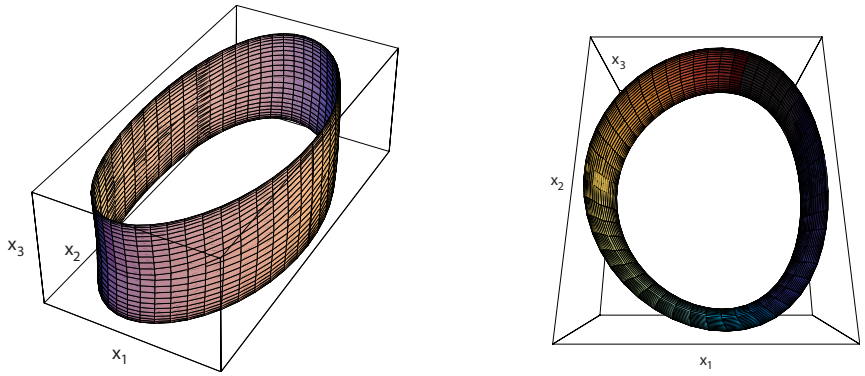


Figure 4. *Left: a torus in the space of coordinates. Right: another view of the same torus.*

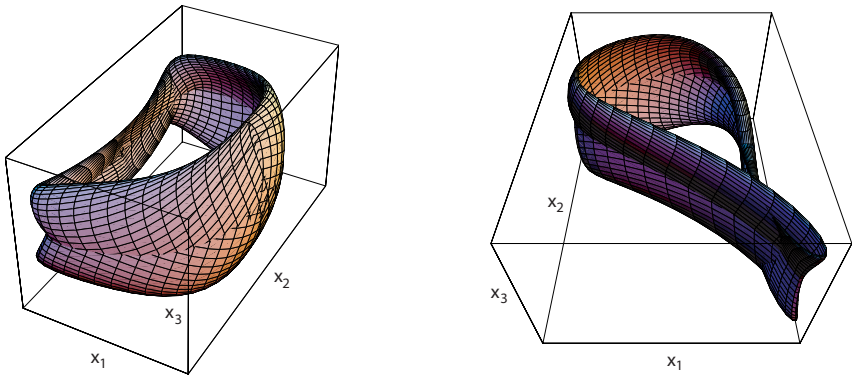


Figure 5. *Two invariant 2-tori for two different values of* (y_3^*, Y_3^*) *(e.g., two roots of* (3.2)*) in the space of coordinates.*

the Lie transformation to a high order L. Two views of a single invariant 2-torus are depicted in Figure 4.

Other invariant 2-tori are drawn in Figure 5.

Next, given a certain invariant 2-torus, fixing either the angle u or the angle t or putting one of them in terms of the other, we obtain some quasi-periodic orbits confined in the torus. We have depicted them in Figure 6.

3.2. Some periodic trajectories. We choose $\mathcal{G} = (X_1^2 + \mu_1^2 x_1^2)/2$ and compute the corresponding generalized normal form $\mathcal{T}$, together with its direct change of coordinates $\boldsymbol{x} = \boldsymbol{X}^{\#}(\boldsymbol{y}^{\#})$, up to terms of degree three, i.e., up to $L = 1$. Alternatively we could have chosen $\mathcal{G} = (X_2^2 + \mu_2^2 x_2^2)/2$, computing its corresponding normal form and analyzing the resulting reduced Hamiltonian. This would lead to the approximation of other periodic trajectories. The techniques are analogous to what we lay out below.

Hamiltonian $\mathcal{T}$, written down in the diagonal complex coordinates (2.2) using now $\boldsymbol{y}^{\#}$, is

$$\mathcal{T} = \sum_{\substack{0\le j\le 1,\, 0\le k,\ell,m,n\le 3\\ 2\le 2j+k+\ell+m+n\le 3}} c(j,k,\ell,m,n)(p_1q_1)^j p_2^k q_2^\ell p_3^m q_3^n, \tag{3.3}$$

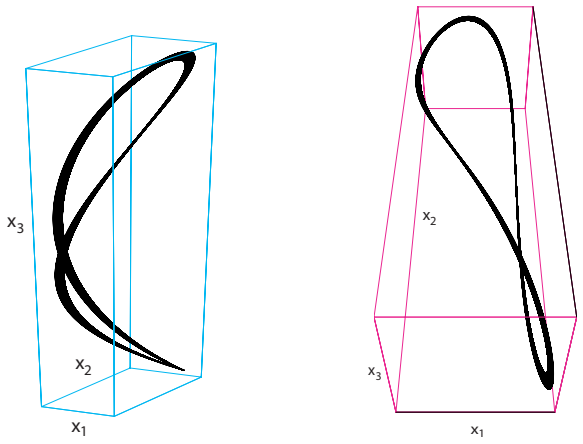


Figure 6. *Two examples of quasi-periodic orbits confined in their corresponding invariant tori in the space of coordinates.*

Table 4
Coefficients of the normal form Hamiltonian $\mathcal{T}$.

$c(1,0,0,0,0)$	$1.125000000000000000000000000000000000000000000000\,\imath$
$c(0,1,1,0,0)$	$1.210582304803311353091514347995139442215099854757 5\,\imath$
$c(0,0,0,1,1)$	$0.335582304803311353091514347995139442215099854757 55$
$c(1,1,0,0,0)$	$-0.491460824834537388029947603913669656105240519171 95$
$c(1,0,1,0,0)$	$0.491460824834537388029947603913669656105240519171 95\,\imath$
$c(1,0,0,1,0)$	$-0.080726447867715688281910958496296389887292109746952\,\imath$
$c(1,0,0,0,1)$	$3.634824707673186493250495102357885829267333915512 4\,\imath$
$c(0,3,0,0,0)$	$1.303744683617272893585911308750873703128426353578 9\,\imath$
$c(0,0,3,0,0)$	$-1.303744683617272893585911308750873703128426353578 9$
$c(0,0,0,3,0)$	$-0.001413592970556240158304896755127595167352354829 6109$
$c(0,0,0,0,3)$	$129.040864005748859507016198556538528233553736784 57$
$c(0,2,1,0,0)$	$-0.575740975073937959063139243927674313362623551551 40$
$c(0,2,0,1,0)$	$-0.16188942482724693423 - 0.27025265010926970431\,\imath$
$c(0,2,0,0,1)$	$7.28930476709861929725 - 12.168515224911591130 07\,\imath$
$c(0,1,2,0,0)$	$0.575740975073937959063139243927674313362623551551 40\,\imath$
$c(0,0,2,1,0)$	$0.16188942482724693423 - 0.27025265010926970431\,\imath$
$c(0,0,2,0,1)$	$-7.28930476709861929725 - 12.168515224911591130 07\,\imath$
$c(0,1,0,2,0)$	$0.026233630285096945 15 + 0.011132608086002999 24\,\imath$
$c(0,0,1,2,0)$	$0.011132608086002999 24 + 0.026233630285096945 15\,\imath$
$c(0,0,0,2,1)$	$-0.093631470141442053571191676928864752587088433271232$
$c(0,1,0,0,2)$	$-53.18554949587127053795 + 22.570032128287318015 73\,\imath$
$c(0,0,1,0,2)$	$22.570032128287318015 73 - 53.18554949587127053795\,\imath$
$c(0,0,0,1,2)$	$4.215891942174566702354984170999338413113368744656 0$
$c(0,1,1,1,0)$	$-0.111775228653366256821364328689788920349112883743 02\,\imath$
$c(0,1,1,0,1)$	$5.032840829077992732754860200161798071774714296709 6\,\imath$
$c(0,1,0,1,1)$	$1.929138488600771446018148152410009161549951453436 1\,\imath$
$c(0,0,1,1,1)$	$1.929138488600771446018148152410009161549951453436 1$

where the coefficients $c(j,k,\ell,m,n)$ are displayed in Table 4.

The reduction induced by the normal form computation is regular and, after fixing the

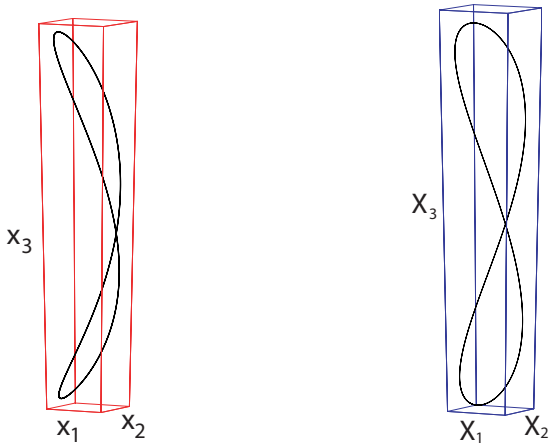


Figure 7. *An approximate periodic orbit in the space of coordinates (left) and in the space of momenta (right). We have fixed the value of integral* $j_1 = 1/50$.

value for the integral, $\mathcal{G} = j_1 \geq 0$, we arrive at a Hamiltonian of two DOF defined on the reduced phase space, which this time is isomorphic to $\mathbf{R}^4$. Now, the corresponding reduced Hamiltonian obtained from $\mathcal{T}$ is called $\mathcal{U}(y_2^\#, y_3^\#, Y_2^\#, Y_3^\#; j_1)$. Hence, the roots of the algebraic system

$$\left(\frac{\partial \mathcal{U}}{\partial y_2^\#}, \frac{\partial \mathcal{U}}{\partial y_3^\#}, \frac{\partial \mathcal{U}}{\partial Y_2^\#}, \frac{\partial \mathcal{U}}{\partial Y_3^\#} \right) = (0,0,0,0) \tag{3.4}$$

give the relative equilibria of $\mathcal{U}$ and are indeed in correspondence with some periodic trajectories of $\mathcal{H}$. Note that system (3.4) depends on the parameter j_1, so we obtain a family of approximate periodic trajectories around the colinear equilibrium point we have chosen parameterized by j_1. Once we have calculated an isolated root of (3.4), say $\boldsymbol{e}^\# = (\bar{y}_2^\#, \bar{y}_3^\#, \bar{Y}_2^\#, \bar{Y}_3^\#)$, we can determine the approximation of a closed orbit computing the direct change of variables, replacing $y_2^\#$, $y_3^\#$, $Y_2^\#$, and $Y_3^\#$ by the particular values they have in $\boldsymbol{e}^\#$, and $(y_1^\#, Y_1^\#)$ by $((\sqrt{2j_1}/\mu_1)\cos t, \sqrt{2j_1}\sin t)$. We present a single periodic orbit of the Hamiltonian $\mathcal{H}$ in Figure 7.

The existence of exact periodic orbits for $\mathcal{H}$ close to the ones calculated above is guaranteed using an approach based on the implicit function theorem [14]. See also the many examples in the book by Meyer and Hall [13], where they prove the existence of families of periodic trajectories through implicit mapping theorems.

Finally, if we choose a ball of radius 10^{-2} centered at the colinear point, the global error up to terms of order $L = 1$ is maintained below 1.6×10^{-3}.

4. Conclusions. We apply (three) different normal form transformations to the problem of a particle orbiting around a straight segment, with the aim of determining some invariant objects in the vicinity of the unstable colinear equilibrium points.

By using the methodology of generalized normal forms [15, 16] and choosing $\mathcal{G} = \mathcal{H}_0$, we have computed high order approximations of the center, the stable, and the unstable manifolds

of one of the colinear equilibrium points. We also have obtained the NHIM associated with such an equilibrium, as well as its stable and unstable manifolds.

Selecting $\mathcal{G} = (X_1^2 + \mu_1^2 x_1^2)/2 + (X_2^2 + \mu_2^2 x_2^2)/2$, we have determined some invariant 2-tori and quasi-periodic orbits of the original Hamiltonian, obtaining furthermore the conditions that some parameters must hold to achieve various bifurcations.

We have also taken $\mathcal{G} = (X_1^2 + \mu_1^2\, x_1^2)/2$, obtaining some periodic orbits of the original Hamiltonian as a result of the computation of the relative equilibria of the corresponding normal form Hamiltonian, for different values of the parameter $\mathcal{G} = g$ introduced through the transformation.

As we give the explicit expressions of the three Hamiltonians, $\mathcal{K}$, $\mathcal{S}$, and $\mathcal{T}$, corresponding to the three normal forms, and since it is straightforward to determine $\mathcal{H}$ (the Hamiltonian previous to the three normalizations) up to terms of degree eight, one could obtain the direct and inverse changes of coordinates. Thus, one can recover the periodic and quasi-periodic orbits, the invariant 2-tori, the NHIM, and its associated manifolds up to a high order of approximation.

We have not described the algorithms applied to compute the three different normal forms and their corresponding changes of coordinates, as they are given in full detail in [6].

Our approach provides a methodology to approximate invariant sets of an equilibrium point for a (polynomial) Hamilton function, using techniques based on generalized normal forms. The technique can be applied for Hamiltonians of m DOF (with $m \geq 1$) and generalized to the computation of the invariant manifolds associated with an invariant d-torus, $d > 0$.

We emphasize that the integral $\mathcal{G}$ of $\mathcal{H}_0$ is chosen previously and it does need to be the part of $\mathcal{H}_0$ whose associated matrix is semisimple. It means that this technique can be applied to calculate formal integrals of polynomial Hamiltonians whose dominant parts are related to nilpotent matrices, A, at the price of introducing nonpolynomial generating functions in many cases.

Acknowledgment. The authors thank the remarks and suggestions made by Dr. Víctor Lanchares (Universidad de La Rioja) on a previous version of the paper.

REFERENCES

[1] A. Deprit, *Canonical transformations depending on a small parameter*, Celestial Mech., 1 (1969), pp. 12–30.

[2] A. Elipe, M. Lara, and A. Riaguas, *Periodic orbits around natural elongated bodies*, Rev. Brasileira Ciências Mecânicas, 21 (1999), pp. 198–207.

[3] A. Elipe and A. Riaguas, *Nonlinear stability under a logarithmic gravity field*, Int. Math. J., 3 (2003), pp. 435–453.

[4] J. Guckenheimer and A. Vladimirsky, *A fast method for approximating invariant manifolds*, SIAM J. Appl. Dyn. Syst., 3 (2004), pp. 232–260.

[5] S. Gutiérrez–Romero, J. F. Palacián, and P. Yanguas, *Classification and reduction of polynomial Hamiltonians with three degrees of freedom*, in Analytic and Numerical Techniques in Orbital Dynamics, Monogr. Real Acad. Ci. Exact. Fís.-Quím. Nat. Zaragoza 22, A. Elipe and M. Palacios, eds., 2003, pp. 101–109.

[6] S. Gutiérrez–Romero, J. F. Palacián, and P. Yanguas, *A universal procedure for normalizing n-degree-of-freedom polynomial Hamiltonian systems*, SIAM J. Appl. Math., 65 (2005), pp. 1130–1152.

[7] S. Gutiérrez–Romero, J. F. Palacián, and P. Yanguas, *The invariant manifolds of a finite straight segment*, in Actas de las VI Jornadas de Mecánica Celeste (Zaragoza, Spain), Monogr. Real Acad. Ci. Exact. Fís.-Quím. Nat. Zaragoza 25, J. F. Palacián and P. Yanguas, eds., 2004, pp. 137–148.

[8] P. Halamek, *Motion in the Potential of a Thin Bar*, Ph.D. thesis, Department of Aerospace Engineering and Engineering Mechanics, University of Texas, Austin, TX, 1988.

[9] M. E. Henderson, *Multiple parameter continuation: Computing implicitly defined k-manifolds*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 12 (2002), pp. 451–476.

[10] J. Henrard, *The algorithm of the inverse for Lie transform*, in Recent Advances in Dynamical Astronomy, B. D. Tapley and V. Szebehely, eds., D. Reidel Publishing Company, Dordrecht, The Netherlands, 1973, pp. 250–259.

[11] À. Jorba and J. Masdemont, *Dynamics in the center manifold of the colinear points of the restricted three body problem*, Phys. D, 132 (1999), pp. 189–213.

[12] B. Krauskopf and H. Osinga, *Two-dimensional global manifolds of vector fields*, Chaos, 9 (1999), pp. 768–774.

[13] K. R. Meyer and G. R. Hall, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Appl. Math. Sci. 90, Springer-Verlag, Berlin, Heidelberg, 1992.

[14] J. F. Palacián, *Invariant manifolds of an autonomous ordinary differential equation from its generalized normal forms*, Chaos, 13 (2003), pp. 1188–1204.

[15] J. F. Palacián and P. Yanguas, *Reduction of polynomial Hamiltonians by the construction of formal integrals*, Nonlinearity, 13 (2000), pp. 1021–1054.

[16] J. Palacián and P. Yanguas, *Reduction of polynomial planar Hamiltonians with quadratic unperturbed part*, SIAM Rev., 42 (2000), pp. 671–691.

[17] A. Riaguas, *Dinámica Orbital Alrededor de Cuerpos Celestes con Forma Irregular*, Ph.D. thesis, Departamento de Matemática Aplicada, Universidad de Zaragoza, Zaragoza, Spain, 1999.

[18] A. Riaguas, A. Elipe, and M. Lara, *Periodic orbits around a massive straight segment*, Celestial Mech. Dynam. Astronom., 73 (1999), pp. 169–178.

[19] A. Riaguas, A. Elipe, and T. López–Moratalla, *Non–linear stability of the equilibria in the gravity field of a finite straight segment*, Celestial Mech. Dynam. Astronom., 81 (2001), pp. 235–248.

[20] T. Uzer, C. Jaffé, J. F. Palacián, P. Yanguas, and S. Wiggins, *The geometry of reaction dynamics*, Nonlinearity, 15 (2002), pp. 957–992.

[21] J. C. van der Meer, *The Hamiltonian Hopf Bifurcation*, Lecture Notes in Math. 1160, Springer-Verlag, Berlin, Heidelberg, 1985.

[22] H. Waalkens, A. Burbanks, and S. Wiggins, *A computational procedure to detect a new type of high-dimensional chaotic saddle and its application to the 3D Hill's problem*, J. Phys. A, 37 (2004), pp. L257–L265.

[23] S. Wiggins, *Normally Hyperbolic Invariant Manifolds in Dynamical Systems*, Appl. Math. Sci. 105, Springer-Verlag, Berlin, Heidelberg, 1994.

[24] S. Wiggins, L. Wiesenfeld, C. Jaffé, and T. Uzer, *Impenetrable barriers in phase-space*, Phys. Rev. Lett., 86 (2001), pp. 5478–5481.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 1, pp. 30–65

Compound Laser Modes of Mutually Delay-Coupled Lasers*

Hartmut Erzgräber†, Bernd Krauskopf‡, and Daan Lenstra§

Abstract. We consider a model of two mutually delay-coupled semiconductor lasers (SLs) in a face to face configuration. The lasers are coherently coupled via their optical fields, where the time delay τ arises from the finite propagation time of the light from one laser to the other. This system is described well by single mode rate equations, which are a system of delay differential equations (DDEs) with one fixed delay.

We study the compound laser modes (CLMs) of the system, where both lasers operate at an identical, time-independent frequency. By making use of numerical continuation applied to the full DDEs, we present a comprehensive geometrical picture of how CLMs depend on the two main physical parameters, namely, the coupling phase and the detuning between the two lasers. The different branches of CLMs are organized by unfoldings of pitchfork bifurcations that exist for zero detuning. As a function of the detuning, different branches of CLMs connect, split, or disappear in transitions through codimension-one singularities in the surface of CLMs.

Key words. mutually delay-coupled lasers, delay differential equations (DDEs), numerical continuation, symmetry breaking, singularity

AMS subject classifications. 15A15, 15A09, 15A23

DOI. 10.1137/040619958

1. Introduction.

In this paper we consider a simple setup of two mutually delay-coupled semiconductor lasers (SLs) in a face to face configuration. The system is sketched in Figure 1 and has recently attracted quite some attention, both experimentally and theoretically. It is seen as a prototype system for understanding the dynamics of coupled SLs. This is crucial because of the present technological trend of integrating SLs on-chip into more complicated optical systems. In particular, coupled lasers are thought to be important devices for use in future all-optical signal processing. Due to the small sizes of SLs and the typical distances between the lasers, one is generally dealing with substantial delay in the coupling.

An SL on its own behaves simply as a damped oscillator, characterized by its *relaxation oscillation* with a typical frequency in the order of a few GHz; note that this damped oscillation is quite harmonic and should not to be confused with relaxation oscillations in slow-fast systems. However, due to a combination of the material properties of the semiconductor

*Received by the editors December 1, 2004; accepted for publication (in revised form) by T. Kaper September 16, 2005; published electronically January 6, 2006.
http://www.siam.org/journals/siads/5-1/61995.html

†Afdeling Natuurkunde en Sterrenkunde, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands (h.erzgraber@few.vu.nl).

‡Department of Engineering Mathematics, University of Bristol, Bristol BS8 1TR, UK (b.krauskopf@bristol.ac.uk).
The work of this author was supported by an Engineering and Physical Sciences Research Council (EPSRC) Advanced Research Fellowship grant.
§Afdeling Natuurkunde en Sterrenkunde, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands, and Research Institute COBRA, Technical University Einhoven, The Netherlands (lenstra@nat.vu.nl).

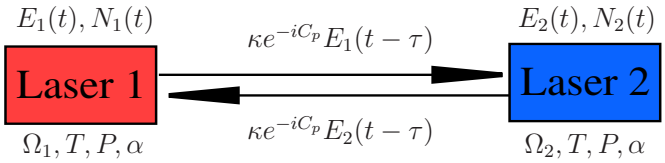


Figure 1. *Sketch of two mutually delay-coupled lasers.*

active material and the low reflectivities of the mirrors, this type of laser is very sensitive to external optical influences [18, 20, 30]. For example, it is well known that delayed optical feedback may destabilize the SL. While this is undesirable in applications such as optical data storage, more recently destabilized SL systems are being considered and studied as broadband light sources for applications such as private communication [8].

The above discussion shows that the SL system we consider here can be seen, more generally, as a prototype of two mutually delay-coupled nonlinear oscillators. There has been quite a lot of interest recently in delay-coupled oscillators in different fields, including arrays of quasi-optical oscillators [43, 39], as well as chemical oscillations, biological clocks, and neural networks; see, for example, [34, 33, 26]. On one hand, delayed coupling of stable subsystems can result in instabilities and even chaos [32]. On the other hand, time delayed coupling can be used to stabilize chaotic systems. Recently studied phenomena in delay-coupled systems include multistabilities, amplitude death, or chaos synchronization in conjunction with symmetry breaking [31, 27, 14, 41].

SLs have the advantage that they can be controlled well, so that different dynamics and bifurcations can be studied experimentally. In the setup of Figure 1 the coupling is achieved by injecting a part of the emitted light of one laser into the respective other laser. We consider here the case of two SLs which are identical, except for a possible *detuning* between the two lasers. The detuning is the difference in their solitary optical frequencies, that is, the frequency the lasers choose when they are not coupled to each other. Note that SLs can be characterized well and then hand selected, so that the assumption that they have identical material properties can be guaranteed to very good approximation in an experiment. Furthermore, the frequencies of the lasers can be controlled and measured very accurately [1].

SL systems are also very fascinating to study theoretically. They can be modeled well by single mode rate equations for the electric field E and the inversion N inside the laser cavity. (The inversion is given by the number of electron-hole pairs, which can recombine to produce one photon each.) External optical influences are modeled by extra terms that lead to a mathematical description in the form of a delay differential equation (DDE), if delay is a feature. Typically, this involves a single fixed delay τ. The best known and most studied laser DDEs are the Lang–Kobayashi equations for an SL subject to optical feedback from a mirror at some fixed distance from the laser [25, 22]. The mutually delay-coupled SLs studied here can also be modeled by rate equations in the same spirit; see section 2 for details. An important feature of the rate equation model is a number of symmetries, especially an S^1-symmetry of rotation of the electric fields of both lasers.

Since DDEs have an infinite-dimensional phase space, they are quite a difficult class of dynamical systems to study. While linear stability analysis and local bifurcation theory for

equilibria of DDEs with fixed delays are well established [2, 12], it is nevertheless quite a challenge to perform a bifurcation analysis of a DDE arising in a specific application. The equations for equilibria and their stability are typically complicated implicit transcendental expressions that need to be solved numerically. The alternative approach is to use numerical continuation techniques—which are well established for the study of ordinary differential equations—directly for the bifurcation study of a given DDE. This is now possible with the recently developed MATLAB package DDE-BIFTOOL [4], which allows one to find and follow equilibria and periodic orbits. Furthermore, DDE-BIFTOOL finds stability information, which is computed by continuing appropriate monodromy operators. In this way, bifurcations of equilibria and periodic orbits can be detected and studied as a function of system parameters. The study of delay effects in SLs has been a major motivation and test case for the further development of numerical continuation tools for DDEs [18].

DDE-BIFTOOL is an essential tool for the case study presented here, as it allows us to study the basic solutions (and their stability), called *compound laser modes* (CLMs), of two mutually delay-coupled SLs as modeled by rate equations with a single fixed delay. CLMs are special types of periodic orbits where the motion is only in the direction of the S^1-symmetry. Physically, they correspond to both lasers lasing with constant, but possibly different, intensities at the same optical frequency. We concentrate on the short delay regime, where the coupling time is of the same order of magnitude as the relaxation oscillation period. From the dynamical systems point of view this intermediate regime between ultrashort delay and long delay times is of particular interest because one can expect rich dynamics due to the competition between the relaxation oscillation frequency and the round-trip frequency. Additionally it has the advantage that the number of CLMs is small, so that their individual behavior can be studied. Also from the experimental point of view this regime is interesting. Reference [7] contains an experimentally observed characteristic bifurcation scenario within the locking region, which exists for small detuning between the two lasers. This scenario involves multistability of CLMs, as was further confirmed with a limited theoretical bifurcation study.

In this paper we take a dynamical systems point of view and present a comprehensive geometric picture of the CLM structure of the system. We show how different types of CLMs coexist, interact, change their stability, and bifurcate over a wide range of the two main parameters, namely, the coupling phase and the detuning between the two lasers. For zero detuning the system has the additional phase-space symmetry of exchanging the two lasers, and we find that certain types of CLMs appear and disappear in pitchfork bifurcations. In fact, the case of zero detuning organizes the bifurcation diagram: when the detuning is "switched on" the pitchfork bifurcations for zero detuning unfold into saddle-node bifurcations. This results, globally, in branches of CLMs in the form of horseshoes. When the detuning is increased further different branches of CLMs connect, split, or disappear in transitions through codimension-one singularities in the surface of CLMs. Finally, a simple limiting situation is reached.

The overall bifurcation diagram of CLMs is an intriguing structure of bifurcations and transitions through singularities of solution surfaces. It forms the "backbone" of constant-intensity solutions, around which the dynamics evolves. Therefore, it forms the basis for understanding the behavior of two mutually delay-coupled SLs. On the other hand, the CLM structure stems from and bears the signature of various underlying symmetries and, as such, is of interest in the more general context of delay-coupled oscillators.

We finish this introduction with a brief overview of the recent literature on the system of two mutually delay-coupled SLs. A theoretical study of the system in the limit of zero delay was performed in [42], while the limit for very large delay is the focus of theoretical studies in [17, 37]. Chaos synchronization and symmetry breaking has been reported for long delay times in [14]. An experimental and numerical study of the onset of chaos synchronization for different coupling strengths and injection currents can be found in [24]. In [28] numerical simulations are performed and an analytical formula is derived for the oscillation frequency in the mode beating regime for short delays. Numerical investigations and an approximate thermodynamic potential are the subject of [24, 36]. In [15, 16, 41] it is shown that for a short delay time τ, in the order of the relaxation oscillation period, frequency locking with continuous wave emission and regular intensity oscillations is dominant. A characteristic scenario, consisting of optical frequency locking leading toward successive states of periodic intensity oscillations, as a function of the detuning between the two lasers, has recently been demonstrated in [41].

The paper is organized as follows. The rate equation model and its properties are discussed in section 2. In section 3 the CLMs are introduced. How CLMs can be continued numerically is explained in section 4. For the case of zero detuning we derive analytical expressions for the CLMs and present a detailed continuation study. In section 6.1 we consider the effect of small detuning. In section 6.2 we further increase the detuning, which results in the restructuring of branches of CLMs due to saddle singularities. We discuss the limit of large detuning in section 6.3. The results are presented in condensed form in section 6.3 as plots of the surface of CLMs over the two-dimensional parameter space. Finally, section 7 contains a discussion of avenues for future work.

2. Rate equation model. Our theoretical analysis is based on a set of Lang–Kobayashi-type rate equations for the normalized complex slowly varying envelope of the optical fields $E_{1,2} = \mathrm{Re}[E_{1,2}] + i\,\mathrm{Im}[E_{1,2}]$ and the normalized inversions $N_{1,2}$. The Lang–Kobayashi equations [22] are an established model to describe a single mode SL that receives conventional optical feedback. In this situation a part of the emitted light is reflected by a mirror at some fixed distance and is then fed back into the laser. The Lang–Kobayashi equations can be extended for the case of two delay-coupled single mode SLs as considered here; for a detailed derivation see [23].

We write the equations in the reference frame of rotation with the averaged optical angular frequency of the two lasers. This means that the optical fields of the lasers are represented by $E_{1,2}(t)e^{i\Omega_0 t}$, where $\Omega_0 = \frac{1}{2}(\Omega_1 + \Omega_2)$ is the average of the lasers' optical angular frequencies $\Omega_{1,2}$, respectively. In nondimensional form the equations can be written as

$$\frac{dE_1(t)}{dt} = (1+i\alpha)N_1(t)E_1(t) + \kappa e^{-iC_p}E_2(t-\tau) - i\Delta E_1(t), \tag{2.1}$$

$$\frac{dE_2(t)}{dt} = (1+i\alpha)N_2(t)E_2(t) + \kappa e^{-iC_p}E_1(t-\tau) + i\Delta E_2(t), \tag{2.2}$$

$$T\frac{dN_1(t)}{dt} = P - N_1(t) - (1+2N_1(t))|E_1(t)|^2, \tag{2.3}$$

$$T\frac{dN_2(t)}{dt} = P - N_2(t) - (1+2N_2(t))|E_2(t)|^2. \tag{2.4}$$

In (2.1)–(2.4) time t is measured in units of the photon lifetime, which for an SL has a typical value on the order of 10 picoseconds. The parameter α is the linewidth enhancement factor. It is typical for SLs and describes the coupling between the phase and the amplitude of the optical fields E. The parameter T is the normalized carrier lifetime. These are material properties of the laser. The parameter P describes the amount of electrical current used to pump the semiconductor active material. For all parameters we adopt the physically meaningful values given in Table 1.

Table 1
Laser parameters and their values.

Symbol	Laser parameter	Value
α	linewidth enhancement factor	2.5
T	electron decay rate	392.0
P	pump parameter	0.23
τ	coupling time	20.0
κ	coupling rate	0.1

Furthermore, the mutual coupling is accounted for by the second term on the right-hand side of (2.1) and (2.2), where τ is the delay time, κ the coupling rate, C_p the coupling phase, and Δ the detuning.

The delay time τ is an intrinsic feature of this coupling, due to the finite propagation time of the light between the spatially separated lasers. We consider here the case of a relatively short distance between the lasers, in the order of centimeters. This is still long compared to the length of the SL cavities of typically a few hundred micrometers. Finally, the coupling rate κ is the fraction of photons coupled into the other laser per unit time.

Our main parameters are the *coupling phase* C_p and the *detuning* Δ. In the reference frame of (2.1)–(2.4) C_p depends on the average optical frequency, namely, $C_p = \Omega_0\tau$. The parameter Δ measures the difference between the optical frequencies of the two uncoupled lasers with respect to the average frequency, that is, $\Delta = \frac{1}{2}(\Omega_2 - \Omega_1)$. We consider C_p and Δ as independent parameters; this is convenient for the analysis and quite common in the field. The coupling phase C_p can be changed accurately in an experiment by changing the distance between the two lasers on the scale of the optical wavelength, or by exploiting the temperature or pump current dependency of Ω_0 [1]; these changes are so small that the other laser parameters remain unchanged within the experimental accuracy. The detuning Δ can be changed by increasing Ω_2 and decreasing Ω_1 by the same magnitude, so that the average frequency and, hence, $C_p = \Omega_0\tau$ remain constant. However, it may be more convenient in an experiment to change the optical frequency of only one of the two lasers. For this case (2.1)–(2.4) can be rewritten in the frame where one laser has fixed optical frequency [42]. However, then the symmetry of exchanging laser 1 with laser 2 has a more complicated expression.

Equations (2.1)–(2.4) are a system of DDEs with a single fixed delay. As such they have an infinite-dimensional phase space, namely, the space $C([-\tau, 0], \mathbb{R}^6)$ of continuous functions over the delay interval $[-\tau, 0]$ with values in (E_1, E_2, N_1, N_2)-space. Thus, in contrast to ordinary differential equations, a single initial condition $x_0 \in \mathbb{R}^6$ is not enough to determine the future evolution of the system. Indeed it is required to prescribe initial data on the entire

interval $[-\tau, 0]$. We refer to [2, 12] as general references on delay equations; see also [10] for a background section on DDEs in the context of an SL with delayed feedback.

Crucial for what follows are a number of symmetries of (2.1)–(2.4). First, there is the continuous S^1-symmetry

$$(E_1, E_2, N_1, N_2) \to (E_1 e^{ib}, E_2 e^{ib}, N_1, N_2). \tag{2.5}$$

This phase-space symmetry is a typical feature of Lang–Kobayashi-type equations, provided that no phase conjugation is involved [21]. Any solution of (2.1)–(2.4) is invariant under any phase shift of both electric fields E_1 and E_2. The S^1-symmetry motivates the ansatz (3.1)–(3.4) of the CLMs of section 3 with a common frequency for both lasers.

Second, there is the reflection symmetry

$$(E_1, E_2, N_1, N_2, \Delta) \to (E_2, E_1, N_2, N_1, -\Delta) \tag{2.6}$$

of interchanging the two lasers, which results in a sign change of Δ. For zero detuning, that is, for $\Delta = 0.0$, this symmetry is a $\mathbb{Z}_2$-symmetry in phase space. When the detuning Δ is then "switched on" this phase-space symmetry is broken, which has important consequences for the organization of the CLMs; see section 6.1.

Third, there is the 2π-translational symmetry

$$(E_1, E_2, N_1, N_2, C_p) \to (E_1, E_2, N_1, N_2, C_p + 2\pi) \tag{2.7}$$

in the feedback phase C_p. As a consequence, the parameter C_p is a circle. We refer to this symmetry as the 2π-translational symmetry. It is often useful to show bifurcation diagrams in the covering space $\mathbb{R}$ of the circle, that is, over several fundamental domains (of length 2π) of the symmetry (2.7).

Fourth, there is the symmetry

$$(E_1, E_2, N_1, N_2, C_p) \to (E_1, -E_2, N_1, N_2, C_p + \pi), \tag{2.8}$$

which is a π-translational symmetry in the feedback phase C_p, combined with a sign change in the optical field of one laser, say, E_2. Due to the S^1-symmetry (2.5) one could alternatively change the sign of E_1. We refer to this symmetry as the π-translational symmetry. As we will see in section 5, it provides a link between different types of CLMs.

3. Compound laser modes. The basic solutions of (2.1)–(2.4) are called the *compound laser modes* (CLMs); they are of the form

$$E_1(t) = R_1^s e^{i\omega^s t}, \tag{3.1}$$

$$E_2(t) = R_2^s e^{i\omega^s t + i\sigma}, \tag{3.2}$$

$$N_1(t) = N_1^s, \tag{3.3}$$

$$N_2(t) = N_2^s, \tag{3.4}$$

where R_i^s, N_i^s, ω^s, and σ are time-independent and real valued. Additionally, R_i^s are taken to be positive without loss of generality. We allow different amplitudes R_i^s and different

steady state inversions N_i^s. However, the lasers must have the same frequency ω^s, which is implied by the S^1-symmetry (2.5). Here ω^s is the deviation between the average solitary laser frequency Ω_0 and the frequency of the coupled laser system. There may also be some time-independent phase shift σ between the lasers. Mathematically, CLMs are periodic orbits, with frequencies that depend on other parameters, where the rotation is in the direction of the symmetry group only. This property of the CLMs must be taken into account when one wants to continue them numerically; see section 4. Physically, CLMs are frequency locked states, in which the lasers operate with constant, but possibly different, intensities.

Note that we consider here the situation that the pump current is sufficiently large, so that the overall system is in the "on-state," that is, the optical fields have nonzero amplitude. Mathematically, this means that the "off-state" given by $(E_1, E_2, N_1, N_2) = (0, 0, P, P)$ is unstable.

Inserting the ansatz (3.1)–(3.4) into (2.1)–(2.4) gives the set of six coupled nonlinear transcendental equations for the six unknowns:

$$0 = R_1^s N_1^s + \kappa R_2^s \cos(-C_p - \omega^s \tau + \sigma), \tag{3.5}$$

$$(\omega^s + \Delta) = \alpha N_1^s + \kappa \frac{R_2^s}{R_1^s} \sin(-C_p - \omega^s \tau + \sigma), \tag{3.6}$$

$$0 = R_2^s N_2^s + \kappa R_1^s \cos(-C_p - \omega^s \tau - \sigma), \tag{3.7}$$

$$(\omega^s - \Delta) = \alpha N_2^s + \kappa \frac{R_1^s}{R_2^s} \sin(-C_p - \omega^s \tau - \sigma), \tag{3.8}$$

$$0 = P - N_1^s - (1 + 2N_1^s)|R_1^s|^2, \tag{3.9}$$

$$0 = P - N_2^s - (1 + 2N_2^s)|R_2^s|^2. \tag{3.10}$$

There is no obvious analytical strategy for solving for the unknowns in some closed form that allows one to create an overall picture of how the CLMs depend on parameters, for example, on C_p and Δ. In fact, the situation is a lot more complicated than for the case of the Lang–Kobayashi equations of an SL with feedback, for which a partial analytical picture is now emerging [29]. It is of course possible to find individual solutions of (3.5)–(3.10) numerically, for example, by root solving with Newton's method. Such roots can then be followed in relevant parameters with standard continuation software, such as AUTO [3].

The approach we take here is in this spirit, but we find and follow CLMs in the full DDE (2.1)–(2.4) by using the package DDE-BIFTOOL [4]. This has the advantage that we obtain stability information along branches of CLMs; see section 4 for details of the numerical procedure.

Furthermore, we gain insight into special cases of CLMs by a study of some special cases of (3.5)–(3.10). Of special interest is the relationship between the frequency ω^s and the phase difference σ. We eliminate the unknown variables R_1, R_2, N_1, N_2 from (3.5)–(3.8), which results in the transcendental equation

$$\begin{aligned}(\omega^s)^2 = \kappa^2(1+\alpha^2)[&\sin(C_p + \omega^s\tau + \sigma + \arctan(\alpha)) \\ &\times \sin(C_p + \omega^s\tau - \sigma + \arctan(\alpha))] - \Delta^2.\end{aligned} \tag{3.11}$$

As we will see in section 5, this equation allows us to identify certain CLMs as solutions of an associated Lang–Kobayashi equation for a laser with optical feedback.

4. Continuation of CLMs. The package DDE-BIFTOOL provides MATLAB routines for the numerical continuation analysis of DDEs [4]. From a starting solution, such as a steady state or a periodic orbit, DDE-BIFTOOL is able to follow a solution branch in one parameter. Furthermore, stability information in the form of eigenvalues or Floquet multipliers can be computed along the branch. In this way, local codimension-one bifurcations can be detected; some of them can be followed in two parameters.

In order to use DDE-BIFTOOL to follow branches of CLMs in (2.1)–(2.4), one needs to realize that CLMs are periodic orbits. However, CLMs are special types of periodic orbits: $R^s_{1,2}$ and $N^s_{1,2}$ are constant, so that the periodic motion is purely in the direction of the continuous S^1-symmetry, with constant speed ω^s and a time-independent phase difference σ; one also speaks of CLMs as group orbits of the S^1-symmetry. The situation is conceptually the same as that for the external cavity modes (ECMs) of the Lang–Kobayashi equations [21, 11].

We can exploit the special nature of the CLMs to continue them as equilibria with DDE-BIFTOOL of an appropriately amended equation; see also [11, 19]. The idea is to move into the frame that is rotating with the speed given by the unknown frequency ω^s. (Recall that ω^s is different for different CLMs, so that the S^1-symmetry cannot be divided out globally [21].) In order to use DDE-BIFTOOL, this can be done by introducing the new parameter b, replacing $E_{1,2}(t)$ in (2.1) and (2.2) with

$$E_{1,2}(t)e^{ibt}. \tag{4.1}$$

This gives the new equations

$$\frac{dE_1}{dt} = (1+i\alpha)N_1E_1 + \kappa e^{-iC_p}E_2(t-\tau) - i(b+\Delta)E_1, \tag{4.2}$$

$$\frac{dE_2}{dt} = (1+i\alpha)N_2E_2 + \kappa e^{-iC_p}E_1(t-\tau) - i(b-\Delta)E_2, \tag{4.3}$$

while (2.3) and (2.4) remain unchanged. During the continuation b is an additional free parameter that is then fixed to $b = \omega^s$, so that the respective CLM is now an equilibrium of (4.2), (4.3), (2.3), and (2.4). Because of the S^1-symmetry, the CLM gives rise to a whole family of nonisolated equilibria in this setup. To obtain an isolated solution one needs to fix the phase, which can be done, for example, by requiring that $\mathrm{Im}(E_1) = 0$.

Along a branch of CLMs stability information is computed using DDE-BIFTOOL in the usual way but with one exception: there is always one extra zero-eigenvalue due to direction of group action, which is neutral. Thus, we can detect codimension-one bifurcations as for any equilibrium, namely, saddle-node bifurcations, pitchfork bifurcations, and Hopf bifurcations. For the interpretation of the results it is important to keep in mind that we are actually dealing with bifurcations of CLMs, which are group orbits of the S^1-symmetry of the respective equilibria. While saddle-node and pitchfork bifurcations simply lead to the creation of different branches of CLMs, Hopf bifurcations actually lead to bifurcating tori in phase space.

In what follows we use DDE-BIFTOOL to derive a comprehensive picture of the CLMs, including stability information, as a function of the feedback phase C_p and the detuning Δ. To this end, we first consider the case of zero detuning in section 5 and then the influence of nonzero detuning in section 6.

5. CLMs for zero detuning. For zero detuning, $\Delta = 0$, the two lasers would operate with the same solitary optical frequency. While the detuning is experimentally easily accessible and can be controlled with good accuracy [13], one may argue that $\Delta = 0$ is not attainable exactly in an experiment. However, it turns out that this special case organizes the dynamics even for small nonzero detuning.

For $\Delta = 0$ the symmetry (2.6) of exchanging the two lasers is a reflectional symmetry in phase space. In particular, (3.11) reduces to

$$(\omega^s)^2 = \kappa^2(1+\alpha^2)[\sin(C_p + \omega^s\tau + \sigma + \arctan(\alpha)) \times \sin(C_p + \omega^s\tau - \sigma + \arctan(\alpha))]. \tag{5.1}$$

For the special choices $\sigma = 0$ and $\sigma = \pi$ we obtain

$$\omega^s = \mp\kappa\sqrt{1+\alpha^2}\sin(C_p + \omega^s\tau \arctan(\alpha)) \tag{5.2}$$

with the respective choice of $\mp$.

We call solutions of (5.2) the *constant-phase CLMs*, where we distinguish further between *in-phase CLMs* with $\sigma = 0$ and *antiphase CLMs* with $\sigma = \pi$. Equation (5.2) is in fact the determining equation for the ECMs of the Lang–Kobayashi equations describing a laser with conventional optical feedback from a mirror at half the distance between the two lasers, except for allowing for both signs $\mp$. As an immediate consequence, the constant-phase CLMs lie on an ellipse in the (ω^s, N^s)-plane. In particular, for constant-phase CLMs one has $R_1^s = R_2^s$ and $N_1^s = N_2^s$, which means that both lasers operate with the same intensity. In-phase CLMs have zero phase difference, $\sigma = 0$, and are described by (5.2) with a minus sign. Physically, this is the case of constructive interference between the optical fields of the two lasers. On the other hand, for antiphase CLMs with a phase difference of $\sigma = \pi$, there is destructive interference between the optical fields of the two lasers, as is expressed by the plus sign in (5.2). Mathematically, the in-phase CLMs are related to the antiphase CLMs by the π-translational symmetry (2.8).

The constant-phase CLMs are the most obvious solutions of (3.5)–(3.10) for zero detuning. However, a bifurcation analysis with numerical continuation shows that even for zero detuning there are additional CLMs where σ is not constant but some function of C_p. We call these solutions *intermediate-phase* CLMs. For this type of CLM one finds that $R_1^s \neq R_2^s$ and $N_1^s \neq N_2^s$, which means that the two lasers operate with different intensities.

Figure 2 shows curves of CLMs in the (ω^s, N^s)-plane. The projection is the representation of choice in the physics literature (also for the Lang–Kobayashi equations) because in effect, it "hides" the 2π-translational symmetry of the parameter C_p. All CLMs lie on closed curves. Furthermore, the frequency ω^s is a quantity that can be measured in an experiment and N^s is directly related to the laser intensity. Each curve is parameterized by the feedback phase C_p, meaning that it represents all CLMs of a given type that exist for any choice of C_p. The symbols in Figure 2 indicate points where saddle-node, pitchfork, and Hopf bifurcations take place along the different branches as C_p is changed. The constant-phase CLMs form the purple ellipse. The bifurcating intermediate-phase CLMs lie on the green closed curve which also has the shape of an ellipse for our choice of parameters. Additionally, there is a separate branch of intermediate-phase CLMs, namely, the two orange islands. For a given C_p, there is

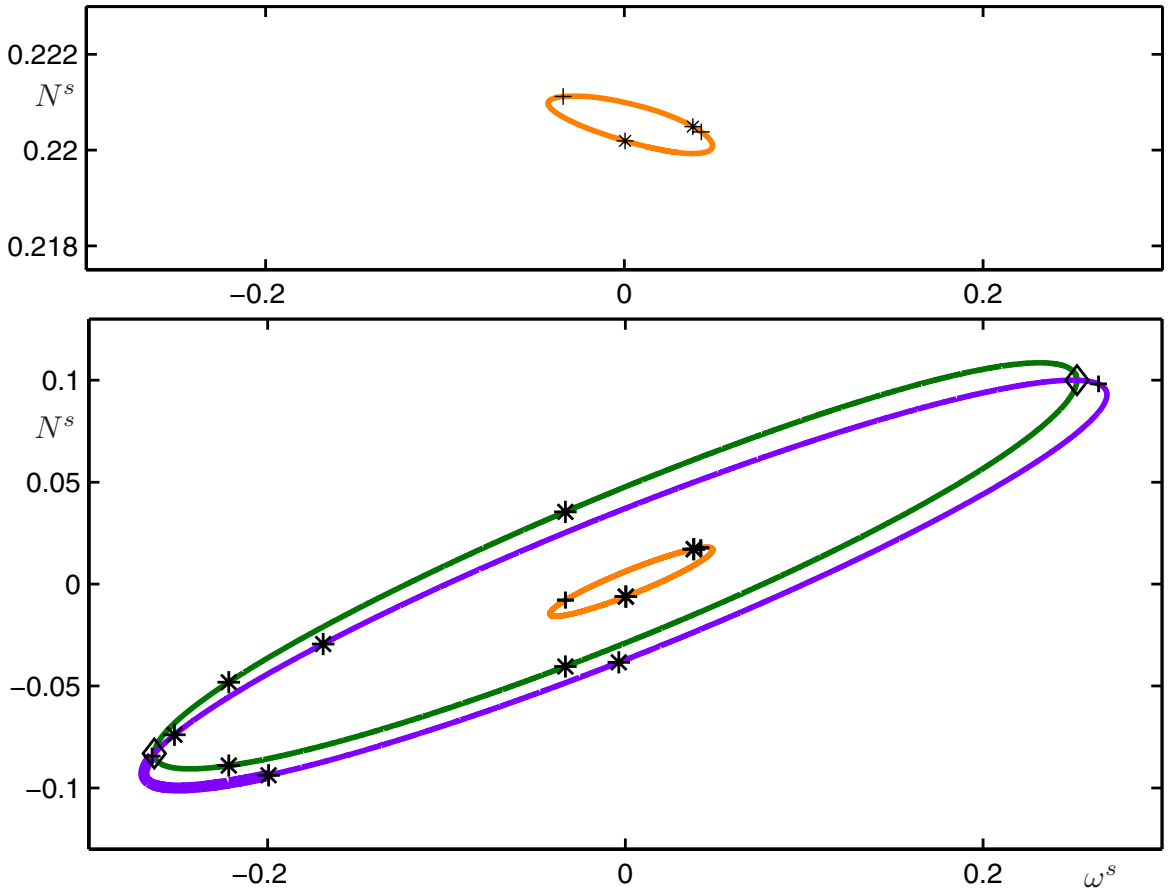


Figure 2. *Curves of CLMs for zero detuning in the (ω^s, N^s)-plane. For the constant-phase CLMs (purple curve) the inversions of both lasers are identical, while for the bifurcating intermediate-phase CLMs (green curve) they are different. There is another disconnected branch of CLMs (orange curves), where the inversion of the two lasers are on two different curves, one for larger values of N^s. (Note the different scale of the vertical axis.) Throughout the paper saddle-node bifurcations are marked by pluses (+), Hopf bifurcations by stars (∗), and pitchfork bifurcations by diamonds (⋄); stable regions are plotted as bold curves. Note that saddle-node bifurcations do not coincide with folds with respect to ω^s.*

a fixed number of CLMs, which lie on the respective curves. This is illustrated in Figure 3 for $C_p = 0$, where circles (∘) mark the inversion of laser 1 and crosses (×) the inversion of laser 2. There is a second set of solutions due to the reflection symmetry (2.6), which can be obtained by interchanging crosses (×) with circles (∘).

We now describe how the different types of CLMs move over the respective curves as C_p is decreased; compare Figures 2 and 3 and the accompanying animation (61995_01.gif). Constant-phase CLMs move over the purple ellipse. In-phase CLMs and antiphase CLMs are born in pairs in the saddle-node bifurcation in the low-inversion region. The saddle-node bifurcations are close to the folds with respect to ω^s, but they do not take place exactly at the folds because ω^s is not a bifurcation parameter. When C_p is decreased, the saddle is moving on the upper half and the node is moving on the lower half of the purple ellipse toward

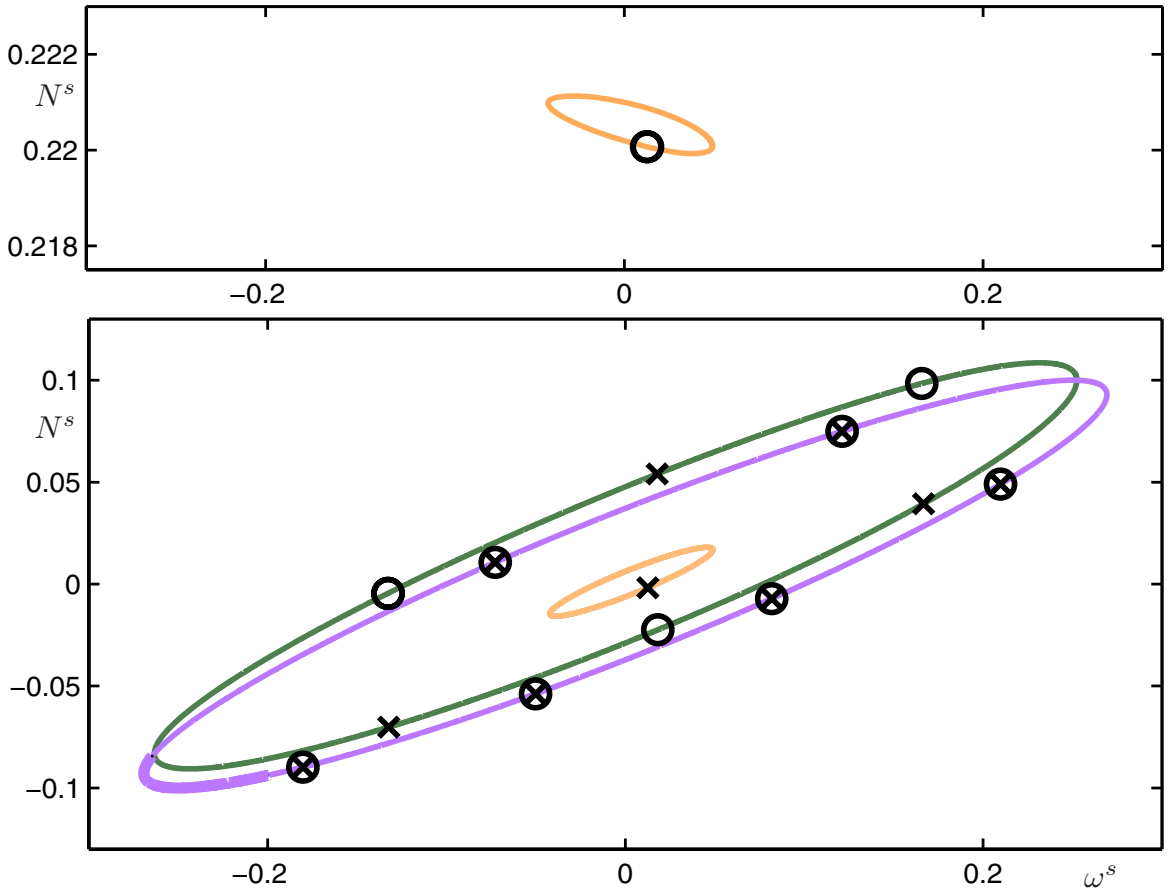


Figure 3. *The curves of CLMs for zero detuning in the (ω^s, N^s)-plane as in Figure 2, but with CLMs for $C_p = 0$. Crosses (×) mark the inversion of laser 1 and circles (○) that of laser 2. A second set of solutions, with crosses and circles interchanged, exists as a result of the second reflection symmetry (2.6). Clicking on the above image displays the accompanying animation (61995_01.gif [1.7MB]).*

the high-inversion region. Eventually they coalesce and disappear in the second saddle-node bifurcation in the high-inversion region. Along the way, the constant-phase CLMs change their stability several times, mostly in Hopf bifurcations. The constant-phase CLMs in the boldfaced region of the purple ellipse are stable. In this region the two lasers show stable emission with the same intensity and, depending on the exact range of C_p, a phase difference of either zero or π. This stable region is bounded by a saddle-node bifurcation on the left and by a Hopf bifurcation on the right.

The pitchfork bifurcation is responsible for the creation of a pair of intermediate-phase CLMs, which lie on the green ellipse in Figures 2 and 3. For intermediate-phase CLMs $N_1^s \neq N_2^s$. In fact, the inversion of, say, laser 1 can be found on the lower half of the green ellipse, whereas the inversion of laser 2 is on the upper half; due to symmetry there is a second solution with laser 1 and laser 2 exchanged. For decreasing C_p, the intermediate-phase CLMs travel along their ellipse from the low-inversion region toward the high-inversion region, where they coalesce and disappear in the second pitchfork bifurcation. We can distinguish between

intermediate-phase CLMs that are born in a pitchfork bifurcation of an in-phase CLM and intermediate CLMs that are born in a pitchfork bifurcation of an antiphase CLM, which we call *increasing-phase CLMs* and *decreasing-phase CLMs*, respectively.

Finally, there is the set of intermediate-phase CLMs on the separate orange islands. Being intermediate-phase CLMs, they also have nonidentical inversions, $N_1^s \neq N_2^s$, and nonidentical amplitudes, $R_1^s \neq R_2^s$. The inversion of, say, laser 1 takes values around its solitary value, corresponding to the origin in Figures 2 and 3, whereas the inversion of laser 2 has significantly higher values of the inversion, as can be seen in the upper panels of these figures. Again due to symmetry there is a second solution with laser 1 and laser 2 exchanged. When decreasing C_p, these intermediate-phase CLMs are born in pairs in a saddle-node bifurcation on the left side of the curve, that is, for negative ω^s. Then they move along the orange, ellipse-like curves, one on the upper ellipse and the other one on the lower ellipse, and eventually coalesce and disappear in the second saddle-node bifurcation.

A disadvantage of the projection from the bifurcation theory point of view is that neither N^s nor ω^s is a bifurcation parameter, so that Figure 2 is not a bifurcation diagram. In particular, we already mentioned that the saddle-node bifurcations do not coincide with the folds with respect to ω^s. Therefore, we now study the CLMs as a function of C_p, which is a main bifurcation parameter we consider here. To this end, we show in Figures 4 and 5 different curves of CLMs as a function of C_p. This representation makes all the symmetries explicit and allows us to discuss in detail how the different types of CLMs depend on the bifurcation parameter C_p and how they interact. The number and location of CLMs for a given C_p can simply be read off by considering all intersections of curves of CLMs with a vertical line corresponding to the value of C_p; for example, the CLMs in Figure 3 correspond to the intersection with the line $\{C_p = 0\}$; in this projection the saddle-node bifurcations are the folds with respect to C_p. It is convenient to show the respective curves of CLMs over several multiples of 2π, meaning that we consider several fundamental domains of the 2π-translational symmetry (2.7).

In Figure 4 we show the constant-phase and bifurcating intermediate-phase CLMs in the (C_p, N^s)-plane. Panel (a) shows the constant-phase CLMs, namely, the in-phase CLMs in pink and the antiphase CLMs in purple. Both form a single, self-intersecting curve, and the image is indeed 2π-translationally invariant. A translation by π transforms the pink into the purple curve and vice versa, which represents the relation between the in-phase and the antiphase CLMs as given by (2.8). Panel (b) of Figure 4 shows the intermediate-phase CLMs that bifurcate in the pitchfork bifurcations from the constant-phase CLMs, where the increasing-phase CLMs are shown in light green and the decreasing-phase CLMs in dark green. There are infinitely many closed curves that appear to have the shape of an ellipse. The image is again invariant under a translation by 2π, while a translation by π transforms the light green into the dark green curves, that is, increasing-phase into decreasing-phase CLMs.

Figure 5 illustrates how the constant-phase and the bifurcating intermediate-phase CLMs interact when seen as a function of C_p. Panel (a) simply shows both sets of curves plotted together in the (C_p, N^s)-plane. The intersections of curves marked by a diamond ($\diamond$) are the pitchfork bifurcations; all other intersections are due to projection. This image clearly shows how the infinitely many ellipse-like curves of intermediate-phase CLMs provide the connection between the two single curves of constant-phase CLMs. This point is further brought out in

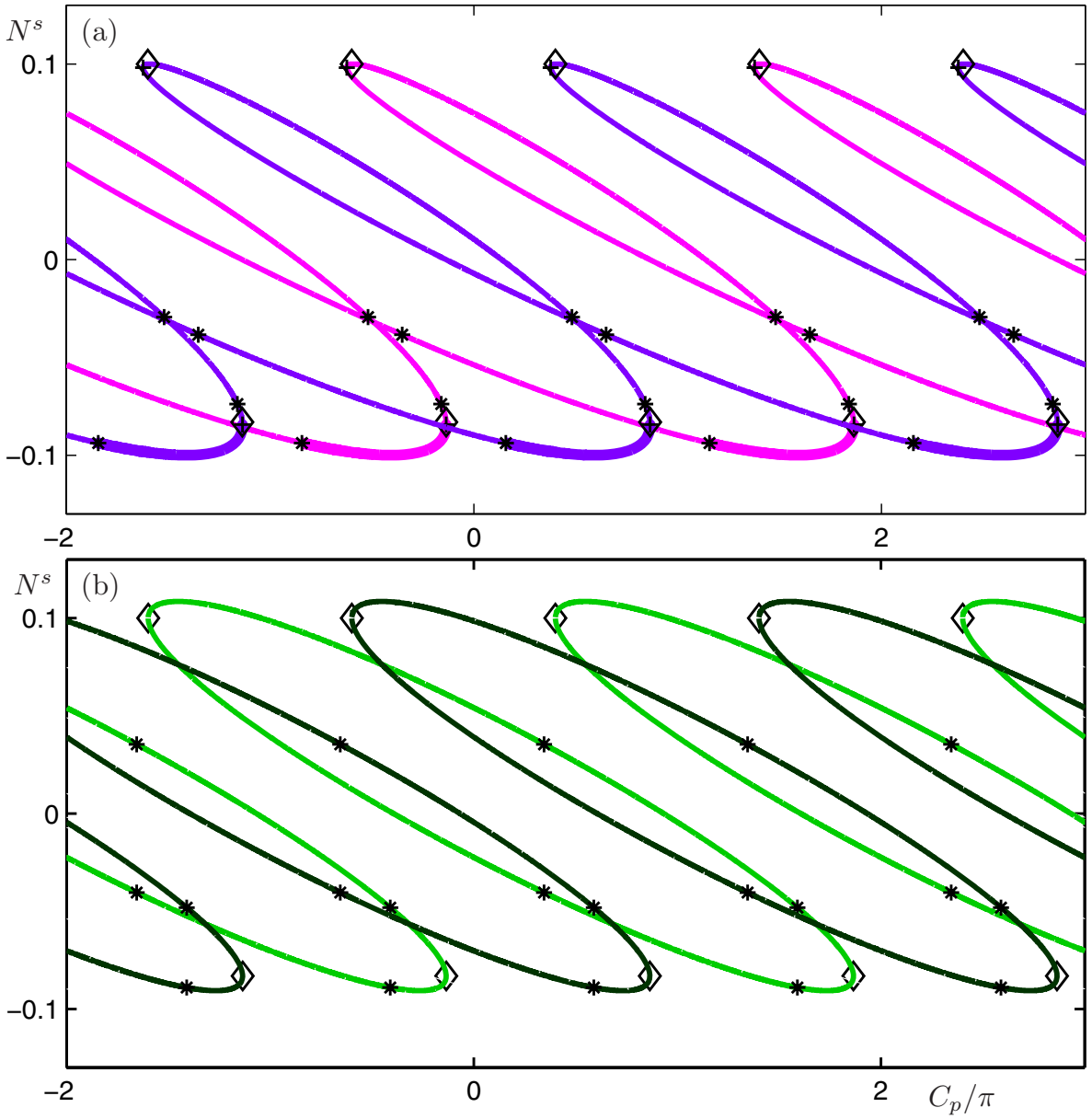


Figure 4. *CLMs for zero detuning from Figure 2 plotted in the (C_p, N^s)-plane. The constant-phase CLMs form two infinitely long, self-intersecting curves* (a). *The in-phase CLMs (pink curve) and the antiphase CLMs (purple curve) are each other's image under the π-translational symmetry* (2.8). *The intermediate-phase CLMs, on the other hand, form infinitely many closed curves* (b). *The increasing-phase CLMs (light green curves) and the decreasing-phase CLMs (dark green curves) are each other's image under the π-translational symmetry* (2.8).

panel (b) of Figure 5, where we show the same CLMs in the (C_p, σ)-plane; notice the additional 2π-symmetry of the figure in the phase difference σ. In this representation the constant-phase CLMs trace out a straight line at $\sigma = 0$ for the in-phase CLMs and at $\sigma = \pm\pi$ for the

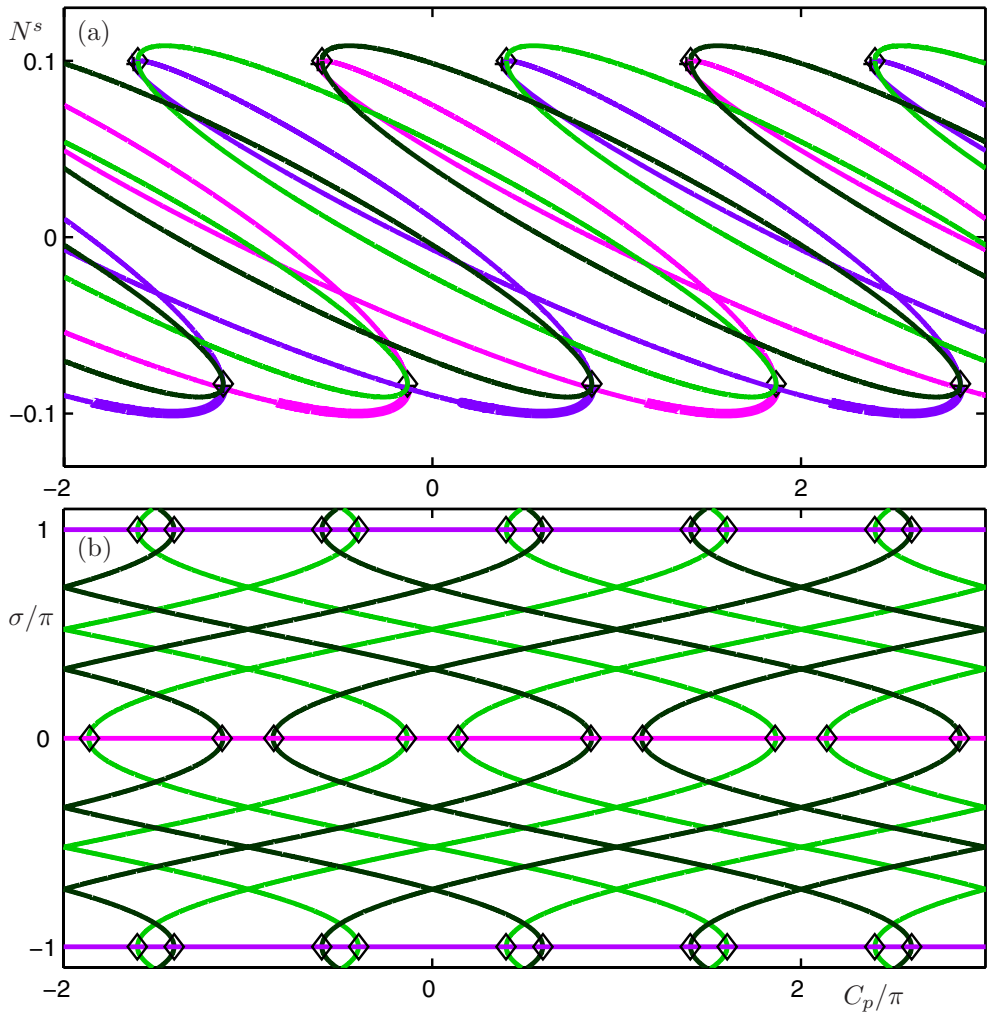


Figure 5. *The constant-phase CLMs and the bifurcating intermediate-phase CLMs from Figure 3 plotted together in the* (C_p, N^s)*-plane* (a) *and in the* (C_p, σ)*-plane* (b)*. The intersections marked by diamonds* ($\diamond$) *are the pitchfork bifurcations.*

antiphase CLMs. The different ellipse-like curves of intermediate-phase CLMs in Figure 4, on the other hand, lead to an intriguing array of additional curves in the (C_p, σ)-plane. Let us concentrate on the rightmost dark green curve with a pitchfork bifurcation of an in-phase CLM at $(C_p, \sigma) \approx (2.8\pi, 0)$. For decreasing C_p, two branches emerge from this point; one is going upward toward higher σ, and the other is going downward toward lower σ. While C_p decreases, the two intermediate-phase CLMs gain an additional phase shift of $\pm\pi$. Eventually they both intersect with the branch of antiphase CLMs at around $(C_p, \sigma) = (-0.5\pi, \pm\pi)$. This is the point of the second pitchfork bifurcation, where this particular pair of intermediate-phase CLMs disappears. Due to the 2π-symmetry in σ the antiphase CLM at $\sigma = -\pi$ is identical to the one at $\sigma = \pi$. All other green branches connect in-phase and antiphase CLMs in a similar way. The relationship between light green and dark green branches is again given

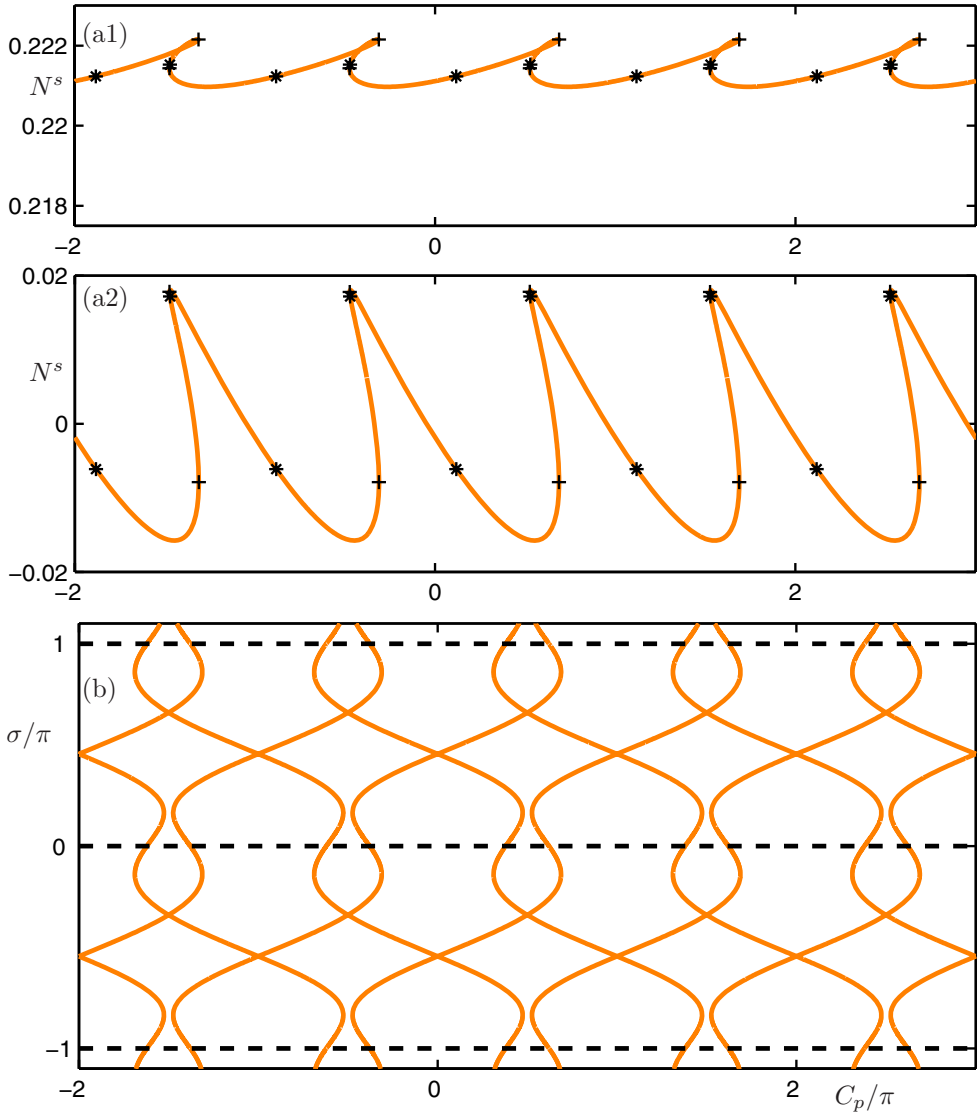


Figure 6. *The other intermediate-phase CLMs from Figure 2 plotted in the* (C_p, N^s)*-plane for one laser* (a1) *and the other laser* (a2)*, and in the* (C_p, σ)*-plane* (b).

by the symmetry (2.8), which in the (C_p, σ)-plane is given by $(C_p, \sigma) \to (C_p + \pi, \sigma + \pi)$.

We end this section by considering in Figure 6 the other intermediate-phase CLMs corresponding to the orange curve in Figure 2. Panels (a1), (a2), and (b) show these CLMs in the (C_p, N^s)-plane. Since the inversions of the two lasers are very different we plot them in different panels; note also the difference in the vertical scales. These intermediate-phase CLMs trace out two curves—one for N^s around zero and one for N^s around 0.222. The image in Figure 6 is not only invariant under translation by 2π but also under translation by π. This means that for this type of intermediate-phase CLM one cannot distinguish two different types that are each other's counterparts under exchanging laser 1 and laser 2. Panel (b) of

Figure 6 shows the phase difference σ of these intermediate-phase CLMs, which are born in pairs in saddle-node bifurcations, corresponding to folds of the branches. The different orange curves are each other's image under the symmetries (2.8) and (2.7), which in the (C_p, σ)-plane are given by $(C_p, \sigma) \to (C_p + \pi, \sigma + \pi)$ and $(C_p, \sigma) \to (C_p + 2\pi, \sigma + 2\pi)$, respectively. Note that, in contrast to the other type of intermediate-phase CLM in Figure 5, these intermediate-phase CLMs do not remain in a finite interval of C_p values. Instead, as a function of C_p they continuously gain (or lose) in phase difference σ.

6. CLMs for nonzero detuning. In this section we study in detail how the structure of the CLMs discussed in the last section changes with the detuning. Due to the symmetry (2.6), we may restrict our attention to $\Delta > 0$. We first consider in section 6.1 the case of relatively small Δ, which can be seen as being "organized" by the CLMs for zero detuning. Indeed $\Delta \neq 0$ breaks the phase-space symmetry of exchanging the two lasers. As a result, the pitchfork bifurcations for $\Delta = 0$ unfold to saddle-node bifurcations leading to an interesting global organization of the different branches of CLMs. Note that, locally, small detuning has only little effect and to some extent one still can speak of identical lasers. In section 6.2 we show that for intermediate Δ there are further interactions between different branches of CLMs due to transitions through saddles and extrema of the surface of CLMs in the three-dimensional (C_p, Δ, N^s)-space. Finally, we consider the limit of very large Δ in section 6.3.

6.1. Perturbation from zero detuning. As an entry point to the analysis of the CLMs for nonzero detuning we present in Figure 7 again the "representation" of the CLMs in the (ω^s, N^s)-plane. The colors now refer to the two different lasers: red corresponds to laser 1, which is "detuned to the red" with respect to the average optical frequency, and blue corresponds to laser 2, which is "detuned to the blue" with respect to the average optical frequency. Recall that (2.1)–(2.4) are written in the reference frame of average optical frequency $\frac{1}{2}(\Omega_1 + \Omega_2)$; the system can be detuned without changing the coupling phase C_p.

For nonzero detuning the inversions of the lasers are no longer identical; that is, $N_1^s \neq N_2^s$. Thus, different curves can be seen for laser 1 and laser 2, respectively. We first concentrate on the two large closed curves in the shape of horseshoes. Their structure can be understood by the unfolding of the pitchfork bifurcations for zero detuning. Depending on the sign of the unfolding parameter, there are two generic possibilities of locally unfolding a pitchfork bifurcation, each consisting of a saddle-node bifurcation and a separated branch [34]. In the coupled laser system both possibilities exist, one in the low-inversion region and one in the high-inversion region. Globally, for the chosen set of parameters, the unfolding of the pitchfork bifurcations leads to the formation of the red and the blue horseshoes.

Furthermore, the small ellipses originate from the two separate ellipses of intermediate-phase CLMs for zero detuning. Each of them is now split up into a red and a blue ellipse. A single CLM for fixed C_p now corresponds to one point on a red curve and another point on the blue curve. This is illustrated in Figure 8 for $C_p = 0$, where circles ($\circ$) mark the inversion of the red laser 1 and crosses ($\times$) that of the blue laser 2.

When C_p is decreased, CLMs move over the different pairs of red and blue branches; compare Figures 7 and 8 and the accompanying animation (61995_02.gif). CLMs are formed in saddle-node bifurcations in the low-inversion region, travel along the horseshoes, and then disappear in the high-inversion region in another saddle-node bifurcation. Similarly, for the

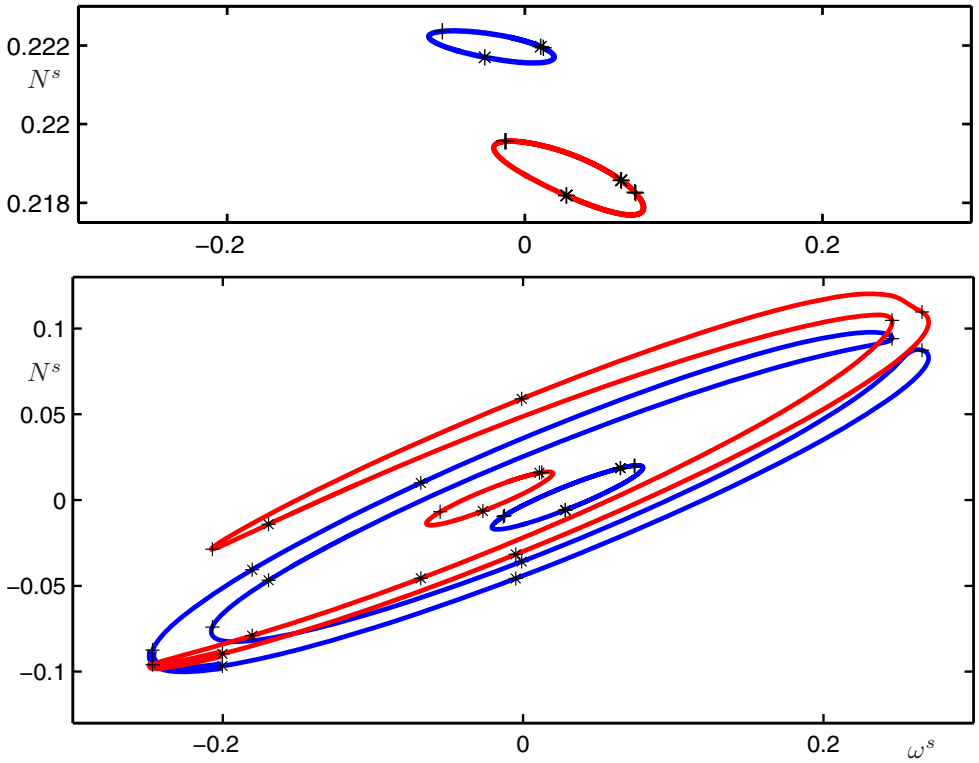


Figure 7. *CLMs for $\Delta = 0.025$ in the (ω^s, N^s)-plane, where we distinguish between the red detuned laser 1 (red curves) and the blue detuned laser 2 (blue curves).*

separate ellipses the associated CLMs are formed in saddle-node bifurcations on the low frequency side of the respective ellipse and disappear in saddle-node bifurcations on the high frequency side. Again, a CLM for fixed C_p corresponds to one point on a red curve and one point on the corresponding blue curve.

The connection with the case of zero detuning is obviously given by decreasing Δ back to zero. Then the two horseshoes move closer together and in the limit form the two ellipses of the constant-phase and bifurcating intermediate-phase CLMs. Similarly, the ellipses of separate intermediate-phase CLMs move together and then form only two ellipses; compare Figures 7 and 2 and see the accompanying animation (61995_03.gif).

We now consider the CLMs for (small) nonzero detuning in dependence on the parameter C_p, that is, in the (C_p, N^s)-plane. Figure 9 shows the curves that correspond to the red and blue horseshoes for three different values of increasing Δ; panel (a) is actually for the data in Figure 7. Both the curve for the red laser and that for the blue laser are closed curves with a single self-intersection. Note that we plot only one such curve each; there are, in fact, infinitely many copies due to the translational symmetries (2.8) and (2.7). As the detuning is increased, the curves of the red and the blue lasers move further apart, but the structure of the branches remains topologically the same in Figure 9.

Figure 10 shows the curves of CLMs of Figure 9 in the (C_p, N^s)-plane (left column) and the (C_p, σ)-plane (right column), but now with the same color coding of these different

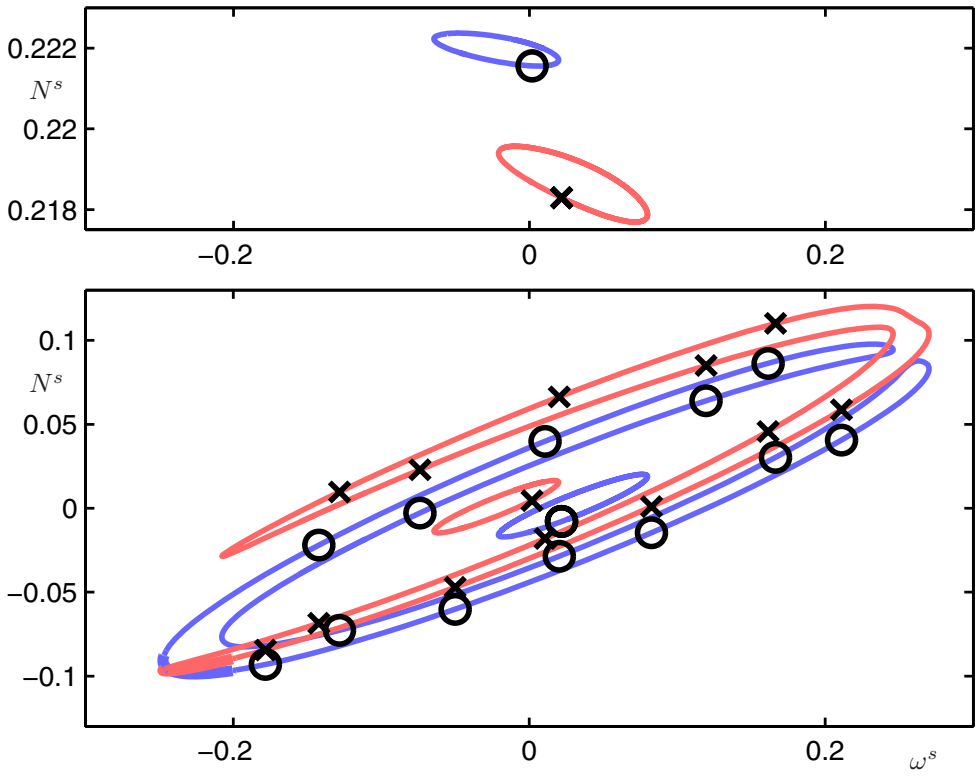


Figure 8. *CLMs for* $\Delta = 0.025$ *in the* (ω^s, N^s)*-plane for fixed value of* $C_p = 0$*. Crosses* (×) *mark the inversion of laser* 1 *and circles* (○) *that of laser* 2*. Clicking on the above image displays the accompanying animation* (61995_02.gif [1.1MB]).

branches as in Figure 6. Note that for nonzero detuning there are no constant-phase CLMs. Nevertheless, we see that there are two substantial sections of the branch where the CLM has an almost constant phase of $\sigma \approx 0$ and $\sigma \approx \pi$. These parts are connected by sections where the phase increases and decreases, respectively. These different sections can be identified as the reminders of corresponding constant-phase and intermediate-phase CLMs for zero detuning, which is clearly brought out by the coloring. Considering the right column and decreasing C_p, a pair of CLMs is born in the saddle-node bifurcation at $C_p \approx 2.7\pi$. The purple branch has an almost constant phase $\sigma \approx \pi$, while the dark green branch has a decreasing phase σ. A second pair of CLMs is born in the saddle node at $C_p \approx 1.8\pi$. The pink branch has an almost constant phase $\sigma \approx 0$, while the light green branch has an increasing phase σ. The decreasing-phase and the increasing-phase branches cross, and the branches come together differently in the other two saddle-node bifurcations on the left. This scenario corresponds to moving over the respective branches of the same color in the left column of Figure 10.

Clearly, Figure 10(a1)–(c1) and Figure 10(a2)–(c2) are a perturbation, resulting in unfoldings of the pitchfork bifurcations, of the respective plots for zero detuning in Figure 5(a) and (b). To show in detail which branches interact in this unfolding of the pitchfork bifurcations, Figure 11 shows enlarged views in the (C_p, N^s)-plane. The left column shows the situation for zero detuning, where the constant-phase CLMs (pink and purple curves) have

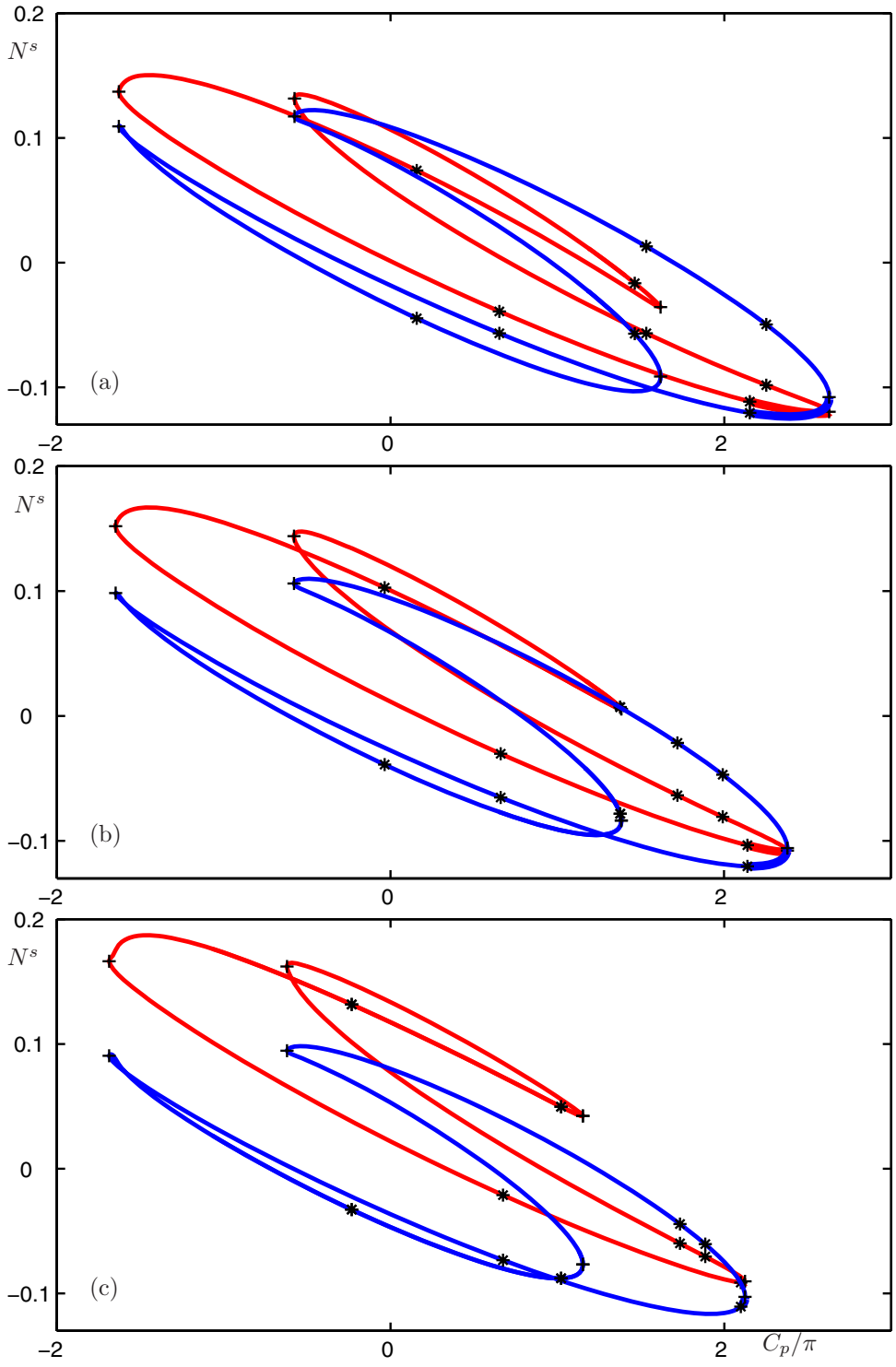


Figure 9. *Detuning sequence of CLMs in the* (C_p, N^s)*-plane. Panel* (a) *shows the CLMs for* $\Delta = 0.025$ *from Figure 2, while in panel* (b) *and* (c) Δ *takes the values* 0.05 *and* 0.075*, respectively. See also the accompanying animations* (61995_03.gif [1.4MB] *and* 61995_04.gif [1.7MB]) *in the* (ω^s, N^s)*-plane.*

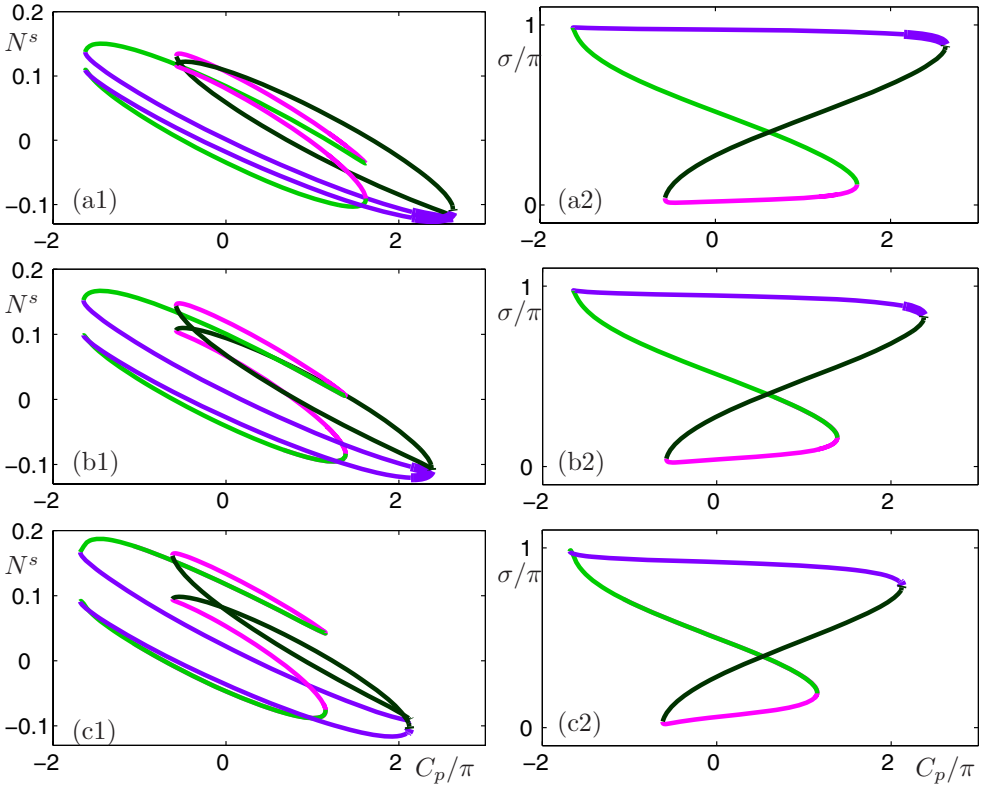


Figure 10. *The detuning sequence in the (C_p, N^s)-plane from Figure 9 shown in the (C_p, N^s)-plane (left column) and in the (C_p, σ)-plane (right column). The different branches are color coded as in Figure 5 to identify (approximate) in-phase, antiphase, increasing-phase, and decreasing-phase CLMs.*

$N_1^s = N_2^s$. The inversions of the intermediate-phase CLMs (dark green and light green curves) lie on the same curve but at different positions because $N_1^s \neq N_2^s$. Consequently, as was discussed earlier, every branch that exists for nonzero detuning "doubles" in the presence of detuning, as can be seen in the right column of Figure 11. The different rows show clearly how different branches connect locally near the pitchfork bifurcations, which globally leads to the structure of CLMs in Figure 10.

6.2. Intermediate values of detuning. As we see now, branches of CLMs connect or disconnect for larger values of the detuning Δ. In Figure 12 all branches of CLMs are shown in the (ω^s, N^s)-plane, for three different intermediate values of Δ. This figure reveals a pinching-off of the red horseshoe, while the blue horseshoe transforms into two concentric circles. Figure 12(b) shows a transversal crossing of the respective parts of the red and blue curves.

The transition can be understood in terms of the corresponding red and blue surfaces in (C_p, Δ, N^s)-space. In fact, Figure 12(b) already shows that we are dealing with a crossing through a saddle singularity of this surface with respect to Δ. This is a classic codimension-one singularity of this surface [9] that leads locally to a different reconnection of the branches involved. In Figure 13 we show how this manifests itself in the (C_p, N^s)-plane in terms

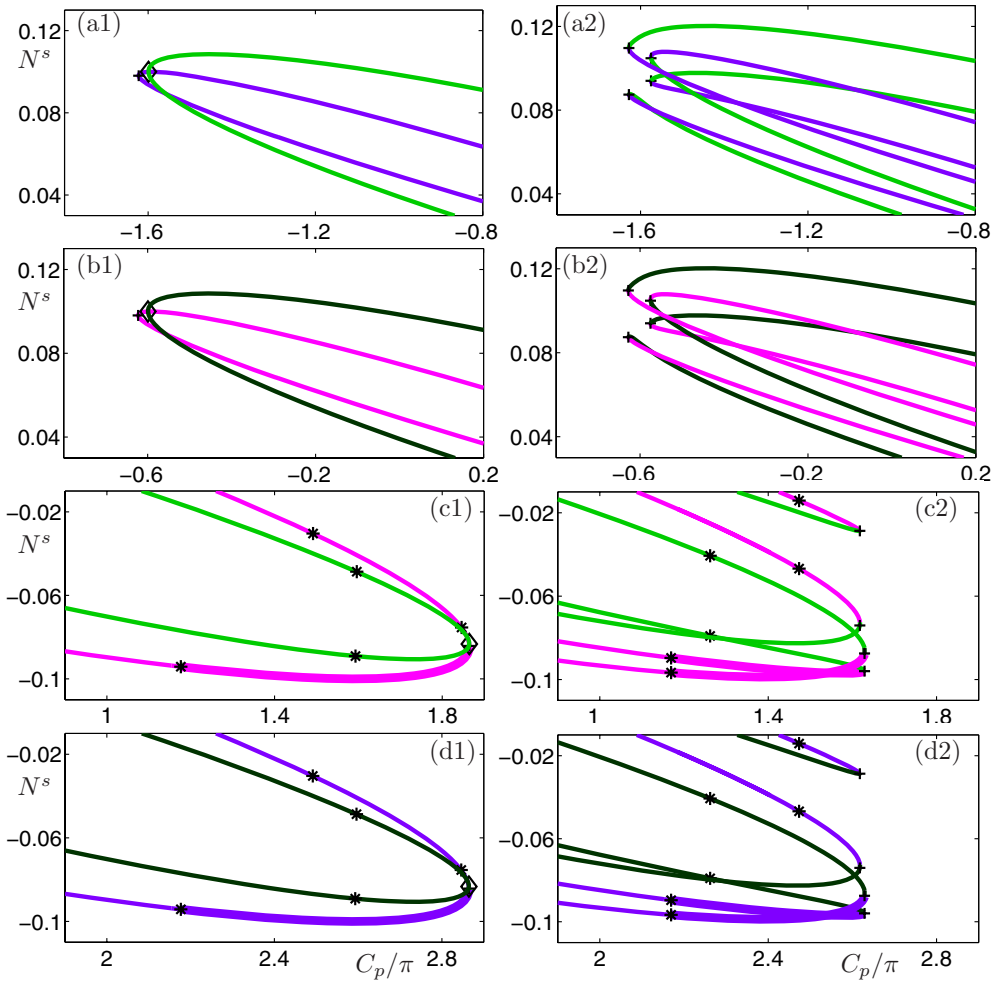


Figure 11. *Enlarged views of branches of CLMs for $\Delta = 0$ (left column) and for $\Delta = 0.025$ (right column) in the (C_p, N^s)-plane near the unfoldings of pitchfork bifurcations showing how different types of branches connect.*

of the inversions of the red and blue lasers. Panel (a) shows the situation for $\Delta = 0.075$, which corresponds to Figure 9(a). As the saddle-singularity is approached, two π-translational symmetry related CLM branches approach each other. After the singularity there are two separate branches of CLMs, that is, pairs of red and blue branches. They are plotted in Figure 13(b) and (c), respectively.

In Figure 14 we plot the curves of CLMs of Figure 13 in the (C_p, N^s)-plane (left column) and the (C_p, σ)-plane (right column), again with the same color coding indicating the different branches in terms of their phase difference σ. Furthermore, we now plot in panel (a) all copies of branches under the translational symmetries. As the saddle-singularity is approached, two π-translational symmetry related CLM branches approach each other; see the region around $(C_p, \sigma) = (-1.8, 0)$ in Figure 14(a1). After the transition we see that the first CLM branch in Figure 14(b) is formed by connecting the upper dark green branch with the lower pink branch,

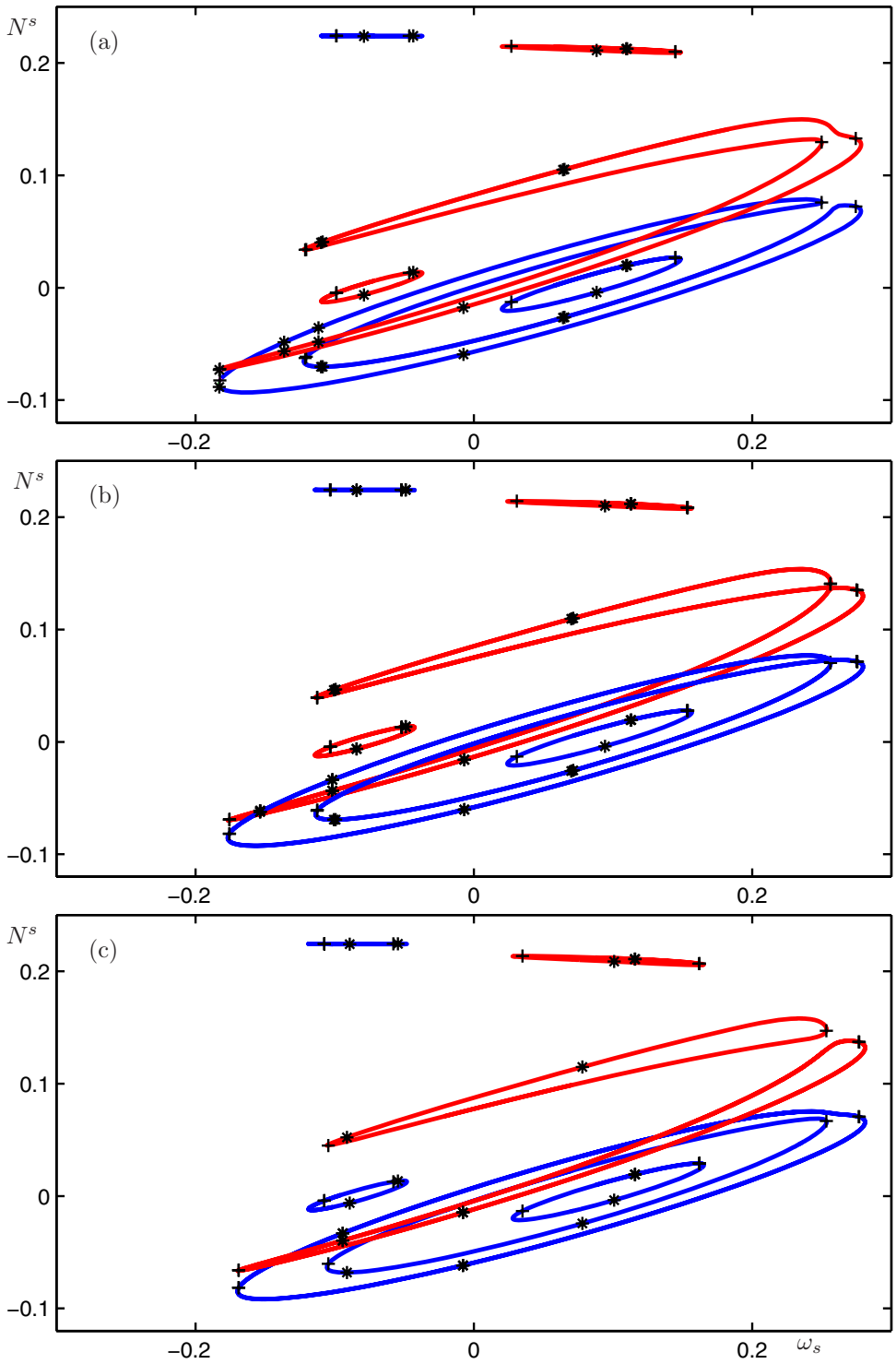


Figure 12. *Detuning sequence of CLMs in the* (ω^s, N^s)*-plane showing the inversions of the red and blue lasers during a transition through a saddle singularity. From* (a) *to* (c) Δ *takes the values* 0.075, 0.080, *and* 0.085. *See also the accompanying animations* (61995_03.gif [1.4MB] *and* 61995_04.gif [1.7MB]).

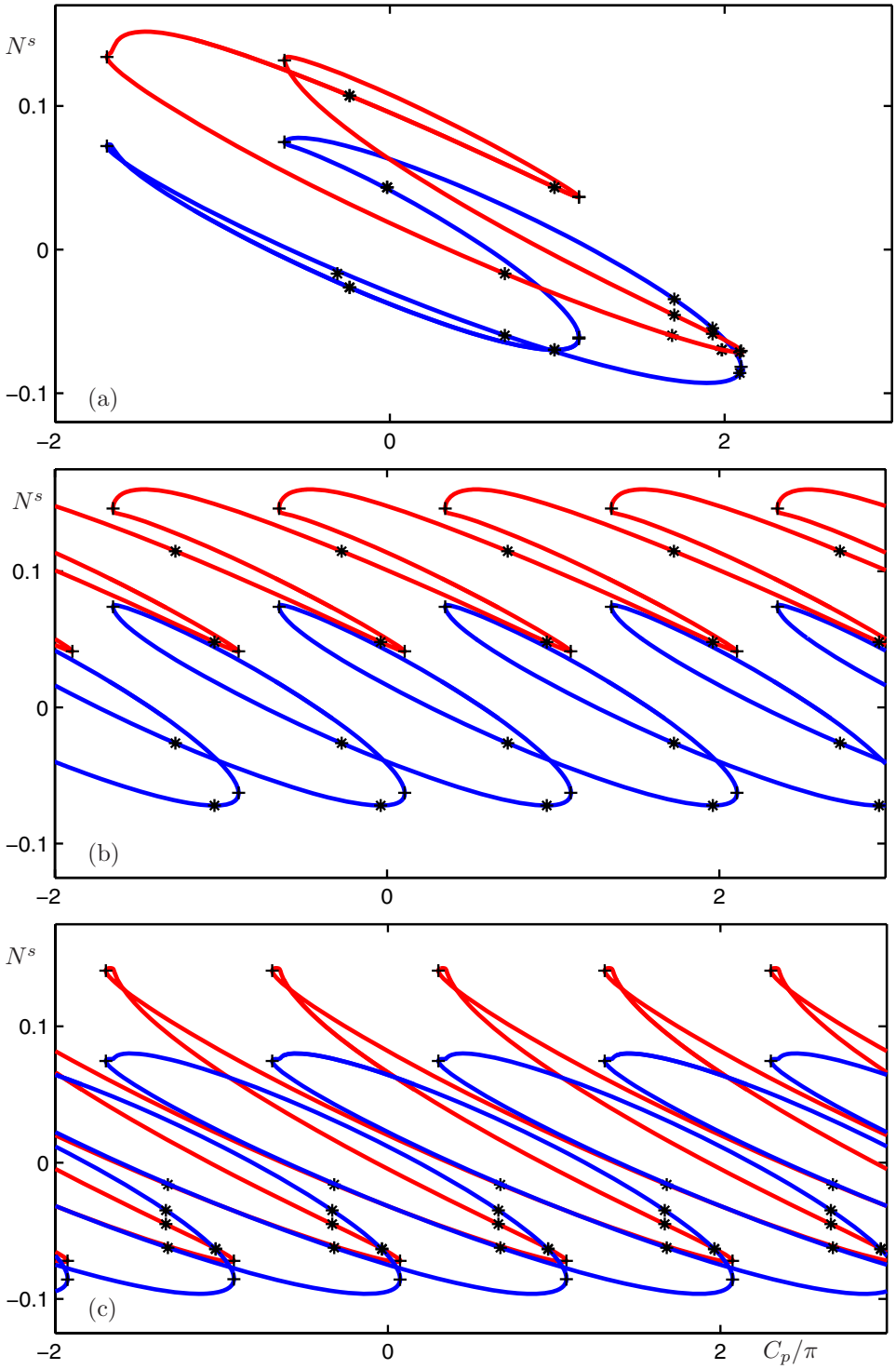


Figure 13. *Panel* (a) *shows the curves of CLMs from Figure* 12 *in the* (C_p, N^s)*-plane before the transition for* $\Delta = 0.075$*, while panels* (b) *and* (c) *show two types of curves of CLMs for* $\Delta = 0.085$ *that are created in the saddle transition.*

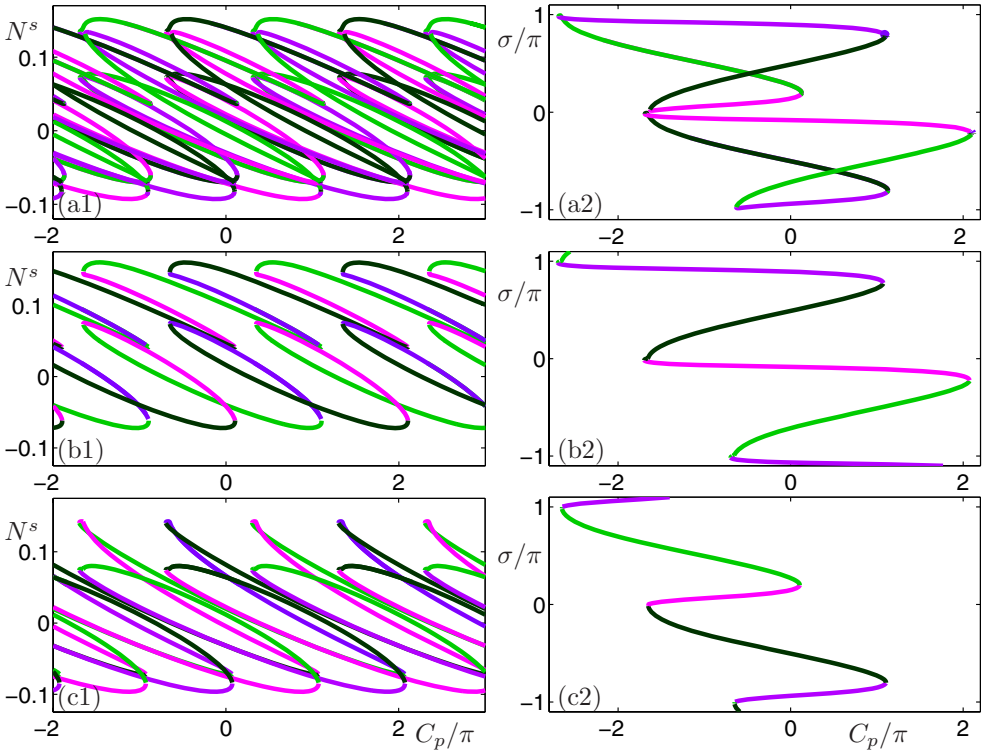


Figure 14. *The curves of CLMs from Figure 13 in the (C_p, N^s)-plane (left column) and in the (C_p, σ)-plane (right column), where the different branches are color coded as in Figure 5 to identify (approximate) in-phase, antiphase, increasing-phase, and decreasing-phase CLMs.*

while the second CLM branch in Figure 14(b) is formed by connecting the lower dark green branch with the upper pink branch. Note that both branches do not have a bounded phase difference as a function of C_p.

The physical interpretation of this transition through a saddle-singularity is as follows. Detuning has the tendency to pull the inversions of the two lasers apart, whereas the coupling ties them together. Before the saddle-singularity the system can unify both tendencies in a single branch of CLMs; see Figures 13(a) and 14(a). However, after the saddle-singularity there are two different branches of CLMs. The branch in Figures 13(b) and 14(b) corresponds to the "pulling-apart tendency" of the detuning, while the one in Figures 13(c) and 14(c) corresponds to the "tying-together tendency" of the coupling. Note that at the saddle-singularity we are dealing with two CLMs with the same inversion, which is also known as the Petermann–Tager condition in the context of a laser with conventional optical feedback [35, 40].

When Δ is increased even further we encounter a second saddle-singularity. However, this time the global organization of the branches of CLMs is such that this results in the merging of two separate branches. In Figure 15 all branches of CLMs are shown in the (ω^s, N^s)-plane, for two different values of detuning, one before and one after this second saddle-singularity. While the exact shape and position of the different branches of CLMs have changed, the situation in Figure 15(a) is topologically the same as that in Figure 12(c). After the bifurcation the

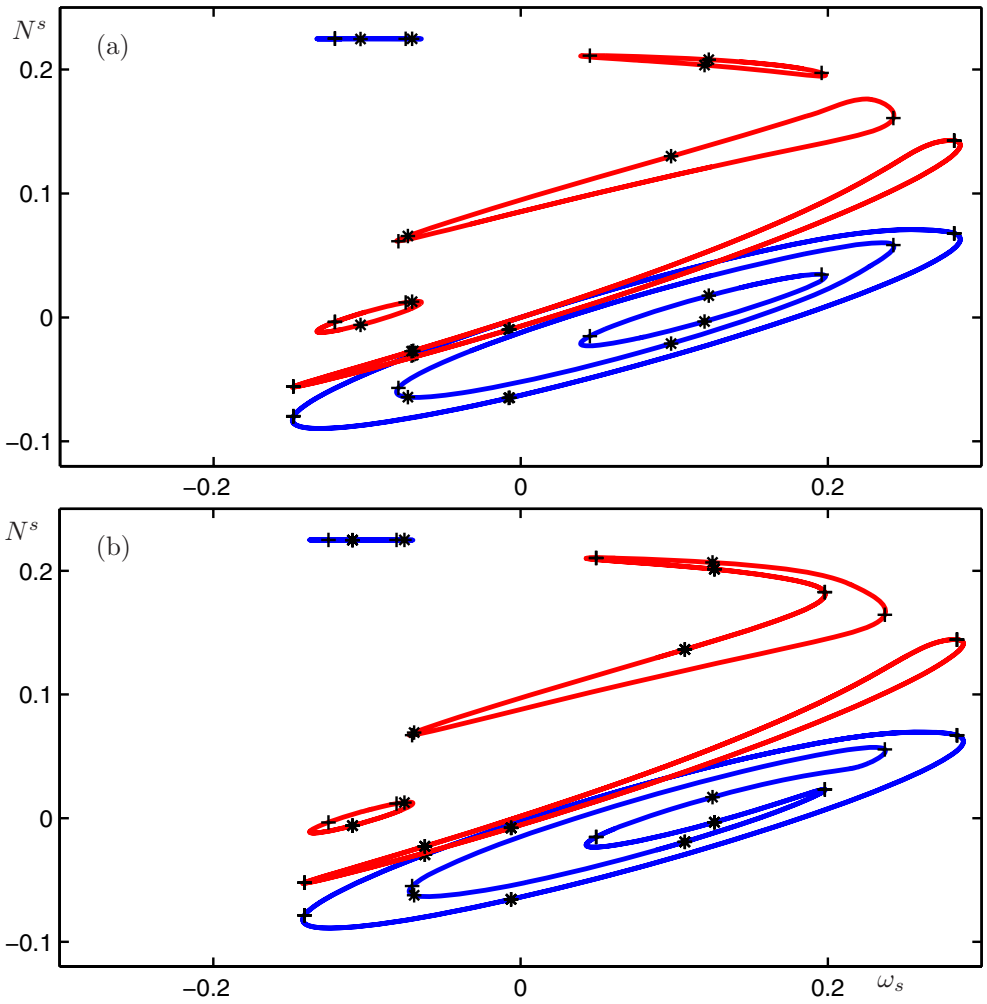


Figure 15. *Detuning sequence of CLMs in the* (ω^s, N^s)*-plane for the red and blue lasers before* (a) *and after* (b) *a further transition through a saddle singularity; the detuning* Δ *takes the values* 0.10 *and* 0.105*, respectively. See also the accompanying animations* (61995_03.gif [1.4MB] *and* 61995_04.gif [1.7MB]).

two separate red branches merge into a "boomerang-like" structure, while the two innermost concentric blue curves form a new horseshoe.

In the (C_p, N^s)-plane this transition manifests itself as shown in Figure 16 in terms of the inversions of the red and blue lasers. Figure 17 again shows the branches of CLMs in the (C_p, N^s)-plane and the (C_p, σ)-plane in the color coding in terms of the phase difference σ. Before the saddle-singularity there are exactly two distinct branches of CLMs, that is, pairs of red and blue curves in Figure 16. As the singularity is approached the low-inversion branch of the red laser hits the high-inversion branch from below. Then these two branches connect differently to form infinitely many closed red curves, and the two blue curves undergo a similar transition to infinitely many bounded, closed curves; see Figure 16(b) and also Figure 17(b1). In the process, the phase difference between the two lasers becomes bounded, as is shown in

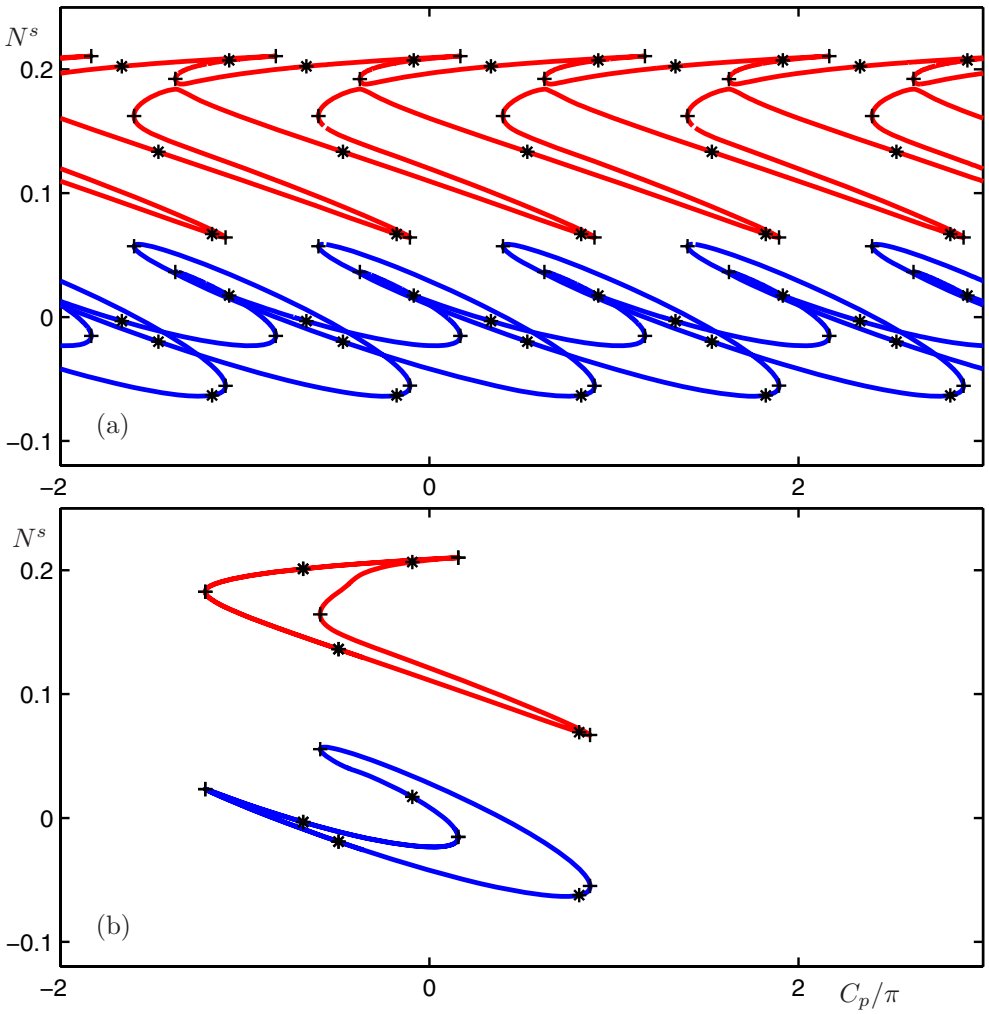


Figure 16. *The curves of CLMs from Figure 15 in the (C_p, N^s)-plane (left column) and in the (C_p, σ)-plane (right column), where the different branches are color coded as in Figure 5 to identify (approximate) in-phase, antiphase, increasing-phase, and decreasing-phase CLMs.*

Figure 17(b2).

6.3. The limit of very large detuning. When Δ is increased even further, as is shown in Figure 18 in the (ω^s, N^s)-plane, a pair of red and blue islands, that is, a separate branch of CLMs, shrinks down to two points and disappears. This happens in another classical codimension-one singularity of the surface of CLMs, namely, a transition through an extremum [9] (specifically, a maximum with respect to Δ). After this last singularity transition the situation remains topologically the same. However, the remaining two pairs of isolas become smaller and they center around the points $(\omega^s, N^s) \approx (\pm\Delta, 0)$ and $(\omega^s, N^s) \approx (\pm\Delta, P)$.

This behavior can be explained by considering the limit of Δ going to infinity. Clearly, $R^s_{1,2}$, $N^s_{1,2}$, ω^s, and σ^s in (3.5)–(3.10) depend on Δ; we denote their limiting values for $\Delta \to \infty$

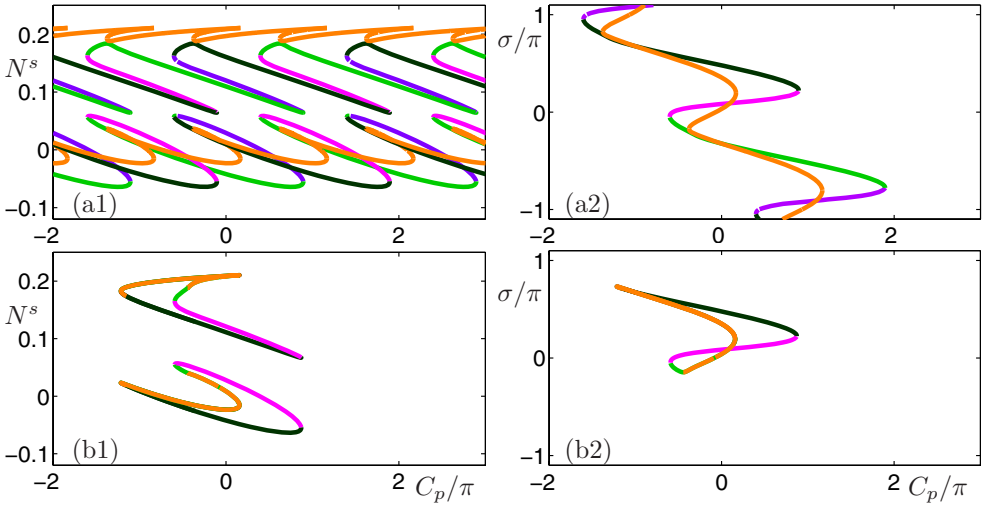


Figure 17. *The curves of CLMs from Figure 15 in the (C_p, N^s)-plane.*

by $\overline{R}^s_{1,2}$, $\overline{N}^s_{1,2}$, $\overline{\omega}^s$, and $\overline{\sigma}^s$, respectively. By solving (3.11) for $1/\Delta^2$ we conclude that $(\omega^s)^2$ grows as Δ^2. In other words, in the limit $\Delta \to \infty$ we have that

$$\overline{\omega}^s = \pm\Delta. \tag{6.1}$$

Due to the symmetry (2.6) it is sufficient to consider only the case $\Delta = \omega^s$. We use (3.5) and (3.6) for $\overline{\omega}^s = \Delta$ to eliminate N_1^s and obtain

$$2\Delta = \kappa\sqrt{(1+\alpha^2)}\frac{R_2^s}{R_1^s}\sin(-C_p - \Delta\tau + \sigma - A_\alpha), \tag{6.2}$$

where $A_\alpha = \arctan\alpha$. Taking reciprocals and letting $\Delta \to \infty$, we conclude that

$$\frac{\overline{R}_1^s}{\overline{R}_2^s} = 0. \tag{6.3}$$

Inserting this limit into (3.7) gives, for $\overline{\omega}^s = \Delta$,

$$\overline{N}_2^s = 0. \tag{6.4}$$

Finally, combining (3.9) and (3.10) gives

$$\frac{|R_1^s|^2}{|R_2^s|^2} = \frac{(1+2N_2^s)(P-N_1^s)}{(1+2N_1^s)(P-N_2^s)}, \tag{6.5}$$

from which we conclude with (6.3) and (6.4) that

$$\overline{N}_1^s = P. \tag{6.6}$$

The amplitudes R_1^s and R_2^s are computed from (3.9) and (3.10) as

$$\overline{R}_1^s = 0 \quad \text{and} \quad \overline{R}_2^s = P. \tag{6.7}$$

Recall that R_1^s and R_2^s are defined to be positive.

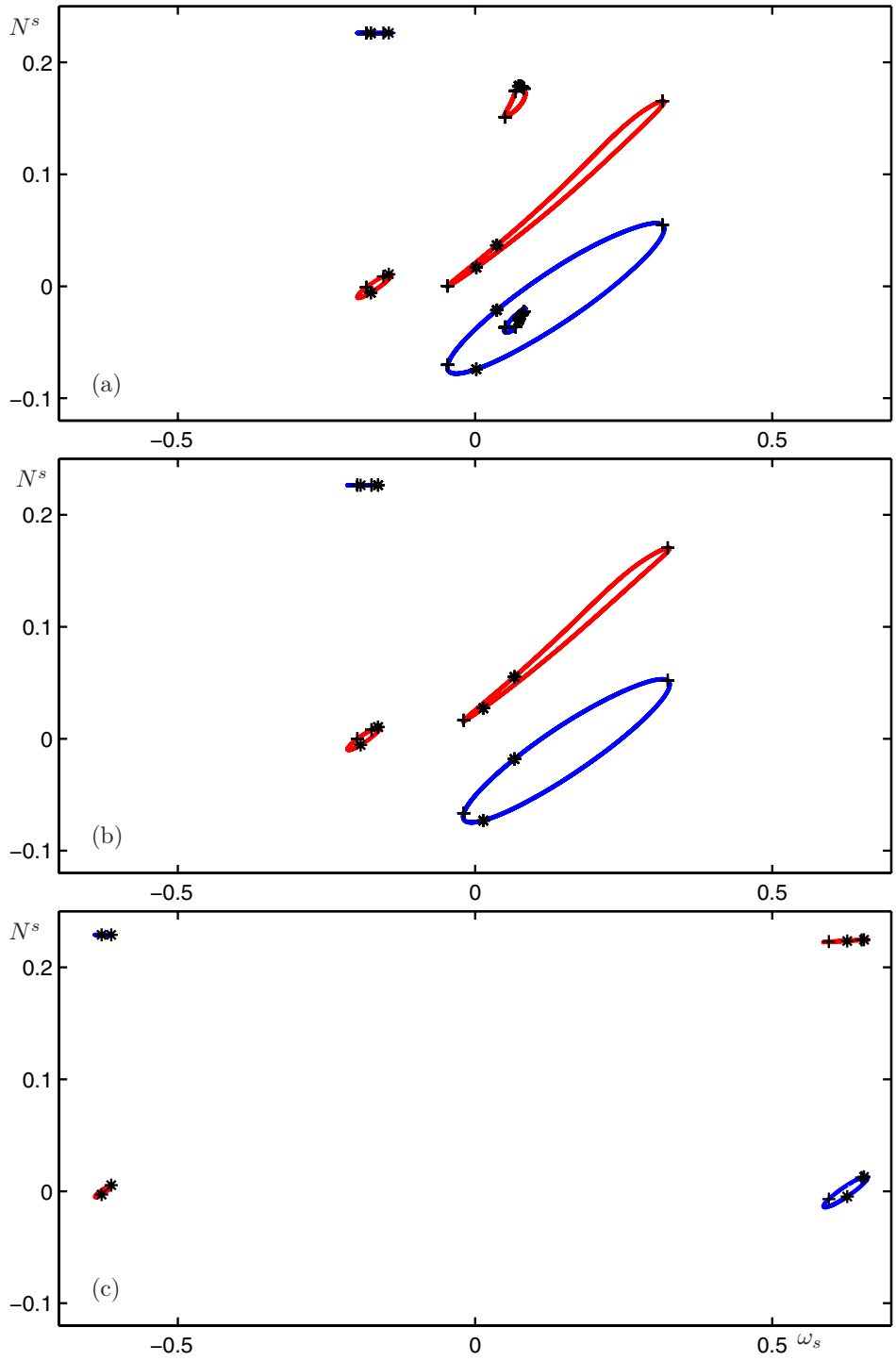


Figure 18. *Detuning sequence of CLMs in the* (ω^s, N^s)*-plane for the red and blue lasers toward the limit of large detuning; from* (a) *to* (c) Δ *takes the values* 0.17, 0.1875, *and* 0.625. *Notice the transition through a maximum between panels* (a) *and* (b) *in which a pair of closed branches disappears. See also the accompanying animations* (61995_03.gif [1.4MB] *and* 61995_04.gif [1.7MB]).

This result on the limit $\Delta \to \infty$ can be interpreted physically as follows. Because of the mismatch in their free-running optical frequencies, the electric field of one laser has a decreasing influence on the electric field of the other laser. When Δ grows, the two lasers become more detuned. Concentrating on the solution on the positive side of the ω^s-axes in Figure 18, it can be seen that the red detuned laser operates around its detuned free-running optical frequency. Due to the coupling the blue laser is forced to also operate at this frequency; that is, it is far away from its detuned free-running optical frequency. Therefore, the blue laser operates around its off-state. In turn this means that the effective coupling between the lasers is small and the red laser is only perturbed a little. In the limit $\Delta \to \infty$ the two lasers are completely independent. This is expressed in (2.1)–(2.4) as two solutions: the red laser is on and the blue laser is off, or vice versa.

6.4. The surface of CLMs. To get an overall impression of the CLM structure we give in Figure 19 an impression of the corresponding surfaces in (Δ, ω^s, N^s)-space. As before, the inversion of laser 1 is plotted in red and that of laser 2 in blue; the projection of the surface of CLMs onto the (Δ, ω^s)-plane is shown in black. The image is built up from 150 Δ-slices that were computed with DDE-BIFTOOL which also form the frames of the accompanying animation (61995_04.gif). Panel (a) shows the whole structure for $\Delta \in [-0.6, 0.6]$. The blue surface is mapped to the red surface and vice versa by the symmetry (2.6) of exchanging the two lasers and changing the sign of Δ. The two surfaces intersect for $\Delta = 0$. For small Δ this surface has a complicated, nested structure. This can be seen in panel (b), where an enlarged view for positive Δ is shown. For increasing (or decreasing) detuning the nested structure starts to disentangle. Indeed this process proceeds as described in section 6.2 with transitions through saddle-singularities. A saddle can be seen clearly on the blue surface in Figure 19(b). Furthermore, one clearly notices the extremum, specifically a maximum with respect to Δ, on the same blue surface where the surface "bends back" to decreasing Δ. The limiting behavior of the CLMs, as discussed in section 6.3, is brought out well by Figure 19(a).

The black projection onto the (Δ, ω^s)-plane shows the parameter region where CLMs exist. As explained in the previous section, ω^s is not a bifurcation parameter, so the boundaries of the projection are not saddle-node bifurcations. Nevertheless, the saddle-node bifurcations are close to the folds with respect to ω^s. Figure 20 shows the curves of saddle-node bifurcations (blue) and the curve of the first Hopf bifurcation (red) that forms a boundary of the region of stable CLMs for small detuning. Panel (a) shows the (Δ, ω^s)-plane for $\Delta \in [-0.7, 0.7]$ and panel (b) is an enlarged view around the stable region (green).

Figure 21 shows the same bifurcation curves in the (Δ, C_p)-plane. In panel (a) we plot the basic structure only once, while in panel (b) we also show all its images under the π-translational symmetry of the system. There is a complicated structure of different regions, especially near the stable region, as can be seen in the further enlargement in panel (c). Overall, there are two closed curves of saddle-node bifurcations. The first branch of saddle-node bifurcations is limited to a detuning interval of $\Delta \approx [-0.23, 0.23]$; it arises in the unfolding of the pitchfork bifurcation of intermediate-phase CLMs (green curves in Figures 2–5). Indeed this shows that the saddle-node bifurcations of the constant-phase CLMs (purple curves in Figures 2–5) are connected with the second set of intermediate-phase CLMs (orange curves in Figures 2, 3, and 6). The second branch of saddle-node bifurcations traces out the large "triangular" curve. There are a number of cusp bifurcations on the saddle-node curves, so

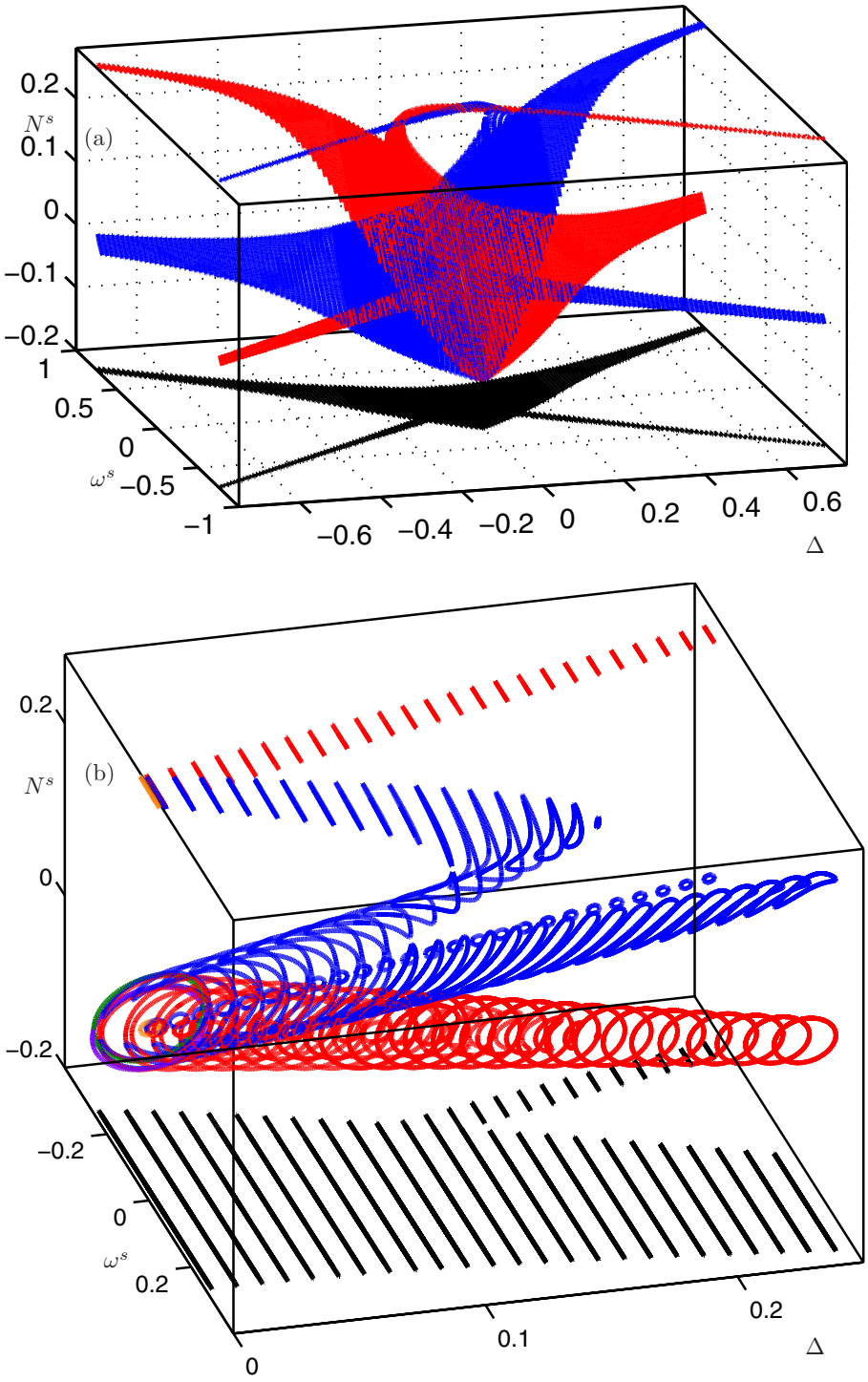


Figure 19. *The surface of CLMs in* (Δ, ω^s, N^s)*-space for* $\Delta \in [-0.6, 0.6]$ *as built up from* 150 *equidistant* Δ*-slices* (a)*, and an enlarged view for positive* Δ (b)*. The inversion of laser* 1 *is shown in red and that of laser* 2 *in blue; the projection of the CLMs onto the* (Δ, ω^s)*-plane is shown in black.*

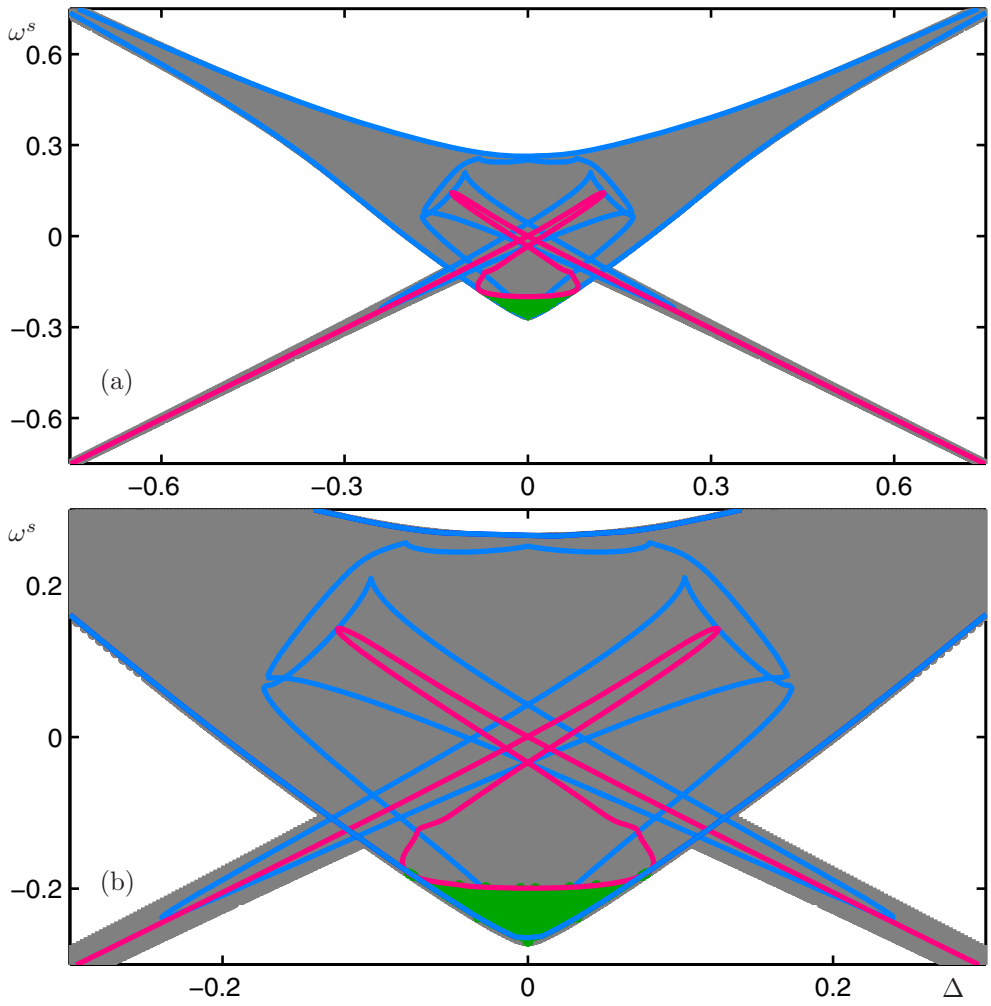


Figure 20. *Two-parameter bifurcation diagram in the* (Δ, ω^s)*-plane with saddle-node bifurcation curves (blue) and the curve of the first Hopf bifurcation (red). The region where CLMs exist is shaded in gray and the region of stable CLMs is plotted in green. Panel* (b) *shows an enlarged view of panel* (a).

that different branches correspond to saddle-node bifurcations of different CLMs. Note that the cusp points do not appear as cusps in the projection onto the (Δ, ω^s)-plane in Figure 20.

7. Outlook. We presented here a comprehensive geometric picture of the CLMs of two identical delay-coupled lasers as a function of the feedback phase and the detuning. This revealed a complicated structure of different types of CLMs, which form the "backbone" of all dynamics in the system. We considered here the case that the SLs are pumped well above their threshold. While this situation is quite typical in experiments, the question of what happens when the lasers are pumped close to threshold arises. It is known that the overall number of CLMs that can be sustained by the system and the frequency of the characteristic relaxation oscillations depend on the pump current. How the CLM structure presented here changes with pump current is discussed in [5].

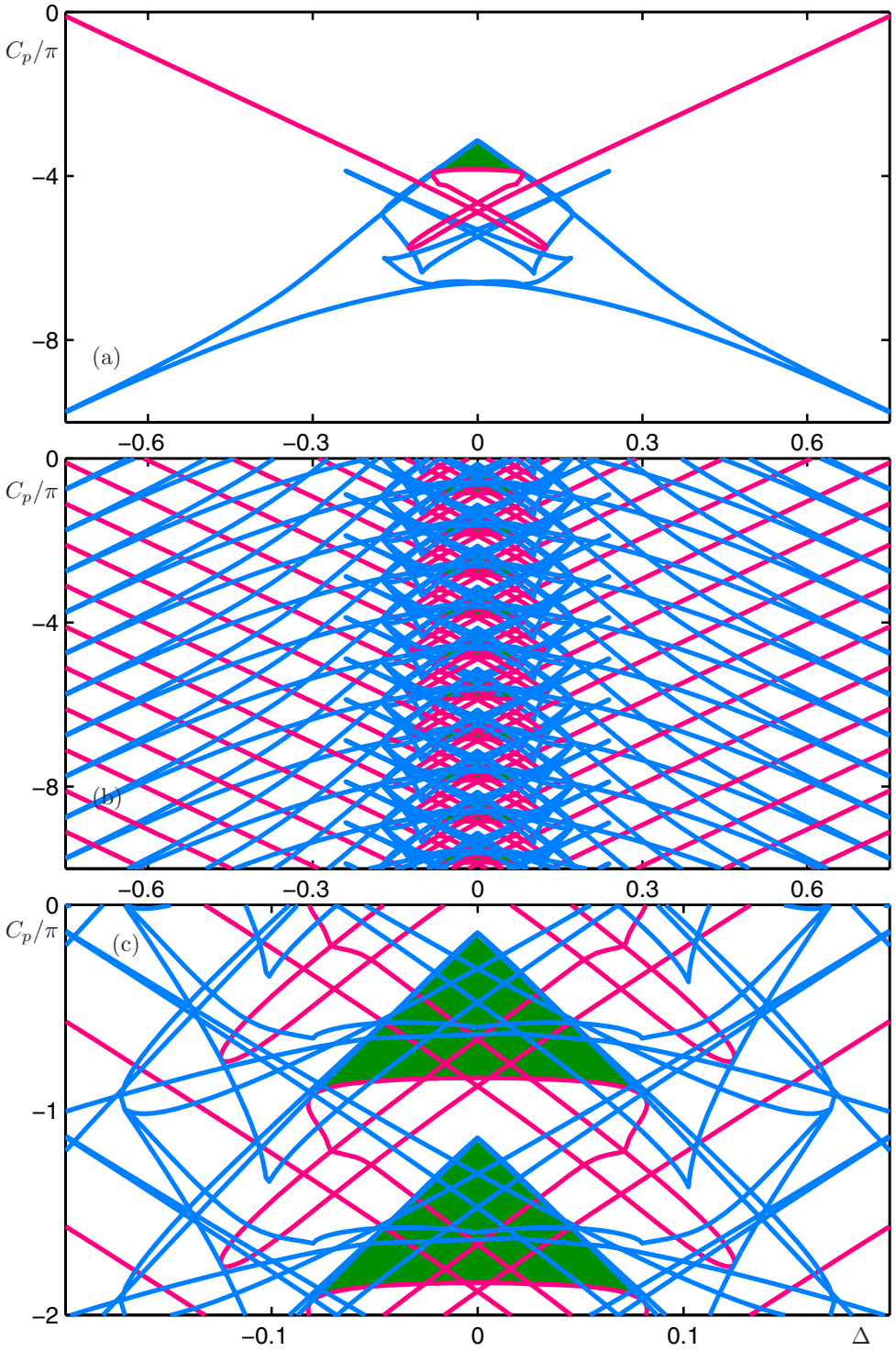


Figure 21. *Two parameter bifurcation diagram in the* (Δ, C_p)*-plane with saddle-node bifurcation in blue and Hopf bifurcation in red. The region of stable CLMs is plotted in green. Panel* (a) *shows the basic structure, panel* (b) *takes the translational symmetries into account, and panel* (c) *shows an enlarged view of panel* (b).

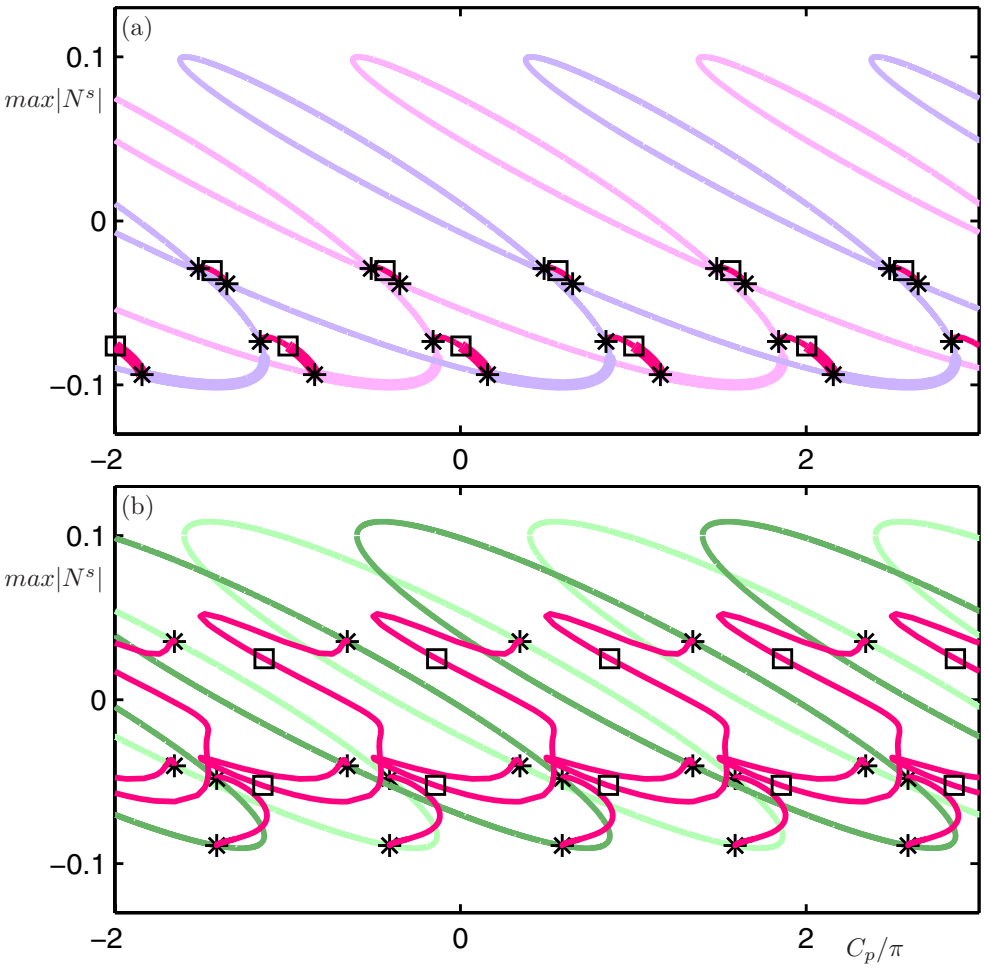


Figure 22. *Branches of periodic solutions (red) in the (C_p, N^s)-plane. Plotted is the amplitude of the periodic solution; stable parts are boldfaced. Periodic solutions emerge from Hopf bifurcations ($*$) of CLMs and destabilize in torus bifurcations ($\square$). The branches of in-phase, antiphase, increasing-phase, and decreasing-phase CLMs are shown in lighter colors; compare with Figure 5.*

After the CLM structure is known, the next logical step is to consider solutions of non-constant intensity of the lasers. First, it is now possible to consider the periodic orbits that are born in Hopf bifurcations. Near the stable region of constant-phase CLMs one finds stable oscillations of the power that can then bifurcate further. To give an idea, we present in Figure 22 branches of periodic orbits in the (C_p, N^s)-plane that bifurcate from branches of constant-phase CLMs, in panel (a), and from branches of intermediate-phase CLMs, in panel (b). The branches of constant-phase and intermediate-phase CLMs are repeated from Figure 5 and plotted in a lighter color. The constant-phase CLMs exhibit four different Hopf bifurcation points in total; the symmetric copies are not taken into account. These Hopf bifurcations are connected in pairs by branches or "bridges" of periodic orbits. Branches of periodic orbits (red curves) are represented by plotting the maximum amplitude of the oscillation; stable parts are boldfaced. A similar picture emerges for the intermediate-phase

CLMs plotted in Figure 22(b). They also exhibit four different Hopf bifurcations, which are pairwise connected by branches of periodic orbits.

The situation for the constant-phase CLMs is similar to the case of the Lang–Kobayashi equations for a laser with conventional optical feedback where one finds "bridges" of periodic orbits that connect different branches of ECMs [11]. However, in the present situation it emerges that such bridges of periodic orbits provide a connection between different types of CLMs. Concentrating on the smaller branch of periodic orbits emerging from the Hopf bifurcation of the constant-phase CLMs, it can be seen that this branch connects two Hopf bifurcations that are located near the intersection of an in-phase CLM branch and an antiphase CLM branch, respectively. This branch of periodic orbits is unstable throughout, but there is a torus bifurcation point on it. The longer branch of periodic orbits connects two Hopf bifurcations near the intersection of an in-phase CLM with an antiphase CLM; one Hopf bifurcation is on the branch of the in-phase CLM and the other one is on the branch of the antiphase CLM. This branch of periodic orbits is initially stable, so that one observes stable oscillations of the power of the two lasers. As C_p is decreased it destabilizes in a torus bifurcation, giving rise to stable quasi-periodic or locked oscillations of the laser power. This torus then breaks up and gives rise to a region of chaotic fluctuations. A more detailed study of connecting bridges and associated routes to chaos is an interesting topic of ongoing research.

Another important question is how similar the two lasers need to be. Indeed, it is practically impossible to produce two identical lasers. On the other hand, present experiments [6, 7, 38] show good agreement with the model as studied here. In other words, apparently it is sufficient that the lasers are similar enough in terms of their material properties. By performing a bifurcation study in the parameters α and T it is possible to study this question systematically.

REFERENCES

[1] G. A. Agrawal and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*, Van Nostrand Reinhold, New York, 1986.

[2] O. Diekmann, S. A. V. Gils, and S. M. V. Lunel, *Delay Equations: Functional-, Complex-, and Nonlinear Analysis*, Springer-Verlag, New York, 1995.

[3] E. J. Doedel, R. Paffenroth, A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. E. Oldeman, B. Sandstede, and X. Wang, *AUTO* 2000*: Continuation and Bifurcation Software for Ordinary Differential Equations (with HomCon)*, 2000, http://www.sourceforge.net/projects/auto2000/.

[4] K. Engelborghs, T. Luzyanina, and G. Samaey, *DDE-BIFTOOL v.* 2.00 *User Manual: A MATLAB Package for Bifurcation Analysis of Delay Differential Equations*, Technical report TW-330, Department of Computer Science, K. U. Leuven, Leuven, Belgium, 2001.

[5] H. Erzgräber, B. Krauskopf, and D. Lenstra, *Mode structure of delay-coupled semiconductor lasers: Influence of the pump current*, J. Opt. B Quantum Semiclass. Opt., 7 (2005), pp. 361–371.

[6] H. Erzgräber, D. Lenstra, B. Krauskopf, and I. Fischer, *Dynamical properties of mutually delayed coupled semiconductor lasers*, in Semiconductor Lasers and Laser Dynamics, Proceedings of SPIE 5452, Strasbourg, France, 2004, pp. 352–361.

[7] H. Erzgräber, D. Lenstra, B. Krauskopf, E. Wille, M. Peil, I. Fischer, and W. Elsässer, *Mutually delay-coupled semiconductor lasers: Mode bifurcation scenarios*, Opt. Commun., 255 (2005), pp. 286–296.

[8] I. Fischer, Y. Liu, and P. Davis, *Synchronization of chaotic semiconductor laser dynamics on subnanosecond time scales and its potential for chaos communication*, Phys. Rev. A (3), 62 (2000), 011801.

[9] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory* I, Springer-Verlag, Heidelberg, 1998.

[10] K. Green, B. Krauskopf, and G. Samaey, *A two-parameter study of the locking region of a semiconductor laser subject to phase-conjugate feedback*, SIAM J. Appl. Dyn. Syst., 2 (2003), pp. 254–276.

[11] B. Haegeman, K. Engelborghs, D. Rose, D. Pieroux, and T. Erneux, *Stability and rupture of bifurcation bridges in semiconductor lasers subject to optical feedback*, Phys. Rev. E (3), 66 (2002), 046216.

[12] J. K. Hale and S. M. V. Lunel, *Introduction to Functional Differential Equations*, Springer-Verlag, New York, 1993.

[13] T. Heil, I. Fischer, W. Elsässer, and A. Gavrielides, *Dynamics of semiconductor lasers subject to delayed optical feedback: The short cavity regime*, Phys. Rev. Lett., 87 (2001), 243901.

[14] T. Heil, I. Fischer, W. Elsässer, J. Mulet, and C. R. Mirasso, *Chaos synchronization and spontaneous symmetry-breaking in symmetrically delay-coupled semiconductor lasers*, Phys. Rev. Lett., 86 (2001), pp. 795–798.

[15] A. Hohl, A. Gavrielides, T. Erneux, and V. Kovanis, *Localized synchronization in two coupled nonidentical semiconductor lasers*, Phys. Rev. Lett., 78 (1997), pp. 4745–4748.

[16] A. Hohl, A. Gavrielides, T. Erneux, and V. Kovanis, *Quasiperiodic synchronization for two delay-coupled semiconductor lasers*, Phys. Rev. A (3), 59 (1999), pp. 3941–3949.

[17] J. Javaloyes, P. Mandel, and D. Pieroux, *Dynamical properties of lasers coupled face to face*, Phys. Rev. E (3), 67 (2003), 036201.

[18] B. Krauskopf, *Bifurcation analysis of lasers with delay*, in Unlocking Dynamical Diversity: Optical Feedback Effects on Semiconductor Lasers, D. M. Kane and K. A. Shore, eds., Wiley, New York, 2005, pp. 147–183.

[19] B. Krauskopf, H. Erzgräber, and D. Lenstra, *Bifurcation analysis of coupled laser modes in mutually delay-coupled lasers*, in the 5th IFAC Workshop on Time-Delay Systems, Leuven, Belgium, 2004.

[20] B. Krauskopf and D. Lenstra, *Fundamental Issues of Nonlinear Laser Dynamics*, American Institute of Physics, College Park, MD, 2000.

[21] B. Krauskopf, G. H. M. Van Tartwijk, and G. R. Gray, *Symmetry properties of lasers subject to optical feedback*, Opt. Commun., 177 (2000), pp. 347–353.

[22] R. Lang and K. Kobayashi, *External optical feedback effects on semiconductor injection laser properties*, IEEE J. Quantum Electron., 16 (1980), pp. 347–355.

[23] J. Mulet, C. Masoller, and C. R. Mirasso, *Modeling bidirectionally coupled single-mode semiconductor lasers*, Phys. Rev. A (3), 65 (2002), 063815.

[24] J. Mulet, C. Mirasso, T. Heil, and I. Fischer, *Synchronization scenario of two distant mutually coupled semiconductor lasers*, J. Opt. B Quantum Semiclass. Opt., 6 (2004), pp. 97–105.

[25] D. Pieroux and P. Mandel, *Bifurcation diagram of a complex delay-differential equation with cubic nonlinearity*, Phys. Rev. E (3), 67 (2003), 056223.

[26] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, UK, 2001.

[27] D. V. Ramana Reddy, A. Sen, and G. L. Johnston, *Time delay induced death in coupled limit cycle oscillators*, Phys. Rev. Lett., 80 (1998), pp. 5109–5112.

[28] F. Rogister and J. Garca-Ojalvo, *Symmetry breaking and high-frequency periodic oscillations in mutually coupled laser diodes*, Opt. Lett., 28 (2003), pp. 1176–1178.

[29] V. Rotschäfer and B. Krauskopf, *A three parameter study of external cavity modes in semiconductor lasers with optical feedback*, in the 5th IFAC Workshop on Time-Delay Systems, Leuven, Belgium, 2004.

[30] J. Sacher, D. Baums, P. Panknin, W. Elsässer, and E. O. Göbel, *Intensity instabilities of semiconductor lasers under current modulation, external light injection, and delayed feedback*, Phys. Rev. A (3), 45 (1992), pp. 1893–1905.

[31] H. Schuster and P. Wagner, *Mutual entrainment of two limit cycle oscillators with time delayed coupling*, Progr. Theoret. Phys., 81 (1989), pp. 939–945.

[32] G. Stepán, *Retarded Dynamical Systems: Stability and Characteristic Functions*, Longman Scientific and Technical, Harlow, UK, 1989.

[33] S. H. Strogatz, *Nonlinear Dynamics and Chaos*, Perseus Books, Reading, MA, 1994.
[34] S. H. Strogatz and I. Stewart, *Coupled oscillators and biological synchronization*, Sci. Amer., 269 (1993), pp. 68–75.
[35] A. A. Tager and K. Petermann, *High-frequency oscillations and self-mode locking in short external-cavity laser diodes*, IEEE J. Quantum Electron., 30 (1994), pp. 1553–1561.
[36] R. Vicente, J. Mulet, M. Sciamanna, and C. Mirasso, *Simple interpretation of the dynamics of mutually coupled semiconductor lasers*, in Proceedings of SPIE 5349, Denver, CO, 2004, pp. 307–318.
[37] E. A. Viktorov, A. M. Yacomotti, and P. Mandel, *Semiconductor lasers coupled face-to-face*, J. Opt. B Quantum Semiclass. Opt., 6 (2004), pp. L9–L12.
[38] E. Wille, M. Peil, I. Fischer, and W. Elsässer, *Dynamical scenarios of mutually delay-coupled semiconductor lasers in the short coupling regime*, in Semiconductor Lasers and Laser Dynamics, Proceedings of SPIE Photonics Europe 2004 5452-05, Strasbourg, France, 2004, pp. 41–50.
[39] S. Wirkus and R. Rand, *The dynamics of two coupled van der Pol oscillators with delay coupling*, Nonlinear Dynam., 30 (2002), pp. 205–221.
[40] M. Wolfrum and D. Turaev, *Instabilities of lasers with moderately delayed optical feedback*, Opt. Commun., 212 (2002), pp. 127–138.
[41] H.-J. Wünsche, S. Bauer, J. Kreissl, O. Ushakov, N. Korneyev, F. Henneberger, E. Wille, H. Erzgräber, M. Peil, W. Elsässer, and I. Fischer, *Synchronization of delay-coupled oscillators: A study of semiconductor lasers*, Phys. Rev. Lett., 94 (2005), 163901.
[42] S. Yanchuk, K. Schneider, and L. Recke, *Dynamics of two mutually coupled semiconductor lasers: Instantaneous coupling limit*, Phys. Rev. E (3), 69 (2004), 056221.
[43] R. A. York and R. C. Compton, *Experimental observation and simulation of mode-locking phenomena in coupled-oscillator arrays*, J. Appl. Phys., 71 (1992), pp. 2959–2965.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 1, pp. 66–83

Block-Diagonalization of the Symmetric First-Order Coupled-Mode System*

Marina Chugunova† and Dmitry Pelinovsky†

Abstract. We consider the Hamiltonian first-order coupled-mode system that occurs in nonlinear optics, photonics, and atomic physics. Spectral stability of gap solitons is determined by eigenvalues of the linearized coupled-mode system, which is equivalent to a four-by-four Dirac system with sign-indefinite metric. In the special class of symmetric nonlinear potentials, we construct a block-diagonal representation of the linearized equations, when the spectral problem reduces to two coupled two-by-two Dirac systems. The block-diagonalization is used in fast numerical computations of eigenvalues with the Chebyshev interpolation algorithm.

Key words. Hamiltonian first-order coupled-mode systems, gap solitons, spectral stability, invariant subspaces, eigenvalues

AMS subject classifications. 34L16, 37K45, 34L40, 35Q51

DOI. 10.1137/050629781

1. Introduction.

Various applications in nonlinear optics [1], photonics band-gap engineering [2], and atomic physics [3] call for systematic studies of the *coupled-mode system*, which is expressed by two first-order semilinear PDEs in one space and one time dimensions. In nonlinear optics, the coupled-mode system describes counter-propagating light waves, which interact with a linear grating in an optical waveguide [4]. In photonics, the coupled-mode system is derived for coupled resonant waves in stop bands of a low-contrast three-dimensional photonic crystal [5]. In atomic physics, the coupled-mode system describes matter-wave Bose–Einstein condensates trapped in an optical lattice [6]. Existence, stability, and nonlinear dynamics of *gap solitons*, which are localized solutions of the coupled-mode system, are fundamental problems of interest in the aforementioned physical disciplines.

In the context of spectral stability of gap solitons, it has been discovered that the linearized coupled-mode system is equivalent to a four-by-four Dirac system with sign-indefinite metric, where numerical computations of eigenvalues represent a difficult numerical task. The pioneer work in [7, 8] showed that spurious unstable eigenvalues originate from the continuous spectrum in the Fourier basis decomposition and the Galerkin approximation. A delicate but time-consuming implementation of the continuous Newton method was developed to differentiate true unstable eigenvalues from the spurious ones [8]. Similar problems were discovered in the variational method [9, 10] and in the numerical finite-difference method [11, 12].

While some conclusions on instability bifurcations of gap solitons in the coupled-mode equations can be drawn on the basis of perturbation theory [7] and Evans function methods

*Received by the editors April 21, 2005; accepted for publication (in revised form) by J. Meiss August 26, 2005; published electronically January 27, 2006. This work was completed with the support of the SharcNet Graduate Scholarship.
http://www.siam.org/journals/siads/5-1/62978.htmll
†Department of Mathematics, McMaster University, Hamilton, Ontario, L8S 4K1, Canada (chugunom@math.mcmaster.ca, dmpeli@math.mcmaster.ca).

[13, 14], the numerical approximation of eigenvalues was an open problem until recently. New progress was made with the use of exterior algebra in the numerical computations of the Evans function [15], when the results of [7] on instability bifurcations of gap solitons were recovered. A similar shooting method was also applied to gap solitons in a more general model of a nonlinear Schrödinger equation with a periodic potential [6].

Our work addresses the problem of numerical approximations of eigenvalues in the linearized coupled-mode system with a different objective. We will show that the linearized system with a symmetric potential function can be block-diagonalized into two coupled two-by-two Dirac systems. The two Dirac systems represent the linearized Hamiltonian of the coupled-mode equations and determine instability bifurcations and unstable eigenvalues of gap solitons.

The main purpose of block-diagonalization is to optimize a numerical algorithm based on Chebyshev interpolation (see a recent application of Chebyshev interpolation to a system of coupled nonlinear Schrödinger equations in [16]). The algorithm computes the entire spectrum of the linearized coupled-mode system. It also allows us to control the spurious eigenvalues at least near the end points of continuous spectrum, where instability bifurcations occur [7, 15]. Due to block-diagonalization, the algorithm requires two times less memory compared to the standard discretization of the full linearized system, and computations of eigenvalues within the same tolerance bound are accelerated approximately twice as much. We report applications of the numerical algorithm to an example of the linearized coupled-mode system with a symmetric quartic potential function.

The paper is organized as follows. Section 2 describes the model and its symmetries. Section 3 gives the construction and properties of gap solitons in the nonlinear coupled-mode system. Section 4 presents block-diagonalization of the linearized coupled-mode system. Section 5 contains numerical computations of the spectrum of the block-diagonalized system. The appendix deals with exact solutions for gap solitons in the coupled-mode system with symmetric homogeneous potential functions.

2. Coupled-mode system.

We consider the *Hamiltonian* first-order coupled-mode system in the form

$$\begin{cases} i(u_t + u_x) + v = \partial_{\bar{u}} W(u, \bar{u}, v, \bar{v}), \\ i(v_t - v_x) + u = \partial_{\bar{v}} W(u, \bar{u}, v, \bar{v}), \end{cases} \tag{2.1}$$

where $(u, v) \in \mathbb{C}^2$, $x \in \mathbb{R}$, $t \geq 0$, and $W(u, \bar{u}, v, \bar{v})$ is real-valued. We assume that the potential function satisfies the following three conditions:

1. W is invariant with respect to the gauge transformation $(u, v) \mapsto e^{i\alpha}(u, v)$ for all $\alpha \in \mathbb{R}$.
2. W is symmetric with respect to the interchange $(u, v) \mapsto (v, u)$.
3. W is analytic in its variables near $u = v = 0$, such that $W = O(4)$.

The first condition is justified by the standard derivation of the coupled-mode system (2.1) with an envelope approximation [5]. The second condition defines a class of symmetric nonlinear potentials. Although it is somewhat restrictive, symmetric nonlinear potentials are commonly met in physical applications of the system (2.1). The third condition is related to the normal form analysis [17], where the nonlinear functions are approximated by Taylor

polynomials. Since the quadratic part of the potential function W is accounted for in the left-hand side of the system (2.1) and the cubic part of W violates the gauge transformation and analyticity assumptions, the Taylor polynomials of W start with quartic terms denoted as $O(4)$.

We find a general representation of the function $W(u,\bar{u},v,\bar{v})$ that satisfies the conditions 1–3 and list all possible (four-parameter) quartic terms of W.

Lemma 2.1. *If $W \in \mathbb{R}$ and condition 1 is satisfied, such that*

$$W(u,\bar{u},v,\bar{v}) = W\left(ue^{i\alpha}, \bar{u}e^{-i\alpha}, ve^{i\alpha}, \bar{v}e^{-i\alpha}\right) \qquad \forall \alpha \in \mathbb{R}, \tag{2.2}$$

then $W = W(|u|^2, |v|^2, \bar{u}v + u\bar{v})$.

Proof. The statement is a special case of Theorem 1.2 on page 450 of [18]. For the readers' convenience, we give a simplified proof based on the symmetry generator for the gauge transformation. By differentiating (2.2) in α and setting $\alpha = 0$, we derive the relation on $W \in \mathbb{R}$:

$$DW \equiv i\left(u\frac{\partial}{\partial u} - \bar{u}\frac{\partial}{\partial \bar{u}} + v\frac{\partial}{\partial v} - \bar{v}\frac{\partial}{\partial \bar{v}}\right) W(u,\bar{u},v,\bar{v}) = 0. \tag{2.3}$$

Consider the set of quadratic variables

$$z_1 = |u|^2, \quad z_2 = |v|^2, \quad z_3 = \bar{u}v + u\bar{v}, \quad z_4 = u^2 + v^2,$$

which is independent for any $u \neq 0$ and $v \neq 0$ in the sense that the Jacobian is nonzero. It is clear that $Dz_{1,2,3} = 0$ and $Dz_4 = 2iz_4$. Therefore, $DW = 2iz_4\partial_{z_4}W = 0$, such that $W = W(z_1, z_2, z_3)$. ■

Lemma 2.2. *If $W \in \mathbb{R}$ and conditions 1–3 are satisfied, then $W = W(|u|^2 + |v|^2, |u|^2|v|^2, u\bar{v} + v\bar{u})$.*

Proof. By Lemma 2.1 and condition 2, we can reorder the arguments of W as $W = W(|u| + |v|, |u||v|, u\bar{v} + v\bar{u})$. By analyticity in condition (3), W may depend only on $|u|^2$ and $|v|^2$ rather than on $|u|$ and $|v|$. ■

Corollary 2.3. *The only quartic potential function $W \in \mathbb{R}$ that satisfies conditions 1–3 is given by*

$$W = \frac{a_1}{2}(|u|^4 + |v|^4) + a_2|u|^2|v|^2 + a_3(|u|^2 + |v|^2)(v\bar{u} + \bar{v}u) + \frac{a_4}{2}(v\bar{u} + \bar{v}u)^2, \tag{2.4}$$

where (a_1, a_2, a_3, a_4) are real-valued parameters. It follows then that

$$\begin{cases} \partial_{\bar{u}}W = a_1|u|^2u + a_2u|v|^2 + a_3\left[(2|u|^2 + |v|^2)v + u^2\bar{v}\right] + a_4\left[v^2\bar{u} + |v|^2u\right], \\ \partial_{\bar{v}}W = a_1|v|^2v + a_2v|u|^2 + a_3\left[(2|v|^2 + |u|^2)u + v^2\bar{u}\right] + a_4\left[u^2\bar{v} + |u|^2v\right]. \end{cases}$$

The potential function (2.4) with $a_1, a_2 \neq 0$, and $a_3 = a_4 = 0$ represents a standard coupled-mode system for a subharmonic resonance, e.g., in the context of optical gratings with constant Kerr nonlinearity [1]. When $a_1 = a_3 = a_4 = 0$, this system is integrable with inverse scattering and is referred to as the massive Thirring model [19]. When $a_1 = a_2 = 0$ and $a_3, a_4 \neq 0$, the coupled-mode system corresponds to an optical grating with varying,

mean-zero Kerr nonlinearity, where a_3 is the Fourier coefficient of the resonant subharmonic and a_4 is the Fourier coefficient of the nonresonant harmonic [5] (see also [4]).

We rewrite the coupled-mode system (2.1) as a Hamiltonian system in complex-valued matrix-vector notation:

$$\frac{d\mathbf{u}}{dt} = J\nabla H(\mathbf{u}), \tag{2.5}$$

where $\mathbf{u} = (u, \bar{u}, v, \bar{v})^T$,

$$J = \begin{bmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix} = -J^T,$$

and $H(u, \bar{u}, v, \bar{v}) = \int_{\mathbb{R}} h(u, \bar{u}, v, \bar{v})dx$ is the Hamiltonian functional with the density

$$h = W(u, \bar{u}, v, \bar{v}) - (v\bar{u} + u\bar{v}) + \frac{i}{2}(u\bar{u}_x - u_x\bar{u}) - \frac{i}{2}(v\bar{v}_x - v_x\bar{v}).$$

The Hamiltonian $H(u, \bar{u}, v, \bar{v})$ is constant in time $t \geq 0$. Due to the gauge invariance, the coupled-mode system (2.1) has another constant of motion $Q(u, \bar{u}, v, \bar{v})$, where

$$Q = \int_{\mathbb{R}} \left(|u|^2 + |v|^2 \right) dx. \tag{2.6}$$

Conservation of Q can be checked by direct computation:

$$\frac{\partial}{\partial t}(|u|^2 + |v|^2) + \frac{\partial}{\partial x}(|u|^2 - |v|^2) = DW = 0, \tag{2.7}$$

where the operator D is defined in (2.3). Due to the translational invariance, the coupled-mode system (2.1) has yet another constant of motion $P(u, \bar{u}, v, \bar{v})$, where

$$P = \frac{i}{2} \int_{\mathbb{R}} \left(u\bar{u}_x - u_x\bar{u} + v\bar{v}_x - v_x\bar{v} \right) dx. \tag{2.8}$$

In applications, the quantities Q and P are referred to as the power and momentum of the coupled-mode system.

3. Existence of gap solitons.

Stationary solutions of the coupled-mode system (2.1) take the form

$$\begin{cases} u_{\rm st}(x,t) = u_0(x+s)e^{i\omega t + i\theta}, \\ v_{\rm st}(x,t) = v_0(x+s)e^{i\omega t + i\theta}, \end{cases} \tag{3.1}$$

where $(s, \theta) \in \mathbb{R}^2$ are arbitrary parameters, while the solution $(u_0, v_0) \in \mathbb{C}^2$ on $x \in \mathbb{R}$ and the domain for parameter $\omega \in \mathbb{R}$ are to be found from the nonlinear ODE system

$$\begin{cases} iu_0' = \omega u_0 - v_0 + \partial_{\bar{u}_0} W(u_0, \bar{u}_0, v_0, \bar{v}_0), \\ -iv_0' = \omega v_0 - u_0 + \partial_{\bar{v}_0} W(u_0, \bar{u}_0, v_0, \bar{v}_0). \end{cases} \tag{3.2}$$

Stationary solutions are critical points of the Lyapunov functional

$$\Lambda = H(u,\bar{u},v,\bar{v}) + \omega Q(u,\bar{u},v,\bar{v}), \tag{3.3}$$

such that variations of Λ produce the nonlinear ODE system (3.2).

Lemma 3.1. *Assume that there exists a decaying solution (u_0, v_0) of the system (3.2) on $x \in \mathbb{R}$. If $W \in \mathbb{R}$ and conditions 1–3 are satisfied, then $u_0 = \bar{v}_0$ (modulo to an arbitrary phase).*

Proof. It follows from the balance equation (2.7) for the stationary solutions (3.1) that

$$|u_0|^2 - |v_0|^2 = C_0 = 0 \qquad \forall x \in \mathbb{R},$$

where the constant $C_0 = 0$ is found from decaying conditions at infinity. Let us represent the solutions (u_0, v_0) in the form

$$\begin{cases} u_0(x) = \sqrt{Q(x)} e^{i\Theta(x)+i\Phi(x)}, \\ v_0(x) = \sqrt{Q(x)} e^{-i\Theta(x)+i\Phi(x)}, \end{cases} \tag{3.4}$$

such that

$$\begin{cases} iQ' - 2Q(\Theta' + \Phi') = 2\omega Q - 2Qe^{-2i\Theta} + 2\bar{u}_0 \partial_{\bar{u}_0} W(u_0,\bar{u}_0,v_0,\bar{v}_0), \\ -iQ' - 2Q(\Theta' - \Phi') = 2\omega Q - 2Qe^{2i\Theta} + 2\bar{v}_0 \partial_{\bar{v}_0} W(u_0,\bar{u}_0,v_0,\bar{v}_0). \end{cases} \tag{3.5}$$

Separating the real parts, we obtain

$$\begin{cases} Q(\cos(2\Theta) - \omega - \Theta' - \Phi') = \mathrm{Re}\left[\bar{u}_0 \partial_{\bar{u}_0} W(u_0,\bar{u}_0,v_0,\bar{v}_0)\right], \\ Q(\cos(2\Theta) - \omega - \Theta' + \Phi') = \mathrm{Re}\left[\bar{v}_0 \partial_{\bar{v}_0} W(u_0,\bar{u}_0,v_0,\bar{v}_0)\right]. \end{cases} \tag{3.6}$$

It follows from Lemma 2.2 that

$$\left(u\frac{\partial}{\partial u} + \bar{u}\frac{\partial}{\partial \bar{u}} - v\frac{\partial}{\partial v} - \bar{v}\frac{\partial}{\partial \bar{v}} \right) W(u,\bar{u},v,\bar{v}) \Big|_{|u|^2=|v|^2} = 0. \tag{3.7}$$

As a result, we have $\Phi' \equiv 0$, such that $\Phi(x) = \Phi_0$. ∎

Corollary 3.2. *Let $u_0 = \bar{v}_0$. The ODE system (3.2) reduces to the planar Hamiltonian form*

$$\frac{d}{dx}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix} \nabla h(p,q), \tag{3.8}$$

where $p = 2\Theta$, $q = Q$, and

$$h = \tilde{W}(p,q) - 2q\cos p + 2\omega q, \qquad \tilde{W}(p,q) = W(u_0,\bar{u}_0,v_0,\bar{v}_0). \tag{3.9}$$

Proof. In variables (Q,Θ) defined by (3.4) with $\Phi(x) = \Phi_0 \equiv 0$, we rewrite the ODE system (3.5) as follows:

$$\begin{cases} Q' = 2Q\sin(2\Theta) + 2\mathrm{Im}\left[\bar{u}_0 \partial_{\bar{u}_0} W(u_0,\bar{u}_0,v_0,\bar{v}_0)\right], \\ Q\Theta' = -\omega Q + Q\cos(2\Theta) - \mathrm{Re}\left[\bar{u}_0 \partial_{\bar{u}_0} W(u_0,\bar{u}_0,v_0,\bar{v}_0)\right]. \end{cases} \tag{3.10}$$

The system (3.10) is equivalent to the Hamiltonian system (3.8)–(3.9) if

$$\left\{ \begin{array}{l} \partial_p \tilde{W}(p,q) = i \left[u_0 \partial_{u_0} - \bar{u}_0 \partial_{\bar{u}_0} \right] W(u_0, \bar{u}_0, v_0, \bar{v}_0), \\ q \partial_q \tilde{W}(p,q) = \left[u_0 \partial_{u_0} + \bar{u}_0 \partial_{\bar{u}_0} \right] W(u_0, \bar{u}_0, v_0, \bar{v}_0). \end{array} \right. \tag{3.11}$$

The latter equation follows from (2.3), (3.4), and (3.7) with the chain rule. ■

Remark 3.3. The family of *stationary* solutions (3.1) can be extended to the family of *traveling* solutions of the coupled-mode system (2.1) by means of the Lorentz transformation [15]. When the boosted variables are applied to the form (3.1),

$$X = \frac{x - ct}{\sqrt{1-c^2}}, \quad T = \frac{t - cx}{\sqrt{1-c^2}}, \quad U = \left(\frac{1-c}{1+c} \right)^{1/4} u, \quad V = \left(\frac{1+c}{1-c} \right)^{1/4} v,$$

where $c \in (-1,1)$, the family of traveling solutions (U_0, V_0) satisfies the constraint $|U_0|^2 = |V_0|^2$ from the balance equation (2.7). However, the representation (3.4) results no longer in the relation $U_0 = \bar{V}_0$, since the relation (3.7) fails for the potential function W in boosted variables $(U, \bar{U}, V, \bar{V})$.

Decaying solutions of the system (3.2) with a homogeneous polynomial function $W(u, \bar{u}, v, \bar{v})$ are analyzed in the appendix. Conditions for their existence are identified for the quartic potential function (2.4). Decaying solutions may exist in the gap of the continuous spectrum of the coupled-mode system (2.1) for $\omega \in (-1, 1)$. We introduce two auxiliary parameters

$$\mu = \frac{1-\omega}{1+\omega}, \qquad \beta = \sqrt{1-\omega^2}, \tag{3.12}$$

such that $0 < \mu < \infty$ and $0 < \beta \leq 1$. When $a_1 = 1$, $a_2 = \rho \in \mathbb{R}$, and $a_3 = a_4 = 0$, we obtain in the appendix the decaying solution $u_0(x)$ in the explicit form

$$u_0 = \sqrt{\frac{2(1-\omega)}{1+\rho}} \frac{1}{(\cosh \beta x + i \sqrt{\mu} \sinh \beta x)}. \tag{3.13}$$

When $\omega \to 1$ (such that $\mu \to 0$ and $\beta \to 0$), the decaying solution (3.13) becomes small in absolute value and approaches the limit of sech-solutions $\mathrm{sech}(\beta x)$. When $\omega \to -1$ (such that $\mu \to \infty$ and $\beta \to 0$), the decaying solution (3.13) remains finite in absolute value and approaches the limit of the algebraically decaying solution:

$$u_0 = \frac{2}{\sqrt{1+\rho}(1+2ix)}.$$

When $a_1 = a_2 = 0, a_3 = 1$, and $a_4 = s \in \mathbb{R}$, the decaying solution $u_0(x)$ exists in two subdomains: $0 < \omega < 1$, $s > -1$ and $-1 < \omega < 0$, $s < 1$. When $0 < \omega < 1$, $s > -1$, the solution takes the form

$$u_0 = \sqrt{\frac{1-\omega}{2}} \frac{(\cosh \beta x - i \sqrt{\mu} \sinh \beta x)}{\sqrt{\Delta_+(x)}}, \tag{3.14}$$

where

$$\Delta_+ = [(s-1)\mu^2 - 2s\mu + (s+1)] \cosh^4(\beta x) + 2[s\mu - (s-1)\mu^2] \cosh^2(\beta x) + (s-1)\mu^2.$$

When $-1 < \omega < 0$, $s < 1$, the solution takes the form

$$u_0 = \sqrt{\frac{1-\omega}{2}} \frac{(\sinh \beta x - i\sqrt{\mu} \cosh \beta x)}{\sqrt{\Delta_-(x)}}, \tag{3.15}$$

where

$$\Delta_- = [(s+1) - 2s\mu - (s-1)\mu^2] \cosh^4(\beta x) + 2[s+1-s\mu] \cosh^2(\beta x) - (s+1).$$

In both limits $\omega \to 1$ and $\omega \to -1$, the decaying solutions (3.14) and (3.15) approach the small-amplitude sech-solution $\text{sech}(\beta x)$. In the limit $\omega \to 0$, the decaying solutions (3.14) and (3.15) degenerate into a nondecaying bounded solution with $|u_0(x)|^2 = \frac{1}{2}$.

4. Block-diagonalization of the linearized system. *Linearization* of the coupled-mode system (2.1) at the stationary solutions (3.1) with $s = \theta = 0$ is defined as follows:

$$\begin{cases} u(x,t) = e^{i\omega t}\left(u_0(x) + U_1(x)e^{\lambda t}\right), \\ \bar{u}(x,t) = e^{-i\omega t}\left(\bar{u}_0(x) + U_2(x)e^{\lambda t}\right), \\ v(x,t) = e^{i\omega t}\left(v_0(x) + U_3(x)e^{\lambda t}\right), \\ \bar{v}(x,t) = e^{-i\omega t}\left(\bar{v}_0(x) + U_4(x)e^{\lambda t}\right), \end{cases} \tag{4.1}$$

where $v_0 = \bar{u}_0$, according to Lemma 3.1. Let $(\mathbf{f}, \mathbf{g})$ be a standard inner product for $\mathbf{f}, \mathbf{g} \in L^2(\mathbb{R}, \mathbb{C}^4)$. Expanding the Lyapunov functional (3.3) into Taylor series near $\mathbf{u}_0 = (u_0, \bar{u}_0, v_0, \bar{v}_0)^T$, we have

$$\Lambda = \Lambda(\mathbf{u_0}) + (\mathbf{U}, \nabla\Lambda|_{\mathbf{u}_0}) + \frac{1}{2}(\mathbf{U}, H_\omega \mathbf{U}) + \cdots, \tag{4.2}$$

where $\mathbf{U} = (U_1, U_2, U_3, U_4)^T$, $\nabla\Lambda|_{\mathbf{u}_0} = 0$, and H_ω is the linearized energy operator in the explicit form

$$H_\omega = D(\partial_x) + V(x), \tag{4.3}$$

where

$$D = \begin{pmatrix} \omega - i\partial_x & 0 & -1 & 0 \\ 0 & \omega + i\partial_x & 0 & -1 \\ -1 & 0 & \omega + i\partial_x & 0 \\ 0 & -1 & 0 & \omega - i\partial_x \end{pmatrix} \tag{4.4}$$

and

$$V = \begin{pmatrix} \partial^2_{\bar{u}_0 u_0} & \partial^2_{\bar{u}_0^2} & \partial^2_{\bar{u}_0 v_0} & \partial^2_{\bar{u}_0 \bar{v}_0} \\ \partial^2_{u_0^2} & \partial^2_{u_0 \bar{u}_0} & \partial^2_{u_0 v_0} & \partial^2_{u_0 \bar{v}_0} \\ \partial^2_{\bar{v}_0 u_0} & \partial^2_{\bar{v}_0 \bar{u}_0} & \partial^2_{\bar{v}_0 v_0} & \partial^2_{\bar{v}_0^2} \\ \partial^2_{v_0 u_0} & \partial^2_{v_0 \bar{u}_0} & \partial^2_{v_0^2} & \partial^2_{v_0 \bar{v}_0} \end{pmatrix} W(u_0, \bar{u}_0, v_0, \bar{v}_0). \tag{4.5}$$

The linearization (4.1) reduces the nonlinear coupled-mode system (2.1) to the linear eigenvalue problem in the form

$$H_\omega \mathbf{U} = i\lambda\sigma\mathbf{U}, \tag{4.6}$$

where σ is a diagonal matrix of $(1,-1,1,-1)$. Due to the gauge and translational symmetries, the energy operator H_ω has a two-dimensional kernel with the eigenvectors:

$$\mathbf{U}_1 = \sigma\mathbf{u}_0(x), \qquad \mathbf{U}_2 = \mathbf{u}_0'(x). \tag{4.7}$$

The eigenvectors $\mathbf{U}_{1,2}$ represent derivatives of the stationary solutions (3.1) with respect to parameters (θ, s).

Due to the Hamiltonian structure, the linearized operator σH_ω has at least a four-dimensional generalized kernel with the eigenvectors (4.7) and two generalized eigenvectors (see [20] for details). The eigenvectors of the linearized operator σH_ω satisfy the σ-orthogonality constraints

$$(\sigma\mathbf{u}_0, \sigma\mathbf{U}) = \int_{\mathbb{R}} (\bar{u}_0 U_1 + u_0 U_2 + \bar{v}_0 U_3 + v_0 U_4)\, dx = 0, \tag{4.8}$$

$$(\mathbf{u}_0', \sigma\mathbf{U}) = \int_{\mathbb{R}} (\bar{u}_0' U_1 - u_0' U_2 + \bar{v}_0' U_3 - v_0' U_4)\, dx = 0. \tag{4.9}$$

The constraints (4.8) and (4.9) represent first variations of the conserved quantities Q and P in (2.6) and (2.8) at the linearization (4.1).

When the constraint $u_0 = \bar{v}_0$ holds, the potential part (4.5) has additional symmetry relations:

$$\partial^2_{u_0\bar{u}_0} W = \partial^2_{v_0\bar{v}_0} W, \quad \partial^2_{\bar{u}_0^2} W = \partial^2_{v_0^2} W, \quad \partial^2_{u_0 v_0} W = \partial^2_{\bar{u}_0\bar{v}_0} W. \tag{4.10}$$

It follows from the explicit form of H_ω and the relations (4.10) that the eigenvalue problem $H_\omega \mathbf{U} = \mu\mathbf{U}$ has two reductions:

$$\text{(i) } U_1 = U_4,\ U_2 = U_3, \qquad \text{(ii) } U_1 = -U_4,\ U_2 = -U_3. \tag{4.11}$$

Our main result on the block-diagonalization of the energy operator H_ω and the linearized operator σH_ω is based on the reductions (4.11).

Theorem 4.1. *Let $W \in \mathbb{R}$ and conditions 1–3 are satisfied. Let (u_0, v_0) be a decaying solution of the system (3.2) on $x \in \mathbb{R}$ with the constraint $v_0 = \bar{u}_0$. There exists an orthogonal similarity transformation S, such that $S^{-1} = S^T$, where*

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{pmatrix},$$

that simultaneously block-diagonalizes the energy operator H_ω,

$$S^{-1} H_\omega S = \begin{pmatrix} H_+ & 0 \\ 0 & H_- \end{pmatrix} \equiv H, \tag{4.12}$$

and the linearized operator σH_ω

$$S^{-1}\sigma H_\omega S = \sigma \begin{pmatrix} 0 & H_- \\ H_+ & 0 \end{pmatrix} \equiv iL, \tag{4.13}$$

where $H_\pm$ *are two-by-two Dirac operators:*

$$H_\pm = \begin{pmatrix} \omega - i\partial_x & \mp 1 \\ \mp 1 & \omega + i\partial_x \end{pmatrix} + V_\pm(x) \tag{4.14}$$

and

$$V_\pm = \begin{pmatrix} \partial^2_{\bar{u}_0 u_0} \pm \partial^2_{\bar{u}_0 \bar{v}_0} & \partial^2_{\bar{u}_0^2} \pm \partial^2_{\bar{u}_0 v_0} \\ \partial^2_{u_0^2} \pm \partial^2_{u_0 \bar{v}_0} & \partial^2_{\bar{u}_0 u_0} \pm \partial^2_{u_0 v_0} \end{pmatrix} W(u_0, \bar{u}_0, v_0, \bar{v}_0). \tag{4.15}$$

Proof. Applying the similarity transformation to the operator $D(\partial_x)$ in (4.4), we have the first term in Dirac operators $H_\pm$. Applying the same transformation to the potential $V(x)$ in (4.5) and using the relations (4.10), we have the second term in Dirac operators $H_\pm$. The same transformation is applied similarly to the linearized operator σH_ω with the result (4.13). ■

Corollary 4.2. (a) *The coupled eigenvalue problem* (4.6) *is equivalent to the block-diagonalized eigenvalue problems*

$$\sigma_3 H_- \sigma_3 H_+ \mathbf{V}_1 = \gamma \mathbf{V}_1, \qquad \sigma_3 H_+ \sigma_3 H_- \mathbf{V}_2 = \gamma \mathbf{V}_2, \qquad \gamma = -\lambda^2, \tag{4.16}$$

where $\mathbf{V}_{1,2} \in \mathbb{C}^2$ *and* σ_3 *is Pauli's diagonal matrix of* $(1, -1)$.

(b) *Let* $\mathbf{u}_0 = (u_0, \bar{u}_0) \in \mathbb{C}^2$ *and* $(\mathbf{f}, \mathbf{g})$ *be a standard inner product for* $\mathbf{f}, \mathbf{g} \in L^2(\mathbb{R}, \mathbb{C}^2)$. *Dirac operators* $H_\pm$ *have simple kernels with the eigenvectors*

$$H_+ \mathbf{u}_0' = 0, \qquad H_- \sigma_3 \mathbf{u}_0 = 0, \tag{4.17}$$

while the vectors $\mathbf{V}_{1,2}$ *satisfy the constraints*

$$(\sigma_3 \mathbf{u}_0, \sigma_3 \mathbf{V}_1) = 0, \qquad (\mathbf{u}_0', \sigma_3 \mathbf{V}_2) = 0. \tag{4.18}$$

Remark 4.3. Block-diagonalization described in Theorem 4.1 has nothing in common with the explicit diagonalization used in the reduction (9.2) of [14] for the particular potential function (2.4) with $a_1 = a_2 = a_4 = 0$ and $a_3 = 1$. Moreover, the reduction (9.2) of [14] does not work for $\omega \neq 0$, while gap solitons do not exist in this particular model for $\omega = 0$.

We apply Theorem 4.1 to the linearized coupled-mode system with the quartic potential function (2.4). When $a_1 = 1$, $a_2 = \rho$, and $a_3 = a_4 = 0$, the decaying solution $u_0(x)$ is given by (3.13) and the potential matrices $V_\pm(x)$ in the Dirac operators $H_\pm$ in (4.14)–(4.15) are found in the explicit form

$$V_+ = (1+\rho) \begin{pmatrix} 2|u_0|^2 & u_0^2 \\ \bar{u}_0^2 & 2|u_0|^2 \end{pmatrix}, \qquad V_- = \begin{pmatrix} 2|u_0|^2 & (1-\rho)u_0^2 \\ (1-\rho)\bar{u}_0^2 & 2|u_0|^2 \end{pmatrix}. \tag{4.19}$$

When $a_1 = a_2 = 0$, $a_3 = 1$, and $a_4 = s$, the decaying solution $u_0(x)$ is given by either (3.14) or (3.15) and the potential matrices $V_\pm(x)$ take the form

$$V_+ = 3\begin{pmatrix} u_0^2 + \bar{u}_0^2 & 2|u_0|^2 \\ 2|u_0|^2 & u_0^2 + \bar{u}_0^2 \end{pmatrix} + s\begin{pmatrix} 2|u_0|^2 & u_0^2 + 3\bar{u}_0^2 \\ \bar{u}_0^2 + 3u_0^2 & 2|u_0|^2 \end{pmatrix}, \tag{4.20}$$

$$V_- = \begin{pmatrix} u_0^2 + \bar{u}_0^2 & -2|u_0|^2 \\ -2|u_0|^2 & u_0^2 + \bar{u}_0^2 \end{pmatrix} + s\begin{pmatrix} 0 & -u_0^2 - \bar{u}_0^2 \\ -u_0^2 - \bar{u}_0^2 & 0 \end{pmatrix}. \tag{4.21}$$

Numerical computations of eigenvalues of the Dirac operators $H_\pm$ and the linearized operator L in (4.12) and (4.13) are developed for the explicit examples (4.19) and (4.20)–(4.21).

5. Numerical computations of eigenvalues. Numerical discretization and truncation of the linearized system (4.6) leads to an eigenvalue problem for large matrices [21]. Parallel software libraries were recently developed for computations of large eigenvalue problems [22]. We shall use the Scalapack library and distribute computations of eigenvalues of the system (4.6) for different parameter values between parallel processors of the SHARCnet cluster Idra using Message Passing Interface.[1]

We implement a numerical discretization of the linearized system (4.6) using the Chebyshev interpolation method [23]. Given a function $u(z)$ defined on the discrete grid of Chebyshev points $z_j = \cos(j\pi/N), j = 0, 1 \ldots N$, we obtain a discretization of the first derivative $u'(z)$ as a multiplication of the vector for values of $u(z)$ on the discrete grid by an $(N+1)$-by-$(N+1)$ matrix, which we denote by $D_N^{(1)}$. If the rows and columns of the differentiation matrix $D_N^{(1)}$ are indexed from 0 to N, the entries of $D_N^{(1)}$ are (see [24] for details)

$$(D_N^{(1)})_{00} = \frac{2N^2+1}{6}, \quad (D_N^{(1)})_{NN} = -\frac{2N^2+1}{6},$$
$$(D_N^{(1)})_{jj} = \frac{-z_j}{2(1-z_j^2)}, \quad j = 1, \ldots, N-1,$$

and

$$(D_N^{(1)})_{ij} = \frac{c_i}{c_j}\frac{(-1)^{i+j}}{(z_i - z_j)}, \quad i \neq j, \quad i, j = 0, \ldots, N,$$

where $c_0 = c_N = 2$ and $c_i = 1$, $i = 1, \ldots, N-1$. To transform the Chebyshev grid from the interval $z \in [-1, 1]$ to the infinite domain $x \in \mathbb{R}$ we will use the exponential map $f(z) = L \tanh^{-1} z$, such that $x_j = f(z_j)$, $j = 0, 1, \ldots, N$. This map is efficient in our case because the potential matrices $V_\pm(x)$ decay exponentially as $|x| \to \infty$. The constant L sets the length scale of the map and we pick up the values of L such that the localization of matrix potentials $V_\pm(x)$ has a sufficient resolution on the discrete grid points.

Using the chain rule, we represent differentiation of $u(x)$ on the discrete grid with the matrix multiplication

$$\mathbf{p} = \left[\left(\frac{\partial f^{-1}(x_i)}{\partial x} D_N^{(1)}\right) u(z_j), j = 0, 1, \ldots, N\right] \equiv D_{N+1}\mathbf{u},$$

[1] Cluster Idra is a part of the SHARCnet network of parallel processors distributed between eight universities in southern Ontario, including McMaster University.

Table 1

Maximum real part of eigenvalues $M = \max|\mathrm{Re}(\lambda)|$ versus the number of Chebyshev polynomials N for two computational intervals $|\mathrm{Im}(\lambda)| < 2$ and $|\mathrm{Im}(\lambda)| < 10$.

N	$M_{\lvert\mathrm{Im}\lambda\rvert<2}$	$M_{\lvert\mathrm{Im}\lambda\rvert<10}$
100	0.085	0.75
200	0.0095	0.52
400	0.0053	0.21
800	$7.12 \cdot 10^{-4}$	0.12
1200	$2.34 \cdot 10^{-4}$	0.09
2500	$3.91 \cdot 10^{-5}$	0.06

where $\mathbf{u}$ is the vector for values of $u(x)$ and $\mathbf{p}$ is the vector for values of $u'(x)$ on the discrete grid. The discretization of the Dirac operators $H_\pm$ is defined by

$$H_\pm = \begin{pmatrix} \omega I_{N+1} - iD_{N+1} & \mp I_{N+1} \\ \mp I_{N+1} & \omega I_{N+1} + iD_{N+1} \end{pmatrix} + \mathrm{diag} V_\pm(x_i), \tag{5.1}$$

where I_{N+1} is the identity $(N+1)$-by-$(N+1)$ matrix.

The continuous spectrum for the linearized coupled-mode system (4.6) can be found from the no-potential case $V(x) \equiv 0$. It consists of two pairs of symmetric branches on the imaginary axis $\lambda \in i\mathbb{R}$ for $|\mathrm{Im}(\lambda)| > 1 - \omega$ and $|\mathrm{Im}(\lambda)| > 1 + \omega$ [7, 15]. In the potential case $V(x) \neq 0$, the continuous spectrum does not move, but the discrete spectrum appears. The discrete spectrum is represented by symmetric pairs or quartets of isolated nonzero eigenvalues and a zero eigenvalue of algebraic multiplicity four for the generalized kernel of σH_ω [7, 15]. We note that symmetries of the Chebyshev grid preserve symmetries of the linearized coupled-mode system (4.6).

The main advantage of the Chebyshev grid is the clustering distribution of the grid points that enables us to control spurious complex eigenvalues. If the eigenvector is analytic in a strip near the interpolation interval, the corresponding Chebyshev spectral derivatives converge geometrically, with an asymptotic convergence factor determined by the size of the largest ellipse in the domain of analyticity [23].

Spurious complex eigenvalues arise from the discretization of the continuous spectrum. When the number of Chebyshev polynomials increases, the real parts of spurious eigenvalues get smaller. Convergence of real parts of eigenvalues to zero is better near the end points of the continuous spectrum $\lambda = \pm i(1-\omega)$ and $\lambda = \pm i(1+\omega)$, from which bifurcations of true unstable eigenvalues are expected to occur (due to analytical results in [7, 13] and numerical results in [7, 15]). Table 1 shows the maximum real part $M = \max|\mathrm{Re}(\lambda)|$ versus N in two computational intervals $|\mathrm{Im}(\lambda)| < 2$ and $|\mathrm{Im}(\lambda)| < 10$ for the linear eigenvalue problem (4.6) with no true unstable eigenvalues. When $N = 2500$, the real parts of spurious eigenvalues in the interval $|\mathrm{Im}(\lambda)| < 2$ are of the order of 10^{-5}. Using more polynomials, we can make the real parts of the eigenvalues of continuous spectrum negligibly small, so that edge bifurcations of unstable eigenvalues can be studied numerically within any required accuracy.

We compute eigenvalues of the energy operator H_ω and the linearized operator σH_ω. It is well known [21, 23] that Hermitian matrices have condition number one, while non-Hermitian matrices may have a large condition number. As a result, numerical computations

Table 2

CPU time T for block-diagonal and full matrices versus the number of Chebyshev points N.

N	T_{block}	T_{full}
100	1.656	1.984
200	11.219	12.921
400	130.953	207.134
800	997.843	$1.583 \cdot 10^3$
1200	$3.608 \cdot 10^3$	$6.167 \cdot 10^3$
2500	$7.252 \cdot 10^3$	$12.723 \cdot 10^3$

for eigenvalues and eigenvectors have better accuracy and faster convergence for self-adjoint operators [21, 23]. We will use the block-diagonalizations (4.12) and (4.13) and compute eigenvalues of H_+, H_-, and L. The block-diagonalized matrix can be stored in a special format which requires two times less memory than a full matrix and it accelerates computations of eigenvalues approximately twice as much. Table 2 shows CPU time T for computations of eigenvalues of σH_ω for block-diagonal and full matrices versus the number of Chebyshev points N. When $N = 2500$, T_{full} is almost twice as large as T_{block}.

Figure 1 displays the pattern of eigenvalues and instability bifurcations for the symmetric quartic potential (2.4) with $a_1 = 1$ and $a_2 = a_3 = a_4 = 0$. The decaying solution $u_0(x)$ and the potential matrices $V_\pm(x)$ are given by (3.13) and (4.19) with $\rho = 0$. Parameter ω of the decaying solution $u_0(x)$ is defined in the interval $-1 < \omega < 1$. The six pictures of Figure 1 show the entire spectrum of L, H_+, and H_- for different values of ω (the continuous multimedia animations that show the transformation of eigenvalues when ω decreases are available as 62978_01.avi and 62978_02.avi).

When ω is close to 1 (the gap soliton is close to a small-amplitude sech-soliton), there exists a single nonzero eigenvalue for H_+ and H_- and a single pair of purely imaginary eigenvalues of L (see subplot (1) on Figure 1). The first set of arrays on the subplot (1) indicates that the pair of eigenvalues of L becomes visible at the same value of ω as the eigenvalue of H_+. This correlation between eigenvalues of L and H_+ can be traced throughout the entire parameter domain on the subplots (1)–(6).

When ω decreases, the operator H_- acquires another nonzero eigenvalue by means of the edge bifurcation [13], with no changes in the number of isolated eigenvalues of L (see subplot (2)). The first complex instability occurs near $\omega \approx -0.18$, when the pair of purely imaginary eigenvalues of L collides with the continuous spectrum and emerges as a quartet of complex eigenvalues, with no changes in the number of isolated eigenvalues for H_+ and H_- (see subplot (3)).

The second complex instability occurs at $\omega \approx -0.54$, when the operator H_- acquires a third nonzero eigenvalue and the linearized operator L acquires another quartet of complex eigenvalues (see subplot (4)). The second set of arrays on the subplots (4)–(6) indicates a correlation between these eigenvalues of L and H_-.

When ω decreases further, the operators H_+ and H_- acquire one more isolated eigenvalue, with no change in the spectrum of L (see subplot (5)). Finally, when ω is close to -1 (the gap soliton is close to the large-amplitude algebraic soliton), the third complex instability occurs, correlated with another edge bifurcation in the operator H_- (see subplot (6)). The

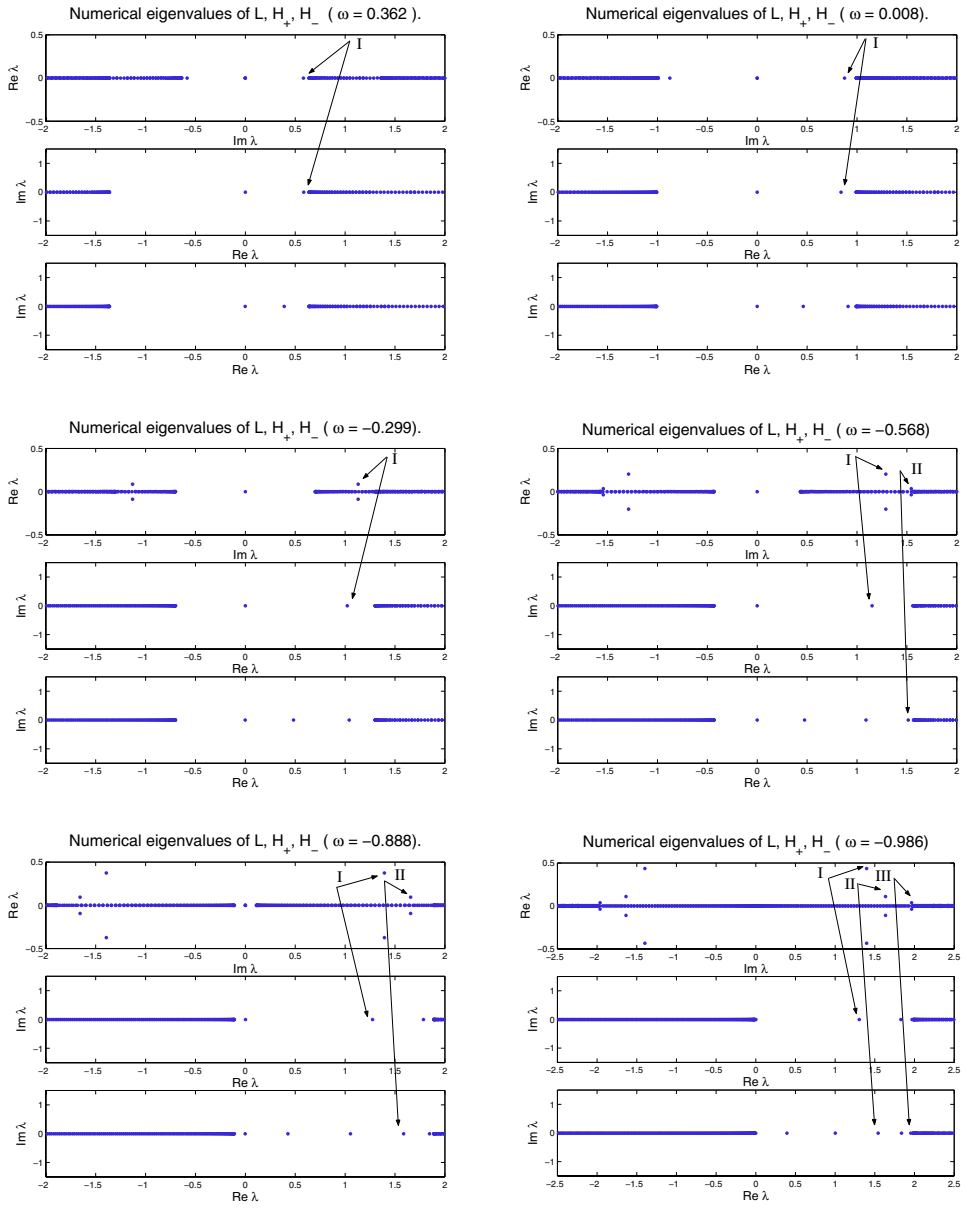


Figure 1. *Eigenvalues and instability bifurcations for the symmetric quartic potential* (2.4) *with* $a_1 = 1$ *and* $a_2 = a_3 = a_4 = 0$. *See also the accompanying animations* (62978_01.avi [724KB] *and* 62978_02.avi [275KB]).

third set of arrays on subplot (6) indicates this correlation. The third complex instability was not detected in the previous numerical studies of the same coupled-mode system [7, 15] (since the previous works did not consider eigenvalues of gap solitons near the limit $\omega = -1$). In a narrow domain near $\omega = -1$, the operator H_+ has two nonzero eigenvalues, the operator H_- has five nonzero eigenvalues, and the operator L has three quartets of complex eigenvalues.

Figure 2 displays the pattern of eigenvalues and instability bifurcations for the symmetric

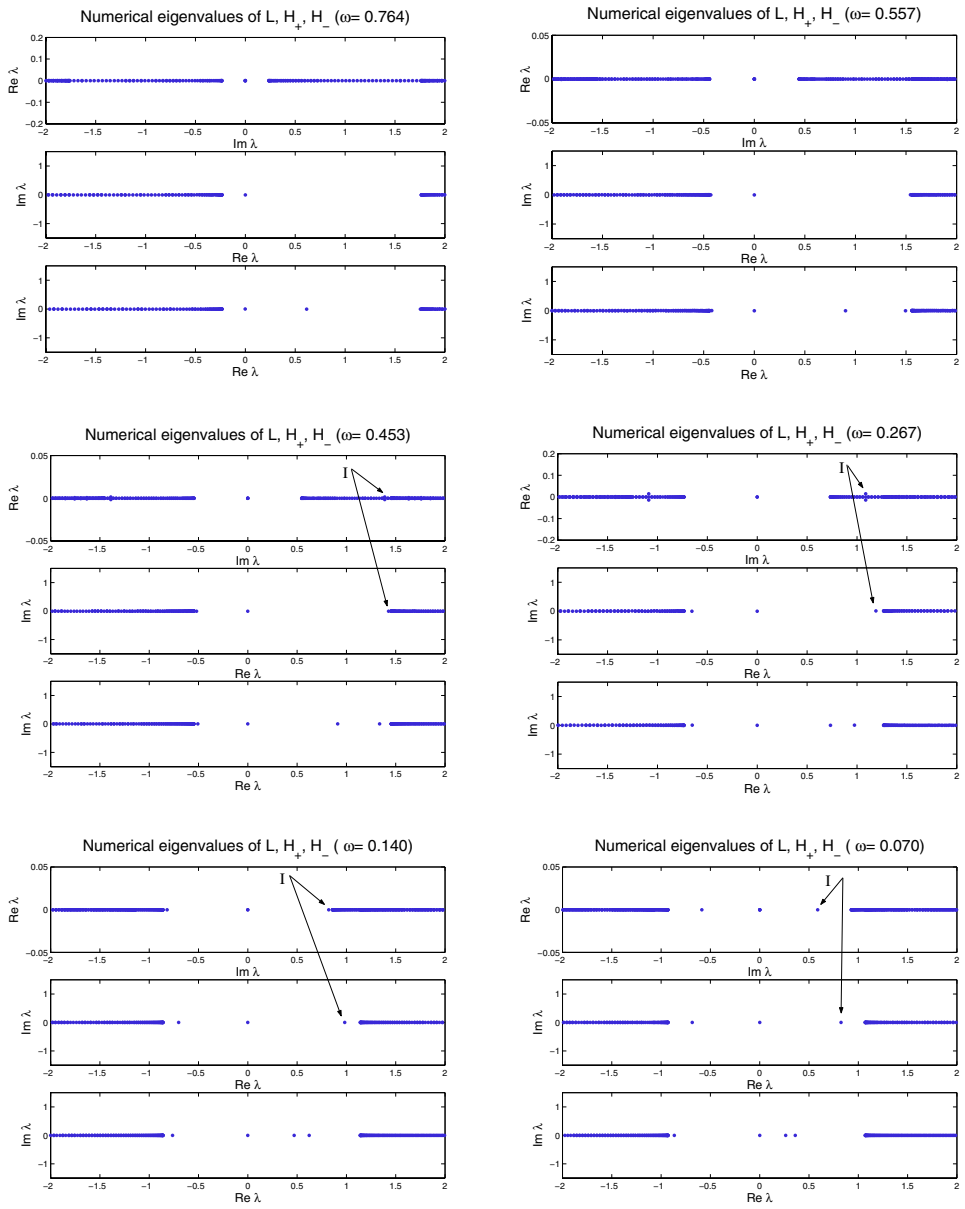


Figure 2. *Eigenvalues and instability bifurcations for the symmetric quartic potential* (2.4) *with* $a_3 = 1$ *and* $a_1 = a_2 = a_4 = 0$.

quartic potential (2.4) with $a_1 = a_2 = a_4 = 0$ and $a_3 = 1$. The decaying solution $u_0(x)$ and the potential matrices $V_\pm(x)$ are given by (3.14) and (4.20) with $0 < \omega < 1$ and $s = 0$. Eigenvalues in the other case $-1 < \omega < 0$ can be found from those in the case $0 < \omega < 1$ by reflections.

When ω is close to 1 (the gap soliton is close to a small-amplitude sech-soliton), there exist one nonzero eigenvalue of H_- and no nonzero eigenvalues of L and H_+ (see subplot

(1)). When ω decreases, two more nonzero eigenvalues bifurcate in H_- from the left and right branches of the continuous spectrum, with no change in the nonzero eigenvalues of L (see subplot (2)). The first complex bifurcation occurs at $\omega \approx 0.45$, when a quartet of complex eigenvalues occurs in L, in correlation with two symmetric edge bifurcations of H_+ from the left and right branches of the continuous spectrum (see subplots (3) and (4)). The first and only set of arrays on the subplots (3)–(6) indicates a correlation between eigenvalues of L and H_+, which is traced through the domain of ω. The inverse complex bifurcation occurs at $\omega \approx 0.15$, when the quartet of complex eigenvalues merges at the edge of the continuous spectrum into a pair of purely imaginary eigenvalues (see subplot (5)). No new eigenvalues emerge for smaller values of ω. When ω is close to 0 (the gap soliton is close to the nondecaying bounded solution), the operator H_+ has two nonzero eigenvalues, the operator H_- has three nonzero eigenvalues, and the operator L has one pair of purely imaginary eigenvalues (see subplot (6)).

We add remarks on two other limiting cases of the symmetric quartic potential (2.4). When $a_1 = a_3 = a_4 = 0$ and $a_2 = 1$, the coupled-mode system is an integrable model and no nonzero eigenvalues of L exist, according to the exact solution of the linearization problem [9]. When $a_1 = a_2 = a_3 = 0$ and $a_4 = \pm 1$, one branch of decaying solutions $u_0(x)$ exists for either sign, according to (3.14) and (3.15). The pattern of eigenvalues and instability bifurcations repeats those in Figure 2.

Our numerical results imply that there exists a correlation between edge bifurcations in the operator L and those in the Dirac operators H_+ and H_-. Analysis of such correlations is beyond the scope of the present paper.

Appendix. Conditions for existence of gap solitons in the homogeneous potential function.

We shall consider the homogeneous potential function $W \in \mathbb{R}$ of the monomial order $2n$ that satisfies conditions 1–3. The general representation of $W(u, \bar{u}, v, \bar{v})$ is given by

$$W = \sum_{s=0}^{n} \sum_{k=0}^{n-s} a_{k,s} \left(u^s \bar{v}^s + \bar{u}^s v^s \right) |u|^{2n-2k-2s} |v|^{2k}, \tag{A.1}$$

where $a_{k,s}$ are real-valued coefficients such that $a_{k_1,s} = a_{k_2,s}$ if $k_1 + k_2 = n - s$ for $s = 0, 1, \ldots, n-1$. Let us introduce new parameters

$$A_s = \sum_{k=0}^{n-s} a_{k,s}, \qquad s = 0, 1, \ldots, n.$$

Using the variables (Q, Θ) defined in (3.4) with $\Phi(x) = \Phi_0 \equiv 0$, we rewrite the ODE system (3.10) in the explicit form

$$\left\{ \begin{array}{c} Q' = 2Q \sin(2\Theta) - 2Q^n \sum_{s=0}^{n} s A_s \sin(2s\Theta), \\ \Theta' = -\omega + \cos(2\Theta) - nQ^{n-1} \sum_{s=0}^{n} A_s \cos(2s\Theta). \end{array} \right. \tag{A.2}$$

There exists a first integral of the system (A.2),

$$-\omega Q + \cos(2\Theta) Q - Q^n \sum_{s=0}^{n} A_s \cos(2s\Theta) = C_0,$$

where $C_0 = 0$ from the zero boundary conditions $Q(x) \to 0$ as $|x| \to \infty$. As a result, the second-order system (A.2) is reduced to the first-order ODE

$$\Theta'(x) = (n-1)(\omega - \cos(2\Theta)), \tag{A.3}$$

while the function $Q(x) \geq 0$ can be found from $\Theta(x)$ as follows:

$$Q^{n-1} = \frac{(\cos(2\Theta) - \omega)}{\sum_{s=0}^{n} A_s \cos(2s\Theta)}. \tag{A.4}$$

We consider the quartic potential function W given by (2.4). Using (A.3) for the case $n = 2$ we obtain

$$\Theta'(x) = \omega - \cos(2\Theta) \tag{A.5}$$

and the correspondence

$$A_0 = \frac{a_1 + a_2 + a_4}{2}, \quad A_1 = 2a_3, \quad A_2 = \frac{a_4}{2}.$$

We rewrite the representation (A.4) for $Q(x)$ as

$$Q = \frac{(t-\omega)}{\phi(t)}, \qquad Q \geq 0, \tag{A.6}$$

where

$$t = \cos(2\Theta), \qquad \phi(t) = a_4 t^2 + 2a_3 t + \frac{a_1 + a_2}{2},$$

such that $t \in [-1, 1]$. Let us consider two cases:

$$\begin{cases} t \geq \omega; & \phi(t) \geq 0 \quad \Rightarrow Q^+, \\ t \leq \omega; & \phi(t) \leq 0 \quad \Rightarrow Q^-. \end{cases} \tag{A.7}$$

We can solve the first-order ODE (A.5) using the substitution $z = \tan(\Theta)$, such that

$$t = \frac{1 - z^2}{1 + z^2}, \qquad z^2 = \frac{1 - t}{1 + t}.$$

After integration with the symmetry constraint $\Theta(0) = 0$, we obtain the solution

$$\left| \frac{(z - \sqrt{\mu})}{(z + \sqrt{\mu})} \right| = e^{2\beta x}, \tag{A.8}$$

where

$$\beta = \sqrt{1 - \omega^2}, \qquad \mu = \frac{1 - \omega}{1 + \omega},$$

and $-1 < \omega < 1$. Two separate cases are considered:

$$|z| \leq \sqrt{\mu}: \quad z = -\sqrt{\mu}\frac{\sinh(\beta x)}{\cosh(\beta x)}, \quad t = \frac{\cosh^2(\beta x) - \mu \sinh^2(\beta x)}{\cosh^2(\beta x) + \mu \sinh^2(\beta x)}, \tag{A.9}$$

where $t \geq \omega$, and

$$|z| \geq \sqrt{\mu}: \quad z = -\sqrt{\mu}\frac{\cosh(\beta x)}{\sinh(\beta x)}, \quad t = \frac{\sinh^2(\beta x) - \mu\cosh^2(\beta x)}{\sinh^2(\beta x) + \mu\cosh^2(\beta x)}, \tag{A.10}$$

where $t \leq \omega$. Let us introduce new parameters

$$\begin{aligned} A &= -2a_3 + a_4 + \frac{a_1 + a_2}{2}, \\ B &= -2a_4 + a_1 + a_2, \\ C &= 2a_3 + a_4 + \frac{a_1 + a_2}{2}. \end{aligned}$$

It is clear that $A = \phi(-1)$ and $C = \phi(1)$. If $t \geq \omega$ and $\phi(t) \geq 0$, it follows from (A.7) and (A.9) that

$$Q^+(x) = \frac{(1-\omega)((\mu+1)\cosh^2(\beta x) - \mu)}{(A\mu^2 + B\mu + C)\cosh^4(\beta x) - (B\mu + 2A\mu^2)\cosh^2(\beta x) + A\mu^2}. \tag{A.11}$$

If $t \leq \omega$ and $\phi(t) \leq 0$, it follows from (A.7) and (A.10) that

$$Q^-(x) = \frac{(\omega-1)((\mu+1)\cosh^2(\beta x) - 1)}{(A\mu^2 + B\mu + C)\cosh^4(\beta x) - (B\mu + 2C)\cosh^2(\beta x) + C}. \tag{A.12}$$

The asymptotic behavior of the function $Q(x)$ at infinity depends on the location of the zeros of the function $\psi(\mu) = A\mu^2 + B\mu + C$. The function $\psi(\mu)$ is related to the function $\phi(t)$; e.g., if $\psi(\mu) = 0$ then $\phi(\omega) = 0$.

A.1. Case $A < 0$, $C > 0$. The quadratic polynomial $\phi(t)$ has exactly one root $\phi(t_1) = 0$ such that $t_1 \in (-1, 1)$. Two branches of decaying solutions with the positive amplitude $Q(x)$ exist. One branch occurs for $t_1 < \omega \leq 1$ with $Q(x) = Q^+(x)$ and the other one occurs for $-1 \leq \omega < t_1$ with $Q(x) = Q^-(x)$. At the point $\omega = t_1$, the solution is bounded and decaying.

A.2. Case $A > 0$, $C > 0$. The quadratic polynomial $\phi(t)$ has no roots or has exactly two roots on $(-1, 1)$. If $\phi(t)$ does not have any roots on $(-1, 1)$, a decaying solution with the positive amplitude $Q(x)$ exists for any $-1 < \omega < 1$ with $Q(x) = Q^+(x)$. If $\phi(t)$ has two roots $\phi(t_1) = 0$ and $\phi(t_2) = 0$ such that $t_1, t_2 \in (-1, 1)$, a decaying solution with $Q(x) = Q^+(x)$ exists only on the interval $\max(t_1, t_2) < \omega \leq 1$. At the point $\omega = \max(t_1, t_2)$, the solution becomes bounded but nondecaying if $t_1 \neq t_2$ and unbounded if $t_1 = t_2$.

A.3. Case $A < 0$, $C < 0$. The quadratic polynomial $\phi(t)$ has no roots or has exactly two roots on $(-1, 1)$. If $\phi(t)$ does not have any roots on $(-1, 1)$, a decaying solution with the positive amplitude $Q(x)$ exists for any $-1 < \omega < 1$ with $Q(x) = Q^-(x)$. If $\phi(t)$ has two roots $\phi(t_1) = 0$ and $\phi(t_2) = 0$ such that $t_1, t_2 \in (-1, 1)$, a decaying solution with $Q(x) = Q^-(x)$ exists only on the interval $-1 \leq \omega < \min(t_1, t_2)$. At the point $\omega = \min(t_1, t_2)$, the solution becomes bounded but nondecaying if $t_1 \neq t_2$ and unbounded if $t_1 = t_2$.

A.4. Case $A > 0$, $C < 0$. No decaying solutions with positive amplitude $Q(x)$ exist.

A.5. Special cases. Two special cases occur when $\phi(1) = 0$ or $\phi(-1) = 0$. If $\phi(1) = 0$, then $Q^+(x)$ has a singularity at $x = 0$ for any $-1 < \omega < 1$. If $\phi(-1) = 0$, then $Q^-(x)$ has a singularity at $x = 0$ for any $-1 < \omega < 1$.

REFERENCES

[1] C. M. de Sterke and J. E. Sipe, *Gap solitons*, Progress in Optics, 33 (1994), pp. 203–260.

[2] Yu. S. Kivshar and G. P. Agrawal, *Optical Solitons: From Fibers to Photonic Crystals*, Academic Press, San Diego, 2003.

[3] L. Pitaevskii and S. Stringari, *Bose-Einstein Condensation*, International Series of Monographs on Physics 116, The Clarendon Press, Oxford University Press, Oxford, UK, 2003.

[4] C. M. de Sterke, D. G. Salinas, and J. E. Sipe, *Coupled-mode theory for light propagation through deep nonlinear gratings*, Phys. Rev. E, 54 (1996), pp. 1969–1989.

[5] D. Agueev and D. Pelinovsky, *Modeling of wave resonances in low-contrast photonic crystals*, SIAM J. Appl. Math., 65 (2005), pp. 1101–1129.

[6] D. E. Pelinovsky, A. A. Sukhorukov, and Yu. S. Kivshar, *Bifurcations and stability of gap solitons in periodic potentials*, Phys. Rev. E, 70 (2004), 036618.

[7] I. V. Barashenkov, D. E. Pelinovsky, and E. V. Zemlyanaya, *Vibrations and oscillatory instabilities of gap solitons*, Phys. Rev. Lett., 80 (1998), pp. 5117–5120.

[8] I. V. Barashenkov and E. V. Zemlyanaya, *Oscillatory instabilities of gap solitons: A numerical study*, Comput. Phys. Comm., 126 (2000), pp. 22–27.

[9] D. J. Kaup and T. I. Lakoba, *The squared eigenfunctions of the massive Thirring model in laboratory coordinates*, J. Math. Phys., 37 (1996), pp. 308–323.

[10] D. J. Kaup and T. I. Lakoba, *Variational method: How it can generate false instabilities*, J. Math. Phys., 37 (1996), pp. 3442–3462.

[11] J. Schollmann, *On the stability of gap solitons*, Phys. A, 288 (2000), pp. 218–224.

[12] J. Schollmann and A. P. Mayer, *Stability analysis for extended models of gap solitary waves*, Phys. Rev. E, 61 (2000), pp. 5830–5838.

[13] T. Kapitula and B. Sandstede, *Edge bifurcations for near integrable systems via Evans function techniques*, SIAM J. Math. Anal., 33 (2002), pp. 1117–1143.

[14] D. E. Pelinovsky and A. Scheel, *Spectral analysis of stationary light transmission in nonlinear photonic structures*, J. Nonlinear Sci., 13 (2003), pp. 347–396.

[15] G. Derks and G. A. Gottwald, *A robust numerical method to study oscillatory instability of gap solitary waves*, SIAM J. Appl. Dyn. Syst., 4 (2005), pp. 140–158.

[16] D. J. B. Lloyd and A. R. Champneys, *Efficient numerical continuation and stability analysis of spatiotemporal quadratic optical solitons*, SIAM J. Sci. Comput., 27 (2005), pp. 759–773.

[17] G. Iooss and M. Adelmeyer, *Topics in Bifurcation Theory and Applications*, 2nd ed., Advanced Ser. Nonlinear Dynam. 3, World Scientific Publishing, River Edge, NJ, 1998.

[18] M. Golubitsky, I. Stewart, and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory, Vol.* 2, Appl. Math. Sci. 69, Springer-Verlag, New York, 1988.

[19] D. J. Kaup and A. C. Newell, *On the Coleman correspondence and the solution of the massive Thirring model*, Lett. Nuovo Cimento, 20 (1977), pp. 325–331.

[20] D. E. Pelinovsky, *Inertia law for spectral stability of solitary waves in coupled nonlinear Schrodinger equations*, Proc. Roy. Soc. London Ser. A, 461 (2005), pp. 783–812.

[21] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, UK, 1992, pp. 60–101.

[22] G. H. Golub and H. A. Van der Vorst, *Eigenvalue computation in the* 20*th century*, J. Comput. Appl. Math., 123 (2000), pp. 35–65.

[23] Y. Saad, *Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems*, Math. Comp., 42 (1984), pp. 567–588.

[24] L. N. Trefethen, *Spectral Methods in MATLAB*, Software Environ. Tools 10, SIAM, Philadelphia, 2000.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 5, No. 1, pp. 84–90

Absolutely Continuous Invariant Measures that Cannot be Observed Experimentally*

Paweł Góra†, Abraham Boyarsky†, MD Shafiqul Islam‡, and Wael Bahsoun§

Abstract. We study an example of a random map where the component maps have absolutely continuous invariant measures (acims), but where computer experiments reveal the surprising fact that all orbits eventually fall into a stable periodic orbit. This is all the more surprising as we prove that this random map admits an acim, μ. We study this phenomenon and explain why μ cannot be observed experimentally.

Key words. random map, absolutely continuous invariant measure, stable periodic orbit, unobservable measure

AMS subject classifications. 37A35, 37H20

DOI. 10.1137/040606478

1. Introduction.

A fundamental problem in dynamical systems is to describe the asymptotic behavior of trajectories. In general, the long term behavior of trajectories of a chaotic dynamical system is unpredictable. Therefore, it is natural to describe the behavior by statistical means. In this approach, one attempts to describe the dynamics by proving the existence of an invariant measure and determining its ergodic properties. In particular, the existence of invariant measures which are absolutely continuous with respect to Lebesgue measure is very important from a physical point of view, because computer simulations of orbits reveal only such invariant measures. In fact in [9] it has been shown that histograms of computer simulations display the invariant measure that is absolutely continuous with respect to Lebesgue measure for a general class of piecewise expanding transformations. An interesting question is, Does a computer simulation always reveal the absolutely continuous invariant measure (acim) of a dynamical system?

We define an invariant measure of a dynamical system to be "observable by computer experiment" if the histograms of almost all sufficiently long computer simulated trajectories approximate the histogram of the acim. It has been common knowledge that most dynamical systems that have an acim reveal this measure in computer simulations. The exceptions to this occur as a result of obvious "computer reasons" such as in the case of the map $x \to 2x \pmod 1$ on a binary computer, where almost all trajectories tend to 0 yet Lebesgue measure is the acim.

*Received by the editors April 7, 2004; accepted for publication (in revised form) by A. Stuart October 13, 2005; published electronically January 27, 2006. This research was supported by NSERC grants.
http://www.siam.org/journals/siads/5-1/60647.html

†Department of Mathematics and Statistics, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec H4B 1R6, Canada (pgora@vax2.concordia.ca, boyar@alcor.concordia.ca).

‡Department of Mathematics and Computer Science, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada (islam@cs.uleth.ca). The work of this author was supported by the Department of Mathematics and Statistics, Concordia University.

§Department of Mathematics and Statistics, University of Victoria, PO BOX 3045 STN CSC, Victoria, B.C., V8W 3P4, Canada (wab@math.uvic.ca). This author is a recipient of a PIMS Postdoctoral Fellowship.

Some maps, as, for example, the logistic map $x \to 4x(1-x)$, have trajectories which land on unstable periodic orbits, but such orbits are in general unobservable in computer experiments. The computer experiment for the logistic map, given sufficient precision and sufficiently long trajectories, will produce an approximation of the acim. In general, any ergodic map with unique acim will theoretically present such behavior. For such a map a typical trajectory is dense in the space so after a long enough time it will come so close to some periodic orbit that the computer will identify both points. Then, this trajectory is trapped in a periodic orbit forever. Fortunately, the typical time before such an occurrence is very long.

In this note we present another mechanism which can result in orbits of a random map with unique acim being trapped in periodic orbits. It is a combination of dynamics of individual component maps and the random controlling process which alternates them. We will reveal a struggle between two opposing dynamical tendencies. One is the ergodic behavior of the acim pushing orbits to wander everywhere (ergodicity), and the other is the strong attraction of periodic points of some combinations of maps constituting the random map, whenever the dense orbit passes close to the attracting periodic points. This struggle between very different dynamical forces is mediated by the precision of the modeling computer.

The dynamical systems we shall work with are generalized maps, referred to as random maps. Consider a collection of transformations $\tau_1, \tau_2, \ldots$ and a set of probabilities $p_1, p_2, \ldots$, $\sum_k p_k = 1$, associated with the transformations. A random map T is a discrete-time dynamical system where one of a number of transformations, $\tau_1, \tau_2, \ldots$, is randomly selected according to fixed or position dependent probabilities, $p_1, p_2, \ldots$ [1, 2, 6, 7, 10, 12], and applied at each iteration of the process. Random maps which consist of continuous and contracting transformations have found interesting applications in the study of fractals [3], in modeling interference effects in quantum mechanics [5], and in computing metric entropy [11]. More general random maps which have some expanding properties have been applied to forecasting in financial markets [2].

In section 2 we present the notation and summarize results we shall need in what follows. In section 3 we present a random map and discuss experimental computer results for this dynamical system. We prove that the random map has an acim and explain why it cannot be observed in a computer experiment.

2. Preliminaries. Let $(X, \mathcal{B}, \lambda)$ be a measure space, where λ is an underlying measure, and $\tau_k : X \to X$, $k = 1, 2, \ldots, K$, be nonsingular transformations. A random map T with constant probabilities is defined as

$$T = \{\tau_1, \tau_2, \ldots, \tau_K; p_1, p_2, \ldots, p_K\},$$

where $\{p_1, p_2, \ldots, p_K\}$ is a set of constant probabilities for any $x \in X$, $T(x) = \tau_k(x)$, with probability p_k. In general, the iterates of T are given by

$$T^N(x) = \tau_{k_N} \circ \tau_{k_{N-1}} \circ \cdots \circ \tau_{k_1}(x)$$

with probability $\Pi_{j=1}^{N} p_{k_j}$. A measure μ is said to be T-invariant if

$$\mu(E) = \sum_{k=1}^{K} p_k \mu(\tau_k^{-1}(E)) \tag{2.1}$$

for any measurable set $E \in \mathcal{B}$ [10]. The Frobenius–Perron operator of T is given by

$$P_T f(x) = \sum_{k=1}^{K} p_k P_{\tau_k}(f(x)),$$

where P_{τ_k} is the Frobenius–Perron operator associated with the transformation τ_k [4]. Pelikan [10] proved the existence of an acim for a random map with constant probabilities under the following condition:

$$\sum_{k=1}^{K} \frac{p_k}{|\tau_k'(x)|} \le \alpha < 1 \tag{2.2}$$

for any $x \in [0,1]$.

Now, we define what it means for a measure to be "observable by computer experiment." First we define a histogram of a trajectory. We assume that $X = [0,1]$ and that λ is the Lebesgue measure.

Definition 2.1 (histogram). *Let $x_0 \in X$ be an initial point and $O_m(x_0) = \{x_0, x_1, x_2, \ldots, x_{m-1}\}$, where $x_{i+1} = T(x_i)$ for $i = 1, 2, \ldots, m-2$, be a finite part of its total trajectory, $O_+(x_0) = \{x_0, x_1, x_2, \ldots\}$. Let $\mathcal{H}_n = \{I_s\}_{s=1}^n$ be a partition of X into n mutually disjoint subintervals of equal measure. The histogram h_{n,m,x_0} of $O_m(x_0)$ is the function*

$$h_{n,m,x_0}(x) = \frac{n}{m} \sum_{s=1}^{n} \left(\sum_{i=0}^{m-1} \chi_{I_s}(x_i) \right) \chi_{I_s}(x). \tag{2.3}$$

The function h_{n,m,x_0} is piecewise constant on the partition $\mathcal{H}_n$, equal on each subinterval I_s to the number of points of $O_m(x_0)$ which fall into I_s. Its integral is equal to 1. For the histogram to be practically useful, we will always assume that m is much larger than n, say, $m > 1000 \cdot n$. In our case, since T is a random map, the trajectory depends also on an additional variable ω which controls the random composition of maps $\{\tau_1, \tau_2, \ldots, \tau_K\}$. If we want to show this dependence explicitly, we will write $h^{\omega}_{n,m,x_0}(x)$ instead of $h_{n,m,x_0}(x)$.

Definition 2.2 (measure observable by computer experiment). *We say that the measure $\mu = f \cdot \lambda$ is observable by computer experiment if, for almost every initial point $x_0 \in X$ and almost every random parameter ω,*

$$h^{\omega}_{n,m,x_0} \to f \quad \text{in } L^1(X, \mathcal{B}, \lambda),$$

as $n \to \infty$, $m > 1000 \cdot n$.

We say that the measure $\mu = f \cdot \lambda$ is not observable by computer experiment if the above definition is not satisfied.

3. An example of a random map which admits an unobservable acim. In this section, we present an example of a random map which admits an unobservable acim. The existence of the acim is proved. At the same time every computer simulated trajectory of this random map eventually falls into the set $\{1/2, 1\}$ and stays there forever. We explain the reasons for such behavior.

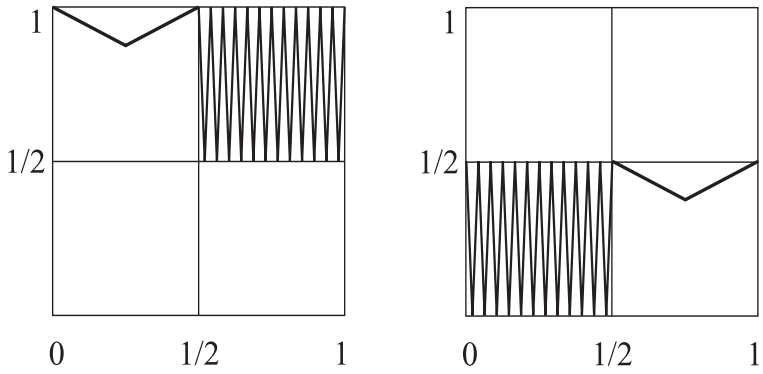


Figure 1. *The graphs of τ_1 and τ_2. Actually, there should be 128 branches of τ_1 on $[1/2,1]$ and of τ_2 on $[0,1/2]$.*

Let $f:[1/2,1]\to[1/2,1]$ and $g:[0,1/2]\to[0,1/2]$ be defined by

$$f(x)=\frac{1}{2}+2\left|x-\frac{3}{4}\right|,\qquad g(x)=2\left|x-\frac{1}{4}\right|.$$

Let $\lambda=1/2+0.00197$. Consider the random map $T=\{\tau_1,\tau_2\}$, as shown in Figure 1, where $\tau_1,\tau_2:[0,1]\to[0,1]$ are defined by

$$\tau_1(x)=\begin{cases}-\lambda x+1, & 0\le x\le \frac{1}{4},\\ 1+\lambda(x-\frac{1}{2}), & \frac{1}{4}<x\le\frac{1}{2},\\ f^7(x), & \frac{1}{2}<x\le 1,\end{cases}\tag{3.1}$$

where $f^7(x)=f\circ f\circ f\circ f\circ f\circ f\circ f(x)$, and

$$\tau_2(x)=\begin{cases}g^7(x), & 0\le x\le\frac{1}{2},\\ -\lambda(x-\frac{1}{2})+\frac{1}{2}, & \frac{1}{2}<x\le\frac{3}{4},\\ \frac{1}{2}+\lambda(x-1), & \frac{3}{4}<x\le 1,\end{cases}\tag{3.2}$$

where $g^7(x)=g\circ g\circ g\circ g\circ g\circ g\circ g(x)$.

These maps have two groups of properties.

Expansive properties—existence of an acim for the random map. The slopes are chosen in such a way that the random map $T=\{\tau_1,\tau_2;\frac{1}{2},\frac{1}{2}\}$ satisfies Pelikan's condition and thus admits an acim. Moreover, since each of the maps τ_1, τ_2 has a unique acim (Lebesgue measure on $[1/2,1]$ and $[0,1/2]$, correspondingly), the random map T also has a unique acim ([10] or [8]).

Attracting properties. $\tau_1(1/2)=1$ and the slope of τ_1 on $[0,1/2]$ is almost equal to $1/2$, so $\tau_1([0,1/2])\simeq[1-1/4,1]$. Similarly, $\tau_2(1)=1/2$ and the slope of τ_2 on $[1/2,1]$ is almost equal to $1/2$, so $\tau_2([1/2,1])\simeq[1/2-1/4,1/2]$.

Why the acim of T cannot be observed by computer simulations. Since we are considering a computer model of a random map it is more convenient to assume that numbers are represented by their binary expansion. Let us consider the computer precision using N digits of

binary expansion. Our remarks below apply directly to the computer implementation used by Maple 9.5, but we believe that they are general enough and other systems use similar implementations. The numbers are represented by sequences of N 0's and 1's starting with a 1 and a number which shows how many zeros should be inserted before the first 1. This causes a nonuniform distribution of computer numbers on the interval $[0, 1]$. For example, at the precision of 6 decimal digits (Maple command `Digits:=6`) or 20 binary digits we have $0.5 - 1/2^{20} = 0.5 - 1/2^{20} + 1/2^{21}$ but $1/2^{20} \neq 1/2^{20} - 1/2^{21}$ for the computer. This property is not important for our further reasoning as we work close to points 0.5 and 1.0 and far away from 0.

Another topic we have to discuss is the computer implementation of the maps τ_1 and τ_2. The most important is the behavior of τ_1 on $[1/4, 1/2)$ and of τ_2 on $[3/4, 1)$. In our example $\lambda \cong 1/2$. Thus, τ_1 on $[1/4, 1/2)$ is almost equal to $x \to 3/4 + x/2$, which in binary means the following: insert 0 in front of the digits of x and add binary 0.11. This in particular means that the binary number of the form $0.011\ldots11???\cdots \in [1/4, 1/2)$ with k 1's is τ_1-mapped into number $0.111\ldots11???\ldots$ with $(k + 2)$ 1's. Similarly, τ_2 on $[3/4, 1)$ is almost equal to $x \to x/2$, which in binary means the following: insert 0 in front of the digits of x. This in particular means that the binary number of the form $0.11\ldots11???\cdots \in [3/4, 1)$ with k 1's is τ_2-mapped into number $0.0111\ldots11???\ldots$ with again k 1's. In particular, at the precision of N binary digits, we have in computer implementation $\tau_1(x) = 1/2$ for any $x \in [1 - 1/2^N, 1]$ and $\tau_2(x) = 1$ for any $x \in [1/2 - 1/2^N, 1/2]$.

The computer is used for two purposes: to produce a sequence of pseudorandom numbers controlling the choice of maps τ_1 and τ_2 and to calculate values of $\tau_1(x)$ or $\tau_2(x)$ for appropriate τ_i. Let us assume that if the sequence of pseudorandom numbers gives 0, then τ_1 is applied, and when it shows 1, τ_2 is applied.

In precision of N binary digits an ideal pseudorandom number generator would produce any 0-1 sequence of length N. Let us assume that ours is not so ideal and produces only every 0-1 sequence of length $N - 1$. We assume that we are at the step in a trajectory such that the following holds.

Condition A. The next $N - 1$ digits in the controlling sequence are $010101\ldots010101$ (for odd N) or $010101\ldots0101010$ (for even N). The current point of the trajectory is $x \in (1/4, 1/2]$; i.e., the first two binary digits of x are 0.01.

Let us see what happens in this situation: because of "0" in the controlling sequence, the map τ_1 is applied; thus, x is attracted to 1 with the contraction factor almost equal to 1/2. This also means that the binary digits of $\tau_2(x)$ start with 0.111. Now, because of "1" in the controlling sequence, the map τ_2 is applied; thus, the current point of the trajectory $\tau_2(x)$ is attracted to 1/2 with the contraction factor almost equal to 1/2. This also means that $\tau_1(\tau_2(x))$ has binary representation starting with 0.0111. Now, the consecutive elements of the controlling sequence are used and the maps τ_1 and τ_2 are applied in alternating fashion. On each second step the current point in the trajectory has two more 1's at the beginning of its binary representation. We can illustrate this as follows:

$$0.01? \to 0.111? \to 0.0111? \to 0.11111?$$
$$\to 0.011111? \to 0.1111111? \to 0.01111111? \to 0.111111111? \to \cdots$$

If N is odd, then after $N-2$ steps the binary representation of the current point is $0.111\ldots111$

with N 1's and the next application of τ_2 (corresponding to the last 1 in the controlling sequence) sends the point to $1/2$. Similarly, if N is even, then after $N-2$ steps the binary representation of the current point is $0.0111\ldots111$ with one 0 and $N-1$ 1's and the next application of τ_1 (corresponding to the last 0 in the controlling sequence) sends the point to 1. It is possible that the trajectory of the point $0.01???\ldots$ falls into $\{1/2, 1\}$ earlier; what we presented above is the worst case scenario.

In conclusion, we showed that whenever Condition A is satisfied the trajectory of the random map ends in the set $\{1/2, 1\}$.

Similarly, many other combinations of closeness of current x to $1/2$ or 1 and a specific sequence in the controlling sequence of pseudorandom numbers lead to the trapping of the T-trajectory in the set $\{1/2, 1\}$.

Proof that acim of T cannot be observed experimentally. How long do we have to wait for Condition A to be satisfied? This depends on the starting point x_0 of the trajectory and the "seed" of the pseudorandom number generator. It can be a few steps or an enormously long time. To estimate the average waiting time for Condition A we can consider T as a skew product map on $[0,1] \times \{0,1\}^{\mathbb{N}}$. Using, for example, results of [8], it can be proved that the acim of T gives a positive weight, say, $w > 0$, to interval $[1/4, 1/2)$. Thus, the skew product invariant measure of the set $S_A = [1/4, 1/2) \times \{010101\ldots010101b\ldots\}$ ($N-1$ fixed digits, with b equal to 0 or 1) is $w \cdot (1/2)^{N-1}$. Also, the skew product map is ergodic so a typical trajectory will visit S_A. This proves that T does not satisfy Definition 2.2, i.e., cannot be observed experimentally, as initial conditions of measure at least $w \cdot (1/2)^{N-1}$ do not produce the histogram of acim.

At the same time we proved that the average waiting time to visit set S_A is $2^{N-1}/w$, which is the average waiting time for Condition A to occur. In practice, it may occur during the few first iterations of the random map T.

We performed computer experiments using Maple 9.5. At decimal precision of 6 digits (Maple command `Digits:=6`) or binary 20 digits we obtained after 1028545 iterations

0.940614
0.469539
0.984375
0.500000
1.

Note that $2^{20} = 1048576$. At decimal precision of 8 digits (Maple command `Digits:=8`) or binary 27 digits we obtained after 1516315 iterations

0.93741313
0.46826846
0.48459072
0.99218750
0.500000
1.

Note that $2^{27} = 134217728$.

Acknowledgments. The authors are grateful to V. Baladi for helpful comments and to anonymous reviewers as well as to the participants of Karol Krzyżewski's 70th Birthday Symposium: From Dynamical Systems to Financial Mathematics, Warsaw, May 12th, 2004.

REFERENCES

[1] W. Bahsoun, P. Góra, and A. Boyarsky, *Stochastic perturbations for position dependent random maps*, Stoch. Dyn., 3 (2003), pp. 545–557.
[2] W. Bahsoun, P. Góra, and A. Boyarsky, *Markov switching for position dependent random maps with application to forecasting*, SIAM J. Appl. Dyn. Syst., 4 (2005), pp. 391–406.
[3] M. Barnsley, *Fractals Everywhere*, Academic Press, London, 1998.
[4] A. Boyarsky and P. Góra, *Laws of Chaos*, Birkhäuser Boston, Boston, 1997.
[5] A. Boyarsky and P. Góra, *A dynamical model interference effects and two slit experiment of quantum physics*, Phys. Lett. A, 168 (1992), pp. 103–112.
[6] J. Buzzi, *Absolutely continuous S.R.B. measures for random Lasota-Yorke maps*, Trans. Amer. Math. Soc., 352 (2000), pp. 3289–3303.
[7] P. Góra and A. Boyarsky, *Absolutely continuous invariant measures for random maps with position dependent probabilities*, J. Math. Anal. Appl., 278 (2003), pp. 225–242.
[8] P. Góra, *Graph theoretic bound on number of acim for random transformations*, Proc. Amer. Math. Soc., 116 (1992), pp. 401–410.
[9] P. Góra and A. Boyarsky, *Why computers like Lebesgue measure*, Comput. Math. Appl., 16 (1988), pp. 321–329.
[10] S. Pelikan, *Invariant densities for random maps of the interval*, Proc. Amer. Math. Soc., 281 (1984), pp. 813–825.
[11] W. Słomczyński, J. Kwapień, and K. Życzkowski, *Entropy computing via integration over fractal measures*, Chaos, 10 (2000), pp. 180–188.
[12] S. Ulam and J. von Neumann, *Random ergodic theorem*, Bull. Amer. Math. Soc., 72 (1952), pp. 46–66.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 5, No. 1, pp. 91–107

Reachable Matrices by a QR Step with Shift*

Delin Chu† and Moody Chu‡

Abstract. One of the most interesting dynamical systems used in numerical analysis is the QR algorithm. An added maneuver to improve the convergence behavior is the QR iteration with shift which is of fundamental importance in eigenvalue computation. This paper is a theoretical study of the set of all isospectral matrices "reachable" by the dynamics of the QR algorithm with shift. A matrix B is said to be reachable by A if $B = RQ + \mu I$, where $A - \mu I = QR$ is the QR decomposition for some $\mu \in \mathbb{R}$. It is proved that in general the QR algorithm with shift is neither reflexive nor symmetric. Examples are given to demonstrate that this relation is neither transitive nor antisymmetric. It is further discovered that the reachable set from a given $n \times n$ matrix A forms 2^{n-1} disjoint open loops if n is even and 2^{n-2} disjoint components each of which is no longer a loop when n is odd.

Key words. QR algorithm, shift, analytic QR decomposition, reachable matrices, reflexivity, symmetry, transitivity

AMS subject classifications. 15A23, 37C10

DOI. 10.1137/040617844

1. Introduction.

Given a matrix $A \in \mathbb{R}^{n\times n}$, by a *$QR$ step with shift* $\mu \in \mathbb{R}$ we mean an operation that involves the following two steps:

(a) Compute the QR factorization of $A - \mu I$,

$$A - \mu I = Q_A(\mu)R_A(\mu), \tag{1.1}$$

where $Q_A(\mu)$ is orthogonal and $R_A(\mu)$ is upper triangular.

(b) Compute

$$A(\mu) = R_A(\mu)Q_A(\mu) + \mu I = Q_A^\top(\mu)AQ_A(\mu). \tag{1.2}$$

Starting with $A^{(1)} := A$, a discrete dynamical system by repeating the above operation with properly chosen shifts $\{\mu_k\}$ applied to the sequence of matrices $A^{(k+1)} := A^{(k)}(\mu_k)$ constitutes what is called the QR algorithm with shift [5]. The QR algorithm with shift is the principal powerhouse in eigenvalue computation. The shifts play a critical role in effecting fast convergence.

Generally speaking, there are two proven strategies for selecting the basic shifts: the Wilkinson shift and the Rayleigh quotient. It is known that if $A^{(1)}$ is an upper Hessenberg

*Received by the editors October 28, 2004; accepted for publication (in revised form) by W. Beyn September 13, 2005; published electronically February 14, 2006. This research was supported in part by the National Science Foundation under grants DMS-0073056 and CCR-0204157.
http://www.siam.org/journals/siads/5-1/61784.html
†Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543 (matchudl@math.nus.edu.sg).
‡Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (chu@math.ncsu.edu).

matrix to begin with, and usually it is made so, then the structure is preserved throughout the QR algorithm. The Wilkinson shift is particularly convenient and useful for Hessenberg matrices in the QR iteration [10]. The Rayleigh quotient updated at each iteration, on the other hand, is known to be effective for inverse iteration [7]. Once basic shift values are determined, they could be incorporated into more advanced computation paradigms, including schemes such as the implicit double shift method or the multishift algorithm [5, 8]. Besides these, there does not seem to exist many other variants of practical shift strategies.

It is well known that the iterates generated by the QR algorithm with *zero* shift are equivalent to the sampling of the Toda lattice (flow) at integer times [1, 9]. While the shift μ may be considered as a "feedback control" ministered step by step to change the convergence behavior, thus far there is no known continuous system modeling the QR algorithm with shift. The purpose of this paper is not to derive any new shift strategies; rather, we find it interesting to understand, at least from a theoretical point of view, how other shifts can affect the dynamics of the QR algorithm. By considering the shift μ as a general parameter, we intend to study the set

$$\mathcal{S}_A := \{A(\mu) \in \mathbb{R}^{n\times n} | \mu \in \mathbb{R}\} \tag{1.3}$$

of "all" possible matrices that are orthogonally similar to A via a QR step with shift. Some clarification of the definitions of both $A(\mu)$ and $\mathcal{S}_A$ is in order before we can proceed further. The QR decomposition in general is not unique, so (1.2) is not well defined. However, because this one-parameter matrix function $A-\mu I$ is real analytic in μ, it is known that a real analytic QR decomposition of this analytic matrix function exists [3, 4]. Once a pair of initial values, say, $Q_A(0)$ and $R_A(0)$, is fixed, we shall restrict our factorization in (1.1) and hence $A(\mu)$ for all other μ to the branch of real analytic QR decomposition that passes through $Q_A(0)$ and $R_A(0)$. Different initial values differ by sign changes and inaugurate different analytic paths. With this in mind, we shall adopt the following notation in this discussion.

Definition 1.1. *Given two matrices $A, B \in \mathbb{R}^{n\times n}$, we say that B is reachable from A, denoted by $A \curvearrowright B$, if $B = A(\mu)$ for some $\mu \in \mathbb{R}$.*

The set $\mathcal{S}_A$ therefore contains all reachable matrices from A. On first glance, the set $\mathcal{S}_A$ as a one-parameter family should represent an isospectral curve in $\mathbb{R}^{n\times n}$. We shall show that it is made of several disjoint real analytic curves, each passing through a certain initial value. We wish to characterize these curves more specifically, preferably in an analytical way. Along the way, it is of theoretical interest to ask the following questions:

- If $A \curvearrowright B$, is it true that $B \curvearrowright A$?
- If $A \curvearrowright B$ and $B \curvearrowright C$, is it true that $A \curvearrowright C$?

In modern algebra terminology, the above questions inquire about whether the QR algorithm with shift enjoys the *symmetry* and *transitivity* properties, respectively. It appears that these questions have not been explored before.

To decipher the set $\mathcal{S}_A$, it suffices to characterize the set

$$\mathcal{H}_A := \{Q_A(\mu) \in \mathcal{O}_n | \mu \in \mathbb{R}\}, \tag{1.4}$$

where $\mathcal{O}_n$ stands for the orthogonal group in $\mathbb{R}^{n\times n}$. Recently there have been considerable efforts in deriving isospectral curves (flows) for a variety of important applications [2, 6]. Our

isospectral curve $\mathcal{S}_A$ is defined via the QR step with shift. Indeed, it is known that any differentiable flow on $\mathcal{O}_n$ satisfies a differential system of the form

$$\dot{Q} = QK, \tag{1.5}$$

where $\dot{Q}$ denotes the derivative with respect to some appropriate parameter and $K \in \mathbb{R}^{n\times n}$ is skew-symmetric. Since both sides of (1.1) are differentiable with respect to μ, we find that

$$\dot{R_A} = -Q_A^\top - KR_A. \tag{1.6}$$

The skew-symmetric matrix K for our purpose therefore must be such that $-Q_A^\top - KR_A$ is upper triangular. We shall justify later that $K = K(\mu)$ can be defined.

Without loss of generality, we shall assume henceforth that no row or column of A can be entirely zero at off-diagonal positions because, otherwise, the matrix A could have been deflated first. For simplicity, we shall also assume that all eigenvalues of A are distinct.

2. The case $n = 2$. We begin with the case $n = 2$ as it provides considerable insight into the general problem. The questions of symmetry for the case $n = 2$ can easily be answered from the following observation.

Theorem 2.1. *Suppose $n = 2$. If $A \curvearrowright B$, then $B \curvearrowright A$.*

Proof. Because $A \curvearrowright B$, there exists a number $\mu_0 \in \mathbb{R}$ such that

$$\begin{aligned} A - \mu_0 I &= Q_A(\mu_0)R_A(\mu_0), \\ B &= Q_A^\top(\mu_0)AQ_A(\mu_0), \end{aligned}$$

where $Q_A(\mu_0)$ is orthogonal and $R_A(\mu_0)$ is upper triangular. We claim that a number $s_0 \in \mathbb{R}$ can be found such that

$$B - s_0 I = Q_A^\top(\mu_0)R_B(s_0), \tag{2.1}$$

where $R_B(s_0)$ is upper triangular. If this claim is true, then we have $Q_B(s_0) = Q_A^\top(\mu_0)$ and

$$R_B(s_0)Q_B(s_0) + s_0 I = Q_A(\mu_0)(B - s_0 I)Q_A^\top(\mu_0) + s_0 I = Q_A(\mu_0)BQ_A^\top(\mu_0) = A,$$

that is, $B \curvearrowright A$. To construct s_0 in (2.1), denote

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \text{and} \quad Q_A(\mu) = \begin{bmatrix} q_{11}(\mu) & q_{12}(\mu) \\ q_{21}(\mu) & q_{22}(\mu) \end{bmatrix}.$$

Note that $q_{12}(\mu) \neq 0$ for every μ because, otherwise, the matrix A would have to be upper triangular and could be deflated. It can quickly be checked that the choice

$$s_0 := b_{11} - \frac{q_{11}(\mu_0)}{q_{12}(\mu_0)} b_{21} \tag{2.2}$$

satisfies (2.1). ∎

The symmetry of the QR step with shift for the case $n = 2$ can be better seen geometrically. Toward that end, we characterize $\mathcal{H}_A$. Write

$$R_A(\mu) = \begin{bmatrix} r_{11}(\mu) & r_{12}(\mu) \\ 0 & r_{22}(\mu) \end{bmatrix}, \quad K(\mu) = \begin{bmatrix} 0 & k(\mu) \\ -k(\mu) & 0 \end{bmatrix};$$

we find that the only choice of $k(\mu)$ making $-Q_A^\top - KR_A$ upper triangular is

$$k(\mu) := \frac{q_{12}(\mu)}{r_{11}(\mu)}. \tag{2.3}$$

We thus obtain from (1.5) and (1.6) the differential systems

$$\begin{cases} \dot{q}_{11} &= -\frac{q_{12}^2}{r_{11}}, \\ \dot{q}_{12} &= \frac{q_{11}q_{12}}{r_{11}}, \\ \dot{q}_{21} &= -\frac{q_{22}q_{12}}{r_{11}}, \\ \dot{q}_{22} &= \frac{q_{21}q_{12}}{r_{11}}, \end{cases} \qquad \begin{cases} \dot{r}_{11} &= -q_{11}, \\ \dot{r}_{12} &= -q_{21} - \frac{q_{12}r_{22}}{r_{11}}, \\ \dot{r}_{22} &= -q_{22} + \frac{q_{12}r_{12}}{r_{11}} \end{cases} \tag{2.4}$$

that describe the dynamics of $Q_A(\mu)$ and $R_A(\mu)$, respectively. Note that $|r_{11}(\mu)|$ is equal to the 2-norm of the first column of $A - \mu I$ and, hence, is never zero. The above differential systems are defined for all $\mu \in \mathbb{R}$. We have thus shown indirectly the differentiability of $A(\mu)$. In fact, because $|r_{11}(\mu)| \to \infty$ as $\mu \to \pm\infty$, it follows that $q_{12}(\mu) \to 0$ and that $Q_A(\mu)$ must converge to a diagonal orthogonal matrix with ± 1 along its diagonal. Together, we have shown that as $|\mu| \to \infty$,

$$A(\mu) \to A_* := \begin{cases} A & \text{if } \det(Q_A(0)) = 1, \\ \begin{bmatrix} a_{11} & -a_{12} \\ -a_{21} & a_{22} \end{bmatrix} & \text{if } \det(Q_A(0)) = -1. \end{cases}$$

The significance of this observation is that $\mathcal{S}_A$ can now be represented symbolically as the union of two disjoint *open* loops. The loop corresponding to $\det(Q_A(0)) = 1$ is depicted in Figure 1. We purposefully use a nonsymmetric loop representation because there is no indication that the curve $A(\mu) \in \mathbb{R}^{2\times 2}$ has any symmetry. The large solid circle denotes the initial point $A(0)$. The arrows indicate the directions $A(\mu)$ move as μ converges to positive or negative infinity, respectively. It is important to note that in both directions $A(\mu)$ converges to A, but generally A does *not* belong to $\mathcal{S}_A$. We denote A by an empty circle. Furthermore, because $\mathcal{O}_2$ is only a one-dimensional manifold, it is not difficult to see that if $Q_A(0)$ and $Q_B(0)$ have the same orientation then the *trajectories* of $Q_A(\mu)$ and $Q_B(\mu)$ are exactly the same. It follows that, if $A \curvearrowright B$, the sets $\mathcal{S}_A$ and $\mathcal{S}_B$ are almost identical except that they exclude A and B, respectively. In particular, we see from the graph that $A \in \mathcal{S}_B$ and hence $B \curvearrowright A$. We will see that the situation is not so simple in higher dimensional cases.

In a similar way, we prove that the relation $\curvearrowright$ is not reflexive in the following sense.

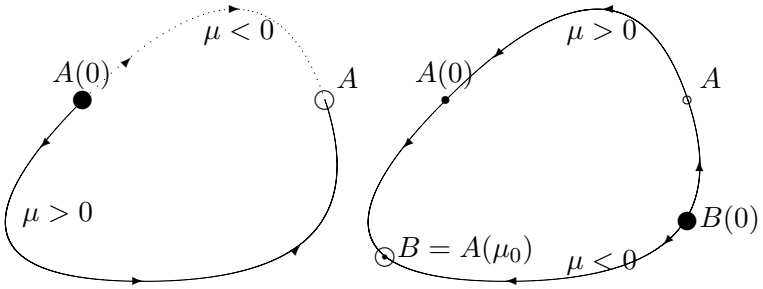


Figure 1. *The representation of $\mathcal{S}_A$ (left) and $\mathcal{S}_B$ (right) when $n = 2$ and $\det(Q_A)(0) = \det(Q_B(0)) = 1$.*

Theorem 2.2. *Suppose $n = 2$. Then $A \curvearrowright A$ if and only if $A = rQ + \mu I$ for some $r, \mu \in \mathbb{R}$ and some orthogonal matrix Q. Hence, generally, $A \not\curvearrowright A$.*

Proof. Suppose $A \curvearrowright A$. Then there exists a $\mu \in \mathbb{R}$ such that

$$A - \mu I = Q_A(\mu)R_A(\mu), \quad A = Q_A^\top(\mu)AQ_A(\mu),$$

with orthogonal $Q_A(\mu)$ and upper triangular $R_A(\mu)$. It follows that

$$Q_A(\mu)R_A(\mu) = R_A(\mu)Q_A(\mu). \tag{2.5}$$

Note that no column of A can be entirely zero at off-diagonal positions, so we must have $q_{21}(\mu) \neq 0$ for all $\mu \in \mathbb{R}$. Consequently, a simple calculation shows that the relationship (2.5) holds if and only if

$$R_A(\mu) = rI$$

for some $r \in \mathbb{R}$.

Conversely, if $A = rQ + \mu I$ for some $r, \mu \in \mathbb{R}$ and orthogonal matrix Q, then

$$A - \mu I = Q(rI), \quad A = Q^\top AQ,$$

that is, $A \curvearrowright A$. ∎

Using Theorems 2.1 and 2.2, it is easy to see that the relation $\curvearrowright$ is not transitive.

Corollary 2.3. *Suppose $n = 2$. Assume that $A \curvearrowright B$ and $B \curvearrowright C$. Then, generally, $A \not\curvearrowright C$.*

3. Algebraic properties in general. For a fixed matrix $A \in \mathbb{R}^{n\times n}$, denote $Q_A(\mu) = [q_{ij}^{[A]}(\mu)]$ and $R_A(\mu) = [r_{ij}^{[A]}(\mu)]$. For simplicity, we shall write $[q_{ij}^{[A]}(\mu)] = [q_{ij}(\mu)]$ and $[r_{ij}^{[A]}(\mu)] = [r_{ij}(\mu)]$ by dropping the references to A without causing ambiguity. Let the spectral decomposition of A be denoted by

$$AW = W\Sigma, \tag{3.1}$$

where $W \in \mathbb{C}^{n\times n}$ is nonsingular and Σ is diagonal.

We first explain how the real analytic QR decomposition can be defined. In order that the vector field $-Q^\top - KR_A$ in (1.6) remains upper triangular, the skew-symmetric matrix $K(\mu) = [k_{ij}(\mu)]$ must satisfy the equations

$$k_{ij}r_{ii} = q_{ij} - \sum_{s=1}^{i-1} k_{sj}r_{si}, \quad i = 1, \ldots, n-1, \quad j = i+1, \ldots, n. \tag{3.2}$$

The entries $k_{ij}(\mu)$ can be uniquely defined only if $r_{ii}(\mu) \neq 0$ for $i = 1, \ldots, n-1$. It is clear, however, that when μ coincides with an eigenvalue of A some $r_{ii}(\mu)$ will be zero. The question is how k_{ij} should be defined in such a situation.

The following example should shed some light on the above concern.

Example 1. *Consider the case where*

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 4 & 5 \\ 3 & 3 & 6 \end{bmatrix}.$$

It is easy to verify that three eigenvalues of A are distinct and $r_{11}(\mu) \neq 0$ for all $\mu \in \mathbb{R}$. However, because

$$A - 2I = \begin{bmatrix} 1 & 1 & 4 \\ 2 & 2 & 5 \\ 3 & 3 & 4 \end{bmatrix},$$

we see that $r_{22}(2) = 0$ and, hence, $k_{23}(\mu)$ cannot be defined directly from the relationship (3.2) at $\mu = 2$. We emphasize that $A - \mu I$ does have a real analytic QR decomposition in μ. Figure 2 illustrates the limiting behavior of

$$r_{22}(\mu) = \pm\sqrt{\frac{\mu^4 - 14\,\mu^3 + 87\,\mu^2 - 212\,\mu + 172}{22 - 6\,\mu + \mu^2}}$$

and

$$k_{23}(\mu) = \frac{q_{23}(\mu) - k_{13}(\mu)r_{12}(\mu)}{r_{22}(\mu)} = \frac{q_{23}(\mu)r_{11}(\mu) - q_{13}(\mu)r_{12}(\mu)}{r_{11}(\mu)r_{22}(\mu)}$$

as μ passes through the eigenvalue $\mu = 2$. To maintain analyticity, it is clear that $r_{22}(\mu)$ has to change sign at $\mu = 2$. In particular, $k_{23}(\mu)$ does have a limit at $\mu = 2$, if the thicker curve is connected to the thinner curve at the discontinuity jump.

The case that $r_{ii}(\mu) = 0$ for some $1 \leq i \leq n-1$ at a certain μ does not happen often. To see this, denote the matrices of the first $n-1$ columns of A, I_n, and $Q_A(\mu)$, respectively, by A_{n-1}, E_{n-1}, and $Q_{n-1}(\mu)$. Observe that

$$A_{n-1} - \mu E_{n-1} = Q_{n-1}(\mu) \begin{bmatrix} r_{11}(\mu) & \cdots & r_{1,n-1}(\mu) \\ & \ddots & \vdots \\ & & r_{n-1,n-1}(\mu) \end{bmatrix}.$$

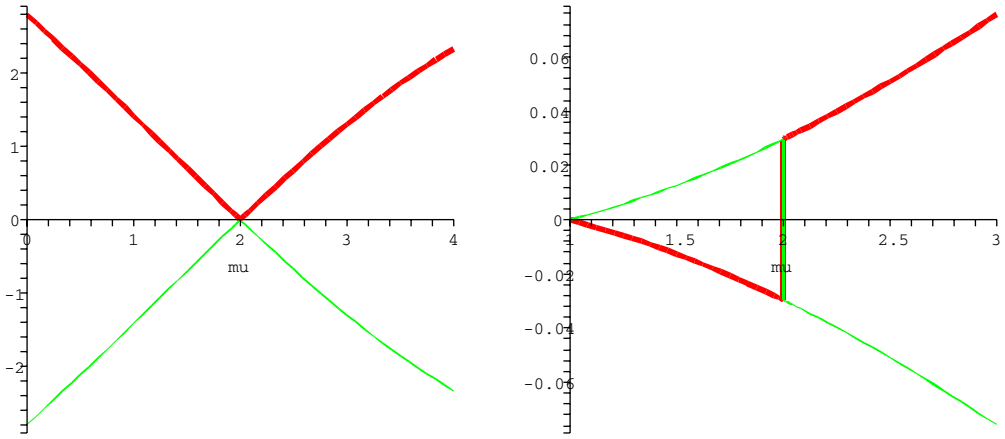


Figure 2. *Evolution of $r_{22}(\mu)$ (left) and $k_{23}(\mu)$ (right) as μ passes through $\mu = 2$. Note that the thicker curves must be continued with the thinner curves at $\mu = 2$ to maintain analyticity.*

It follows that $r_{ii}(\mu)$, $i = 1, \ldots, n-1$, is never zero if and only if

$$\operatorname{rank}(A_{n-1} - \mu E_{n-1}) = n - 1 \quad \text{for all } \mu \in \mathbb{R}. \tag{3.3}$$

We mention that the rank condition (3.3) is closely related to the so-called *observability* in control theory [11]. It can easily be argued that the condition (3.3) holds for almost all matrices $A \in \mathbb{R}^{n\times n}$, and this will be assumed to be the case in our discussion.

It is important to realize that $R(\mu)$ can have at most one zero diagonal element because eigenvalues of A are distinct. By following the analytic curve of $R(\mu)$ under the condition (3.3), the only diagonal entry of $R(\mu)$ that could ever vanish is $r_{nn}(\mu)$ at real eigenvalues of A. The real analytic QR decomposition therefore can be defined by the differential equations (1.5) and (1.6) with

$$k_{ij} = \frac{q_{ij} - \sum_{s=1}^{i-1} k_{sj} r_{si}}{r_{ii}}, \quad i = 1, \ldots, n-1, \quad j = i+1, \ldots, n. \tag{3.4}$$

With a real analytic curve in mind, it remains to see how the set $\mathcal{S}_A$ would look in higher dimensional space.

In contrast to Theorem 2.1 for the case $n = 2$, the following example illustrates the asymmetry when $n = 3$.

Example 2. *Let*

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

It is not difficult to verify the facts that the spectrum of either A or B consists of distinct eigenvalues, that the condition (3.3) holds for both A and B, that the QR factorization of A

is given by

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and that

$$B = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^{\top} A \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

that is, $A \curvearrowright B$.

For any $s \in \mathbb{R}$, *let*

$$B - sI = Q_B(s)R_B(s)$$

denote the QR decomposition of $B - sI$. *If we write* $Q_B(s) = [q_{ij}^{[B]}(s)]$, *then it must be the case that for all* $s \in \mathbb{R}$,

$$q_{31}^{[B]}(s) = 0, \quad q_{21}^{[B]}(s) \neq 0.$$

On the other hand, let $\widetilde{Q} = [\widetilde{q}_{ij}]$ *denote an arbitrary matrix with* $\widetilde{q}_{31} = 0$. *Upon computing the products* $\widetilde{Q}A$ *and* $B\widetilde{Q}$, *we find that* $\widetilde{Q}A = B\widetilde{Q}$ *if and only if* $\widetilde{Q}$ *is of the form*

$$\widetilde{Q} = \begin{bmatrix} \widetilde{q}_{11} & \widetilde{q}_{11} & \widetilde{q}_{21} \\ \widetilde{q}_{21} & \widetilde{q}_{11} & 0 \\ 0 & \widetilde{q}_{21} & \widetilde{q}_{11} - \widetilde{q}_{21} \end{bmatrix}.$$

However, $\widetilde{Q}$ *of the structure above cannot be orthogonal. We thus conclude that* $A \neq Q_B(s)^T B Q_B(s)$ *for any* $s \in \mathbb{R}$ *so long as* $Q_B(s)$ *is orthogonal. That is,* $B \not\curvearrowright A$.

Indeed, the asymmetry illustrated above is generically true; that is, we can prove that the QR step with shift in general is not symmetric.

Theorem 3.1. *Assume that* A *has* n *distinct eigenvalues and* $A \curvearrowright B$. *If* $n > 2$, *then generally* $B \not\curvearrowright A$.

Proof. We shall prove by contradiction. Assume $B \curvearrowright A$. Then for some $s \in \mathbb{R}$ we should have

$$\begin{aligned} B - sI &= Q_B(s)R_B(s), \\ A &= Q_B^{\top}(s)BQ_B(s), \end{aligned}$$

with $Q_B(s) \in \mathcal{O}_n$ and $R_B(s)$ upper triangular. On the other hand, because $A \curvearrowright B$, there exists a number $\mu_0 \in \mathbb{R}$ such that

$$\begin{aligned} A - \mu_0 I &= Q_A(\mu_0)R_A(\mu_0), \\ B &= Q_A^{\top}(\mu_0)AQ_A(\mu_0), \end{aligned}$$

with $Q_A(\mu_0) \in \mathcal{O}_n$ and $R_A(\mu_0)$ upper triangular. By rewriting the relationship

$$W\Sigma W^{-1} = A = (Q_A(\mu_0)Q_B(s))^\top A Q_A(\mu_0)Q_B(s),$$

we see another spectral decomposition of A,

$$A(Q_A(\mu_0)Q_B(s)W) = (Q_A(\mu_0)Q_B(s)W)\Sigma. \tag{3.5}$$

Since all eigenvalues of A are distinct, the corresponding eigenvectors of the same eigenvalue can differ only by a scalar multiplication. That is, we must have

$$Q_A(\mu_0)Q_B(s)W = W\Lambda, \tag{3.6}$$

with

$$\Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\} \in \mathbb{C}^{n\times n}, \quad |\lambda_i| = 1, \quad i = 1, \ldots, n. \tag{3.7}$$

Rewriting

$$Q_B(s) = Q_A^\top(\mu_0)W\Lambda W^{-1},$$

we obtain

$$Q_A^\top(\mu_0)(A - sI)Q_A(\mu_0) = B - sI = Q_B(s)R_B(s) = Q_A^\top(\mu_0)W\Lambda W^{-1}R_B(s).$$

Equivalently, we have the equation

$$(A - sI)W\bar{\Lambda}W^{-1} = W\bar{\Lambda}W^{-1}(A - sI) = R_B(s)Q_A^\top(\mu_0),$$

where $\bar{\Lambda}$ denotes the complex conjugate of Λ. It follows that

$$(A - sI)W\bar{\Lambda}W^{-1}(A - \mu_0 I) = R_B(s)R_A(\mu_0), \tag{3.8}$$

$$W\bar{\Lambda}W^{-1}(A - sI)Q_A(\mu_0) = R_B(s). \tag{3.9}$$

In particular, note that both matrices on the right-hand sides of (3.8) and (3.9) are upper triangular. We claim the mere fact that the strictly lower triangular part of the two matrices on the left-hand sides of (3.8) and (3.9) is an overdetermined system for s. To see this point, write A and W in rows,

$$A = \begin{bmatrix} \mathfrak{a}_1^\top \\ \mathfrak{a}_2^\top \\ \vdots \\ \mathfrak{a}_n^\top \end{bmatrix}, \quad W = \begin{bmatrix} \mathfrak{w}_1^\top \\ \mathfrak{w}_2^\top \\ \vdots \\ \mathfrak{w}_n^\top \end{bmatrix}, \quad \text{and} \quad I = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}.$$

The strictly lower triangular part of (3.8) being zero means that

$$(\mathfrak{a}_j^\top - s\mathbf{e}_j^\top)W\bar{\Lambda}W^{-1}(A - \mu_0 I)\begin{bmatrix} I_{j-1} \\ 0 \end{bmatrix} = 0, \quad j = 2, \ldots, n,$$

or equivalently,

$$\mathfrak{w}_j^\top \Sigma\bar{\Lambda}\Sigma(\Sigma-\mu_0 I)W^{-1}\begin{bmatrix} I_{j-1} \\ 0 \end{bmatrix} = s\mathfrak{w}_j^\top \bar{\Lambda}(\Sigma-\mu_0 I)W^{-1}\begin{bmatrix} I_{j-1} \\ 0 \end{bmatrix}, \quad j=2,\ldots,n. \tag{3.10}$$

There are $n(n-1)/2$ equations for the single variable s in order to satisfy (3.8). Similarly, (3.9) leads to another overdetermined system for s. Because W, Σ, and μ_0 are independent, and Λ is of the particular diagonal form, we conclude that the system (3.8) and (3.9) generally does not have a solution s. This contradicts the assumption that $B \curvearrowright A$. ■

Using similar arguments, we now show that the QR step with shift is not reflexive.

Theorem 3.2. *Assume that A has n distinct eigenvalues. Then in general $A \not\curvearrowright A$; that is,*

$$Q_A^\top(\mu)AQ_A(\mu) \neq A \quad \textit{for all } \mu \in \mathbb{R}. \tag{3.11}$$

Proof. Again, we prove by contradiction. Assume $A \curvearrowright A$. By (1.1), we see the relationship that

$$Q_A^\top(\mu)AQ_A(\mu) = A \iff R_A(\mu)Q_A(\mu) = Q_A(\mu)R_A(\mu) \iff Q_A(\mu)A = AQ_A(\mu). \tag{3.12}$$

Assuming that the spectral decomposition of A is given by (3.1), we should also have

$$Q_A^\top(\mu) = W\Lambda W^{-1} \tag{3.13}$$

for some Λ in the form (3.7). Upon comparing this with the strictly lower triangular part of the equation $Q_A^\top(\mu)(A-\mu I) = R_A(\mu)$, we conclude that

$$\mathfrak{w}_j^\top \Lambda W^{-1}(A-\mu I)\begin{bmatrix} I_{j-1} \\ 0 \end{bmatrix} = 0, \quad j=2,\ldots,n,$$

or equivalently,

$$\mathfrak{w}_j^\top \Lambda\Sigma W^{-1}\begin{bmatrix} I_{j-1} \\ 0 \end{bmatrix} = \mu\mathfrak{w}_j^\top \Lambda W^{-1}\begin{bmatrix} I_{j-1} \\ 0 \end{bmatrix}, \quad j=2,\ldots,n. \tag{3.14}$$

There are $n(n-1)/2$ equations in (3.14) for the scalar μ. The system (3.14) in general has no solution for μ. ■

The following example demonstrates the antireflexivity asserted in Theorem 3.2.

Example 3. *Consider the matrix A in Example 2. For any $\mu \in \mathbb{R}$, let*

$$A-\mu I = Q_A(\mu)R_A(\mu)$$

denote the QR decomposition of $A-\mu I$. By the structure of A, it is easy to see that for all $\mu \in \mathbb{R}$, we must have

$$q_{31}(\mu) = 0, \quad q_{21}(\mu) \neq 0.$$

Using the property $q_{31}(\mu) = 0$, we find that $Q_A(\mu)A = AQ_A(\mu)$ for some $\mu \in \mathbb{R}$ if and only if

$$Q_A(\mu) = \begin{bmatrix} q_{11}(\mu) & 0 & q_{21}(\mu) \\ q_{21}(\mu) & q_{11}(\mu)+q_{21}(\mu) & 0 \\ 0 & q_{21}(\mu) & q_{11}(\mu) \end{bmatrix}.$$

For $Q_A(\mu)$ of such a structure to be orthogonal, it must be such that $q_{21}(\mu) = 0$. This is a contradiction. Therefore, $A \not\curvearrowright A$.

For the case $n = 2$, we can easily use the symmetry and the antireflexivity to show that the QR step with shift is generally not transitive. For the case $n > 2$, the relation $\curvearrowright$ is no longer symmetric by Theorem 3.1. However, we offer the following example to demonstrate that the transitivity remains failed.

Example 4. *Let A and B be the same matrices as in Example 2 so that $A \curvearrowright B$. Let*

$$C = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Observe that the QR factorization of $B - I$ is given by

$$B - I = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix},$$

while

$$C = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^{\top} B \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

It follows that $B \curvearrowright C$. Using the same argument as that in the proof for $B \not\curvearrowright A$ in Example 2, we can show that $A \not\curvearrowright C$.

It is worth clarifying one more scenario. We already know that the operation of a QR step with shift is generally not symmetric, as shown in Theorem 3.1. Could it be possible that it is antisymmetric in the sense that both $A \curvearrowright B$ and $B \curvearrowright A$ imply that $A = B$? Clearly, this is not necessarily the case when $n = 2$ because the relation is symmetric. The following example shows that the relation $\curvearrowright$ is not antisymmetric for $n > 2$ either.

Example 5. *Define*

$$A = \begin{bmatrix} \sqrt{2} & 0 & 3+\sqrt{2} \\ 2 & -\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 2 & 1+\sqrt{2} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 & 2\sqrt{2} \\ -\sqrt{2} & 2+\sqrt{2} & -2-\sqrt{2} \\ -1 & \sqrt{2} & -1 \end{bmatrix}.$$

Then matrices in the QR decompositions of

$$A = Q_A R_A, \quad B - I = Q_B R_B$$

are given by

$$Q_A = \begin{bmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{2} \end{bmatrix}, \quad Q_B = Q_A^T$$

and

$$R_A = \begin{bmatrix} 2\sqrt{2} & -2 & 0 \\ 0 & \sqrt{2} & 2+2\sqrt{2} \\ 0 & 0 & 2 \end{bmatrix}, \quad R_B = \begin{bmatrix} -2 & 2+\sqrt{2} & -2 \\ 0 & \sqrt{2}-1 & 2+\sqrt{2} \\ 0 & 0 & -2\sqrt{2} \end{bmatrix},$$

respectively. It is easy to verify that $B = Q_A^T A Q_A$ *and* $A = Q_B^T B Q_B$. *Equivalently, we have* $A \curvearrowright B$ *and* $B \curvearrowright A$. *But, it is clear that* $A \neq B$.

4. Reachable set. In this section, we first describe the limit points of $A(\mu)$ as μ goes to infinity. By studying carefully the inertia of these limit points, together with the fact that we are following an analytic moving frame, we are able to ultimately characterize the set $\mathcal{S}_A$ of all reachable matrices from a given matrix A.

Theorem 4.1. *The asymptotic behavior of* $A(\mu)$ *where* $Q_A(\mu)$ *and* $R_A(\mu)$ *follow from the dynamical systems* (1.5) *and* (1.6) *with* K *specified as in* (3.4) *is given by*

$$\lim_{\mu \to \pm\infty} A(\mu) = D_\pm A D_\pm, \tag{4.1}$$

where

$$D_\pm = \lim_{\mu \to \pm\infty} Q_A(\mu) \tag{4.2}$$

is a diagonal matrix with diagonal entries either 1 *or* -1. *The inertia of the signature matrix* $D_\pm$ *is predetermined by the initial value, say,* $Q_A(0)$.

Proof. Write A and $Q_A(\mu)$ in columns,

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}, \quad Q_A(\mu) = \begin{bmatrix} \mathbf{q}_1(\mu) & \mathbf{q}_2(\mu) & \cdots & \mathbf{q}_n(\mu) \end{bmatrix},$$

and denote

$$R_A(\mu) = \begin{bmatrix} r_{11}(\mu) & r_{12}(\mu) & \cdots & r_{1n}(\mu) \\ & r_{22}(\mu) & \cdots & r_{2n}(\mu) \\ & & \ddots & \vdots \\ & & & r_{nn}(\mu) \end{bmatrix}.$$

Observe first that

$$\frac{r_{11}(\mu)}{\mu}\mathbf{q}_1(\mu) = \frac{\mathbf{a}_1}{\mu} - \mathbf{e}_1, \tag{4.3}$$

and

$$\lim_{|\mu|\to\infty} \left| \frac{r_{11}(\mu)}{\mu} \right| = \lim_{|\mu|\to\infty} \left\| \frac{\mathbf{a}_1}{\mu} - \mathbf{e}_1 \right\|_2 = 1. \tag{4.4}$$

Because $A - \mu I$ is nonsingular for all $\mu \in \mathbb{R}$ with $|\mu| > \|A\|_2$, the matrix $R_A(\mu)$ is also nonsingular for all $\mu \in \mathbb{R}$ with $|\mu| > \|A\|_2$. Consequently, $r_{11}(\mu) \neq 0$ for all $\mu \in \mathbb{R}$ with $|\mu| > \|A\|_2$. Together with the property that $R_A(\mu)$ is analytic, we see that $r_{11}(\mu)$ is analytic

and does not change its sign for all $\mu \in \mathbb{R}$ with $|\mu| > \|A\|_2$. Thus, it follows from (4.3) and (4.4) that $\lim_{\mu\to\pm\infty} \frac{r_{11}(\mu)}{\mu}$ exists and

$$\lim_{\mu\to\pm\infty} \mathbf{q}_1(\mu) = \pm\mathbf{e}_1, \tag{4.5}$$

where the sign in front of $\mathbf{e}_1$ depends on the initial value $\mathbf{q}_1(0)$.

We now proceed by induction. Assume it has been proven that

$$\lim_{|\mu|\to\infty} \begin{bmatrix} \mathbf{q}_1(\mu) & \cdots & \mathbf{q}_k(\mu) \end{bmatrix} = \begin{bmatrix} \pm\mathbf{e}_1 & \cdots & \pm\mathbf{e}_k \end{bmatrix}, \tag{4.6}$$

with appropriate signs attached to each limit points. Consider what will happen to $\mathbf{q}_{k+1}(\mu)$ as $|\mu| \to \pm\infty$. Observe that

$$\begin{bmatrix} \mathbf{q}_1(\mu) & \cdots & \mathbf{q}_k(\mu) \end{bmatrix}^\top \left(\frac{\mathbf{a}_{k+1}}{\mu} - \mathbf{e}_{k+1} \right) = \frac{1}{\mu} \begin{bmatrix} r_{1,k+1}(\mu) \\ \vdots \\ r_{k,k+1}(\mu) \end{bmatrix}, \tag{4.7}$$

$$\left\| \frac{1}{\mu} \begin{bmatrix} r_{1,k+1}(\mu) \\ \vdots \\ r_{k,k+1}(\mu) \\ r_{k+1,k+1}(\mu) \end{bmatrix} \right\|_2 = \left\| \frac{\mathbf{a}_{k+1}}{\mu} - \mathbf{e}_{k+1} \right\|_2. \tag{4.8}$$

Regardless of the signs involved in (4.6), we see from (4.7) that

$$\lim_{|\mu|\to\infty} \frac{1}{\mu} \begin{bmatrix} r_{1,k+1}(\mu) \\ \vdots \\ r_{k,k+1}(\mu) \end{bmatrix} = \begin{bmatrix} \pm\mathbf{e}_1 & \cdots & \pm\mathbf{e}_k \end{bmatrix}^\top \mathbf{e}_{k+1} = \mathbf{0},$$

and thus

$$\lim_{|\mu|\to\infty} \left| \frac{r_{k+1,k+1}(\mu)}{\mu} \right| = 1.$$

Similar to $r_{11}(\mu)$, $r_{k+1,k+1}(\mu)$ is analytic and does not change its sign for all $\mu \in \mathbb{R}$ with $\mu > \|A\|_2$, and since $R_A(\mu)$ is analytic and nonsingular for all $\mu \in \mathbb{R}$ with $\mu > \|A\|_2$, we get that $\lim_{\mu\to\pm\infty} \frac{r_{k+1,k+1}(\mu)}{\mu}$ exists. Furthermore, because

$$\frac{\mathbf{a}_{k+1}}{\mu} - \mathbf{e}_{k+1} = \begin{bmatrix} \mathbf{q}_1(\mu) & \cdots & \mathbf{q}_k(\mu) & \mathbf{q}_{k+1}(\mu) \end{bmatrix} \left(\frac{1}{\mu} \begin{bmatrix} r_{1,k+1}(\mu) \\ \vdots \\ r_{k,k+1}(\mu) \\ r_{k+1,k+1}(\mu) \end{bmatrix} \right), \tag{4.9}$$

we conclude that

$$\lim_{\mu\to\pm\infty} \mathbf{q}_{k+1}(\mu) = \pm\mathbf{e}_{k+1}, \tag{4.10}$$

with appropriate sign. By the induction principle, we have proved (4.2) and, hence, (4.1). ■

The flow $Q_A(\mu)$ defined by (1.5) is an analytic moving frame of n mutually orthogonal vectors. The orientation of the coordinate system, therefore, cannot be altered throughout the transformation. The issue of inertia involved in $D_\pm$ therefore is worth further investigation.

First, we note the symmetry of the differential equations that characterize the analytic QR field.

Lemma 4.2. *The skew-symmetric matrix $K(Q,R)$ defined by (3.4) is an even function in (Q,R). Consequently, the vector field given in (1.5) and (1.6) with K defined by (3.4) is an odd function in $Q_A(\mu)$ and $R_A(\mu)$.*

Proof. It is clear that $k_{1j} = \frac{q_{1j}}{r_{11}}$, $j = 2,\dots,n$, is even. Assume that k_{sj} is even for all $s = 1,\dots,i-1$ and $j = s+1,\dots,n$. By the definition (3.4), it is easy to see that k_{ij} is even for all $j = i+1,\dots,n$. By the induction principle, the result is proved. ∎

The symmetry described in Lemma 4.2 implies that the analytic QR factorization starting with $-Q_A(0)$ and $-R_A(0)$ is precisely the negative of the analytic QR factorization starting with $Q_A(0)$ and $R_A(0)$. Both positive and negative analytic paths produce the same $A(\mu)$. Consequently, the $\mathcal{S}_A$ is made of at most 2^{n-1} disjoint analytic segments.

Second, we note that from given initial values $Q_A(0)$ and $R_A(0)$ the limit point D_-AD_- is not necessarily the same as D_+AD_+, even if $\det(Q_A(0)) = 1$. More specifically, recall the fact

$$\lim_{\mu\to\pm\infty}\left|\frac{r_{ii}(\mu)}{\mu}\right| = 1, \quad i = 1,\dots,n-1, \tag{4.11}$$

which has been proved in Theorem 4.1. Under the assumption (3.3), $r_{ii}(\mu)$, $i = 1,\dots,n-1$, never vanishes for all μ. We conclude that

$$\lim_{\mu\to\infty}\frac{r_{ii}(\mu)}{\mu} = -\lim_{\mu\to-\infty}\frac{r_{ii}(\mu)}{\mu}, \quad i = 1,\dots,n-1, \tag{4.12}$$

together with the fact that

$$\lim_{\mu\to\pm\infty}\frac{r_{ii}(\mu)}{\mu}\mathbf{q}_i(\mu) = -\mathbf{e}_i, \quad i = 1,\dots,n-1. \tag{4.13}$$

We now see that

$$\lim_{\mu\to\infty}\mathbf{q}_i(\mu) = -\lim_{\mu\to-\infty}\mathbf{q}_i(\mu), \quad i = 1,\dots,n-1. \tag{4.14}$$

It remains only to examine what happens to $\mathbf{q}_n(\mu)$ as $\mu\to\pm\infty$.

Under assumption (3.3), $r_{ii}(\mu)\neq 0$ $(i = 1,\dots,n-1)$ for all $\mu\in\mathbb{R}$. But

$$|\det(A-\mu I)| = |\det(R_A(\mu))| = |r_{11}(\mu)\cdots r_{n-1,n-1}(\mu)r_{nn}(\mu)|,$$

so $r_{nn}(\mu) = 0$ for some $\mu\in\mathbb{R}$ if and only if this μ is a real eigenvalue of A. If n is even, then A has an even number, possibly zero, of real eigenvalues. That is, $r_{nn}(\mu)$ has an even number of real roots. Note that all eigenvalues of A are simple; we obtain that all real roots of $r_{nn}(\mu)$

are simple. This yields that $r_{nn}(\mu)$ changes sign at the real eigenvalues of A, and so it changes sign by an even number of times. It follows that

$$\lim_{\mu\to\infty} \frac{r_{nn}(\mu)}{\mu} = -\lim_{\mu\to-\infty} \frac{r_{nn}(\mu)}{\mu},$$

and hence

$$\lim_{\mu\to\infty} \mathbf{q}_n(\mu) = -\lim_{\mu\to-\infty} \mathbf{q}_n(\mu). \tag{4.15}$$

Together with (4.14), we have proved the following result.

Corollary 4.3. *Assume that the condition* (3.3) *holds. If n is even, then it is always true that $D_- = -D_+$. In this case, $\mathcal{S}_A$ is made of 2^{n-1} disjoint open loops with endpoint at $A_* = D_-AD_- = D_+AD_+$. (See Figure 3.) There is one and only one possibility such that $D_-AD_- = D_+AD_+ = A$, that is, D_+ is of one sign.*

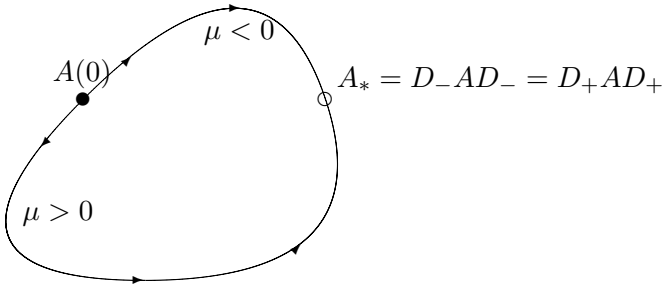


Figure 3. *A typical analytic component of $\mathcal{S}_A$ when n is even.*

When n is odd, the situation is different. Because $r_{nn}(\mu)$ changes its sign at least once and generally by an odd number of times, we have

$$\lim_{\mu\to\infty} \frac{r_{nn}(\mu)}{\mu} = \lim_{\mu\to-\infty} \frac{r_{nn}(\mu)}{\mu},$$

and hence

$$\lim_{\mu\to\infty} \mathbf{q}_n(\mu) = \lim_{\mu\to-\infty} \mathbf{q}_n(\mu). \tag{4.16}$$

We therefore have the following observation.

Corollary 4.4. *Assume that the condition* (3.3) *holds. If n is odd, then D_- agrees with D_+ only at the last diagonal entry. In this case, $\mathcal{S}_A$ is made of 2^{n-2} disjoint components, each of which is no longer a loop, but rather always has a positive distance between the endpoints D_-AD_- and D_+AD_+. (See Figure 4.)*

Be aware that the number of disjoint components of $\mathcal{S}_\mathcal{A}$ is reduced to 2^{n-2} if n is odd because the endpoints of each component are distinct.

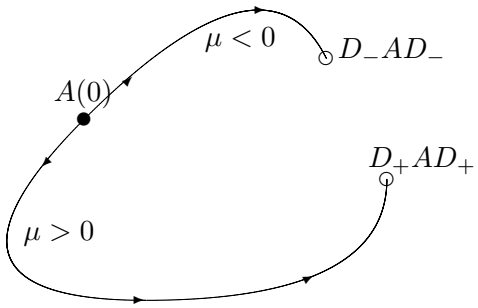


Figure 4. *A typical analytic component of $\mathcal{S}_A$ when n is odd.*

5. Conclusion. Employing the notion of analytic QR decomposition, we have explored the set of all isospectral matrices reachable via a QR step with shift. Several interesting facts are observed. One main discovery is that the reachable set from a given $n \times n$ matrix forms finitely many disjoint components, each of which can be characterized by a differential equation. Furthermore, each of the 2^{n-1} components when n is even appears as an open loop, whereas each of the 2^{n-2} components when n is odd does not form a loop and has a positive gap at the endpoints. This paper does not address the issue of more effective shift strategies but does offer some interesting insight into the structure of all reachable matrices. For example, it is shown in this paper that such a process is not symmetric; that is, in general $B \not\curvearrowleft A$ even if $A \curvearrowright B$.

The conventional shift strategy used in the eigenvalue computation gives rise to a sequence of isospectral matrices that are possibly leaping from one component of $A^{(k)}$ with a certain kind of orientation to another component of $A^{(k+1)} = A^{(k)}(\mu_k)$ with a different orientation, since no preservation of orientation is required in the QR algorithm. One possible further point of research might be to find the optimal μ so that $A^{(k)}(\mu)$ is closest to the desirable, say, triangular, form.

Acknowledgment. The problem of identifying all isospectral matrices reachable via QR iterations with shift was suggested to the second author by Graham Gladwell.

REFERENCES

[1] M. T. Chu, *The generalized Toda flow, the QR algorithm and the center manifold theorem*, SIAM J. Algebraic Discrete Methods, 5 (1984), pp. 187–210.
[2] M. T. Chu and G. H. Golub, *Structured inverse eigenvalue problems*, Acta Numer., 11 (2002), pp. 1–71.
[3] L. Dieci and T. Eirola, *On smooth decomposition of matrices*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 800–819.
[4] H. Gingold and P. F. Hsieh, *Globally analytic triangularization of a matrix function*, Linear Algebra Appl., 169 (1992), pp. 75–101.
[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[6] A. Iserles, H. Z. Munthe-Kass, S. P. Nørsett, and A. Zanna, *Lie-group methods*, in Acta Numerica 2000, Acta Numer. 9, Cambridge University Press, Cambridge, UK, 2000, pp. 215–365.
[7] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Classics Appl. Math. 20, SIAM, Philadelphia, 1997.

[8] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
[9] W. W. Symes, *The QR algorithm and scattering for the finite nonperiodic Toda lattice*, Phys. D, 4 (1981/82), pp. 275–280.
[10] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford University Press, New York, 1988.
[11] W. W. Wonham, *Linear Multivariable Control: A Geometric Control*, 3rd ed., Springer-Verlag, New York, 1985.
[12] K. Wright, *Differential equations for the analytic singular value decomposition of a matrix*, Numer. Math., 63 (1992), pp. 283–295.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 1, pp. 108–139

The Effects of Varying the Timing of Inputs on a Neural Oscillator*

Christina Ambrosio-Mouser†, Farzan Nadim‡, and Amitabha Bose§

Abstract. The gastric mill network of the crab *Cancer borealis* is an oscillatory neural network with frequency ~ 0.1 Hz. Oscillations in this network require neuromodulatory synaptic inputs as well as rhythmic inputs from the faster (~ 1 Hz) pyloric neural oscillator. We study how the frequency of the gastric mill network is determined when it receives rhythmic input from two different sources but where the timing of these inputs may differ. We find that over a certain range of the time difference one of the two rhythmic inputs plays no role what so ever in determining the network frequency, while in another range, both inputs work together to determine the frequency. The existence and stability of periodic solutions to model sets of equations are obtained analytically using geometric singular perturbation theory. The results are validated through numerical simulations. Comparisons to experiments are also presented.

Key words. synapse, stomatogastric ganglion, periodic orbit, Poincaré map

AMS subject classifications. 34A26, 37C25, 92B05

DOI. 10.1137/050625795

1. Introduction.

Many rhythmically active biological systems require input from extrinsic sources to produce their activity. Such extrinsic inputs may arrive as a trigger signal thereby switching on the oscillation, or be continuously present as forcing or feedback for the duration of the oscillation [17, 18, 27]. The extrinsic input itself is often rhythmic and its frequency may or may not match that of the target oscillator. Often, the oscillating network receives multiple inputs simultaneously—for example, a central command input and a sensory feedback input. Such inputs are readily identified in central pattern generators (CPGs), neural networks within the central nervous system (CNS) that are responsible for generating rhythmic motor behaviors such as locomotion, swimming, or breathing [13, 15, 17]. However, even in cases where the extrinsic inputs to a biological oscillator are known, the significance of the rhythmicity of such inputs or the consequences of having multiple inputs is largely unknown; often a nonrhythmic (tonic) input or only one of multiple inputs is sufficient to produce the biological oscillation. For instance, the CPG responsible for swimming in lower species of fish consists of chains of oscillators in the spinal cord [14]. This CPG receives rhythmic command input from the brain as well as multiple rhythmic and tonic sensory feedback inputs from the body. Without these inputs the CPG is inactive and the animal does not swim. However, the

*Received by the editors March 2, 2005; accepted for publication (in revised form) by D. Terman October 12, 2005; published electronically February 14, 2006. This work was supported in part by the National Science Foundation DMS-0315862 (AB, CA) and the National Institutes of Health MH-60605 (FN).
http://www.siam.org/journals/siads/5-1/62579.html
†Department of Mathematics, Medgar Evers College, Brooklyn, NY 11225 and Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 (cmouser@mec.cuny.edu).
‡Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 and Department of Biological Sciences, Rutgers University at Newark, Newark, NJ 07102 (farzan@njit.edu).
§Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 (bose@njit.edu).

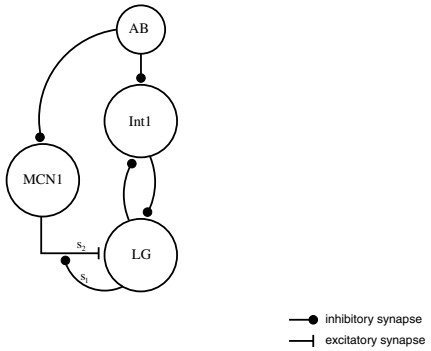


Figure 1. *Synaptic architecture of neurons associated with the gastric mill rhythm.*

isolated spinal cord, in the presence of a chemical stimulant (tonic input) produces "fictive swimming," rhythmic patterns which appear identical to those responsible for swimming in the intact animal [12].

In this paper, we use mathematical analysis to investigate the significance of the timing of multiple inputs to a biological oscillator. We focus on the rhythmically active crustacean gastric mill CPG located within the stomatogastric ganglion (STG). The STG is one of several ganglia in the crustacean CNS that control feeding and digestion behaviors [24]. As in many invertebrate systems, the small number of neurons in the STG makes it amenable to both experimental and modeling studies of neuronal networks responsible for generating behaviors.

The gastric mill rhythm is generated by a subset of neurons in the STG. At the heart of this rhythm are two neurons, lateral gastric (LG) and interneuron 1 ($Int1$), which make reciprocally inhibitory synaptic connections and (when the gastric mill rhythm is activated) oscillate in antiphase with a frequency of 0.1 Hz (see Figure 1). The gastric mill CPG, however, is a conditional oscillator; its activity depends on (modulatory) input from central command neurons located in other CNS ganglia [18, 24]. One such neuron, the modulatory commissural neuron 1 ($MCN1$), when excited, elicits a sustained gastric mill rhythm [8]. Previous studies have shown that input from the much faster pyloric CPG (frequency 1 Hz; also located in the STG) is crucial in setting the frequency of the $MCN1$-elicited gastric mill rhythm [3, 21]. This input comes in the form of an inhibitory synapse to $Int1$ from the pyloric pacemaker neuron anterior burster (AB). Although tonic excitation of $MCN1$ is sufficient for eliciting a gastric mill rhythm [3], it is known that, in the intact CNS, $MCN1$ is itself rhythmically active [30]. The rhythmicity of $MCN1$ is also due to a synaptic input it receives from AB. Thus, there are two pathways by which the pyloric pacemaker neuron AB influences the gastric mill network: the direct synaptic connection from AB to $Int1$ and the indirect synaptic influence exerted by inhibiting $MCN1$ which, in turn, excites the gastric mill neurons. Recent experimental findings show that the presence of either or both of the two pathways of AB influences results in a gastric mill rhythm of similar frequency [30]. However, when both pathways are removed (and thus the $MCN1$ activity is tonic) the gastric mill rhythm slows down significantly. Although the mechanism through which the direct pyloric input influences the gastric mill frequency is understood, it is not known how the rhythmic activity of $MCN1$ provides an apparently redundant mechanism for maintaining the gastric

mill frequency.

We expand on the techniques of Manor et al. [16] and build a simplified, biophysically based model of the gastric mill network incorporating the effects of both $MCN1$ and AB. Using this model, we show that the time difference between the two distinct AB influences on the gastric mill is critical in determining the gastric mill frequency. When m, the time delay between the AB inhibition of $Int1$ and $MCN1$, is small, the gastric mill frequency is determined solely by direct modulatory effects of $MCN1$ on the gastric mill neurons. However, when m is large, the gastric mill rhythm operates at a higher frequency and is determined by both the $MCN1$ and AB inputs. Throughout this work, we use geometric singular perturbation theory to examine the behavior of our model on low-dimensional manifolds. We also use phase plane analysis as a means of geometrically understanding the behavior of the network. Using these mathematical tools, we prove the existence and stability of periodic solutions of the model when the activity of $MCN1$ is either tonic or rhythmic. Furthermore, we confirm our findings through numerical simulations in which we change the delay parameter m and numerically calculate the period. We then compare our results with the experimental findings of [30]. The results from our model match the experimental results only when m is either 0 or small. We note that in the work of Wood et al., the delay between the pyloric and modulatory inputs is never explicitly measured. Our analysis, therefore, gives a possible explanation for the experimental results. The findings could be tested in future experiments by artificially introducing different delays between the timing of the modulatory and pyloric inputs to the gastric mill network and determining the effect on the gastric mill rhythm frequency.

The remainder of the paper is organized as follows. In section 2, we derive a model set of equations and show how it can be reduced into sets of fast and slow equations. The effect of various synaptic inputs on relevant nullclines are illustrated with the goal of showing how fixed points of the fast set of equations depend on these inputs. In section 3, we prove the existence, uniqueness, and stability of periodic solutions for four different cases of the gastric mill rhythm. These cases are considered so that we may parallel the study of Wood et al. [30]. In section 4, we derive an analytic formula for the period of the solutions for the four different cases. We then check our analytic results against simulations. In section 5, we discuss the impact of making a certain synapse (that from $MCN1$ to LG) voltage dependent. Section 6 contains a discussion to conclude the paper.

2. Model.

Our model consists of the gastric mill network composed of $Int1$ and LG, the pacemaker neuron AB of the pyloric network, and the modulatory commissural neuron $MCN1$; see Figure 1. LG and $Int1$ are modeled as passive neurons (LG having a subthreshold resting potential and $Int1$ having a suprathreshold resting potential). In the absence of AB input, $MCN1$ is tonically active. Therefore, we also model $MCN1$ as a passive neuron with a suprathreshold resting potential. LG and $Int1$ have reciprocally inhibitory synapses between them. $Int1$ and $MCN1$ receive inhibitory input from the pacemaker neuron AB. AB and $MCN1$ both fire in pyloric time with a period denoted P_{AB}. AB sends an inhibitory synapse to $Int1$. $MCN1$ sends a slow excitatory synapse to LG. In addition, LG presynaptically inhibits $MCN1$ each time it fires, thus removing the excitation from $MCN1$ to LG. Through this circuitry, the voltage of LG is able to increase above threshold (due to the excitation it receives from $MCN1$), causing $Int1$ to become suppressed, and then decay back below

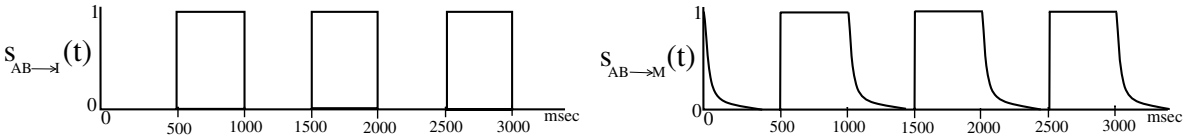


Figure 2. *The synaptic variables $s_{AB\to I}(t)$ and $s_{AB\to M}(t)$. Note that $s_{AB\to M}(t)$ decays with time constant τ_{AB}.*

threshold (due to the presynaptic inhibition), thereby producing the antiphase oscillations of the gastric mill rhythm [9, 16].

2.1. Equations. We do not explicitly model the pacemaker neuron AB but instead incorporate its effect on $MCN1$ and $Int1$ through the synaptic variables $s_{AB\to I}(t)$ and $s_{AB\to M}(t)$. $s_{AB\to I}(t)$ is a square wave with amplitude 1 and period, P_{AB}, which has experimentally been found to be approximately 1 *sec*. Let D_c denote the duty cycle of AB (the ratio of its active time to its period). During one period of AB, the variable $s_{AB\to I}$ is equal to 1 for a time $D_c P_{AB}$ and equal to 0 for a time $P_{AB}[1 - D_c]$. $s_{AB\to M}(t)$ is similar in form to $s_{AB\to I}(t)$ in that $s_{AB\to M}$ oscillates between 0 and 1. The jump in $s_{AB\to M}$ from 0 to 1 is instantaneous. However, $s_{AB\to M}$ decreases from 1 to 0 with time constant τ_{AB} (see Figure 2).

The dynamics of the system evolve along two distinct timescales. One is a slow timescale corresponding to the effect of presynaptic inhibition from LG to the slow excitatory component of the $MCN1$ synapse. The other is a fast timescale along which all other synapses and intrinsic properties evolve. We use a small parameter, ϵ, to demarcate these two timescales.

The equations to describe the activity of LG, $Int1$, and $MCN1$ are

(1)
$$\epsilon \frac{dV_L}{dt} = -g_{leak,L}[V_L - E_{leak,L}] - \bar{g}_{I\to L} n_\infty(V_I)[V_L - E_{I\to L}] - g_s(V_L)s(t)[V_L - E_{exc}],$$

(2)
$$\epsilon \frac{dV_I}{dt} = -g_{leak,I}[V_I - E_{leak,I}] - \bar{g}_{L\to I} n_\infty(V_L)[V_I - E_{L\to I}] - \bar{g}_{AB\to I} s_{AB\to I}(t)[V_I - E_{AB\to I}],$$

(3)
$$\epsilon \frac{dV_M}{dt} = -g_{leak,M}[V_M - E_{leak,M}] - \bar{g}_{AB\to M} s_{AB\to M}(t)[V_M - E_{AB\to M}],$$

where V_L is the voltage of LG, V_I is the voltage of $Int1$, and V_M is the voltage of $MCN1$. $g_{leak,L}$, $g_{leak,I}$, and $g_{leak,M}$ are the conductances of the leak currents in LG, $Int1$, and $MCN1$. $E_{leak,L}$, $E_{leak,I}$, and $E_{leak,M}$ are the reversal potentials of the leak currents in LG, $Int1$, and $MCN1$. Denote the right-hand sides of (1) and (2) by $f(V_L, V_I, s)$ and $g(V_I, V_L, s_{AB\to I})$, respectively.

The parameters of the reciprocally inhibitory synapses between $Int1$ and LG are $\bar{g}_{I\to L}$ and $\bar{g}_{L\to I}$ (the maximal conductances) and $E_{I\to L}$ and $E_{L\to I}$ (the reversal potentials). $n_\infty(V_I)$ and $n_\infty(V_L)$ are sigmoidally shaped gating functions lying between 0 and 1:

(4)
$$n_\infty(V_x) = \left(1 + \exp\frac{v_x - V_x}{k_x}\right)^{-1},$$

where v_x is the half-activation voltage and k_x is inversely related to the slope at this point.

The fast, periodic inhibitory input from AB to $MCN1$ is described by $\bar{g}_{AB\to M}s_{AB\to M}(t)[V_M - E_{AB\to M}]$, where $\bar{g}_{AB\to M}$ is the conductance of the synapse and $E_{AB\to M}$ is the reversal potential which is chosen to be less than $E_{leak,M}$ so that the input from AB to $MCN1$ is inhibitory. The equation to describe the activity of $s_{AB\to M}(t)$ with respect to AB is

$$\epsilon\frac{ds_{AB\to M}}{dt} = \begin{cases} [1 - s_{AB\to M}]/\tau_{M_1}, & V_{AB} \geq V_{Th(AB)}, \\ -\epsilon s_{AB\to M}/\tau_{M_2}, & V_{AB} < V_{Th(AB)}, \end{cases} \tag{5}$$

where V_{AB} is a square wave with period P_{AB} and duty cycle $= D_c$. Thus, when $s_{AB\to M} = 0$, V_M lies at a maximum voltage of $E_{leak,M}$. When $s_{AB\to M} = 1$, V_M lies at a minimum voltage of V_M^*, where

$$V_M^* = \frac{g_{leak,M}E_{leak,M} + \bar{g}_{AB\to M}E_{AB\to M}}{g_{leak,M} + \bar{g}_{AB\to M}}. \tag{6}$$

The periodic, inhibitory input from AB to $Int1$ is given by $\bar{g}_{AB\to I}s_{AB\to I}(t)[V_I - E_{AB\to I}]$. $\bar{g}_{AB\to I}$ is the conductance and $E_{AB\to I}$ is the reversal potential. An important aspect of this work is to highlight the fact that different timing relationships of AB input to $MCN1$ and $Int1$ lead to dramatically different frequencies of the gastric mill rhythm. To this end, we will use the parameter, m, to delay the AB input to $Int1$ relative to $MCN1$. In other words, if the AB input to $MCN1$ turns on at $t = 0$, the input from AB to $Int1$ does not turn on until $t = m$. The parameter m is a constant which can range between 0 and P_{AB} (the period of AB).

The effect of the excitation that LG receives from $MCN1$ is given in (1) by $g_s(V_L)s(t)[V_L - E_{exc}]$, where $g_s(V_L) = \bar{g}_s s_\infty(V_L)$ is the voltage dependent conductance of the synapse, E_{exc} is the reversal potential, and $s(t)$ models the amount of excitation LG receives. The function s_∞ is a sigmoidal gating function similar in form to n_∞; its exact form will be discussed later in section 5. We express $s(t) = s_1(t)s_2(t)$ as the product of two different effects. $s_1(t)$ models the effect of the presynaptic inhibition of the slow excitatory component of the $MCN1$ to LG synapse. $s_2(t)$ models the effect of the AB inhibition of $MCN1$ on the fast excitatory component of the $MCN1$ to LG synapse. The relevant equations are

$$\frac{ds_1}{dt} = \begin{cases} [1 - s_1]/\tau_{r_1}, & V_L \leq V_T, \\ -s_1/\tau_{f1}, & V_L > V_T, \end{cases} \tag{7}$$

$$\epsilon\frac{ds_2}{dt} = \begin{cases} [1 - s_2]/\tau_{r2}, & V_M \geq V_{Th(M)}, \\ [s_{2min} - s_2]/\tau_{f2}, & V_M < V_{Th(M)}. \end{cases} \tag{8}$$

The time constants τ_{M_1}, τ_{M_2}, τ_{r1}, τ_{f1}, τ_{r2}, and τ_{f2} in (5), (7), and (8) are $O(1)$ with respect to ϵ. When V_L goes above threshold, the presynaptic inhibition turns on, which causes $s_1(t)$ to decrease on the slow timescale. When V_L goes below threshold, the presynaptic inhibition turns off, which causes $s_1(t)$ to increase on the slow timescale (see Figure 1). When $MCN1$ is inhibited by AB, s_2 decreases on the fast timescale. Once the inhibition from AB is removed, s_2 increases on the fast timescale. The parameters $V_{Th(M)}$ and V_T denote the activation thresholds for these two synapses.

To understand which parameters are important in controlling the gastric mill frequency, we use phase plane analysis along with geometric singular perturbation theory to reduce the full flow to a study of flow on lower-dimensional slow manifolds. From (1)–(3) and (7)–(8), we see that V_L, V_I, V_M, and s_2 evolve on a faster timescale than s_1. Setting $\epsilon = 0$ in the equations yields the slow equations

$$0 = f(V_L, V_I, s), \tag{9}$$

$$0 = g(V_I, V_L, s_{AB\to I}), \tag{10}$$

$$0 = -g_{leak,M}[V_M - E_{leak,M}] - \bar{g}_{AB\to M} s_{AB\to M}(t)[V_M - E_{AB\to M}], \tag{11}$$

$$0 = [1 - s_{AB\to M}]/\tau_{M_1}, \quad V_{AB} \geq V_{Th(AB)}, \tag{12}$$

$$\frac{ds_{AB\to M}}{dt} = -s_{AB\to M}/\tau_{M_2}, \quad V_{AB} < V_{Th(AB)}, \tag{13}$$

$$\frac{ds_1}{dt} = \begin{cases} [1 - s_1]/\tau_{r1}, & V_L \leq V_T, \\ -s_1/\tau_{f1}, & V_L > V_T, \end{cases} \tag{14}$$

$$0 = \begin{cases} [1 - s_2]/\tau_{r2}, & V_M \geq V_{Th(M)}, \\ [s_{2min} - s_2]/\tau_{f2}, & V_M < V_{Th(M)}. \end{cases} \tag{15}$$

The set of points satisfying $f(V_L, V_I, s) = 0$ and $g(V_I, V_L, s_{AB\to I}) = 0$ are called the V_L and V_I nullclines, respectively. In slow time, (9) and (10) imply that any trajectory is forced to lie on the V_L and V_I nullclines while s_1 slowly evolves and s_2 instantaneously jumps between s_{2min} and 1 whenever V_M crosses the threshold $V_{TH(M)}$. In slow time, $s_{AB\to M}$ jumps to 1 whenever V_{AB} increases above $V_{Th(AB)}$ and decays exponentially to 0 whenever V_{AB} decreases below $V_{Th(AB)}$. Note that this slow decay of $s_{AB\to M}$ implies from (11) that the voltage V_M slowly increases from V_M^* toward $E_{leak,M}$. We choose $V_{TH(M)}$ to lie between these two values such that the time for V_M to increase from V_M^* to $V_{TH(M)}$ equals a predetermined time called T_C.

Note that when $\bar{g}_{AB\to M} = 0$, then V_M is always greater than $V_{Th(M)}$, and we refer to $MCN1$ as being tonically active. In this case, $s_2 = 1$. Alternatively, when $\bar{g}_{AB\to M}$ is sufficiently large, then V_M goes above and below $V_{Th(M)}$ in pyloric time and we say that $MCN1$ is rhythmically active. In this case, s_2 jumps between s_{2min} and 1 each time V_M crosses $V_{Th(M)}$. By choosing τ_{r_1} and τ_{f_1} small relative to P_{AB}, we note that V_M may cross $V_{Th(M)}$ several times before V_L crosses V_T.

To define fast equations, let $\zeta = t/\epsilon$ in (1)–(3), (5), and (7)–(8); then set $\epsilon = 0$ to obtain

$$\frac{dV_L}{d\zeta} = f(V_L, V_I, s), \tag{16}$$

$$\frac{dV_I}{d\zeta} = g(V_I, V_L, s_{AB\to I}), \tag{17}$$

$$\frac{dV_M}{d\zeta} = -g_{leak,M}[V_M - E_{leak,M}] - \bar{g}_{AB\to M} s_{AB\to M}(t)[V_M - E_{AB\to M}], \tag{18}$$

$$\frac{ds_{AB\to M}}{d\zeta} = \begin{cases} [1 - s_{AB\to M}]/\tau_M, & V_{AB} \geq V_{Th(AB)}, \\ 0, & V_{AB} < V_{Th(AB)}, \end{cases} \tag{19}$$

$$\frac{ds_2}{d\zeta} = \begin{cases} [1 - s_2]/\tau_{r2}, & V_M \geq V_{Th(M)}, \\ [s_{2min} - s_2]/\tau_{f2}, & V_M < V_{Th(M)}, \end{cases} \tag{20}$$

$$\frac{ds_1}{d\zeta} = 0. \tag{21}$$

Therefore, in fast time V_L and V_I evolve according to the dynamics of $f(V_L, V_I, s)$ and $g(V_I, V_L, s_{AB\to I})$, and s_2 increases and decreases between 1 and s_{2min} while s_1 remains constant. These equations govern transitions between the different branches of the V_I and V_L nullclines. These transitions occur instantaneously with respect to the slow flow.

2.2. Geometry of nullclines. In the previous subsection, we derived reduced fast and slow equations which govern the flow of trajectories in relevant phase spaces. For the slow equations (9)–(15), the slow variable s_1 evolves according to (14), while the activity of the fast variables V_L, V_I, V_M, and s_2 is constrained through the algebraic equations (9), (10), (11), and (15). Note that $s_{AB\to M}$ is a fast variable whenever $V_{AB} \geq V_{Th(AB)}$ according to (13) and is a slow variable whenever $V_{AB} < V_{Th(AB)}$ according to (12). For the fast equations (16)–(21), the slow variables act as parameters. Fixed points of the fast equations correspond to situations where the V_I and V_L nullclines intersect. We will be interested in situations where the existence and stability of fixed points of the fast subsystem change. Generally speaking, this may occur because the slow variable s_1 causes the fast system to undergo a saddle-node bifurcation, or if s_{AB} changes, causing the position of the nullclines to change on the fast timescale in phase space. Below, we show geometrically how these situations may arise.

We shall first consider the case when the synapse from $MCN1$ to LG is not voltage dependent. We do this by letting $s_\infty(V_L) \equiv 1$. The effect of the voltage dependency of this synapse is considered in section 5. We find the explicit equations for the nullclines by solving (9) for V_L and (10) for V_I to find that

$$V_L = F(V_I, s) = \frac{g_{leak,L}E_{leak,L} + \bar{g}_{I\to L}n_\infty(V_I)E_{I\to L} + \bar{g}_s(V_L)sE_{exc}}{g_{leak,L} + \bar{g}_{I\to L}n_\infty(V_I) + \bar{g}_s(V_L)s} \tag{22}$$

and

$$V_I = G(V_L, s_{AB\to I}) = \frac{g_{leak,I}E_{leak,I} + \bar{g}_{L\to I}n_\infty(V_L)E_{L\to I} + \bar{g}_{AB\to I}s_{AB\to I}(t)E_{AB\to I}}{g_{leak,I} + \bar{g}_{L\to I}n_\infty(V_L) + \bar{g}_{AB\to I}s_{AB\to I}(t)}. \tag{23}$$

A simultaneous solution to (9) and (10) can be found graphically by plotting $F(V_I, s)$ versus $G(V_L, s_{AB\to I})$. An intersection of these two nullclines corresponds to a fixed point of the fast equations. However, the position of the nullclines in $V_I - V_L$ phase space changes as a function of s and s_{AB}; see Figure 3. In general, increases (decreases) in s move the V_L nullcline to the right (left), either in slow time due to changes in s_1 or in fast time due to changes in s_2. The V_I nullcline has two possible positions in phase space depending on whether $s_{AB\to I} = 0$ or 1. The nullcline corresponding to $s_{AB\to I} = 1$ is lower in phase space than the one corresponding to $s_{AB\to I} = 0$. We note that the left branch of the V_I nullcline shifts down much more than the right branch since on the right branch V_I is already close to $E_{AB\to I}$ independent of $s_{AB\to I}(t)$. The number and stability of fixed points also changes as a function of s and $s_{AB\to I}$. We identify four important values of $(s, s_{AB\to I})$ as $(s_L^{off}, 0)$, $(s_L^{on}, 1)$,

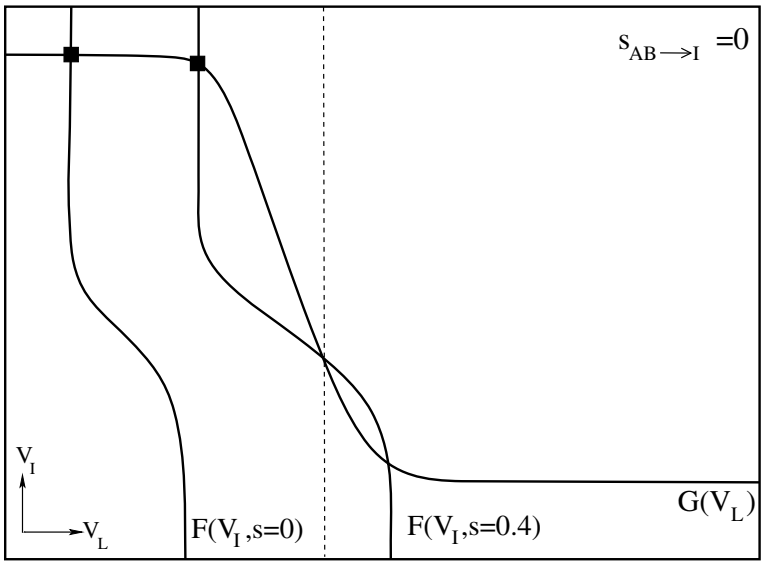


Figure 3. *The V_L and V_I nullclines plotted in phase space for two different values of $s(t)$. The V_L nullcline is labeled F, the V_I nullcline is labeled G, and the dashed vertical line is $V_L = V_T$. When $s = 0$, the V_L nullcline is to the far left. As s increases, the V_L nullcline shifts to the right. Solid squares denote fixed points of the fast equations (16)–(21).*

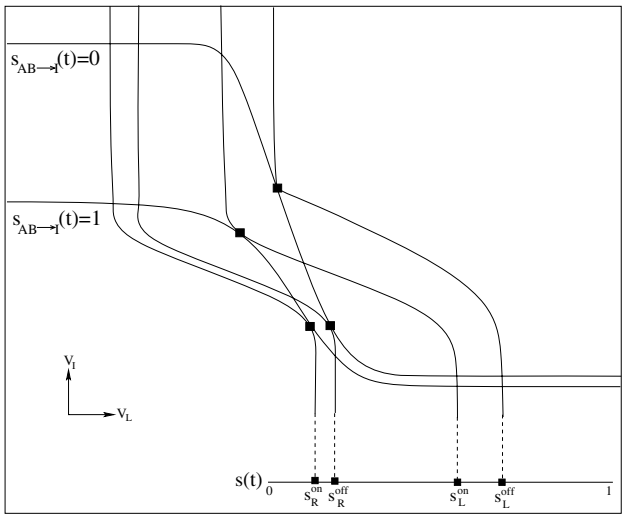


Figure 4. *Position of the V_L and V_I nullclines for $(s, s_{AB\to I}) = (s_R^{on}, 1)$, $(s_R^{off}, 0)$, $(s_L^{on}, 1)$, and $(s_L^{off}, 0)$. At these four points, the nullclines intersect tangentially, resulting in the loss (or gain) of two fixed points through a saddle-node bifurcation.*

$(s_R^{off}, 0)$, and $(s_R^{on}, 1)$; see Figure 4. The superscripts *off* and *on* refer to the AB input to $Int1$ which can be either absent (*off*) or present (*on*). These points correspond to values when the two nullclines intersect tangentially resulting in the loss (or gain) of two fixed points through a saddle-node bifurcation. Because s_1 is increasing on the left branches and decreasing on

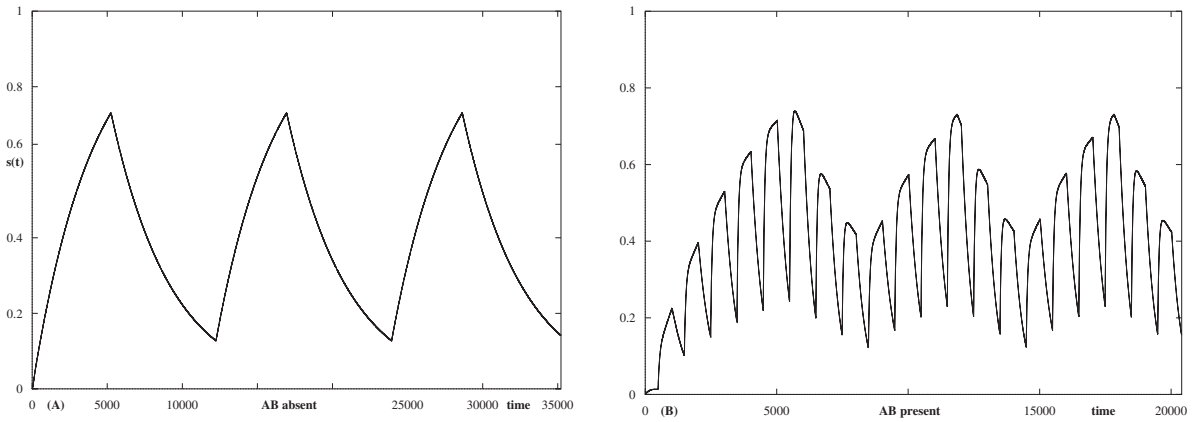


Figure 5. (A) *s versus t for the case when the $MCN1$ to LG excitation is constant ($s_2 = 1$). s increases to 1 with rate $1/\tau_{r1}$ when $V_L \leq V_T$ and s decreases to 0 with rate $1/\tau_{f1}$ when $V_L > V_T$.* (B) *s versus t for the case when the $MCN1$ to LG excitation is rhythmic. When $V_L \leq V_T$, s_1 increases to 1 with rate $1/\tau_{r1}$ and s_2 jumps between 1 and s_{2min} when $s_{AB\to M}(t)$ jumps between 0 and 1. When $V_L > V_T$, s_1 decreases to 0 with rate $1/\tau_{f1}$ and s_2 jumps between 1 and s_{2min} when $s_{AB\to M}(t)$ jumps between 0 and 1.*

the right, the ordering of these bifurcation points is $s_R^{on} < s_R^{off} < s_L^{on} < s_L^{off}$. These values can be calculated analytically; see the appendix. For our numerical simulations, we chose the parameters such that $s_R^{on} = .11$, $s_R^{off} = .127$, $s_L^{on} = .31$, and $s_L^{off} = .73$.

On the slow timescale, the solution trajectory must lie at the intersection of the V_I and V_L nullclines, i.e., at a fixed point of the fast subsystem. Thus to understand the evolution of trajectories in the $V_I - V_L$ phase space, we need to understand how the position of fixed points changes as a function of s and s_{AB}.

Let us first consider the case when $s_{AB\to M}(t) \equiv 0$. Then $MCN1$ is tonically active and sits at a value of $E_{leak,M}$. $E_{leak,M}$ is chosen to be larger than $V_{Th(M)}$ which, as we see from (15), allows $s_2 = 1$. As a result, we have $s(t) = s_1(t) * 1$ which means that $s(t)$ increases toward 1 with time constant τ_{r1} when $V_L < V_T$ and decreases toward 0 with time constant τ_{f1} when $V_L > V_T$ (see Figure 5(A)).

From (22), as s slowly increases, the V_L nullcline slowly shifts to the right, thus causing the position of the stable fixed point to shift to the right. This continues until the V_L nullcline shifts far enough to the right so that the stable fixed point on the left branches of the nullclines is lost through a saddle-node bifurcation when $s = s_L^{off}$; see Figure 6. Once the fixed point is lost, the trajectory is forced to jump on the fast timescale (see (16)–(17)) to the only remaining stable fixed point which is on the right branches of the nullclines. This jump pushes V_L above V_T causing $s(t)$ to begin to decrease. When s decreases, the V_L nullcline slowly shifts to the left until the stable fixed point on the right branches of the V_L and V_I nullclines similarly undergoes a saddle-node bifurcation at $s = s_R^{off}$. The trajectory then makes a jump back to the left branches of the nullclines which forces V_L below V_T. Similar dynamics occur when $s_{AB\to I}(t) \equiv 1$, except now the trajectory would pass through the bifurcation points s_L^{on} and s_R^{on} during its transition between left and right branches.

When the $MCN1$ excitation to LG is rhythmic instead of tonic, s_2 changes on the fast timescale between 1 and s_{2min} as V_M crosses over $V_{Th(M)}$ while s_1 increases toward 1 when

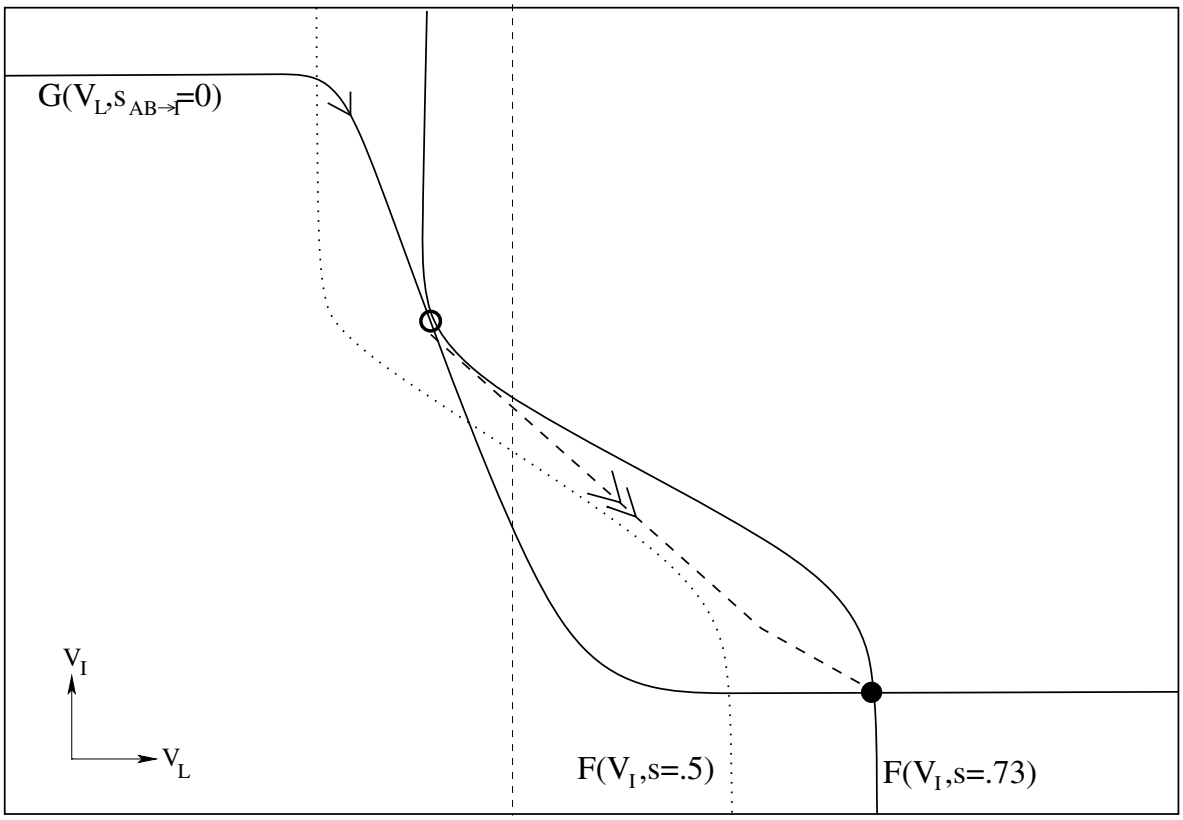


Figure 6. *V_L and V_I nullclines for different values of s when $s_{AB\to I}(t) = 0$. The dashed vertical line is $V_L = V_T$. When $s = 0.5$, the fixed point is on the left branches of the V_L and V_I nullclines (V_L nullcline dotted). When $s = s_L^{off}$, the fixed point undergoes a saddle-node bifurcation at $\circ$. The trajectory is therefore forced to jump (dashed line with double arrows) to the stable fixed point (shown by $\bullet$) on the right branches of the nullclines where $V_L > V_T$.*

$V_L \le V_T$ and decreases toward 0 when $V_L > V_T$ on the slow timescale. This causes $s(t)$ to generally have a shape as shown in Figure 5(B). Notice that the envelope of $s(t)$ activity is the same as in the tonic excitability case seen in Figure 5(A), but now there are rapid changes in $s(t)$ due to the rapid changes in $s_2(t)$. The jump of s_2 between s_{2min} and 1 causes the V_L nullcline to instantaneously jump to the right when s_2 jumps to 1 and instantaneously jump to the left when s_2 jumps to s_{2min}. The distance of these jumps in the V_L nullcline, calculated from (22), is $F(V_I, s_1 * 1) - F(V_I, s_1 * s_{2min})$. Note that in the $MCN1$ rhythmic case, fixed points can be lost in two different ways. They may be lost as before through a saddle-node bifurcation as s is slowly changing due to changes in s_1; see Figure 7(A). Or they may be lost when s_2 changes on the fast timescale. For example, on the left branches, it may be that $s_1 * s_{2min} < s_L^{off}$, but $s_1 * 1 > s_L^{off}$. In this case, the fixed point would be lost if s_{2min} changed to 1 due to a change in $MCN1$ activity; see Figure 7(B). On the right branches, it may be that $s_1 * 1 > s_R^{off}$, but $s_1 * s_{2min} < s_R^{off}$. In this case, the fixed point would be lost when s_2 changes from 1 to s_{2min}.

In the case where $s_{AB\to I}(t)$ is a square wave, the trajectory will always lie on a nullcline

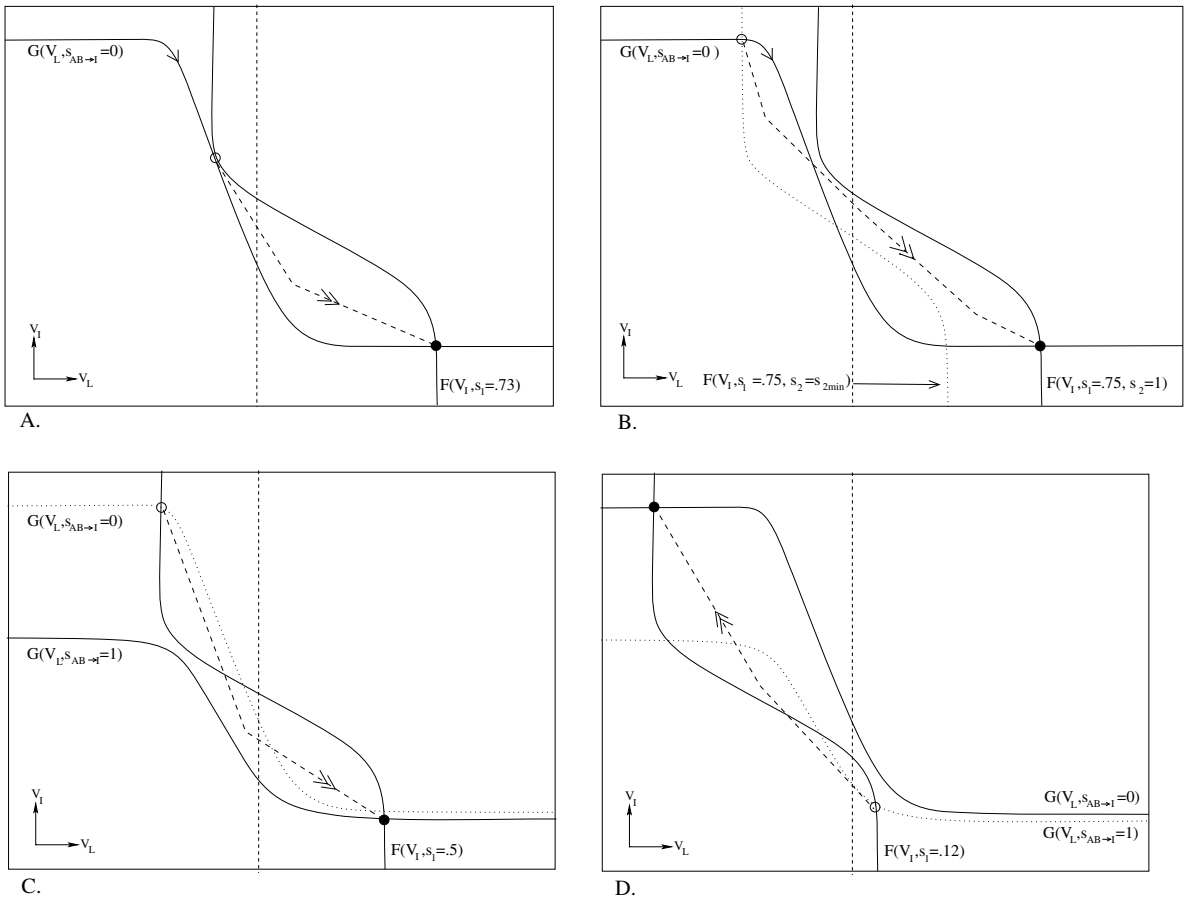


Figure 7. *For $MCN1$ rhythmic, fixed points can be lost in two ways: through a saddle-node bifurcation as s slowly changes due to s_1 or when s_2 changes on the fast timescale. In all figures, the dashed vertical line is $V_L = V_T$.* (A) *On the left branches of the nullclines, $s_2 = 1$ and s_1 moves the V_L nullcline to the right resulting in a saddle-node bifurcation of the fixed point at $\circ$ once $s = s_L^{off}$. The trajectory is forced to jump to the stable fixed point on the right branches of the nullclines (shown by $\bullet$). A similar transition can also occur when $s_{AB\to I} = 1$ and s_1 reaches s_L^{on}.* (B) *When s_2 jumps from s_{2min} to 1, the V_L nullcline jumps from the left (dotted nullcline) to the right (solid nullcline). Thus the fixed point on the left branches of the nullclines is instantaneously lost because $s > s_L^{off}$ and the trajectory will jump to the fixed point on the right branches of the nullclines (shown by $\bullet$). A similar transition can also occur when $s_{AB\to I} = 1$ and s_2 jumps to 1 such that $s > s_L^{on}$.* (C) *When $MCN1$ is rhythmic and $s_{AB\to I}(t)$ oscillates between 0 and 1, the fixed point can be lost in another way. While $s_2 = 1$ and $s_{AB\to I}(t)$ jumps from 0 to 1 (the position of the V_I nullcline jumps from the upward dotted nullcline to the lower solid nullcline), the fixed point on the left branches of the nullclines (shown by $\circ$) is instantaneously lost because $s > s_L^{on}$ and the solution trajectory is forced to jump to the fixed point on the right branches of the nullclines (shown by $\bullet$).* (D) *On the right branches of the nullclines for $MCN1$ tonic, when $s_{AB\to I}$ jumps from 1 to 0 (the position of the V_I nullcline jumps from the lower dotted nullcline to the upward solid nullcline), the fixed point (shown by $\circ$) is instantaneously lost because $s_R^{on} < s < s_R^{off}$ and the solution trajectory must jump to the stable fixed point on the left branches of the nullclines (shown by $\bullet$).*

with either $s_{AB\to I} = 0$ or $s_{AB\to I} = 1$. Now fixed points can be lost in three different ways. Consider the left branches. As before, a fixed point can be lost as s increases slowly through a bifurcation point or instantaneously as s_2 changes from s_{2min} to 1. The third way it can be

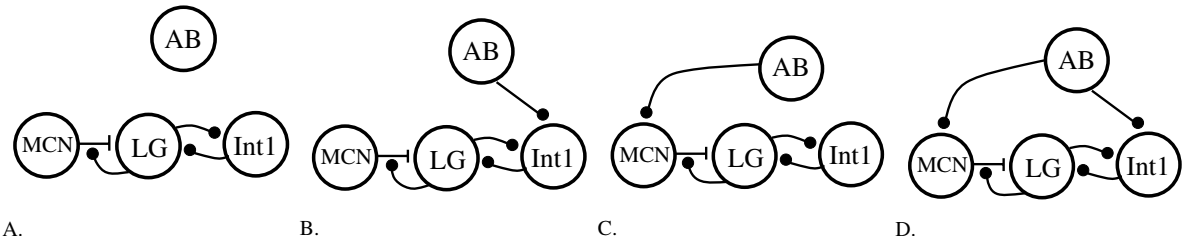


Figure 8. *Circuitry for cases* 1–4. (A) *The AB inputs to* $MCN1$ *and to* $Int1$ *are absent.* (B) *The AB input to* $Int1$ *is present but the AB input to* $MCN1$ *is absent.* (C) *The AB input to* $MCN1$ *is present but the AB input to* $Int1$ *is absent.* (D) *The AB inputs to* $Int1$ *and to* $MCN1$ *are present.*

lost is if $s_L^{off} > s > s_L^{on}$ and $s_{AB\to I}$ switches from 0 to 1; see Figure 7(C) (see Figure 7(D) for the analogous loss of a fixed point on the right branches).

3. Different cases for the gastric mill frequency. To understand how the two different inputs of AB and $MCN1$ modulate the gastric mill frequency, we parallel the study of Wood et al. [30] by considering four different cases (see Figure 8):

Case 1. Tonic $MCN1$ excitation with the AB input to $Int1$ absent.

Case 2. Tonic $MCN1$ excitation with the AB input to $Int1$ present.

Case 3. Rhythmic $MCN1$ excitation with the AB input to $Int1$ absent.

Case 4. Rhythmic $MCN1$ excitation with the AB input to $Int1$ present.

In each case, we shall prove the existence, local uniqueness, and stability of a periodic solution and then calculate the period of this solution. The proofs of existence, local uniqueness, and stability of periodic solutions exploit the different timescales. In Cases 2 through 4, this will reduce to finding fixed points of appropriate one-dimensional maps. The proofs construct singular periodic solutions which are valid at $\epsilon = 0$, whose existence, local uniqueness, and stability can be extended to the ϵ sufficiently small case [20].

In the biological circuit it is known that the synapse from $MCN1$ to LG is dependent on the voltage of LG. For mathematical clarity, we shall postpone discussing the voltage dependent case until section 5. Instead, we shall first concentrate on the voltage independent case where we set $s_\infty(V_L) \equiv 1$.

Case 1: *Tonic* $MCN1$ *excitation with the* AB *input to* $Int1$ *absent.* When considering Case 1, we set $\bar{g}_{AB\to I} = 0$ and $\bar{g}_{AB\to M} = 0$ in (2) and (3) so that all input from AB is absent. When $\bar{g}_{AB\to M} = 0$, $V_M > V_{Th(M)}$ for all t and $MCN1$ is tonically active. This allows us to set $s_2 \equiv 1$ so that $s = s_1 * s_2$ will only follow the dynamics of s_1.

In Case 1, the only way a fast transition between branches can occur is by s_1 passing through the bifurcation points s_L^{off} or s_R^{off}. To construct a periodic solution, let $s_1(0) = s_R^{off}$ such that the trajectory at $t = 0^-$ is at the bifurcation point on the right branches at the intersection of the V_I and V_L nullclines; see Figure 9. At $t = 0^+$, the trajectory jumps back to the left branches at the intersection of the nullclines; see Figure 9(A). On these branches, $V_L < V_T$ and thus s_1 will increase until it reaches the bifurcation point s_L^{off} at $t = T_1$; see Figure 9(B)–(C). The trajectory will jump back to the right branch and since $V_L > V_T$, s_1 will now decrease until it comes back to s_R^{off} at $t = T_2$; see Figure 9(D). Thus the value of s_1 will have returned to its original value at time T_2. Since all the fast variables are slaved

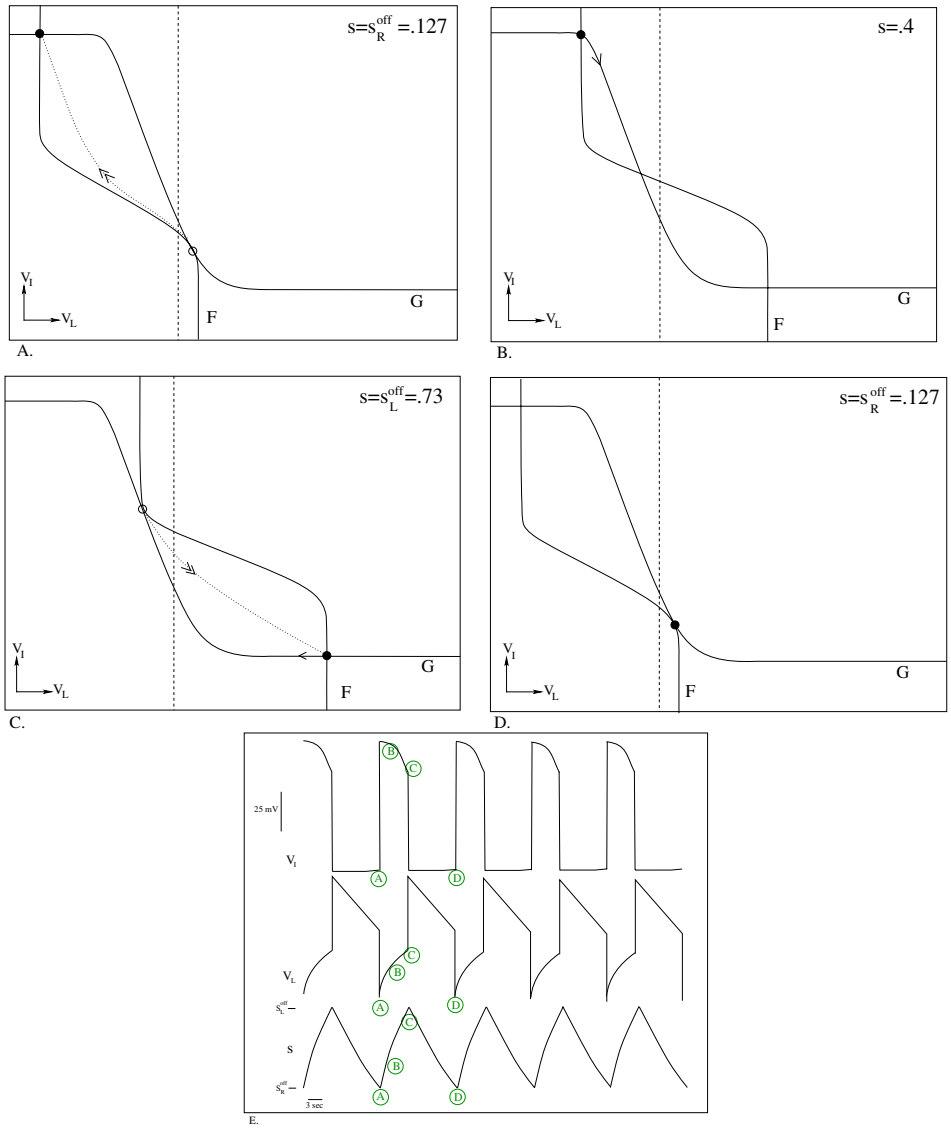


Figure 9. *Plots of the V_I and V_L nullclines for different values of s in Case* 1. • *marks the position of the trajectory when it is at a stable fixed point,* ∘ *marks the point from which the trajectory will jump when the stable fixed point bifurcates, and the dotted lines indicate the position of the trajectory during the jumps. The arrows show the direction of flow. The dashed vertical line marks the threshold, V_T. In* (A), *$s = s_R^{off}$, which is the point at which the saddle-node bifurcation occurs on the right branches of the nullclines. Thus, the trajectory will be forced to lie on the stable fixed point on the left branches of the nullclines.* (B) *On the left branches of the nullclines, $V_L < V_T$, which means that s will begin to increase toward* 1. (C) *s continues to increase until it reaches the value s_L^{off}, where the stable fixed point on the left branches of the nullclines undergoes a saddle-node bifurcation. The trajectory is, therefore, forced to jump to the fixed point on the right branches of the nullclines. This jump causes s to cross above V_T so that s begins to decrease.* (D) *s has decreased to $s = s_R^{off}$, at which time the fixed point on the right branches is again lost through a saddle-node bifurcation and the trajectory is forced to return to the upper left branches of the nullclines. Therefore, the solution trajectory lies on a periodic orbit.* (E) *The voltage traces of LG and Int*1 *are plotted as s increases and decreases between s_L^{off} and s_R^{off}. The labels* A, B, C, *and* D *indicate where the trajectory is in the phase plane at the given values of $s(t)$, V_I, and V_L.*

through (9) and (10) to the behavior of s_1, we do not explicitly need to check their evolution during the time interval $[0, T_2]$. In this sense, proving the existence of this periodic solution has been reduced to proving that the single variable s_1 is periodic. Thus it is seen that there exists a singular periodic solution whose period is T_2. In section 4, we will both analytically and numerically calculate T_2.

The solution is unique and stable since if $s_1(0) > s_R^{off}$ and the trajectory was on the right branches of the nullclines, for example, then the solution can be flowed forward at time $\hat{t}$ such that $s_1(\hat{t}) = s_R^{off}$. From here the solution trajectory would follow the dynamics described above and return to s_R^{off} at time $t = T_2 + \hat{t}$. Thus, by flowing backward in time, it is seen that $s_1(T_2) = s_1(0)$.

Case 2: *Tonic* $MCN1$ *excitation with the* AB *input to* $Int1$ *present.* In Case 2, $\bar{g}_{AB\to M} = 0$ so that $s_2 \equiv 1$ and thus the $MCN1$ to LG excitation is tonic. Now the AB to $Int1$ inhibition is present ($\bar{g}_{AB\to I} > 0$). Without loss of generality, let $m = 0$ in (2). Hence, s causes the V_L nullcline to slowly shift to the right and left as in Case 1 and $s_{AB\to I}(t)$ causes the V_I nullcline to instantaneously jump down when $s_{AB\to I}(t)$ goes to 1 and to jump back up when $s_{AB\to I}(t)$ returns to 0.

To understand the control of frequency in Case 2, we again consider the nullclines in the phase plane. When $V_L > V_T$, the V_L nullcline moves to the left slowly because τ_{f1} is large. The AB input to $Int1$, on the other hand, is fast and periodic so that the V_I nullcline shifts up and down repeatedly and instantaneously compared with the shift of the V_L nullcline. Thus, on the right branches of the nullclines, three cases arise for the loss of the fixed point. The first possibility is that while $s_R^{on} < s < s_R^{off}$, $s_{AB\to I}(t)$ switches from 1 to 0, forcing the V_I nullcline to jump up causing the stable fixed point to be immediately lost. This forces the trajectory to jump directly to the stable fixed point on the left branches of the V_I and V_L nullclines; see Figure 7(D).

The second possibility is that when $s_{AB\to I}(t) = 1$, s decreases until the fixed point is lost through the saddle-node bifurcation at $s = s_R^{on}$. The third possible way for the fixed point on the right branches to be lost is as in Case 1. That is, while $s_{AB\to I} = 0$, s decreases to s_R^{off}; see Figure 9(D). Which of these cases occurs depends upon the amount of time that $s_{AB\to I}(t)$ spends in its active and inactive phases and the timing of the AB input to $Int1$. In other words, the timing of the periodic jumps in $s_{AB\to I}$ affects the timing of the shifts in the V_I nullcline, which in turn determines which case occurs. The fixed point on the left branch can be lost similarly to the ways discussed above. Let us say that a periodic solution obeys property A if the associated trajectory jumps from the right to left branches when $s_{AB\to I}(t)$ switches from 1 to 0 and from left to right branches through the bifurcation point s_L^{on}; see Figure 10.

Recall that the pyloric period is much smaller than the gastric mill period. Thus, while LG is inactive ($V_L < V_T$), s_{AB} can oscillate several times, say, j times, between 0 and 1. The exact number of times depends on the time constant τ_{r1}. Similarly when LG is active ($V_L > V_T$), the number of oscillations, k, of s_{AB} depends on the time constant τ_{f1}. This implies that the periodic solution in case 2 depends on the relationship between τ_{r1}, τ_{f1} and the pyloric input frequency of AB. In the following theorem we will derive a relationship which τ_{r1} and τ_{f1} need to satisfy in order to find a periodic solution with property A. This

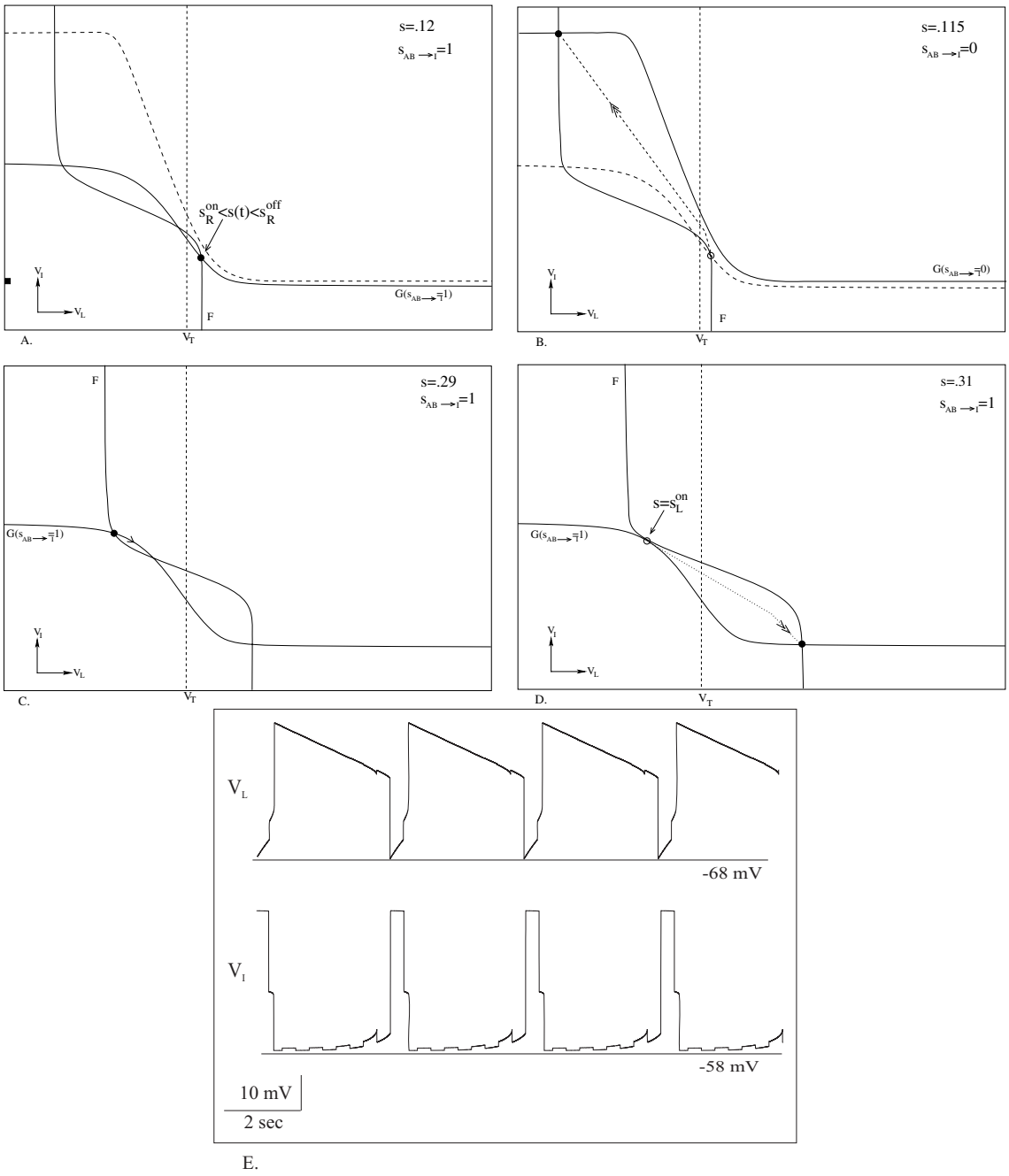


Figure 10. *In Case* 2*, a periodic orbit obeying property A jumps from the left to right branches and from the right to left branches of the nullclines in the following way.* (A) *When* s_1 *lies between* s_R^{off} *and* s_R^{on}*, the stable fixed point lies on the right branches of the nullclines when* $s_{AB\to I} = 1$*.* (B) *As soon as* $s_{AB\to I}$ *jumps back to* 0*, the fixed point on the right branches of the nullclines is instantaneously lost because* $s_1 < s_R^{off}$ *and the trajectory jumps back to the left branches.* (C) *The trajectory lies at the stable fixed point* (•) *on the left branches of the nullclines where* $s_{AB\to I} = 1$ *and* s_1 *increases toward* 1*.* (D) *While* $s_{AB\to I}$ *remains equal to* 1*,* s_1 *increases and becomes sufficiently large for a saddle-node bifurcation to occur through* s_L^{on}*.* (E) *Voltage traces of* V_L *and* V_I *for a periodic orbit obeying property A.*

involves fixing the integers j and k first. To that end define

$$h(k) = \frac{D_c}{\ln\left(\frac{1-s_R^{off}(s_R^{off}/s_L^{on})^{D_c/k}}{1-s_R^{off}}\right)} \ln\left(\frac{1-s_R^{off}}{1-s_L^{on}}\right) + D_c - 1. \tag{24}$$

Theorem. *Let k and j be integers which satisfy $j < h(k)$. There exist values $\tau_{r1}(j)$, $\tau_{f1}(k)$, and $\bar{g}_{AB\to I}$ large enough such that (1)–(3) and (7)–(8) possess a locally unique, asymptotically stable periodic solution obeying property A with period $P = (j + k + 1)P_{AB}$, where j is the number of times s_{AB} oscillates between 0 and 1, while $V_L < V_T$ and k is the number of time s_{AB} oscillates between 0 and 1 when $V_L > V_T$.*

Proof. We shall construct a Poincaré map $\mathcal{P}$ of a certain interval $\mathcal{I}$ into itself. Existence and stability of the periodic solution are determined by showing that $\mathcal{P}$ is a contraction on $\mathcal{I}$, thereby also yielding local uniqueness of the periodic solution. To construct the periodic solution in question we will show that the associated trajectory will jump from the left to the right branches from the bifurcation point s_L^{on}. It will jump from the right to the left branches from a point $s^* \in \mathcal{I}$ at one of the times when $s_{AB\to I}$ switches from one to zero.

To construct $\mathcal{I}$, consider the points s_R^{off} and s_R^{on} corresponding to the bifurcation points along the right branches of the V_I-V_L nullclines when $s_{AB\to I} = 0$ (AB off) and $s_{AB\to I} = 1$ (AB on), respectively. By choosing $\bar{g}_{AB\to I}$ and τ_{f1} sufficiently large, we can guarantee that the time distance from s_R^{off} to s_R^{on} under the dynamics $s' = -s/\tau_{f1}$ is larger than $D_c P_{AB}$. Indeed the time Δt between these two points on the right branches is $\tau_{f1}\ln(s_R^{off}/s_R^{on})$, where s_R^{on} is a decreasing function of $\bar{g}_{AB\to I}$. Moreover, there exists $\hat{s}$ such that $\hat{s} = s_R^{off}\exp(-D_c P_{AB}/\tau_{f1})$. Thus the time distance from s_R^{off} to $\hat{s}$ on the right branches is exactly $D_c P_{AB}$. We let $\mathcal{I} = [\hat{s},\ s_R^{off}]$; see Figure 11. Note that at this point, we are only stating that we need τ_{f1} sufficiently large. Below, we will be more specific.

We next show that $\mathcal{I}$ maps into itself under the flow if τ_{r1} and τ_{f1} are chosen appropriately. We flow the endpoints of the interval $\mathcal{I}$, s_R^{off}, and $\hat{s}$, through one cycle of the V_I and V_L oscillation and show that these endpoints are mapped into $\mathcal{I}$. Thus by continuous dependence on initial conditions, all points in $\mathcal{I}$ will map into $\mathcal{I}$.

First consider a trajectory $s_a(t)$ where $s_a(0) = s_R^{off}$ and the trajectory is on the right branch. Next let $s_{AB\to I}(0^-) = 1$ and $s_{AB\to I}(0^+) = 0$, so that the trajectory jumps back to the left branch at $t = 0^+$. The dynamics of s_a on the left branches obey $s' = (1 - s)/\tau_{r1}$. By choosing

$$\tau_{r1}(j) = (j + 1 - D_c)P_{AB}/\ln([1 - s_R^{off}]/[1 - s_L^{on}]), \tag{25}$$

we can guarantee that $s_a([j+1-D_c]P_{AB}) = s_L^{on}$. This means that the trajectory which starts at s_R^{off} will leave the left branches of the nullclines through the bifurcation point s_L^{on} along the $s_{AB\to I} = 1$ nullcline at time $t = (j + 1 - D_c)P_{AB}$.

Next consider a trajectory $s_b(t)$ with the initial condition given by $s_b(0) = \hat{s}$. Recall that the time distance from s_R^{off} to $\hat{s}$ on the right branches is given by $\tau_{f1}\ln(s_R^{off}/\hat{s})$. On the left branches, the time between these points is governed by τ_{r1} and is equal to $\tau_{r1}\ln([1 - \hat{s}]/[1 -$

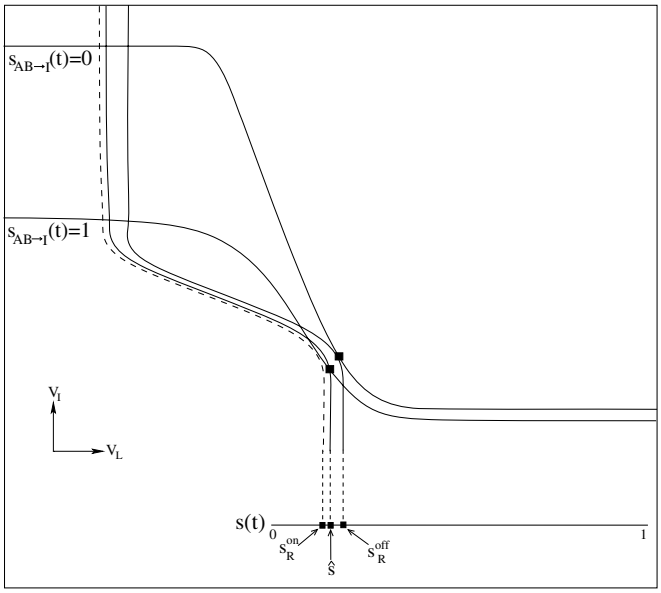


Figure 11. *To define the Poincaré map for Case 2, let* $I = [\hat{s}, s_R^{off}]$*, where* $\hat{s} = s_R^{off}\exp(-D_c P_{AB}/\tau_{f1})$*. Note that* τ_{f1} *is chosen sufficiently large to ensure that* $\hat{s} > s_R^{on}$*.*

$s_R^{off}]$). Thus if

$$\tau_{r1} < \tau_{f1}\frac{\ln(s_R^{off}/\hat{s})}{\ln([1-\hat{s}]/[1-s_R^{off}])}, \tag{26}$$

then the time between these points on the left branches will be less than that on the right branches and, in particular, will be less than $D_c P_{AB}$. This type of time compression between cells across a jump is analogous to fast threshold modulation [26]. Thus the trajectory starting with initial condition at $\hat{s}$ at $t = 0$ will reach s_L^{on} when $s_{AB\to I} = 1$ at a time T_1 bounded between $(j+1-D_c)P_{AB}$ and $(j+1)P_{AB}$. Therefore, any trajectory with $s(0) \in \mathcal{I}$ will do the same. We also note that once trajectories with initial conditions $s_a(0)$ and $s_b(0)$ jump from the right branches of the nullclines to the left branches, the interval $\mathcal{I}$ becomes inverted so that the trajectory with initial condition $s_a(0)$ becomes the leading cell.

Note that the time between any two trajectories remains invariant while they both evolve on the left branches and even across the jump back to the right branches. That the trajectories remain the same time distance apart on the left branches follows from the fact that they both obey the same differential equation ($s' = (1-s)/\tau_{r1}$). Moreover, since they leave the left branches through the same point s_L^{on}, the time distance between them when the leading cell reaches the bifurcation point is the same as the time distance apart when the trailing cell reaches this point. When both trajectories are on the right branches, the time distance again remains invariant since both trajectories evolve under $s' = -s/\tau_{f1}$. In particular, the time distance between $s_a(T_1)$ and $s_b(T_1)$ is less than $D_c P_{AB}$.

Consider again the trajectory $s_a(t)$ which had $s_a(0) = s_R^{off}$ and $s_a([j+1-D_c]P_{AB}) = s_L^{on}$. We want this trajectory to spend k oscillations of s_{AB} on the right branches. We also want

the trajectory to be in a position to jump back to the left branches when $s_{AB\to I}$ switches from one to zero. Finally, since we want $\mathcal{I}$ to map into itself, we choose

$$\tau_{f1}(k) = \frac{(k + D_c)P_{AB}}{\ln(s_L^{on}/\hat{s})}, \tag{27}$$

such that $s_a([j + k + 1]P_{AB}) = \hat{s}$. Note that by substituting $\hat{s} = s_R^{off} \exp(-D_c P_{AB}/\tau_{f1})$ into (27) and solving for τ_{f1}, we obtain

$$\tau_{f1}(k) = \frac{kP_{AB}}{\ln(s_L^{on}/s_R^{off})}. \tag{28}$$

In other words, with the choices of τ_{r1} and τ_{f1} that we have made, the trajectory with initial condition s_R^{off} is mapped back to $\hat{s}$ at a time $T_2^- = (j + k + 1)P_{AB}$.

Next consider the trajectory $s_b(t)$ where $s_b(0) = \hat{s}$ and $s_b(T_1) = s_L^{on}$. To construct the Poincaré map, we need $s_b(T_2^-) \in \mathcal{I}$. Thus we need $\hat{s} < s_b(T_2^-) < s_R^{off}$. That $\hat{s} < s_b(T_2^-)$ follows by continuity since $\hat{s} = s_a(T_2^-) < s_b(T_2^-)$. The value $s_b(T_2^-) < s_R^{off}$ since the time distance from s_R^{off} to $s_a(T_2^-) = \hat{s}$ is $D_c P_{AB}$, whereas the time distance from $s_b(T_2^-)$ to $s_a(T_2^-)$ is less than $D_c P_{AB}$.

We have just shown that the trajectories whose s values are associated with the endpoints of the interval $\mathcal{I}$ have s values mapped back to $\mathcal{I}$ after a time T_2^-. Thus by continuity with respect to initial conditions, any trajectory with $s(0) \in \mathcal{I}$ will end up with $s(T_2^-) \in \mathcal{I}$ at a time when $s_{AB\to I}$ will switch from one to zero. Therefore we can define a one-dimensional Poincaré map $\mathcal{P} : \mathcal{I} \to \mathcal{I}$ where $\mathcal{P}(s) = s(T_2^-)$.

To show that $\mathcal{P}$ is a contraction mapping on $\mathcal{I}$ let $s_a(0) > s_b(0) \in \mathcal{I}$ be arbitrary. Let Δt denote the time on the right branch between them. At $t = 0^+$, the trajectories jump back to the left branch. Because of our choice of time constants $\tau_{r1}(j)$ and $\tau_{f1}(k)$, the new time between these points is less than Δt. As before the time distance between these trajectories remains invariant as they evolve along the left branches, across the jump to the right branches, and then back to $\mathcal{I}$. Since this new time is less than the original time, $s_a(T_2^-) - s_b(T_2^-) < \alpha[s_a(0) - s_b(0)]$, where $\alpha < 1$ is dependent on τ_{r1} and τ_{f1}. Therefore $\mathcal{P}$ is a contraction. As a result, there exists a unique value $s^*_{Case\ 2} \in \mathcal{I}$ such that $\mathcal{P}(s^*_{Case\ 2}) = s^*_{Case\ 2}$. This value is asymptotically stable and corresponds to a locally unique singular periodic solution of equations (9)–(21). For ϵ small, results in [20] show that an actual solution to (1)–(3) and (7)–(8) exists within an $O(\epsilon)$ neighborhood of the singular one.

Equation (26) provides a condition on the time constants τ_{f1} and τ_{r1} for which the theorem holds. This condition can be translated into a relationship between the integers j and k. Namely, by substituting $\hat{s} = s_R^{off} \exp(-D_c P_{AB}/\tau_{f1})$ into the fraction on the right-hand side of (26), we see that the numerator of that expression reduces to $D_c P_{AB}/\tau_{f1}$, while the denominator reduces to

$$\ln\left(\frac{1 - s_R^{off} \exp(-D_c P_{AB}/\tau_{f1})}{1 - s_R^{off}}\right). \tag{29}$$

Now substituting (28) into (29) and then substituting the resulting expression into (26), we obtain

$$\tau_{r1} < \tau_{f1} \frac{(D_c P_{AB}/\tau_{f1})}{\ln\left(\frac{1-s_R^{off}(s_R^{off}/s_L^{on})^{D_c/k}}{1-s_R^{off}}\right)}. \tag{30}$$

Canceling τ_{f1} from the right-hand side, substituting τ_{r1} from (25), and solving for j, we obtain

$$j < \frac{D_c}{\ln\left(\frac{1-s_R^{off}(s_R^{off}/s_L^{on})^{D_c/k}}{1-s_R^{off}}\right)} \ln\left(\frac{1-s_R^{off}}{1-s_L^{on}}\right) + D_c - 1. \tag{31}$$

The right-hand side of (31) is what we called $h(k)$ in the statement of the theorem.

Remark. Note that if (31) is not satisfied, then we cannot find time constants τ_{r1} and τ_{f1} for which a periodic solution satisfying property A exists. However, by choosing τ_{r1} and τ_{f1} differently, we could instead have easily constructed a periodic solution whose s_1 value passed through the bifurcation point s_R^{off} on the transition from right to left branches and which jumped back to the right branches with $s_1 \in (s_L^{on}, s_L^{off})$ when $s_{AB\to I}$ switched from one to zero.

Case 3: *Rhythmic MCN*1 *excitation with AB input to Int*1 *absent.* In Case 3, the input from AB to $MCN1$ is present ($\bar{g}_{AB\to M} > 0$) so that the $MCN1$-elicited excitation to LG is rhythmic. Once again, we set $\bar{g}_{AB\to I} = 0$ so that the V_I nullcline remains at a fixed position. In this case, s_2 jumps instantaneously between s_{2min} and 1, while s_1 increases with rate $1/\tau_{r1}$ and decreases with rate $1/\tau_{f1}$, thus causing the activity of s to be rhythmic.

Again, consider the nullclines. Suppose the trajectory is at the stable fixed point on the right branches of the nullclines. Thus, s_1 is decreasing and slowly pushing the V_L nullcline to the left. Recall that s_2 jumps between 1 and s_{2min} as V_M changes. The jumps in s_2 cause the V_L nullcline to shift to the right and left on the fast timescale. The size of the shift in the V_L nullcline depends on the value of s_{2min}. The trajectory eventually jumps to the fixed point on the left branches of the nullclines when (a) $s_2 = 1$ and s_1 moves the fixed point to the position where the nullclines are tangent, i.e., $s_1 = s_R^{off}$, or (b) when $s_2 = s_{2min}$ and s_1 decreases to s_R^{off}/s_{2min} moving the fixed point to the position where the nullclines are tangent, or (c) when s_2 jumps to s_{2min} resulting in an instantaneous shift of the V_L nullcline past the point of bifurcation of the fixed points. Which case occurs depends on the speed at which s_1 decreases (τ_{f1}), the amount of time the V_L nullcline spends being shifted to the left by s_2 (the amount of time that $s_{AB\to M}(t)$ spends in its active or inactive phases), and the timing of the instantaneous shifting of the V_L nullcline (timing of the AB input to $MCN1$).

Once the trajectory has jumped to the left branches of the nullclines, the V_L nullcline slowly shifts to the right due to s_1 and then instantaneously jumps to the left when $V_M < V_{Th(M)}$. Again, the trajectory eventually jumps to the fixed point on the right branches of the nullclines when (a′) $s_2 = 1$ and s_1 moves the fixed point to the position where the nullclines are tangent (i.e., $s_1 = s_L^{off}$; see Figure 7(A)), or when (b′) $s_2 = s_{2min}$ and s_1 moves the fixed point to the position where the nullclines are tangent, or when (c′) s_2 jumps to 1 resulting in an

instantaneous shift of the V_L nullcline past the point of bifurcation of the fixed points; see Figure 7(B).

To have a one-dimensional map, we need one of the jumps of the trajectory from the left to right branches or from the right to left branches of the nullclines to occur through a saddle-node bifurcation (where the value of s_1 is known) and the other to occur instantaneously when s_2 jumps between 1 and s_{2min} (where the value of s_1 will be defined as a fixed point of the map). Therefore, we shall establish the existence of the periodic solution which follows the subcases (c) and (a′) above. Namely, the trajectory will jump from the left to the right branches through the bifurcation point s_L^{off}, i.e., s_1 and s_2 known, and from the right to the left branches when s_2 jumps down from 1 to s_{2min}, i.e., s_2 known and s_1 to be determined by the map; see Figure 12.

Recall that $s(t) = s_1(t) * s_2(t)$, where we consider $s_2(t) = 1$ when $V_M \geq V_{Th(M)}$ and $s_2(t) = s_{2min}$ when $V_M < V_{Th(M)}$. When $s_{AB\to M}(t)$ jumps to 1, V_M instantaneously jumps below $V_{Th(M)}$. Thus, $s_2(t)$ instantaneously jumps to s_{2min}. However, $s_{AB\to M}(t)$ does not instantaneously jump from 1 to 0, but slowly decays with time constant $1/\tau_{M_2}$. Thus, V_M requires a small amount of time, T_C, to go above $V_{Th(M)}$. In our model, we chose $V_{Th(M)}$ such that T_C is approximately $P_{AB}/20$. Hence, for one cycle of AB activity, $s_2(t) = s_{2min}$ for time $D_c P_{AB} + T_C$ and $s_2(t) = 1$ for time $(1 - D_c)P_{AB} - T_C$.

As in Case 2, we can construct a Poincaré map $\mathcal{P}$ of an interval of s values on the right branch, $\mathcal{I} = [\tilde{s}, s_R]$, into itself. Here $s_R = s_1^R * 1$, where $s_1^R = s_R^{off}/s_{2min}$. We let $\tilde{s} = [s_1^R \exp([[D_c - 1]P_{AB} + T_C]/\tau_{f1})] * 1$ so that the time distance between s_R and $\tilde{s}$ is $[1 - D_c]P_{AB} - T_C$. Let $\tilde{s}_1 = [s_1^R \exp([[D_c - 1]P_{AB} + T_C]/\tau_{f1})]$; see Figure 13.

We consider a trajectory $s_a(t)$ where $s_a(0^-) = s_1^R * s_2(0^-)$ and the trajectory is on the right branch of the nullclines. Let $s_{AB\to M}(0^-) = 0$ and $s_{AB\to M}(0^+) = 1$ so that $s_2(0^-) = 1$ and $s_2(0^+) = s_{2min}$. Thus the trajectory jumps back to the left branch at $t = 0^+$ when s_2 jumps from 1 to s_{2min}. We then choose $\tau_{r1}(j) = ((j + D_c)P_{AB} + T_C)/ln([1 - s_1^R]/[1 - s_L^{off}])$ which guarantees that $s_a((j + D_c)P_{AB} + T_C) = s_L^{off}$ so that the trajectory which starts at s_R at $t = 0^-$ will leave the left branches of the nullclines through the bifurcation point s_L^{off} at $t = (j + D_c)P_{AB} + T_C$. Using the same argument as in Case 2 with an equivalent condition on τ_{r1} as in (25), the trajectory $s_b(t)$ with initial condition $s_b(0) = \tilde{s}_1 * s_2(0)$ will be forced to reach s_L^{off} at a time T_1 bounded between $(j + D_c)P_{AB} + T_C$ and $(j + 1)P_{AB}$. Therefore, any trajectory with $s(0) \in \mathcal{I}$ will also reach s_L^{off} during these times.

Next, we choose $\tau_{f1}(k)$ so that $s_a(t)$ gets mapped back to $\mathcal{I}$, particularly to $\tilde{s}$ at the instant before $s_{AB\to M}(t)$ jumps from 0 to 1. Therefore, we let $\tau_{f1}(k) = [[k + [1 - D_c]]P_{AB} - T_C]/ln(s_1^L/\tilde{s}_1)$. Thus, at $s_a(T_2^-)$, where $T_2 = (k + j + 1)P_{AB}$, $s_a(t)$ lies in $\mathcal{I}$. In a similar argument to that of Case 2, the trajectory $s_b(t)$ with $s_b(0) = \tilde{s}_1 * s_2(0)$ and $s_b((j+1)P_{AB}) = s_L^{off}$ will also lie in $\mathcal{I}$ at $t = T_2^-$ with $\tilde{s} = s_a(T_2^-) < s_b(T_2^-) < s_R$.

We define a one-dimensional Poincaré map $\mathcal{P} : \mathcal{I} \to \mathcal{I}$ where $\mathcal{P}(s) = s(T_2)$. The argument showing that $\mathcal{P}$ is a contraction mapping on $\mathcal{I}$ is the same as in Case 2. Consequently, there exists a locally unique, asymptotically stable value $s^*_{Case\ 3} \in \mathcal{I}$ such that $\mathcal{P}(s^*_{Case\ 3}) = s^*_{Case\ 3}$.

Remark. The periods of the constructed solutions in Cases 2 and 3 are both $(j+k+1)P_{AB}$. Note that this occurs since the values of $\tau_{r1}(j)$ and $\tau_{f1}(k)$ are chosen to be different in both cases. In general, if, a priori, τ_{r1} is chosen to have the same value for both Cases 2 and 3, and

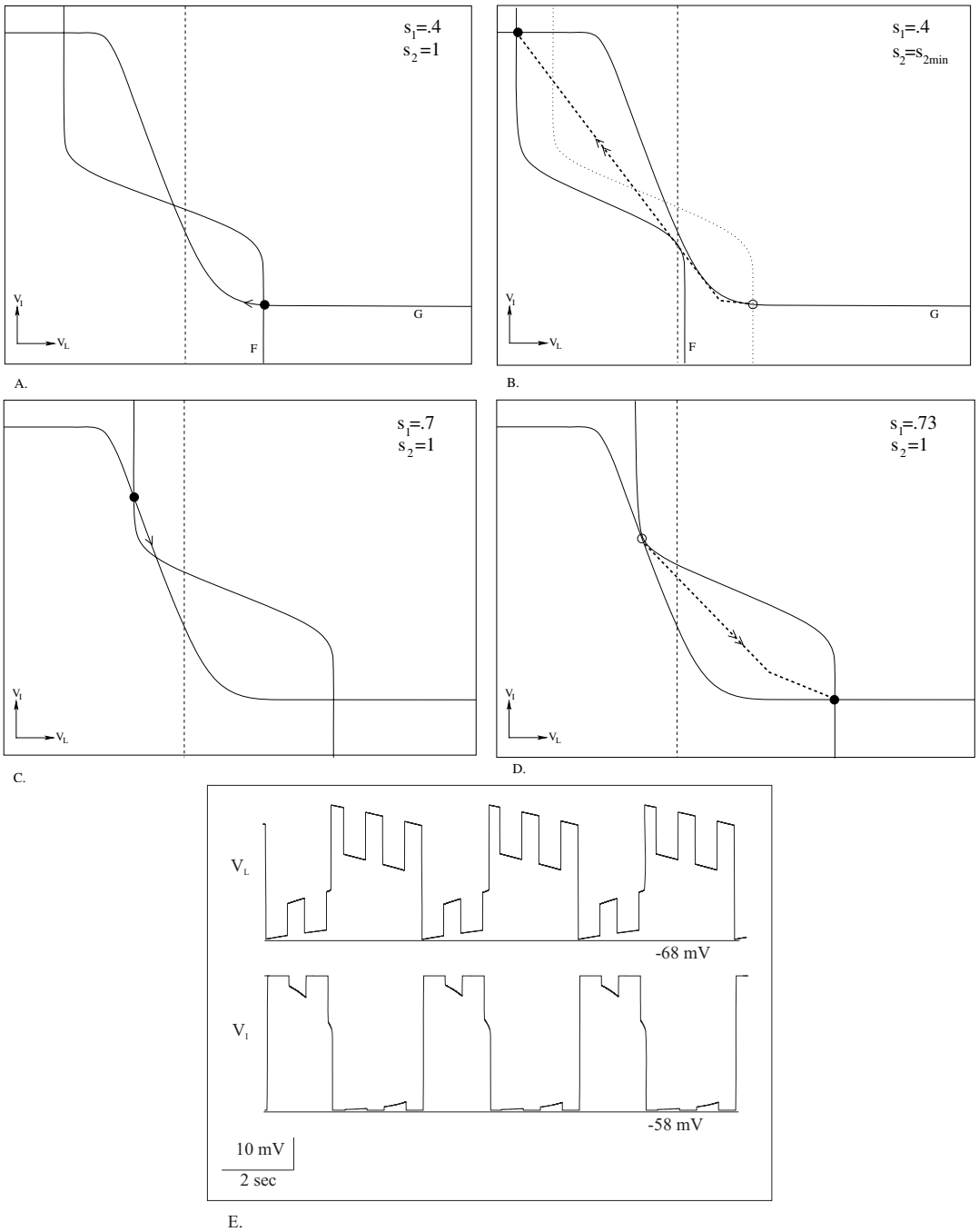


Figure 12. *Case* 3 *solution trajectory and nullclines.* (A) *The trajectory lies at the stable fixed point* (•) *on the right branches of the nullclines where* $s_{AB\to I} = 0$ *and* $s_2 = 1$. (B) *When* s_2 *jumps to* s_{2min}, s *instantaneously goes below* s_R^{off} *and the fixed point is lost. Thus, the solution trajectory is forced to jump to the stable fixed point on the left branches of the nullclines (shown by the dashed line with double arrows).* (C)–(D) *On the left branches of the nullclines, the fixed point is lost while* $s_2 = 1$ *and* s_1 *becomes large enough for the saddle-node bifurcation to occur through* s_L^{off}. *Now, the solution trajectory is forced to jump to the right branches of the nullclines.* (E) *Voltage traces of LG and Int*1 *for* s *satisfying the above conditions.*

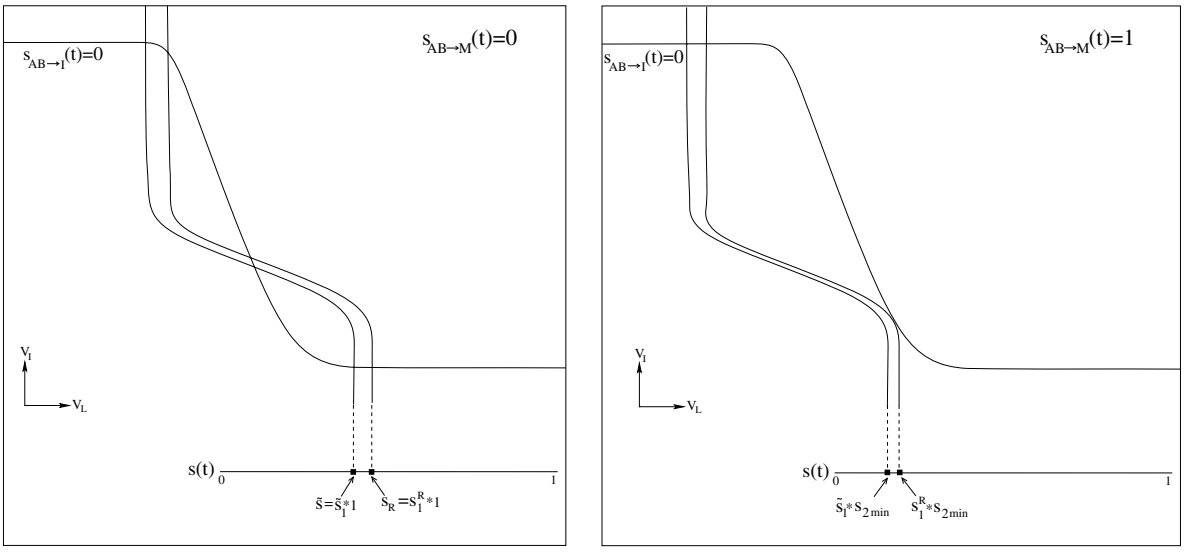


Figure 13. (A) *To define the Poincaré map for Case* 3*, we let* $I = [\tilde{s}, s_R]$ *where* $s_R = s_1^R * 1$, $s_1^R = s_R^{off}/s_{2min}$, *and* $\tilde{s} = [s_1^R \exp([[D_c - 1]P_{AB} + T_C]/\tau_{f1}] * 1$. (B) *For* $s = s_R$*, when* s_2 *jumps from* 1 *to* s_{2min}*,* s *jumps to* s_R^{off}. *Similarly, for* $s = \tilde{s}$*, when* s_2 *jumps to* s_{2min}*,* s *jumps to* $\tilde{s}_1 * s_{2min}$.

similarly for τ_{f1}, then the periods of Cases 2 and 3 need not be the same.

Case 4: *Rhythmic* $MCN1$ *excitation with* AB *input to* $Int1$ *present.* In Case 4, the $MCN1$ to LG excitation is rhythmic ($\bar{g}_{AB\to M} > 0$) and AB inhibits $Int1$ ($\bar{g}_{AB\to I} > 0$). Consequently, the V_L nullcline shifts to the right and left with slow changes in s_1 and with quick jumps of s_2 between 1 to s_{2min}. The V_I nullcline jumps up and down instantaneously due to the $s_{AB\to I}$ oscillations between 0 and 1.

Suppose the trajectory lies at the stable fixed point on the left branches of the V_L and V_I nullclines. Here, $V_L < V_T$ which allows s_1 to increase. When $s_{AB\to M}(t) \to 1$, the V_L nullcline jumps to the left. Similarly, when $s_{AB\to I}(t) \to 1$, the V_I nullcline jumps down. Recall that the time difference between when the V_L and V_I nullclines shift is controlled by the parameter m. For example, if $m = 0$, the V_L nullcline jumps to the left at the same time that the V_I nullcline jumps down. However, if $m = D_c P_{AB}$, then when the V_L nullcline jumps to the left, the V_I nullcline jumps up. This creates several possibilities for the length of the period.

To provide more insight into the role of m, suppose again that $m = 0$. In Case 2, the jumping down of the V_I nullcline allowed the fixed point to bifurcate at smaller values of s_1 relative to Case 1. However, in Case 4 for $m = 0$, whenever the V_I nullcline jumps down, the V_L nullcline jumps back to the left. If this jump to the left is large enough ($1 - s_{2min}$ is large) and the jump down in the V_I nullcline is not extremely large, there will still exist a stable fixed point on the left branches. In this case, the trajectory will have to wait until $s_2 = 1$ and s_1 has grown large enough so that the fixed point occurs where the nullclines intersect tangentially for $s_{AB\to I}(t) = 0$ (Figure 7A) or when s_2 jumps to 1 ($s_{AB\to M}(t)$ jumps to 0) and the fixed point is instantaneously lost (Figure 7B). This is equivalent to (a′) and (c′) in Case 3. If the jump to the left of the V_L nullcline is not large and/or the jump down in the V_I nullcline is extremely large, the fixed point will be lost and the trajectory will immediately jump to

the stable fixed point on the right branches of the nullclines. This is equivalent to Case 2. The above result also extends to the situation in which $m \in (0, T_C]$ (where T_C is the time it takes for V_M to increase under (11) and (13) from V_M^* to $V_{TH(M)}$) because for $0 < m \leq T_C$, each time the V_I nullcline is in the downward position, the V_L nullcline is shifted to the left. Therefore, the fixed point cannot be lost until s_1 grows large enough for the bifurcation to occur while $s_2 = 1$.

Next suppose that $m = D_c P_{AB}$. For $m = D_c P_{AB}$, each time the V_I nullcline is shifted in the downward position, the V_L nullcline remains to the right. Therefore, as opposed to the situation in which $m = 0$, the fixed point on the left branches of the nullclines can be lost due to the jump down of the V_I nullcline as in Case 2; see Figure 7C. This same idea extends to values of m lying in a neighborhood, $[R_1, R_2]$, of $D_c P_{AB}$ where $R_1 > T_C$, $R_2 < P_{AB}$. For $m \in [R_1, R_2]$, there is always some amount of time for which the V_L nullcline is to the right while the V_I nullcline is shifted downward, thus allowing the fixed point to be lost at an earlier time than in Case 3.

We now consider the existence of a periodic solution for Case 4 with $\tau_{r1}(j)$ and $\tau_{f1}(k)$ defined as in Case 3. For $m \in (0, T_C]$, the periodic orbit will be defined in exactly the same way as Case 3. Consider the interval $\mathcal{I}$ on the right branches as defined in Case 3 with $s_{AB\to M}(0^-) = 0$ and $s_{AB\to M}(0^+) = 1$. Let $s_a(t)$ be a trajectory with $s_a(0) = s_1^R * s_2(0)$. At $t = 0^+$, $s_{AB\to M}$ jumps to 1 and $s_{AB\to I}$ remains equal to 0 because of the small delay m. Thus, s_2 instantaneously jumping to s_{2min} forces the trajectory to the left branches in the same way as Case 3 because the V_I nullcline remains in the upward position at $t = 0^+$.

On the left branches with $m \in (0, T_C]$, each time $s_{AB\to I} = 1$, pushing the V_I nullcline down, $s_2 = s_{2min}$ and the V_L nullcline is forced to the left. Thus, the trajectory with initial conditions as stated above cannot jump from s_L^{on} as in Case 2 because each time the V_I nullcline is in its downward position, the V_L nullcline is shifted too far to the left for s to reach s_L^{on}. Therefore, the only time at which the fixed point of the left branches can be lost is when $s_2 = 1$. At $t = (j + D_C)P_{AB}$, $s_{AB\to M}(t)$ will jump to 0. However, s_2 will not jump to 1, returning the V_L nullcline to the right, until $t = (j + D_C)P_{AB} + T_C$. Hence, at $t = (j + D_C)P_{AB} + T_C$, $s_{AB\to I}$ will already be equal to 0 since $m \leq T_C$ and the trajectory will reach s_L^{off} tangentially as in Case 3. By choosing $\tau_{f1}(k)$ as in Case 3, $s_a(t)$ gets mapped to $\tilde{s}$ at $t = T_2^-$.

With the same argument as in Case 3, the trajectory $s_b(t)$ with $s_b(0) = \tilde{s}_1 * s_2(0)$ and $s_b(T_1) = s_L^{off}$ (where $T_1 \in ((j + D_c)P_{AB} + T_C, (j + 1)P_{AB} + T_C)$) will also be mapped back to $\mathcal{I}$ at $t = T_2^-$ with $\tilde{s} = s_a(T_2^-) < s_b(T_2^-) < s_R$. Therefore, the one-dimensional Poincaré map $\mathcal{P} : \mathcal{I} \to \mathcal{I}$ where $\mathcal{P}(s) = s(T_2)$ is established exactly as in Case 3. Thus for $m \in (0, T_C]$, the same arguments apply to show that there exists a unique, asymptotically stable periodic orbit in Case 4, and the periodic orbits of Cases 3 and 4 have the same period.

For $m \in [R_1, R_2]$, the period of the solution trajectories in Case 4 is locked to the period of AB oscillations and is, therefore, much shorter than the period of solution trajectories in Case 3. For $m \in [R_1, R_2]$, as stated above, on the left branches of the nullclines, the inhibition from AB to $MCN1$ and to $Int1$ is timed such that while the V_I nullcline is shifted downward, $s_2 = 1$ which places the V_L nullcline to the right. Thus, s_1 does not need to grow very large for the LG interburst to end. Once on the right branches, the burst of LG is ended during the first time s_2 jumps to s_{2min} because s_1 is sufficiently small (due to the fact that the LG

interburst was ended for a small value of s_1) to push the V_L nullcline far enough to the left to cause a loss in the fixed point. Similarly, once the solution trajectory is back to the left branches, s_1 is large enough (because s_1 did not decay a long time on the right branches) so that the first jump in s_2 back to 1 causes a loss of the fixed point and an end to the LG interburst. In the regions $T_C < m < R_1$ and $R_2 < m \leq P_{AB}$, the solution trajectories remain periodic but are slightly more complicated to describe than those outside of these regions. For example, with the parameters fixed as above, when $m \in (0, T_C]$ or $m \in [R_1, R_2]$ consecutive LG bursts have exactly the same length as do consecutive $Int1$ bursts. However, when m is not in these regions, consecutive LG bursts and consecutive $Int1$ bursts need not have the same length. Instead, several cycles of LG and $Int1$ oscillations may be required before the LG (and $Int1$) burst duplicates its length. We further explain this in the next section.

4. Determining the frequency of solutions. The period of the gastric mill rhythm can be computed as the sum of the LG burst and the LG interburst. During the interburst, s_1 increases toward a maximum value which we shall denote s_{max}. Similarly, during the burst, s_1 decreases toward a minimum value s_{min}. The periodic solutions in Cases 1–4 are then computed by finding out how much time is needed for s_1 to evolve between the values of s_{max} and s_{min} on the left (during the interburst) and right (during the burst) branches. Using (7), it is straightforward to see that

$$P = \tau_{r1} ln\left(\frac{1 - s_{min}}{1 - s_{max}}\right) + \tau_{f1} ln\left(\frac{s_{max}}{s_{min}}\right). \tag{32}$$

The main question now is to determine the values s_{min} and s_{max} for each of the four cases. However, these values have already been determined in the construction of the periodic solutions above. In particular, for Case 1, $s_{min} = s_R^{off}$ and $s_{max} = s_L^{off}$. For Case 2, $s_{min} = s^*_{Case\ 2}$ and $s_{max} = s_L^{on}$. Note that $s_L^{on} < s_L^{off}$. Since the AB inhibition does not affect the right branches of the V_I nullcline too much, $s^*_{Case\ 2} \approx s_R^{off}$. Thus from (32), it is seen that the period of Case 2 is smaller than the period of Case 1 since the interburst of LG is shorter. This result is consistent with what was found by Manor et al. [16].

In Case 3, $s_{min} = s^*_{Case\ 3}$ and $s_{max} = s_L^{off}$. Here $s^*_{Case\ 3} \in [s_1^R \exp([[D_c - 1]P_{AB} + T_C]/\tau_{f1}), s_1^R]$. Finally, for Case 4, when $m \in (0, T_c]$, the periodic solution is different than the one constructed in Case 3 due to the shifting of the V_I nullcline. However, the period of the solutions in Cases 3 and 4 are the same because the values of s_{min} and s_{max} are the same. When $m > T_c$, $s_{min} \in [s_1^R \exp([[D_c - 1]P_{AB} + T_C]/\tau_{f1}), s_1^R]$ and $s_{max} \in [s_L^{on}, 1 + (s_L^{on} - 1)\exp(-(D_C P_{AB} + T_C)/\tau_{r1})]$. Calculations of $s_L^{off}, s_R^{off}, s_L^{on}$, and s_R^{on} can be found in the appendix. Using (32), we calculated values for the period for Cases 1 through 4. The analytic results are shown in Table 1 for Cases 1 through 3. There, we assumed for Case 2 that $s_{min} = s^*_{Case\ 2}$ equals the average of s_R^{off} and s_R^{on}, while for Case 3, $s_{min} = s^*_{Case\ 3} = s_R^{off}/s_{smin}$.

To confirm the validity of our analytic computations of the period using (32), we also numerically solved the set of equations in our model using XPP [11]. See the appendix for parameter values. For Cases 1 through 3, these results are also given in Table 1 and show a close correlation between the calculation of the periods obtained analytically and numerically.

In Figure 14, we show the results of numerically calculating the period in Case 4 as a function of the delay parameter m with $\tau_{r1} = 7200$ and $\tau_{f1} = 5500$. We see that for

Table 1

A comparison of the analytic versus numerical determinations of period in Cases 1–3.

Period calculation for $\tau_{r1} = 4900$ msec, $\tau_{f1} = 4000$ msec		
Case	XPP simulation	Analytic formula
Case 1	period=10,140 msec	period=10,075 msec
Case 2	period=5,000 msec	period=4,688 msec
Case 3	period=4,000 msec	period=3794 msec

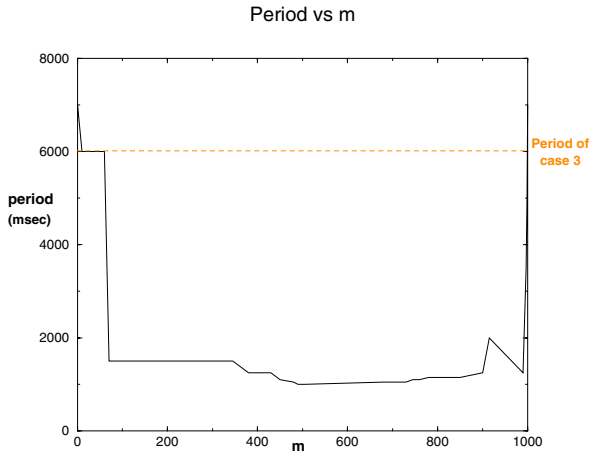


Figure 14. *The period of the gastric mill cycle is plotted for different delays, m, in Case* 4. *The period of Case* 3 *is marked by the dashed line. There is a small range of delays,* $0 < m \leq 60$*, for which the period of the gastric mill rhythm is equal for Cases* 3 *and* 4.

$0 < m \leq 60$, the period of the gastric mill rhythm for Case 3 is equal to that of Case 4. For $470 < m \leq 740$, the period of the gastric mill is equal to the period of AB activity. For $60 < m \leq 470$, there is a transition between having a period equal to that of Case 3 to the much shorter period of AB activity. Similarly, for $740 < m \leq 1000$, the period begins to increase from 1 *sec* up to the period found in Case 3. As stated in the previous section, for $60 < m \leq 470$ and for $740 < m \leq 1000$, it may take several cycles of LG and $Int1$ oscillations before the LG and $Int1$ burst lengths duplicate themselves where $R_1 = 470$ and $R_2 = 740$. In these situations, the period is calculated as the time it takes to have two duplicate LG burst lengths divided by the number of cycles of LG oscillations occurring in that time. Figure 15 shows the gastric mill rhythm frequency as found in experiments by Wood et al. [30] for Cases 1–4. In this work, Wood et al. [30] artificially replicate the effect of AB activity on $MCN1$ through computer controlled stimulation of $MCN1$. We see that for $MCN1$ tonic, the frequency of the network is much higher when the AB input to $Int1$ is present. For $MCN1$ rhythmic, the frequency of the network is higher than when $MCN1$ is tonic; however, there is no change in frequency when the AB input to $Int1$ is added to the network. Figure 16 shows that our model accurately replicates the behavior of the actual gastric mill when $0 < m \leq 60$. Thus, the time mismatch between the pyloric and modulatory inputs to the gastric mill network is critical in establishing the correct frequency of the system.

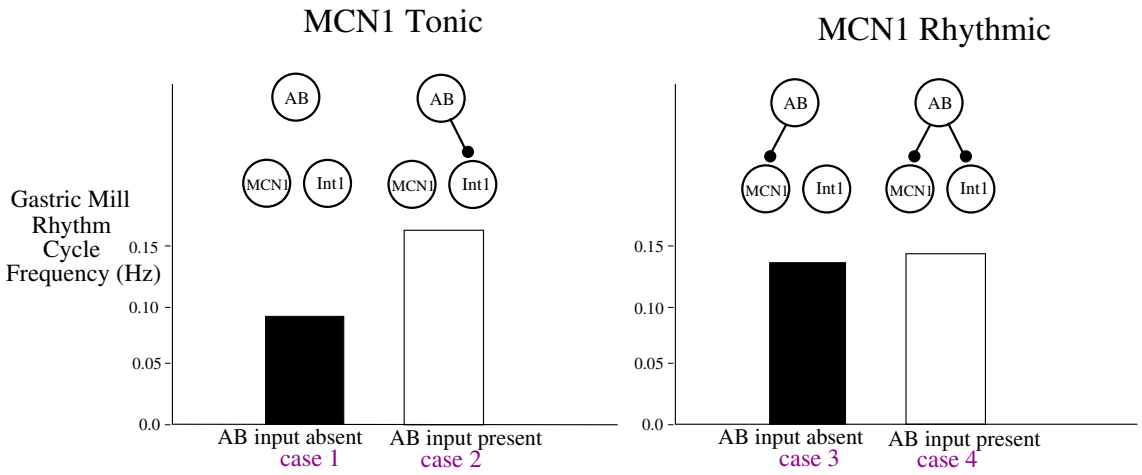


Figure 15. *Experimental findings of the gastric mill rhythm cycle frequency for Cases* 1–4 [30].

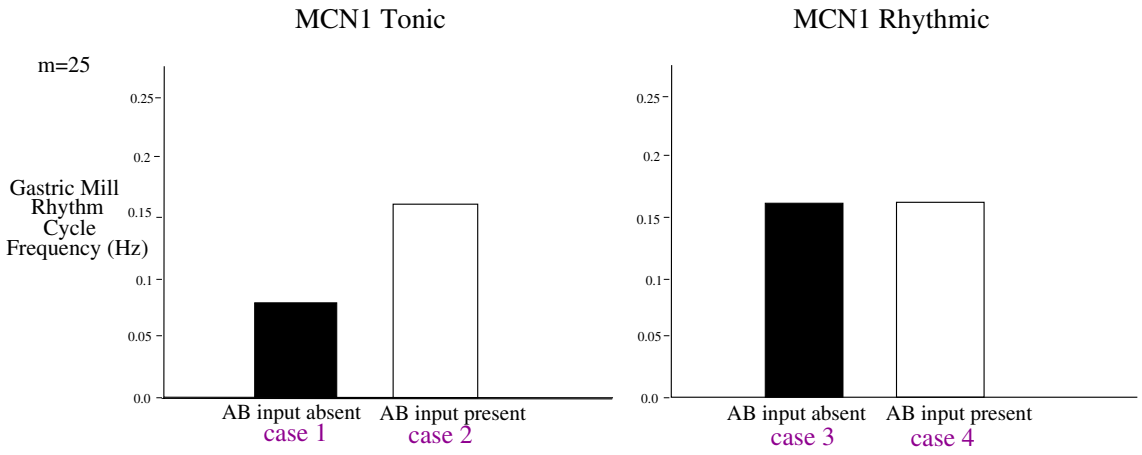


Figure 16. *Calculations of the gastric mill rhythm cycle frequency for Cases* 1–4 *using our model with* $\tau_{r1} = 7200msec$, $\tau_{f1} = 5500msec$, $\tau_{r2} = 1msec$, $\tau_{f2} = 1msec$, *period of* $s_{AB} = 1sec$, $m = 25msec$, *and* $\bar{g}_s = 6mS/cm^2$ *in Cases* 1 *and* 2 *and* $\bar{g}_s = 7mS/cm^2$ *in Cases* 3 *and* 4.

5. Voltage dependent $MCN1$ to LG synapse. We now consider the effect of having voltage dependent coupling between $MCN1$ and LG as opposed to a constant conductance synapse. Therefore, we return to (1) with $g_s(V_L) = \bar{g}_s s_\infty(V_L)$, where $s_\infty(V_L)$ is a sigmoidal gating function varying between 0 and 1 of the form

$$s_\infty(V) = \left(1 + \exp\frac{v_k - V_k}{k}\right)^{-1}. \tag{33}$$

Due to this voltage dependency, the amount of excitation that LG receives from $MCN1$ will depend upon the voltage of LG causing LG to receive less excitation when it, itself, is at a low voltage. Thus, when LG is in its interburst, the strength of the synapse will be weaker than in the voltage independent case. When LG is in its burst, the strength of the synapse

will increase to a value near to that of the previous section. As a consequence of the weaker conductance, in all of the Cases 1–4, the voltage dependency will increase the LG interburst duration because s will be required to grow to a larger value of s_{max} for the fixed point to be lost on the left branches of the nullclines. On the right branches of the nullclines, where LG is in its burst, $s_\infty(V_L)$ is closer to 1, and, therefore, the burst duration of LG will not be directly affected as significantly as the interburst duration. However, the value of s_{min} will be slightly larger than when the conductance is not voltage dependent because as V_L decreases on the right branches, s_∞ also decreases. Therefore, s will not need to decrease as much to cause a loss of the fixed point and an end to the LG burst. The increase in the interburst duration in all Cases 1–4, however, is larger than the decrease in burst duration, resulting in an increase in the period of the solutions.

Upon relaxing the conditions that s_2 and $s_{AB\to I}$ jump between their minimum and maximum values instantaneously, in addition to increasing the period in Cases 1–4, the voltage dependency also increases the range of m over which the period of Case 3 equals the period of Case 4. When the rise and fall of s_2 is not instantaneous but occurs on the slow timescale, s_2 moves continuously between its maximum and minimum values. Regardless of whether the conductance of the $MCN1$ synapse to LG is constant or voltage dependent, the same condition must be satisfied for the period of Case 3 to equal the period of Case 4. This condition is that the fixed point on the left branches of the nullclines must be lost through s_L^{off}. That is, m must be chosen to live in a certain interval, say, $[M_1, M_2]$, such that once s_1 has grown large enough for s to reach s_L^{on} while $s_{AB\to I} = 1$, the V_L nullcline must be shifted far enough to the left by s_2 when $s_{AB\to M} = 1$ so that the saddle-node bifurcation does not occur at s_L^{on}. This situation persists until m becomes just larger than M_2. For $M_2 < m < D_c P_{AB}$, when $s_{AB\to I}$ jumps to 1, $s_{AB\to M}$ will already equal 1 so the V_L nullcline will already be to the left. However, before $s_{AB\to I}$ returns 0, $s_{AB\to M}$ will return 0. Consequently, the V_L nullcline will move to the right and the fixed point will not be lost through s_L^{on}.

When s_2 changes on the slow timescale and m is slightly too large as described above, the loss of the fixed point through s_L^{on} often occurs while s_2 is increasing toward 1 but has not yet reached its maximum value of 1. As a specific example, for $\tau_{r1} = \tau_{f1} = 4000$, $\tau_{r2} = \tau_{f2} = 325$, and conductance of the $MCN1$ to LG synapse constant ($s_\infty(V_L) = 1$), the periods of Cases 3 and Cases 4 are the same for $80 \le m \le 275$. Note that the lower bound on the interval is 80 (not 0) since we have relaxed the condition that s_2 and $s_{AB\to I}$ change instantaneously. Once $m > 275$, the saddle-node bifurcation occurs through s_L^{on} instead of s_L^{off}, thus causing the period of Case 4 to be smaller than the period of Case 3.

Now let us consider the effect of the voltage dependent conductance on the position of the V_L nullcline and, therefore, the role it plays in altering the interval of m over which the period in Case 3 equals that in Case 4. When LG is at a low voltage, $s_\infty(V_L)$ is close to 0. Thus, on the left branches of the nullclines, when s_2 decreases to its minimum, there is a much larger jump to the left of the V_L nullcline than when the conductance is constant. Furthermore, even as s_2 increases back to 1, the V_L nullcline remains significantly far to the left until s_2 gets very close to 1. Therefore, even as m increases to the range $[M_2, D_c P_{AB}]$ where $s_{AB\to M}$ jumps back to 0 (forcing s_2 to increase back to 1) just before $s_{AB\to I}$ returns to 0, the V_L nullcline will remain too far to the left (because it takes some amount of time for s_2 to grow close enough to 1) for the saddle-node bifurcation to occur at s_L^{on}. Hence, the fixed point cannot

be lost through s_L^{on} while s_2 is growing toward 1, as occurs when the conductance is constant. Furthermore, when $s_{AB\to I}$ is a half-sine function as in Manor et al. [16], the V_I nullcline spends less time in the downward position. Consequently, there is again a smaller range of time for s to reach s_L^{on} when the V_I nullcline is in the downward position. Accordingly, a larger interval of m will exist for the fixed point on the left branches of the nullclines to be lost from s_L^{off} as in Case 3 when the conductance is voltage dependent than when the conductance is constant. Returning to the example in the above paragraph but now allowing the conductance of the $MCN1$ to LG synapse to be voltage dependent, the interval of m for which the period of Case 3 equals the period of Case 4 extends to $80 \leq m \leq 350$. When s_2 has instantaneous kinetics, there is no significant difference in the range of m between the voltage dependent and non–voltage dependent cases because s_2 is always equal to either 1 or 0 so the fixed point on the left branches cannot be lost while s_2 is increasing toward 1.

6. Discussion. Networks involved in the generation of rhythmic movements often involve sets of reciprocally inhibitory neurons that rely on external stimuli to trigger oscillations or to set the appropriate frequency of the rhythm [2, 6, 25]. It has been observed that while tonic stimulation may often be sufficient to elicit the network activity, the synaptic inputs driving these circuits are themselves rhythmic [5, 29]. An example of this is the pyloric network of the lobster stomatogastric nervous system which receives rhythmic excitatory input. The same effects of this rhythmic input, however, can be achieved through tonic firing of the input cells [19].

Furthermore, it has been noted in many cases that although one source of input is sufficient to produce oscillations in the target network, multiple inputs act together to generate and set the frequency of the network. The heartbeat of the leech, for example, is controlled by pairs of reciprocally inhibitory neurons. These oscillators receive inhibitory input from interneurons that act to coordinate the activity of the separate oscillators [7]. Einum and Buchanan [10] also recently showed that reticulospinal neurons of the lamprey brain stem receive both excitatory and inhibitory rhythmic inputs from neurons in the spinal cord during locomotor activity.

The stomatogastric nervous system of the crab, made up of an asymmetric half-center oscillator, provides a nice example of a system that receives multiple rhythmic synaptic inputs in order to oscillate. The interactions between the gastric mill network and pyloric network have been extensively studied to show how each network acts to influence one another's frequency [3, 4, 28]. In their work, Nadim et al. [21] and Manor et al. [16] considered how the frequency of the gastric mill rhythm is generated and controlled in the presence of both a slow modulatory input and a much faster periodic input.

In this paper, we continued upon the work of Nadim et al. [21] and Manor et al. [16] with the aim of mathematically explaining the experimental results of Wood et al. [30]. Specifically, we addressed the effect of having a rhythmic modulatory input versus a tonic input drive the network oscillations and then how two simultaneous rhythmic inputs work together to determine the network frequency. In order to do this, we incorporated the rhythmicity of the modulatory projection neuron on the existing model of Manor et al. [16]. We then derived conditions on the parameters that dictate the strength and rise and decay rates of the synaptic currents to ensure the existence, local uniqueness, and stability of periodic solutions. Once periodic orbits were established, we derived a formula to estimate the period of such orbits

in the presence and absence of pyloric input to the gastric mill network for both tonic and rhythmic modulatory input.

Using geometric, singular perturbation theory, the multidimensional system is reduced to studying the position of the nullclines in the $V_L - V_I$ phase plane with the variable controlling the amount of excitation provided from the modulatory projection neuron to the gastric mill treated as a parameter. The model shows that the rhythmicity of the projection neuron speeds the gastric mill rhythm by allowing the loss of a relevant stable fixed point to occur at an earlier time than when the input is tonic. Thus, although tonic stimulation of the gastric mill network can generate the gastric mill rhythm, the rhythmicity of the input speeds the frequency of the gastric mill rhythm as in seen by Wood et al. [30].

In the presence of the rhythmic modulatory excitation and fast pyloric inhibition, the timing of the jump of the V_L and V_I nullclines in response to the AB input to $Int1$ and to $MCN1$ determines how the loss of the stable fixed point on either the left or right branches of the nullclines will occur. This, in turn, determines the length of the LG and $Int1$ interburst and burst durations. The frequency calculated from this model matches the experimental results of Wood et al. [30] only when there is either a short delay or no delay in the timing of the two pyloric inputs. In this case, the position of the V_L nullcline in response to the AB inhibition of $MCN1$ prevents the AB disinhibition of LG from ending the LG interburst. Thus, it is as if there is only one source of synaptic input to the gastric mill network. Therefore the analysis gives a possible biological mechanism by which the effect of the two simultaneous synaptic inputs can overlap to result in a frequency equivalent to that of having only one of the inputs present. If the delay is chosen differently, however, the gastric mill rhythm has a higher frequency than when only one of the inputs is present because the position of the V_L nullcline does not prevent LG from getting disinhibited by AB. Thus, the timing of the inputs can be used as a tool to switch between different modes of firing frequency. This may serve as a means by which different chewing patterns are elicited.

The biological and mathematical reductions of the full, compartmental model of Nadim et al. [21] implemented by Manor et al. [16] and extended to this work have proven to be instrumental in understanding the frequency regulation of the gastric mill rhythm and intercircuit coordination with the pyloric network. The reduced model neglects all intrinsic currents and models the neurons in this network as having only leak currents. Despite the severity of the reductions, the reduced network is able to accurately model the gastric mill rhythm and its response to the slow, modulatory, and fast pyloric inputs. In [1], Ambrosio shows that the results found through analysis of the reduced model do extend to the full model. The reduced model also clarifies the relationship between the synaptic rise and decay times of the AB inhibition of $MCN1$ in the full model necessary to obtain the experimentally observed behavior. Furthermore, because the reduced model consisting of passive neurons is able to accurately reproduce the qualitative behavior of the full model, it is clear that the synaptic currents and their timing with respect to one another are the primary components responsible for the dynamics of the gastric mill rhythm. This is important because the ability to ignore the intrinsic dynamics of each of the neurons results in significantly simpler equations. This makes mathematical analysis much more accessible. For example, in this network, we were able to reduce the study of our system to the study of a one-dimensional map. This then allowed us to define a Poincaré map to prove the existence and stability of periodic orbits,

which would have been much more difficult if working in higher dimensions.

Such techniques can be extended to numerous other models whose intrinsic and synaptic currents act on multiple timescales. The leech heartbeat mentioned above, for example, is controlled by a network of reciprocally inhibitory neurons that are dependent upon both synaptic and intrinsic currents. These currents exhibit both fast and slow dynamics and a biophysically detailed model of this network exists [22, 23]. Although this model was shown to accurately reproduce many of the behaviors of the real network, some properties have not yet been able to be reproduced and the significance of certain currents is not yet clearly understood. In particular, a reduced model may give some insight into the extreme sensitivity of the oscillations to the leak current parameters seen in the more detailed models. More generally, a reduced version of this model that is more amenable to mathematical investigation in terms of allowing for a reduction to lower dimensions and phase plane analysis is likely to more clearly reveal many of the underlying properties responsible for such things as the network oscillations and sensitivity to synaptic and intrinsic inputs.

Appendix A. We describe how to calculate the bifurcation points s_L^{off}, s_L^{on}, s_R^{off}, and s_R^{on}. On the left branch, when the two nullclines intersect tangentially for $s_{AB\to I}(t) = 0$, then $s = s_L^{off}$. Similarly, on the right branch when $s = s_R^{off}$ and $s_{AB\to I}(t) = 0$, the fixed point occurs when the two nullclines intersect tangentially. Thus, to calculate s_L^{off} and s_R^{off}, we use the equations for the V_L and V_I nullclines:

$$V_L = F(V_I, s) = \frac{g_{leak,L}E_{leak,L} + \bar{g}_{I\to L}n_\infty(V_I)E_{I\to L} + \bar{g}_s s E_{exc}}{g_{leak,L} + \bar{g}_{I\to L}n_\infty(V_I) + \bar{g}_s s} \tag{34}$$

and

$$V_I = G(V_L, s_{AB\to I}) = \frac{g_{leak,I}E_{leak,I} + \bar{g}_{L\to I}n_\infty(V_L)E_{L\to I} + \bar{g}_{AB\to I}s_{AB\to I}(t)E_{AB\to I}}{g_{leak,I} + \bar{g}_{L\to I}n_\infty(V_L) + \bar{g}_{AB\to I}s_{AB\to I}(t)}. \tag{35}$$

We rewrite (35):

$$\begin{aligned} V_L &= -k_L ln\left[-1 + \frac{\bar{g}_{L\to I}(V_I - E_{L\to I})}{-g_{leak,I}(V_I - E_{leak,I}) - \bar{g}_{AB\to I}s_{AB\to I}(t)(V_I - E_{AB\to I})}\right] + v_L \\ &\doteq \tilde{G}(V_I, s_{AB\to I}). \end{aligned} \tag{36}$$

We find the equation for the tangent point by solving

$$\frac{dF(V_I, s)}{dV_I} = \frac{d\tilde{G}(V_I, s_{AB\to I})}{dV_I}. \tag{37}$$

From (37), we obtain a quadratic equation for s:

$$\begin{aligned} &\bar{g}_s^2 s^2 + s\left[2\bar{g}_s[g_{leak,L} + \bar{g}_{I\to L}n_\infty(V_I)] - \bar{g}_{I\to L}E_{I\to L}\frac{dn_\infty(V_I)}{dV_I}g_s\frac{dV_I}{d\tilde{G}}\right. \\ &\left. + \bar{g}_{I\to L}\frac{dn_\infty(V_I)}{dV_I}\bar{g}_s E_{exc}\frac{dV_I}{d\tilde{G}})\right] + [g_{leak,L} + \bar{g}_{I\to L}n_\infty(V_I)]^2 \\ &- \frac{dV_I}{d\tilde{G}}\left[g_{leak,L}\bar{g}_{I\to L}E_{I\to L}\frac{dn_\infty(V_I)}{dV_I} - \bar{g}_{I\to L}\frac{dn_\infty(V_I)}{dV_I}[g_{leak,L}E_{leak,L}\right. \\ &\left. + \bar{g}_{I\to L}n_\infty(V_I)E_{I\to L}]\right] = 0. \end{aligned} \tag{38}$$

Next, we use the restriction that the tangency of the nullclines must occur at a fixed point. Therefore, we use the equations for the V_L and V_I nullclines to determine the fixed points for different values of s. We rewrite (34) as

$$s = (-g_{leak,L}(V_L - E_{leak,L}) - \bar{g}_{I\to L} n_\infty(V_I)(V_L - E_{I\to L}))/g_s(V_L)(V_L - E_{exc}). \tag{39}$$

We then plug (36) into (39) to obtain an equation for $s = S(V_I)$. This equation says that for each value of V_I there exists a unique value of s which will cause the nullclines to intersect. We then check to see if this value of s also satisfies the quadratic equation (38). If it does, we have found a bifurcation point of the fast subsystem. There are two values of s which satisfy the equations above. The smaller valued one corresponds to s_L^{off}, and the larger corresponds to s_R^{off}. To calculate s_L^{on} and s_R^{on}, we follow the same steps as above but with $s_{AB\to I}(t) = 1$ in (35).

We used the analytically calculated values of the bifurcation points to obtain the results given in section 4. In Case 2, we assume that $s_{max} = s_L^{on}$ and s_{min} is the average of s_R^{off} and s_R^{on}. In Case 3, we assume the maximum value that s takes is s_L^{off} while $s_2 = 1$, which implies that $s_{max} = s_L^{off}$ and the minimum value s assumes is s_R^{off} when $s_2 = s_{2min}$. Therefore, s_{min} is $\frac{s_R^{off}}{s_{2min}}$. In Case 4, the values of s_{max} and s_{min} depend on m. For m near 0, for example, s_{max} and s_{min} are calculated as in Case 3. We also numerically solved (1)–(8) and (33). The parameter values used in the numerical calculations are given in Table 2.

Table 2
Parameters of the reduced model.

$g_{leak,L} = 1mS/cm^2$	$E_{leak,L} = -60mV$	$\bar{g}_{I\to L} = 5mS/cm^2$	$E_{I\to L} = -80mV$
$g_{leak,I} = .75mS/cm^2$	$E_{leak,I} = 10mV$	$\bar{g}_{L\to I} = 2mS/cm^2$	$E_{L\to I} = -80mV$
$g_{leak,M} = 2mS/cm^2$	$E_{leak,M} = 10mV$	$\bar{g}_{AB\to M} = 15mS/cm^2$	$E_{AB\to M} = -60mV$
$g_s = 4mS/cm^2$	$E_{exc} = 43mV$	$\bar{g}_{AB\to I} = .9mS/cm^2$	$E_{AB\to I} = -60mV$
$V_T = -30mV$	$v_x = -30mV$	$k_x = 4mV$	

REFERENCES

[1] C. Ambrosio, *The Control of Frequency of an Excitable Network Simultaneously Subjected to Multiple Oscillatory Inputs*, Ph.D. thesis, New Jersey Institute of Technology, Newark, NJ, 2005.

[2] S. P. Aiken, F. M. Kuenzi, and N. Dale, *Xenopus embryonic spinal neurons recorded in situ with patch-clamp electrodes—conditional oscillator after all?*, Eur. J. Neurosci., 18 (2003), pp. 333–343.

[3] M. Bartos, Y. Manor, F. Nadim, E. Marder, and M. P. Nusbaum, *Coordination of fast and slow rhythmic neuronal circuits*, J. Neurosci., 19 (1999), pp. 6650–6660.

[4] M. Bartos and M. Nusbaum, *Intercircuit control of motor pattern modulation by presynaptic inhibition*, J. Neurosci., 17 (1997), pp. 2247–2256.

[5] M. J. Bourque and A. Kolta, *Properties and interconnections of trigeminal interneurons of the lateral pointine reticular formation in the rat*, J. Neurophys., 86 (2001), pp. 2583–2596.

[6] R. L. Calabrese, *Oscillations in motor pattern-generating networks*, Curr. Opin. Neurobiol., 5 (1995), pp. 816–823.

[7] R. Calabrese, F. Nadim, and O. Olsen, *Heartbeat control in the medicinal leech: A model system for understanding the origin, coordination, and modulation of rhythmic motor patterns*, J. Neurobiol., 27 (1995), pp. 390–402.

[8] M. J. Coleman, P. Meyrand, and M. P. Nusbaum, *A switch between two modes of synaptic transmission mediated by presynaptic inhibition*, Nature, 378 (1995), pp. 502–505.
[9] M. J. Coleman and M. P. Nusbaum, *Functional consequences of compartmentalization of synaptic input*, J. Neurosci., 14 (1994), pp. 6544–6552.
[10] J. F. Einum and J. T. Buchanan, *Reticulospinal neurons receive direct spinobulbar inputs during locomotor activity in lamprey*, J. Neurophys., 92 (2004), pp. 1384–1390.
[11] B. Ermentrout, *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*, Software Environ. Tools 14, SIAM, Philadelphia, 2002.
[12] S. Grillner and T. Matsushima, *The neural network underlying locomotion in lamprey—synaptic and cellular mechanisms*, Neuron, 7 (1991), pp. 1–15.
[13] S. Grillner and P. Wallen, *Central pattern generators for locomotion, with special reference to vertebrates*, Ann. Rev. Neurosci., 8 (1985), pp. 233–261.
[14] O. Kiehn and O. Kjaerulff, *Distribution of central pattern generators for rhythmic motor outputs in the spinal cord of limbed vertebrates*, Ann. NY Acad. Sci., 860 (1998), pp. 110–129.
[15] S. P. Lieske, M. Thoby-Brisson, P. Telgkamp, and J. M. Ramirez, *Reconfiguration of the neural network controlling multiple breathing patterns: Eupnea, sighs and gasps*, Nat. Neurosci., 3 (2000), pp. 600–607.
[16] Y. Manor, F. Nadim, S. Epstein, J. Ritt, E. Marder, and N. Kopell, *Network oscillations generated by balancing graded asymmetric reciprocal inhibition in passive neurons*, J. Neurosci., 19 (1999), pp. 2765–2779.
[17] E. Marder and D. Bucher, *Central pattern generators and the control of rhythmic movements*, Curr. Biol., 11 (2001), pp. R986–R996.
[18] E. Marder and R. L. Calabrese, *Principles of rhythmic motor pattern generation*, Physiol. Rev., 76 (1996), pp. 687–717.
[19] J. Miller and A. Selverston, *Mechanisms underlying pattern generation in lobster stomatogastric ganglion as determined by selective inactivation of identified neurons.* IV. *Network properties of pyloric system*, J. Neurophys., 48 (1982), pp. 1416–1432.
[20] E. Mishchenko and N. Rozov, *Differential Equations with Small Parameters and Relaxation Oscillations*, Plenum Press, New York, 1980.
[21] F. Nadim, Y. Manor, M. P. Nusbaum, and E. Marder, *Frequency regulation of a slow rhythm by a fast periodic input*, J. Neurosci., 18 (1998), pp. 5053–5067.
[22] F. Nadim, O. H. Olsen, E. De Schutter, and R. L. Calabrese, *Modeling the leech heartbeat elemental oscillator* I. *Interactions of intrinsic and synaptic currents*, J. Comput. Neurosci., 2 (1995), pp. 215–235.
[23] O. H. Olsen, F. Nadim, and R. L. Calabrese, *Modeling the leech heartbeat elemental oscillator* II. *Exploring the parameter space*, J. Comput. Neurosci., 2 (1995), pp. 237–257.
[24] M. P. Nusbaum and M. P. Beenhakker, *A small-systems approach to motor pattern generation*, Nature, 417 (2002), pp. 343–350.
[25] R. A. Satterlie, *Reciprocal inhibition and postinhibitory rebound produce reverberation in a locomotor pattern generator*, Science, 229 (1985), pp. 402–404.
[26] D. Somers and N. Kopell, *Rapid synchronization through fast threshold modulation*, Biol. Cyber., 68 (1993), pp. 393–407.
[27] P. S. G. Stein, S. Grillner, A. I. Selverston, and D. G. Stuart, eds., *Neurons, Networks, and Motor Behavior*, MIT Press, Cambridge, MA, 1997.
[28] J. B. Thuma, L. G. Morris, A. L. Weaver, and S. L. Hooper, *Lobster (Panulirus interruptus) pyloric muscles express the motor patterns of three neural networks, only one of which innervates the muscles*, J. Neurosci., 23 (2003), pp. 8911–8920.
[29] J. C. Weeks, *Neuronal basis of leech swimming: Separation of swim initiation, pattern generation, and intersegmental coordination by selective lesions*, J. Neurophys., 45 (1981), pp. 698–723.
[30] D. E. Wood, Y. Manor, F. Nadim, and M. P. Nusbaum, *Intercircuit control via rhythmic regulation of projection neuron activity*, J. Neurosci., 24 (2004), pp. 7455–7463.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 5, No. 1, pp. 140–156

Sufficient Conditions for Fast Switching Synchronization in Time-Varying Network Topologies*

Daniel J. Stilwell†, Erik M. Bollt‡, and D. Gray Roberson†

Abstract. In previous work [J. D. Skufca and E. Bollt, *Mathematical Biosciences and Engineering*, 1 (2004), pp. 347–359], empirical evidence indicated that a time-varying network could propagate sufficient information to allow synchronization of the sometimes coupled oscillators, despite an instantaneously disconnected topology. We prove here that if the network of oscillators synchronizes for the static time-average of the topology, then the network will synchronize with the time-varying topology if the time-average is achieved sufficiently fast. Fast switching, fast on the time-scale of the coupled oscillators, overcomes the desynchronizing decoherence suggested by disconnected instantaneous networks. This result agrees in spirit with that of [J. D. Skufca and E. Bollt, *Mathematical Biosciences and Engineering*, 1 (2004), pp. 347–359] where empirical evidence suggested that a moving averaged graph Laplacian could be used in the master-stability function analysis [L. M. Pecora and T. L. Carroll, *Phys. Rev. Lett.*, 80 (1998), pp. 2109–2112]. A new fast switching stability criterion herein gives sufficiency of a fast switching network leading to synchronization. Although this sufficient condition appears to be very conservative, it provides new insights about the requirements for synchronization when the network topology is time-varying. In particular, it can be shown that networks of oscillators can synchronize even if at every point in time the frozen-time network topology is insufficiently connected to achieve synchronization.

Key words. nonlinear dynamics, complex networks, synchronization, controllability and observability, self-organization, communication in complex networks

AMS subject classifications. 37A25, 37D45, 37N25, 93B05, 94C15, 05C80, 92D30

DOI. 10.1137/050625229

1. Introduction.

Since Huygen's early observations of weakly coupled clock pendula [31], synchronization has been found in a wide variety of phenomena, ranging from biological systems that include fireflies in the forest [14, 41], animal gates [16], descriptions of the heart [29, 59, 27], and improved understanding of brain seizures [43] to chemistry [37], nonlinear optics [60, 61, 62], and meteorology [20]. See one of the many excellent reviews now available, including [11, 50, 15, 57, 26, 42]. In particular, it has been known for more than 20 years that chaotic oscillators can synchronize under suitable coupling [24, 3, 46]. Meanwhile, in recent years, the study of large scale, random networks has become an extremely active area with the advent of advances in both theory and scientific application across numerous disciplines,

*Received by the editors February 24, 2005; accepted for publication (in revised form) by B. Ermentrout September 13, 2005; published electronically March 3, 2006.
http://www.siam.org/journals/siads/5-1/62522.html

†Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 (stilwell@vt.edu, roberson@vt.edu). The work of the first author was supported by the National Science Foundation under IIS0238092 and OCE0354810 and the Office of Naval Research under N00014-03-1-0444.

‡Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY 13699-5815 (bolltem@clarkson.edu). The work of this author has been supported by the National Science Foundation under DMS0404778.

as reviewed in [44, 19, 4, 5, 63], for example. Recent investigations have sought to characterize how oscillator elements coupled according to a large scale network architecture are impacted by the choice of architecture and corresponding spectral properties of the network [6, 34, 33, 30, 25, 38]. In particular, the master-stability function formalism [47, 48] relates spectral properties of the graph Laplacian of the network to synchrony of supported oscillators, and this has been used in the study of synchronization stability on arbitrary network architecture [6].

Despite the very large volume of literature to be found, the great majority of research activities have been focused on static networks whose connectivity and coupling strengths are constant in time. For example, static networks are assumed for the analysis of [47, 48, 6]. However, there are applications where the coupling strengths and even the network topology can evolve in time. Recent work such as [56, 32, 65] are among the few to consider time-dependent couplings. See also [35] in which a so-called *function dynamics* gives rise to networks that evolve according to a dynamical system, somewhat similarly to our networks. The recent work of Blykh, Belykh, and Hasler on "blinking-systems" [9] has complementary results to ours, which we review at the end of this introduction.

In prior work [54], we were motivated by applications that include how a disease might occur in a network of agents in which the agents move, but the disease itself has its own time-scale. We describe a competition of two time-scales. Said plainly, the disease has a natural typical incubation rate and a natural infections rate (for example in a susceptible→exposed→infected→recovered (SEIR) model), so if a susceptible agent does not come in contact with an infected agent in the disease time-scale, then there should be no new infection. Mathematically, we constructed a "moving neighborhood network" (MNN), a network of agents which move ergodically and connect when in close proximity to each other. Such a network was shown to lead to a Poisson distributed degree distribution instantaneously, and hence the network was typically instantaneously disconnected. It consists of typically many small subcomponents at each instant. Such a description alone would suggest that there could be no global synchronization of the oscillators carried by each agent which are coupled according to the disconnected network. However, it was found that if the agents move quickly enough, then roughly described, in any recent time window, a given agent might be likely to have had some amount of coupling to most other agents. It turns out that for fast enough moving agents, these random time-varying connections were enough to overcome even chaotic oscillators' sensitive dependence tendency to drift apart asynchronously. We formalized this idea by introducing a new description of the connectivity, a "moving averaged" graph Laplacian. We showed empirically that the spectrum of this construction works quite well together with the master-stability formalism to accurately predict synchronization.

Besides our original motivation in mathematical epidemiology, it can be argued that this work has strong connections to ad hoc communication systems and control systems on time-varying networks. Fundamental connections between chaotic oscillations and proof of synchronization through symbolic dynamics [55, 49] and control [17, 12, 28] and even definition of chaos through symbolic dynamics suggest this work is rooted in a description of information flow in the network.

Coordinated control for platoons of autonomous vehicles can also be addressed using network concepts [18, 51, 21]. Each vehicle is represented by a node, and communication or

mutual sensing is represented by connections between nodes. In [21] the average position of a platoon of vehicles is regulated, and the graph Laplacian is used to describe communication between vehicles. It is shown that the spectrum of the graph Laplacian can be used to indicate stability of the controlled system. As pointed out in [51], the use of a graph Laplacian is not entirely common since it appears naturally for only a limited class of control objectives. The simplified model form explored in this paper, (2.1), is morally inspired by these problems where there seems to be a notion of average information propagation in a network.

These considerations have led us in this work to consider a simplified version of the moving agents of our MNN model. Considering certain time-varying coupled network architectures, we can now make rigorous but sufficient statements concerning fast switching, and we use mathematical machinery not so far typically used in the synchronization community. The main result of this work comes from the fields of switched systems, and specifically builds on the concept of *fast switching*. Switched systems are a class of systems whose coefficients undergo abrupt change. For example, consider the linear state equation

$$\dot{x}(t) = A_{\rho(t)}x(t), \tag{1.1}$$

where $\rho(t) : \mathbb{R} \mapsto \mathbb{Z}_+$ is a switching sequence that selects elements from a family of matrix-valued coefficients $\Theta = \{A_1, A_2, \dots\}$. When each element of Θ is Hurwitz, stability of (1.1) is guaranteed if $\rho(t)$ switches sufficiently slowly. Further restrictions on elements of Θ, such as existence of a common Lyapunov function, can guarantee stability for arbitrary switching functions, including those that are not slow. An excellent overview of the field of switched systems and control is presented in [40] and in the book [39].

Even when the elements of Θ are not all Hurwitz, stability of (1.1) is still possible, although the class of switching functions is further restricted. For example, in [64] a stabilizing switching sequence is determined by selecting elements of Θ based on the location of the state $x(t)$ in the state space. This is essentially a form of state feedback.

When no elements of Θ are Hurwitz, which is the case that is considered herein, stability of (1.1) can sometimes be guaranteed if the switching sequence is sufficiently fast. Loosely speaking, it can be shown that

$$\dot{x}(t) = A_{\rho(t/\varepsilon)}x(t) \tag{1.2}$$

is asymptotically stable if there exists a constant T such that the time-average

$$\frac{1}{T}\int_t^{t+T} A_{\rho(\tau)}d\tau$$

is Hurwitz for all t, and if ε is sufficiently small. This fact has been established in [36, 7, 58] for several classes of linear systems related to (1.2). Similar results have been presented in [2, 1] for classes of nonautonomous nonlinear systems where time is parameterized by t/ε as in (1.2). In this case, stability of a specific average system implies stability of the original system if ε is sufficiently small. In addition, this work requires the existence of a Lyapunov function that is related to a certain average of the system but which is not a function of time. This requirement is too restrictive for the class of linear time-varying systems considered herein.

A new fast switching stability condition, presented in section 3, is derived in order to assess local stability of networked oscillators about the synchronization manifold.

Recent complementary results have been authored by Belykh, Belykh, and Hasler, in [8] and [9]. They developed a method called the "connection graph stability method" whereby even for networks of time-varying connections, a bound is established based on explicitly considering the total length of all paths through edges on the network connection graph. Their bound links average path length in a way which allows them to consider a small-world regular $2k$-nearest lattice with long range connections which are switched on and off with a certain probability p during short time intervals τ. They also conclude synchronization thresholds not dissimilar to ours relating the switching time of the necessary (long range) connections which must be small relative to the synchronization time. However the specifics of their methods, rooted explicitly in graph theory, are different from ours, which are rooted in the field of switched systems from control engineering. Consequently, the specific details of both hypothesis and conclusions in our work are not the same as those in [8, 9].

2. Preliminaries. We consider a network of coupled oscillators consisting of r identical oscillators,

$$\dot{x}_i(t) = f(x_i(t)) + \sigma B \sum_{j=1}^{r} l_{ij}(t) x_j(t), \quad i = 1, \dots, r, \tag{2.1}$$

where $x_i(t) \in \mathbb{R}^n$ is the state of oscillator i, $B \in \mathbb{R}^{n \times n}$, and the scalar σ is a control variable that sets the coupling strength between oscillators. This model is inspired by the applied questions discussed in the introduction in that it has time-varying connections which still allow for enough connectionism for global synchronization, and it is of a sufficiently simplified form to admit a complete and rigorous analysis. The scalars $l_{ij}(t)$ are elements of the graph Laplacian of the network graph and describe the interconnections between individual oscillators. Let $G(t)$ be the time-varying graph consisting of r vertices v_i together with a set of ordered pairs of vertices $\{v_i, v_j\}$ that define the edges of the graph. In this work, we assume that $\{v_i, v_i\} \in G(t)$ for $i = 1, \dots, r$. Let $\tilde{G}(t)$ be the $r \times r$ adjacency matrix corresponding to $G(t)$; then $\tilde{G}_{i,j}(t) = 1$ if $\{v_i, v_j\}$ is an edge of the graph at time t and $\tilde{G}_{i,j}(t) = 0$ otherwise. The graph Laplacian is defined as

$$L(t) = \operatorname{diag}(d(t)) - \tilde{G}(t), \tag{2.2}$$

where the ith element of $d(t) \in \mathbb{R}^r$ is the number of vertices that vertex i is connected to, including itself. Note that solutions of (2.1) must be interpreted in the weak sense of Carathéodory. Indeed, the presence of a switching network leads to nonsmooth solutions, i.e., piecewise differentiable solutions which are smooth only between switching instants. For existence and uniqueness theorems for such nonlinear systems, one may refer to [22, 53].

Synchronization can be assessed by examining local asymptotic stability of the oscillators along the synchronization manifold. Linearizing each oscillator (2.1) about the trajectory $x^o(t)$, which is assumed to be on the synchronization manifold, yields

$$\dot{z}_i(t) = F(t) z_i(t) + \sigma B \sum_{j=1}^{r} l_{ij}(t) z_j(t), \tag{2.3}$$

where

$$z_i(t) = x_i(t) - x^o(t), \tag{2.4}$$

and $F(t) = Df$ evaluated at $x^o(t)$. Let $L(t)$ be the $r \times r$ matrix composed of entries $l_{ij}(t)$; then the system of linearized coupled oscillators is written

$$\begin{aligned} \dot{z}(t) &= (I_r \otimes F(t) + \sigma(I_n \otimes B)(L \otimes I_r))\, z(t) \\ &= (I_r \otimes F(t) + \sigma L \otimes B)\, z(t), \end{aligned} \tag{2.5}$$

where "$\otimes$" is the Kronecker product and $z(t) = [z_1^T(t), \ldots, z_r^T(t)]^T$. Standard properties of the Kronecker product are utilized here and in what follows, including: for conformable matrices A, B, C, and D, $(A \otimes B)(C \otimes D) = AC \otimes BD$. Notation throughout is standard, and we assume that $\|\cdot\|$ refers to an induced norm.

It has been shown in [47, 48] that the linearized set of oscillators (2.5) can be decomposed into two components: one that evolves along the synchronization manifold and another that evolves transverse to the synchronization manifold. If the latter component is asymptotically stable, then the set of oscillators will synchronize.

The claimed decomposition is achieved using a Schur transformation. We briefly describe the decomposition since it plays a central role in our assessment of synchronization under time-varying network connections. Let $P \in \mathbb{R}^{n\times n}$ be a unitary matrix such that $U = P^{-1}LP$, where U is upper triangular. The eigenvalues $\lambda_1, \ldots, \lambda_r$ of L appear on the main diagonal of U. The transformation is not unique, in that the triangular structure of U can be obtained with the eigenvalues of L in any order along the diagonal. A change of variables $\xi(t) = (P \otimes I_n)^{-1} z(t)$ yields

$$\begin{aligned} \dot{\xi}(t) &= (P \otimes I_n)^{-1} (I_r \otimes F(t) + \sigma L \otimes B)\, (P \otimes I_n)\, \xi(t) \\ &= \left(I_r \otimes F(t) + \sigma P^{-1} L P \otimes B\right) \xi(t) \\ &= (I_r \otimes F(t) + \sigma U \otimes B)\, \xi(t). \end{aligned} \tag{2.6}$$

Due to the block-diagonal structure of $I_r \otimes F(t)$ and the upper triangular structure of U, stability of (2.6) is equivalent to stability of the subsystems

$$\dot{\xi}_i(t) = (F(t) + \sigma\lambda_i B)\xi_i(t), \qquad i = 1, \ldots, r, \tag{2.7}$$

where $\lambda_1, \ldots, \lambda_r$ are the eigenvalues of L. Note that since the row sums of L are zero, the spectrum of L contains at least one zero eigenvalue. We assign $\lambda_1 = 0$, which is consistent with particular choices of the transformation matrix P. Thus

$$\dot{\xi}_1(t) = F(t)\xi_1(t)$$

evolves along the synchronization manifold, while (2.7) with $i = 2, \ldots, r$ evolves transverse to the synchronization manifold [47]. Since the oscillators are assumed identical, the (identity) synchronization manifold is invariant for all couplings, the question being its stability. The set of coupled oscillators will synchronize if the synchronization manifold is stable, or, equivalently, if (2.7) with $i = 2, \ldots, r$ is asymptotically stable.

3. Main result. For a given static network, the master-stability function characterizes values of σ for which a set of coupled oscillators (2.5) synchronizes [47, 6, 23]. The graph Laplacian matrix L has r eigenvalues, which we label

$$0 = \lambda_1 \leq \cdots \leq \lambda_r = \lambda_{max}. \tag{3.1}$$

The stability question reduces by linear perturbation analysis to a constraint upon the eigenvalues of the Laplacian,

$$\sigma\lambda_i \in (\alpha_1, \alpha_2) \quad \forall i = 2, \ldots, r, \tag{3.2}$$

where α_1, α_2 are given by the *master-stability function* (MSF), a property of the oscillator equations. For σ small, synchronization is unstable if $\sigma\lambda_2 < \alpha_1$; as σ is increased, instability arises when

$$\sigma\lambda_{max} > \alpha_2. \tag{3.3}$$

By algebraic manipulation of (3.2), if

$$\frac{\lambda_{max}}{\lambda_2} < \frac{\alpha_2}{\alpha_1} =: \beta, \tag{3.4}$$

then there is a coupling parameter, σ_s, that will stabilize the synchronized state. For some networks, no value of σ satisfies (3.2). In particular, since the multiplicity of the zero eigenvalue defines the number of completely reducible subcomponents, if $\lambda_2 = 0$, the network is not connected, and synchronization is not stable. However, even when $\lambda_2 > 0$, if the spread of eigenvalues is too great, then synchronization may still not be achievable.

For the case of a time-varying network topology, represented by $L(t)$, our principal contribution is to show that the network can synchronize even if the static network for any frozen value of t is insufficient to support synchronization. Specifically, we show that the time-average of $L(t)$, not the frozen values of $L(t)$, is an indicator of synchronization. If the time-average of $L(t)$ is sufficient to support synchronization, then the time-varying network will synchronize if the time-average is achieved sufficiently fast.

Theorem 3.1. *Suppose a set of coupled oscillators with linearized dynamics*

$$\dot{z}_s(t) = \left(I_r \otimes F(t) + \sigma\bar{L} \otimes B\right) z_s(t) \tag{3.5}$$

has an asymptotically stable synchronization manifold, regarding $z(t) \to 0$ in (2.4). Then there exists a positive scalar ε^* such that the set of oscillators with linearized dynamics

$$\dot{z}_a(t) = \left(I_r \otimes F(t) + \sigma L(t/\varepsilon) \otimes B\right) z_a(t) \tag{3.6}$$

and time-varying network connections $L(t)$ are also asymptotically stably synchronized, again regarding $z(t) \to 0$ in (2.4), for all fixed $0 < \varepsilon < \varepsilon^*$, if there exists a constant T such that $L(t)$ satisfies

$$\frac{1}{T}\int_t^{t+T} L(\tau)d\tau = \bar{L} \tag{3.7}$$

and the column sums of $L(t)$ are all zero for all t.

Remark 3.2. *Since $L(t)$ represents a time-varying network, we may assume that for each value of t, $L(t)$ is a graph Laplacian as defined in (2.2). Thus the time-average $\bar{L}$ in (3.7) is not a graph Laplacian. In other words, $\bar{L}$ does not necessarily correspond to a particular network topology and arises only as the time-average of $L(t)$. However, $\bar{L}$ does inherit the zero row and column sum property of $L(t)$.*

A preliminary lemma is required to prove Theorem 3.1, the proof of which appears in the appendix.

Lemma 3.3. *Suppose there exists a constant T for which the matrix-valued function $E(t)$ is such that*

$$\frac{1}{T}\int_t^{t+T} E(\tau)d\tau = \bar{E} \tag{3.8}$$

for all t and

$$\dot{x}(t) = (A(t) + \bar{E})x(t), \qquad x(t_o) = x_o, \qquad t \geq t_o, \tag{3.9}$$

is uniformly exponentially stable. Then there exists $\varepsilon^* > 0$ such that for all fixed $\varepsilon \in (0, \varepsilon^*)$,

$$\dot{z}(t) = (A(t) + E(t/\varepsilon))z(t), \qquad z(t_o) = z_o, \qquad t \geq t_o, \tag{3.10}$$

is uniformly exponentially stable.

Proof of Theorem 3.1. First we show that the Schur transformation that decomposes the set of oscillators (3.5) with static $\bar{L}$ also induces a similar decomposition for (3.6) with time-varying $L(t)$. Then we apply Lemma 3.3 to show that the modes of the system that evolve transverse to the synchronization manifold are stable if ε is sufficiently small.

Let $P \in \mathbb{R}^{r\times r}$ be a unitary matrix such that $\bar{U} = P^{-1}\bar{L}P$, where

$$\bar{U} = \begin{bmatrix} 0 & \bar{U}_1 \\ 0_{(r-1)\times 1} & \bar{U}_2 \end{bmatrix}$$

is the Schur transformation of $\bar{L}$, and $\bar{U}_2 \in \mathbb{R}^{(r-1)\times(r-1)}$ is upper triangular. Without loss of generality, we have assumed that the leftmost column of P is the unity norm eigenvector $[\sqrt{1/r}, \ldots, \sqrt{1/r}]^T$ corresponding to a zero eigenvalue. The change of variables $\xi_s(t) = (P \otimes I)^{-1} z_s(t)$ yields the decomposition $\xi_s = [\xi_{s1}, \xi_{s2}]^T$, where $\xi_{s1} \in \mathbb{R}^n$, $\xi_{s2} \in \mathbb{R}^{n(r-1)}$, and ξ_{s2} satisfies

$$\dot{\xi}_{s2}(t) = \left(I_{r-1} \otimes F(t) + \sigma\bar{U}_2 \otimes B\right)\xi_{s2}(t). \tag{3.11}$$

As discussed in section 2, (3.11) is asymptotically stable by hypothesis.

We now consider the same change of variables applied to (3.6). First, note that

$$U(t) = P^{-1}L(t)P = \begin{bmatrix} 0 & U_1(t) \\ 0_{(r-1)\times 1} & U_2(t) \end{bmatrix}$$

since the column sums for $L(t)$ are zero for all t. The change of variables $\xi_a(t) = (P \otimes I)^{-1} z_a(t)$ yields the decomposition $\xi_a = [\xi_{a1}, \xi_{a2}]^T$, where $\xi_{a1} \in \mathbb{R}^n$ evolves along the synchronization

manifold and $\xi_{a2} \in \mathbb{R}^{n(r-1)}$ evolves transverse to the synchronization manifold. To verify that the oscillators synchronize, it is sufficient to show that

$$\dot{\xi_{a2}}(t) = (I_{r-1} \otimes F(t) + \sigma U_2(t/\varepsilon) \otimes B)\, \xi_{a2}(t) \tag{3.12}$$

is asymptotically stable when ε is sufficiently small. Since

$$\begin{aligned} \bar{U} &= P^{-1}\bar{L}P \\ &= \frac{1}{T}\int_t^{t+T} P^{-1}L(\tau)P d\tau \\ &= \frac{1}{T}\int_t^{t+T} U(\tau) d\tau \end{aligned}$$

we conclude that

$$\bar{U}_2 = \frac{1}{T}\int_t^{t+T} U_2(\tau) d\tau. \tag{3.13}$$

Thus the desired result is obtained by direct application of Lemma 3.3 along with (3.11), (3.12), and (3.13). ■

4. Illustration. To illustrate fast switching concepts applied to synchronization of a set of oscillators, we consider a set of r Rössler attractors

$$\begin{aligned} \dot{x}_i(t) &= -y_i(t) - z_i(t) - \sigma \sum_{j=1}^{r} l_{ij}(t/\varepsilon) x_j(t), \\ \dot{y}_i(t) &= x_i(t) + a y_i(t), \\ \dot{z}_i(t) &= b + z_i(t)(x_i(t) - c), \end{aligned} \tag{4.1}$$

where $i = 1, \ldots, r$, $a = 0.165$, $b = 0.2$, $c = 10$, and $\sigma = 0.3$. Oscillators are coupled through the x_i variables via $l_{ij}(t)$. Coupling between subsystems (nodes) is defined by a time-varying graph $G(t)$, with corresponding adjacency matrix $\tilde{G}(t)$. The graph Laplacian $L(t)$, with entries $l_{ij}(t)$, is defined as in (2.2).

For the purposes of illustration, we choose a set of five graphs and corresponding adjacency matrices $\tilde{G}_1, \ldots, \tilde{G}_5$ with the property that none of them are fully connected. That is, each graph contains pairs of nodes that do not have a path between them. However, the union of vertices over all five graphs yields a fully connected graph with the longest path between nodes containing no more than two other nodes. All five graphs and the union of graph vertices are shown in Figure 1.

A simple strategy is chosen for switching among graph Laplacians associated with the set of graphs. We choose the T-periodic $L(t)$ defined over one period by

$$L(t) = \sum_{i=1}^{5} L_i \chi_{[(i-1)T/5,\, iT/5)}(t),$$

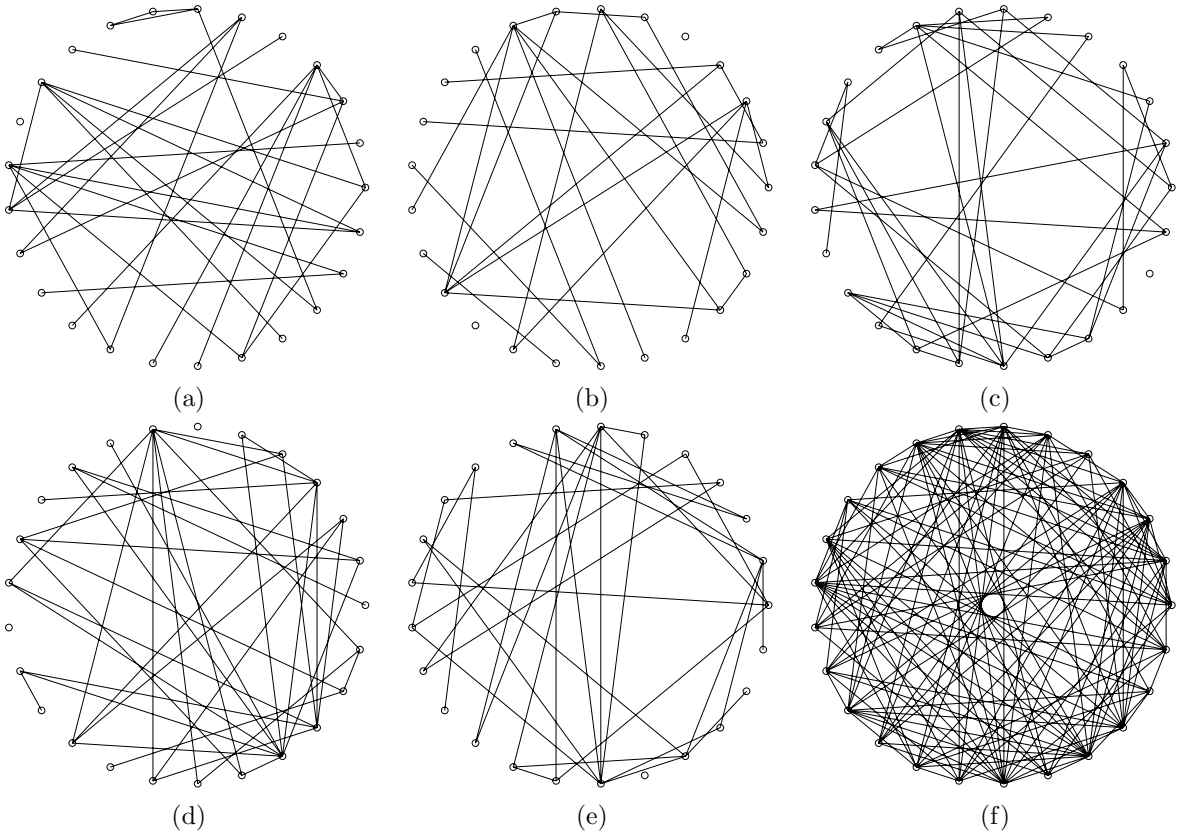


Figure 1. (a)–(e) *are graphs* G_1 *through* G_5*, respectively, while* (f) *is the union of graphs.*

where $\chi_{[t_1,\,t_2)}(t)$ is the indicator function with support $[t_1,\,t_2)$. The time-average of $L(t)$ is

$$\begin{aligned}\bar{L} &= \frac{1}{\varepsilon T}\int_0^{\varepsilon T} L(t/\varepsilon)dt \\ &= \frac{1}{5}\sum_{i=1}^{5} L_i.\end{aligned} \tag{4.2}$$

Toward computing the upper bound for ε given by (A.11), the set of coupled oscillators (4.1) with coupling defined by (4.2) are integrated. The x-coordinate for each oscillator is shown in Figure 2. The x-coordinates clearly synchronize. Asymptotic stability of the oscillators with respect to the synchronization manifold is suggested by plotting the sum-square deviation of the states

$$\sum_{i=1}^{r}(x_i(t)-\mu_x(t))^2+(y_i(t)-\mu_y(t))^2+(z_i(t)-\mu_z(t))^2 \tag{4.3}$$

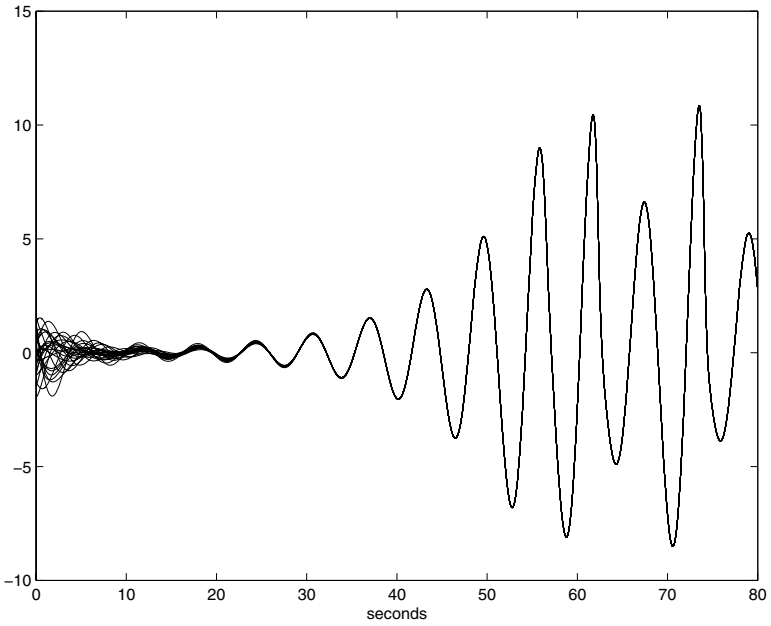


Figure 2. *The x-coordinate for the set of coupled Rössler attractors using the average graph Laplacian.*

about the averages

$$\mu_x(t) = \frac{1}{r}\sum_{i=1}^{r} x_i(t),$$

where $\mu_y(t)$ and $\mu_z(t)$ are defined similarly. Approximately exponential decay of (4.3) is evident in Figure 3, indicating that the oscillators synchronize.

The linear time-varying system (2.5) corresponding to the set of coupled Rössler attractors is computed from the Jacobian of the right-hand side of (4.1) evaluated at the solutions shown in Figure 2.

As described in the proof of Lemma 3.3, a Schur transformation U that diagonalizes $\bar{L}$ is computed and used as a state transformation to decompose the linear time-varying system (2.5) into a component that evolves along the synchronization manifold and another component that evolves transverse to the synchronization manifold. The upper bound for ε given in Theorem 3.1 is computed from the component of the linear system that evolves transverse to the synchronization manifold,

$$\dot{\xi}_{a2}(t) = (I_{r-1} \otimes F(t) + \sigma U_2 \otimes B)\xi_{a2}(t).$$

We now estimate the constants α, ρ, η, and μ needed to compute the right-hand side of (A.11) in the proof of Lemma 3.3 (see the appendix). This is used to compute a maximum value of ε. The constant α is computed from (A.4), while the transition matrix is computed from

$$\dot{\Phi}(t,\tau) = (I_{r-1} \otimes F(t) + \sigma U_2 \otimes B)\Phi(t,\tau), \qquad \Phi(\tau,\tau) = I.$$

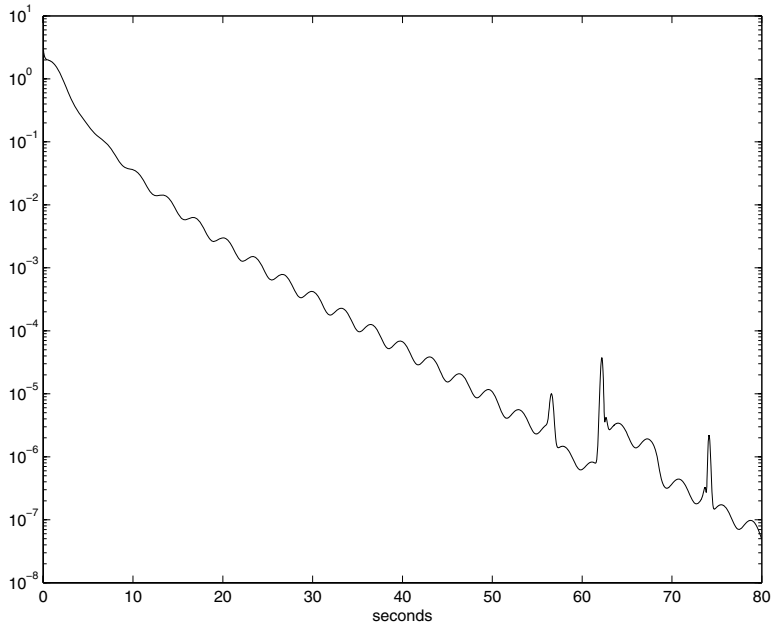


Figure 3. *Sum-square deviation in* (4.3) *for the set of coupled Rössler attractors using the average network* $\bar{L}$.

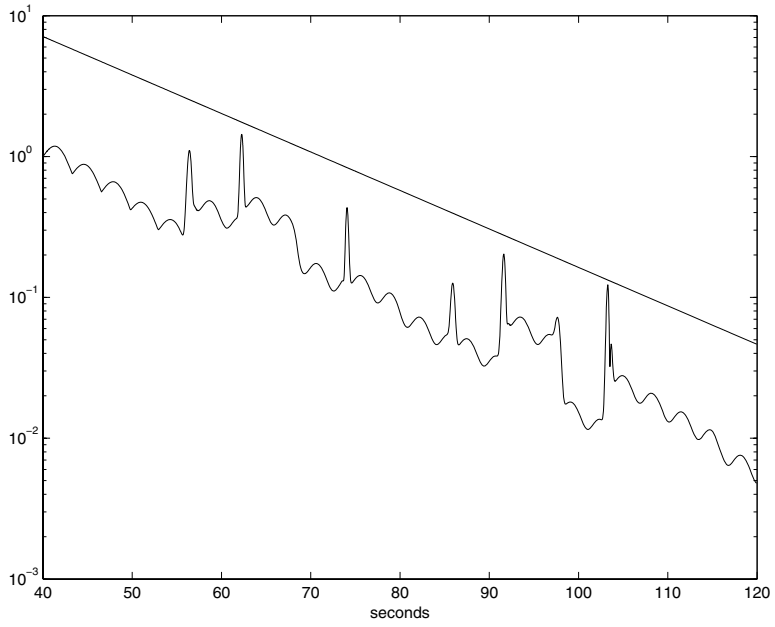


Figure 4. *Norm of the transition matrix* $\Phi(t,\tau)$ *along with an exponentially decaying upper bound.*

The norm of the transition matrix $\|\Phi(t,\tau)\|$ is shown in Figure 4. The initial time τ is chosen to be 40 seconds to ensure that the states of (4.1) are reasonably close to the synchronization manifold. An upper bound that satisfies $\|\Phi(t,\tau)\| \leq \gamma e^{-\lambda(t-\tau)}$ is also shown in Figure 4. The

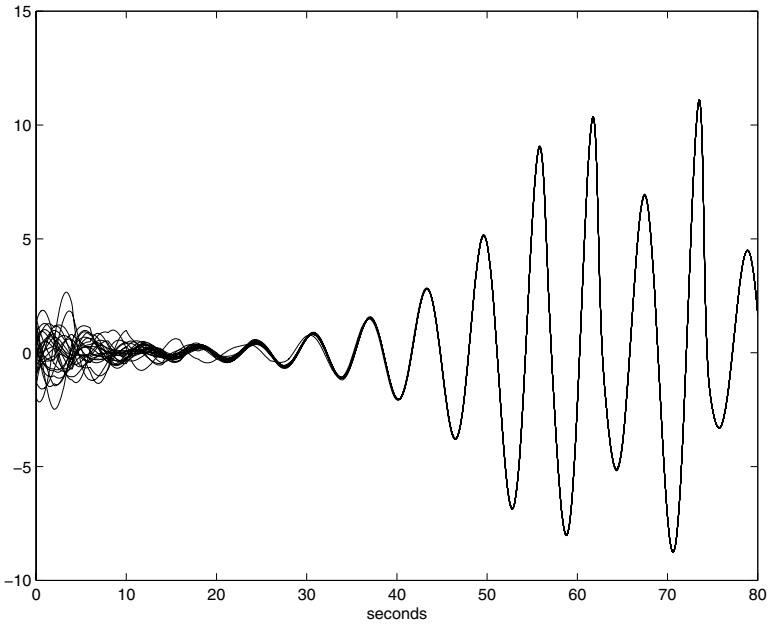


Figure 5. *The x-coordinate for the set of coupled Rössler attractors using the switched network where $\varepsilon = 1$.*

coefficients ρ, μ, and η in (A.8) are computed from γ and λ when evaluating the right-hand side of (A.11). Choosing $T = 1$, the right-hand side of (A.11) is evaluated for this example, and we determine that the set of coupled oscillators will synchronize if $\varepsilon < 3.3 \times 10^{-7}$. This shows that our bound is exceedingly conservative. For example, empirically the oscillators will synchronize with $\varepsilon = 1$, as shown in Figure 5.

Appendix.

Proof of Lemma 3.3. Since (3.9) is uniformly exponentially stable, there exist a symmetric matrix function $Q(t)$ and positive scalars η, ρ, and μ such that the Lyapunov function

$$v(x(t), t) = x^T(t)Q(t)x(t)$$

satisfies

$$\eta\|x(t)\|^2 \leq v(x(t), t) \leq \rho\|x(t)\|^2, \tag{A.1}$$

$$\frac{d}{dt}v(x(t), t) \leq -\mu\|x(t)\|^2 \tag{A.2}$$

for all t. To establish uniform exponential stability of (3.10), we will show that $v(z(t), t)$ is also a Lyapunov function for (3.10) if ε is sufficiently small. This claim is achieved by showing that for sufficiently small values of ε,

$$\Delta v(z, t + \varepsilon T, t) \equiv v(z(t + \varepsilon T), t + \varepsilon T) - v(z(t), t) \tag{A.3}$$

is negative definite for all t. Expanding (A.3) yields

$$\begin{aligned}\Delta v(z,t+\varepsilon T,t) &= z^T(t+\varepsilon T)Q(t+\varepsilon T)z(t+\varepsilon T) - z^T(t)Q(t)z(t)\\ &= z^T(t)\Big(\Phi_E^T(t+\varepsilon T,t)Q(\varepsilon T+t)\Phi_E(t+\varepsilon T,t) - Q(t)\Big)z(t),\end{aligned}$$

where $\Phi_E(t,t_0)$ is the transition matrix corresponding to $A(t)+E(t/\varepsilon)$, i.e.,

$$z(t) = \Phi_E(t,t_0)z_0$$

is the solution to (3.10), as discussed, for example, in [52]. Similarly denoting the transition matrix for $A(t)+\bar{E}$ as $\Phi_{\bar{E}}(t,t_0)$, we use the Peano–Baker series representation of the transition matrix to define

$$\begin{aligned}H(t+\varepsilon T,t) &= \Phi_E(t+\varepsilon T,t) - \Phi_{\bar{E}}(t+\varepsilon T,t)\\ &= I + \int_t^{t+\varepsilon T} A(\sigma_1)+E(\sigma/\varepsilon)d\sigma\\ &\quad + \sum_{i=2}^{\infty}\int_t^{t+\varepsilon T} A(\sigma_1)+E(\sigma_1/\varepsilon)\int_t^{\sigma_1}\cdots\int_t^{\sigma_{i-1}} A(\sigma_i)+E(\sigma_i/\varepsilon)d\sigma_i\cdots d\sigma_1\\ &\quad - I - \int_t^{t+\varepsilon T} A(\sigma_1)+\bar{E}d\sigma - \sum_{i=2}^{\infty}\int_t^{t+\varepsilon T} A(\sigma_1)+\bar{E}\int_t^{\sigma_1}\cdots\int_t^{\sigma_{i-1}} A(\sigma_i)\\ &\quad + \bar{E}d\sigma_i\cdots d\sigma_1.\end{aligned}$$

By hypothesis,

$$\int_t^{t+\varepsilon T} E(\sigma/\varepsilon)d\sigma = \varepsilon T\bar{E},$$

which implies that

$$\begin{aligned}H(t+\varepsilon T,t) &= \sum_{i=2}^{\infty}\int_t^{t+\varepsilon T} A(\sigma_1)+E(\sigma_1/\varepsilon)\int_t^{\sigma_1}\cdots\int_t^{\sigma_{i-1}} A(\sigma_i)+E(\sigma_i/\varepsilon)d\sigma_i\cdots d\sigma_1\\ &\quad - \sum_{i=2}^{\infty}\int_t^{t+\varepsilon T} A(\sigma_1)+\bar{E}\int_t^{\sigma_1}\cdots\int_t^{\sigma_{i-1}} A(\sigma_i)+\bar{E}d\sigma_i\cdots d\sigma_1.\end{aligned}$$

Defining

$$\alpha \equiv \sup_{t\geq 0}\big(\max\big(\|A(t)+\bar{E}\|,\|A(t)+E(t/\varepsilon)\|\big)\big) \tag{A.4}$$

a bound for $H(t+\varepsilon T,t)$ is computed:

$$\|H(t+\varepsilon T,t)\| \leq 2\big(e^{\varepsilon T\alpha} - 1 - \varepsilon T\alpha\big). \tag{A.5}$$

Noting that $\Phi_E = \Phi_{\bar{E}} + H$, Δv is expressed as

(A.6)
$$\begin{aligned}\Delta v(z, t+\varepsilon T, t) &= z^T(t)\Big(\Phi_{\bar{E}}^T(t+\varepsilon T, t)Q(t+\varepsilon T)\Phi_{\bar{E}}(t+\varepsilon T, t) - Q(t)\Big)z(t)\\ &+ z^T(t)\Big(\Phi_{\bar{E}}^T(t+\varepsilon T, t)Q(t+\varepsilon T)H(t+\varepsilon T, t) + H^T(t+\varepsilon T, t)Q(t+\varepsilon T)\Phi_{\bar{E}}(t+\varepsilon T, t)\\ &+ H^T(t+\varepsilon T, t)Q(t+\varepsilon T, t)H(t+\varepsilon T, t)\Big)z(t).\end{aligned}$$

The task now is to compute an upper bound for $\Delta v(z, t+\varepsilon T, t)$ and show that this bound is negative if ε is sufficiently small. Several well-known relationships that are consequences of (A.1), (A.2), and uniform exponential stability of (3.9) are utilized (see, for example, [52, p. 101, 117] or [13, p. 202]). Namely,

$$\|Q(t)\| \le \rho, \tag{A.7}$$

$$\|\Phi_{\bar{E}}(t, t_o)\| \le \sqrt{\rho/\eta}\, e^{-\frac{\mu}{2\rho}(t-t_o)}, \tag{A.8}$$

$$v(x(t), t) \le e^{-\frac{\mu}{\rho}(t-t_o)} v(x(t_o), t_o) \tag{A.9}$$

for $t \ge t_o$.

To compute an upper bound for the first term on the right-hand side of (A.6) we note that if $x(t) = z(t)$ is chosen as the initial condition of (3.9) at time t, then

$$z^T(t)\Big(\Phi_{\bar{E}}^T(t+\varepsilon T, t)Q(t+\varepsilon T)\Phi_{\bar{E}}(t+\varepsilon T, t) - Q(t)\Big)z(t) = v(x(t+\varepsilon T), t+\varepsilon T) - v(x(t), t).$$

From (A.9) and (A.1),

$$\begin{aligned}v(x(t+\varepsilon T), t+\varepsilon T) - v(x(t), t) &\le (e^{-\mu\varepsilon T/\rho} - 1)v(x(t), t)\\ &\le \rho(e^{-\mu\varepsilon T/\rho} - 1)\|x(t)\|^2.\end{aligned}$$

Thus,

$$z^T(t)\Big(\Phi_{\bar{E}}^T(t+\varepsilon T, t)Q(t+\varepsilon T)\Phi_{\bar{E}}(t+\varepsilon T, t) - Q(t)\Big)z(t) \le \rho(e^{-\mu\varepsilon T/\rho} - 1)\|z(t)\|^2. \tag{A.10}$$

Combining (A.5), (A.7), (A.8), and (A.10) yields the desired upper bound

$$\begin{aligned}\Delta v(z, t+\varepsilon T, t) \le \Big(&\rho(e^{-\mu\varepsilon T/\rho} - 1) + 4\rho(\sqrt{\rho/\eta}\, e^{-\frac{\mu\varepsilon T}{2\rho}})(e^{\varepsilon T\alpha} - 1 - \varepsilon T\alpha)\\ &+ 4\rho(e^{\varepsilon T\alpha} - 1 - \varepsilon T\alpha)^2\Big)\|z(t)\|^2.\end{aligned} \tag{A.11}$$

Defining the continuously differentiable function $g(\varepsilon, x)$ to be the right-hand side of (A.11), it can be shown that $g(0, z) = 0$ and $\frac{\partial}{\partial\varepsilon} g(0, z) = -\mu T\|z\|^2 < 0$. Thus since $g(\varepsilon, z) \to \infty$ as $\varepsilon \to \infty$, there exists ε^* such that $g(\varepsilon^*, z) = 0$ and $g(\varepsilon, z) < 0$ for all $\varepsilon \in (0, \varepsilon^*)$ and $z \ne 0$. Thus $\Delta v(z, t+\varepsilon T, t) < 0$ for all $\varepsilon \in (0, \varepsilon^*)$ and $z \ne 0$.

To show that negative-definiteness of $\Delta v(z, t+\varepsilon T, t)$ is sufficient to establish stability of (3.10), choose ε and $\gamma > 0$ that satisfy

$$\Delta v(z, t_o+\varepsilon T, t_o) = v(z(t_o+\varepsilon T), t_o+\varepsilon T) - v(z(t_o), t_o) \le -\gamma\|z(t_o)\|^2$$

for all t_o. From (A.1), $v(z(t_o), t_o) \leq \rho \|z(t_o)\|^2$, which implies that

$$v(z(t_o + \varepsilon T), t_o + \varepsilon T) - v(z(t_o), t_o) \leq -(\gamma/\rho) v(z(t_o), t_o).$$

Thus

$$v(z(t_o + \varepsilon T), t_o + \varepsilon T) \leq (1 - \gamma/\rho) v(z(t_o), t_o).$$

Repeating this argument yields

$$v(z(t_o + k\varepsilon T), t_o + k\varepsilon T) \leq (1 - \gamma/\rho)^k v(z(t_o), t_o)$$

for any positive integer k. Thus $v(z(t_o + k\varepsilon T), t_o + k\varepsilon T) \to 0$ as $k \to \infty$ which implies that $z(t_o + k\varepsilon T) \to 0$ as $k \to \infty$. Since the limiting behavior is valid for any t_o, uniform exponential stability of (3.10) is established. ■

REFERENCES

[1] D. Aeyels and J. Peuteman, *A new aymptotic stability criteria for nonlinear time-variant differential equations*, IEEE Trans. Automat. Control, 43 (1998), pp. 968–971.

[2] D. Aeyels and J. Peuteman, *On exponential stability of nonlinear time-varying differential equations*, Automatica J. IFAC, 35 (1999), pp. 1091–1100.

[3] V. S. Affraimovich, N. N. Verichev, and M. I. Rabinovich, *Stochastic synchronization of oscillation in dissipative systems*, Izv. Vyssh. Uchebn. Zaved. Radiofiz., 29 (1986), pp. 1050–1060.

[4] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Modern Phys., 74 (2002), pp. 47–97.

[5] A.-L. Barabási and E. Bonabeau, *Scale-free networks*, Scientific American, 288 (2003), pp. 60–69.

[6] M. Barahona and L. M. Pecora, *Synchronization in small-world systems*, Phys. Rev. Lett., 89 (2002), 054101; nlin.CD/0112023.

[7] R. Bellman, J. Bentsman, and S. M. Meerkov, *Stability of fast periodic systems*, IEEE Trans. Automat. Control, 30 (1985), pp. 289–291.

[8] V. N. Belykh, I. V. Belykh, and M. Hasler, *Connection graph stability method for synchronized coupled chaotic systems*, Phys. D, 195 (2004), pp. 159–187.

[9] I. V. Belykh, V. N. Belykh, and M. Hasler, *Blinking model and synchronization in small-world networks with a time-varying coupling*, Phys. D, 195 (2004), pp. 188–206.

[10] I. I. Blekman, *Synchronization in Science and Technology*, Nauka, Moscow, 1981 (in Russian); ASME Press, New York, 1988 (in English).

[11] S. Boccaletti, J. Kurths, G. Osipov, D. L. Valladares, and C. S. Zhou, *The synchronization of chaotic systems*, Phys. Rep., 366 (2002), pp. 1–101.

[12] E. M. Bollt, *Review of chaos communication by feedback control of symbolic dynamics*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 13 (2003), pp. 269–285.

[13] R. W. Brockett, *Finite Dimensional Linear Systems*, John Wiley and Sons, New York, 1970.

[14] J. Buck and E. Buck, *Synchronous fireflies*, Scientific American, 234 (1976), p. 74.

[15] G. Chen and Y. Xinghuo, eds., *Chaos Control. Theory and Applications*, Lecture Notes in Control and Inform. Sci. 292, Springer-Verlag, New York, 2003.

[16] J. J. Collins and I. Stewart, *Coupled nonlinear oscillators and the symmetries of animal gaits*, Science, 3 (1993), pp. 349–392.

[17] N. J. Corron, S. D. Pethel, and K. Myneni, *Synchronizing the information content of a chaotic map and flow via symbolic dynamics*, Phys. Rev. E, 66 (2002), 036204.

[18] J. P. Desai, J. P. Ostrowski, and V. Kumar, *Modeling and control of formations of nonholonomic mobile robots*, IEEE Trans. Robotics and Automation, 17 (2001), pp. 905–908.

[19] S. Dorogovtsev and J. F. F. Mendes, *Evolution of networks*, Advances in Physics, 51 (2002), pp. 1079–1187.
[20] G. S. Duane, P. J. Webster, and J. B. Weiss, *Go-occurrence of northern and southern hemisphere blocks as partially synchronized chaos*, J. Atmospheric Sci., 56 (1999), pp. 4183–4205.
[21] J. A. Fax and R. M. Murry, *Information flow and cooperative control of vehicle formations*, IEEE Trans. Automat. Control, 49 (2004), pp. 1465–1476.
[22] A. F. Filippov, *Differential Equations with Discontinuous Right-Hand Sides*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.
[23] K. Fink, G. Johnson, T. Carroll, D. Mar, and L. Pecora, *Three coupled oscillators as a universal probe of synchronization stability in coupled oscillator arrays*, Phys. Rev. E, 61 (2000), pp. 5080–5090.
[24] H. Fujisaka and T. Yamada, *Stability theory of synchronized motion in coupled-oscillator systems*, Prog. Theoret. Phys., 69 (1983), pp. 32–47.
[25] P. Garcia, A. Parravano, M. G. Cosenza, J. Jimenez, and A. Marcano, *Coupled map networks as communication schemes*, Phys. Rev. E, 65 (2002), 015205; nlin.CD/0201042.
[26] L. Glass and M. C. Mackey, *From Clocks to Chaos: The Rythms of Life*, Princeton University Press, Princeton, NJ, 1988.
[27] M. R. Guevara, A. Shrier, and L. Glass, *Phase-locked rhythms in periodically stimulated heart cell aggregates*, American Journal of Physiology, 254 (Heart Circ. Physiol. 23) (1988), pp. H1–H10.
[28] S. Hayes, C. Grebogi, and E. Ott, *Communicating with chaos*, Phys. Rev. Lett., 70 (1993), pp. 3031–3034.
[29] J. Honerkamp, *The heart as a system of coupled nonlinear oscillators*, J. Math. Biol., 19 (1983), pp. 69–88.
[30] H. Hong, M. Y. Choi, and B. J. Kim, *Synchronization on small-world networks*, Phys. Rev. E, 65 (2002), 026139; cond-mat/0110359.
[31] C. Hugenii, *Horoloquium Oscilatorium*, Apud F. Muguet, Parisiis, 1673.
[32] J. Ito and K. Kaneko, *Spontaneous structure formation in a network of chaotic units with variable connection strengths*, Phys. Rev. Lett., 88 (2002), 028701.
[33] S. Jalan and R. E. Amritkar, *Self-organized and driven phase synchronization in coupled map scale free networks*, Phys. Rev. Lett., submitted; nlin.AO/0201051.
[34] J. Jost and M. P. Joy, *Spectral properties and synchronization in coupled map lattices*, Phys. Rev. E, 65 (2002), 016201; nlin.CD/0110037.
[35] N. Kataoka and K. Kaneko, *Dynamical networks in function dynamics*, Phys. D, 181 (2003), pp. 235–251.
[36] R. L. Kosut, B. D. O. Anderson, and I. M. Y. Mareels, *Stability theory for adaptive systems: Method of averaging and persistency of excitation*, IEEE Trans. Automat. Control, 32 (1987), pp. 26–34.
[37] Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence*, Springer-Verlag, Berlin, 1984.
[38] D.-S. Lee, *Synchronization transition in scale-free networks: Clusters of synchrony*, Phys. Rev. E, 72 (2005), 026208; cond-mat/0410635.
[39] D. Liberzon, *Switching in Systems and Control*, Birkhäuser Boston, Boston, 2003.
[40] D. Liberzon and A. S. Morse, *Basic problems in stability and design of switched systems*, IEEE Control Systems, 19 (1999), pp. 59–70.
[41] R. E. Mirollo and S. H. Strogatz, *Synchronization of pulsed-coupled biological oscillators*, SIAM J. Appl. Math., 50 (1990), pp. 1645–1662.
[42] E. Mosekilde, Y. Maistrenko, and D. Postnov, *Chaotic Synchronization: Applications to Living Systems*, World Sci. Ser. Nonlinear Sci. Ser. A Monogr. Treatises 42, World Scientific, River Edge, NJ, 2002.
[43] T. I. Netoff and S. J. Schiff, *Decreased neuronal synchronization during experimental seizures*, J. Neuroscience, 22 (2002), pp. 7297–7307.
[44] M. E. J. Newman, *The structure and function of complex networks*, SIAM Rev., 45 (2003), pp. 167–256.
[45] T. Nishikawa, A. E. Motter, Y.-C. Lai, and F. C. Hoppensteadt, *Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize?*, Phys. Rev. Lett., 91 (2003), 014101; cond-mat/0306625.
[46] L. M. Pecora and T. L. Carroll, *Synchronization in chaotic systems*, Phys. Rev. Lett., 64 (1990), pp. 821–825.

[47] L. M. Pecora and T. L. Carroll, *Master stability function for synchronized coupled systems*, Phys. Rev. Lett., 80 (1998), pp. 2109–2112.
[48] L. Pecora, T. Carroll, G. Johnson, D. Mar, and K. S. Fink, *Synchronization stability in coupled oscillator arrays: Solution for arbitrary configurations*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 10 (2000), pp. 273–290.
[49] S. D. Pethel, N. J. Corron, Q. R. Underwood, and K. Myneni, *Information flow in chaos synchronization: Fundamental tradeoffs in precision, delay, and anticipation*, Phys. Rev. Lett., 90 (2003), 254101.
[50] A. Pikovsky, M. Rosemblum, and J. Kurths, *Synchronization, A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, UK, 2001.
[51] D. G. Roberson and D. J. Stilwell, *Control of an autonomous underwater vehicle platoon with a switched communication network*, in Proceedings of the American Control Conference, Portland, OR, 2005, pp. 4333–4338.
[52] W. J. Rugh, *Linear Systems Theory*, 2nd ed., Prentice-Hall, Upper Saddle River, NJ, 1996.
[53] G. Sansone and R. Conti, *Non-linear Differential Equations*, Pergamon Press, Oxford, England, 1964.
[54] J. D. Skufca and E. Bollt, *Communication and synchronization in disconnected networks with dynamic topology: Moving neighborhood networks*, Math. Biosci. Eng., 1 (2004), pp. 347–359; nlin.CD/0307010.
[55] T. Stojanovski, L. Kocarev, and R. Harris, *Applications of symbolic dynamics in chaos synchronization*, IEEE Trans. Circuits Systems I Fund. Theory Appl., 44 (1997), pp. 1014–1018.
[56] T. Stojanovski, L. Kocarev, U. Parlitz, and R. Harris, *Sporadic driving of dynamical systems*, Phys. Rev. E, 55 (1997), pp. 4035–4048.
[57] S. Strogatz, *Sync: The Emerging Science of Spontaneous Order*, Hyperion, New York, 2003.
[58] J. Tokarzewski, *Stability of periodically switched linear systems and the switching frequency*, Internat. J. Systems Sci., 18 (1987), pp. 697–726.
[59] V. Torre, *A theory of synchronization of two heart pacemaker cells*, J. Theoret. Biol., 61 (1976), pp. 55–71.
[60] G. D. Van Wiggeren and R. Roy, *Communication with chaotic lasers*, Science, 279 (1998), pp. 1198–1200.
[61] R. Roy and K. S. Thornburg, Jr., *Experimental synchronization of chaotic lasers*, Phys. Rev. Lett., 72 (1994), pp. 2009–2012.
[62] J. Ohtsubop, *Feedback induced instability and chaos in semiconductor lasers and their applications*, Optical Review, 6 (1999), pp. 1–15.
[63] D. J. Watts, *Six Degrees: The Science of a Connected Age*, Norton, New York, 2003.
[64] M. Wicks, P. Peleties, and R. DeCarlo, *Switched controller synthesis for the quadratic stabilization of a pair of unstable linear systems*, European Journal of Control, 4 (1998), pp. 140–147.
[65] D. H. Zanette and A. S. Mikhailov, *Dynamical systems with time-dependent coupling: Clustering and critical behavior*, Phys. D, 194 (2004), pp. 203–218.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 1, pp. 157–177

Computation of Spiral Spectra*

Paul Wheeler† and Dwight Barkley†

Abstract. A computational linear stability analysis of spiral waves in a reaction-diffusion equation is performed on large disks. As the disk radius R increases, eigenvalue spectra converge to the absolute spectrum predicted by Sandstede and Scheel. The convergence rate is consistent with $1/R$, except possibly near the edge of the spectrum. Eigenfunctions computed on large disks are compared with predicted exponential forms. Away from the edge of the absolute spectrum the agreement is excellent, while near the edge computed eigenfunctions deviate from predictions, probably due to finite-size effects. In addition to eigenvalues associated with the absolute spectrum, computations reveal point eigenvalues. The point eigenvalues and associated eigenfunctions responsible for both core and far-field breakup of spiral waves are shown.

Key words. spiral wave, excitable media, oscillatory media, eigenvalues, breakup

AMS subject classifications. 37G99, 74J30, 92C99

DOI. 10.1137/050624273

1. Introduction.

Rotating spiral waves are found in many chemical and biological systems and have been the subject of intense study for many years [10, 14, 16, 28]. The equations governing these systems are typically of reaction-diffusion type. Although each system is modeled in detail by specific equations—which are often very complex—generic features of the spiral waves can be understood from reaction-diffusion equations with simple nonlinearities. Figure 1 shows a spiral wave in a generic model reaction-diffusion system described in detail in section 2. For the model parameters in Figure 1 the spiral wave rotates with constant frequency and shape; i.e., it is a rotating wave.

The focus of our work is a computational study of the linear stability spectra of rotating spiral waves such as those shown in Figure 1. To explain the motivation behind this study it is necessary to first recall the recent analysis by Sandstede and Scheel [21, 22, 23, 24] on the spectra of rotating spiral waves. Their work examines spectra on large bounded disks and on unbounded domains. The results can be summarized as follows (see Figure 2). On large bounded disks, the linear stability spectrum consists of point eigenvalues and what is called the absolute spectrum. The absolute spectrum is not actually part of the stability spectrum. However, all but possibly a finite number of point eigenvalues converge to the absolute spectrum as the domain size tends to infinity. That is, except for finitely many eigenvalues that are created through the underlying pattern as a whole, or possibly by the boundary conditions, all eigenvalues on large bounded domains are expected to be close to the

*Received by the editors February 11, 2005; accepted for publication (in revised form) by T. Kaper November 10, 2005; published electronically March 3, 2006.
http://www.siam.org/journals/siads/5-1/62427.html
†Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK (wheeler@maths.warwick.ac.uk, barkley@maths.warwick.ac.uk).

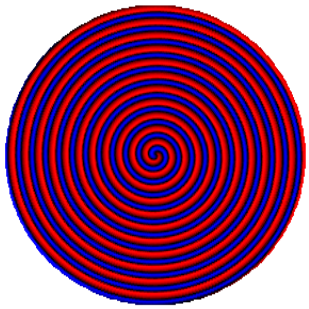


Figure 1. *Rotating spiral wave solution of reaction-diffusion equations described in section 2. Colors indicate the level of the u field, with blue used for u near zero and red used for u near 1. The wave rotates counterclockwise. The domain radius, $R = 80$, is approximately 10 times the spiral wavelength. Homogeneous Neumann boundary conditions, corresponding to zero chemical flux, are imposed at the domain boundary. Model parameters are $a = 0.75$, $b = 0.0006$, and $\epsilon = 0.0741$.*

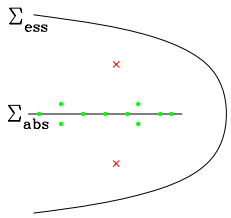


Figure 2. *Illustration of spectra in the complex plane for spirals on bounded and unbounded domains. Σ_{abs} and Σ_{ess} represent the absolute and essential spectra, respectively. Points represent eigenvalues on a large bounded domain, which approach Σ_{abs} as the domain size tends to infinity. Crosses represent the point spectrum, which does not approach Σ_{abs} as the domain size tends to infinity.*

absolute spectrum. The point eigenvalues have well-defined limits as the domain size tends to infinity.

In practice the absolute spectrum must be computed numerically for any given reaction-diffusion equation; see, e.g., [22]. Such computations require discretization in only one space dimension and thus are relatively simple compared with computing eigenvalues of the full stability problem on a large domain, such as in Figure 1.

For spiral waves on the unbounded plane, the linear stability spectrum consists of point eigenvalues and the essential spectrum. The essential spectrum is a continuous spectrum and is determined only by the far-field wave trains of the spiral. It too is relatively easy to compute numerically in one space dimension. The point eigenvalues again depend on the underlying spiral pattern as a whole.

To see how these linear stability spectra may be relevant in practice, we show in Figure 3 simulations of two instabilities of rotating waves (62427_01.gif and 62427_02.gif) on relatively large domains, and the corresponding absolute and essential spectra obtained by Sandstede and Scheel [22]. In each case a rotating wave becomes unstable in a rather dramatic fashion, and the spiral breaks up. Multiple spiral waves appear in each of these simulations shortly after the time shown. In Figure 3(a) the breakup initiates in the central region of the spiral and is referred to as core breakup [3, 15, 22], whereas in Figure 3(b) the breakup first takes

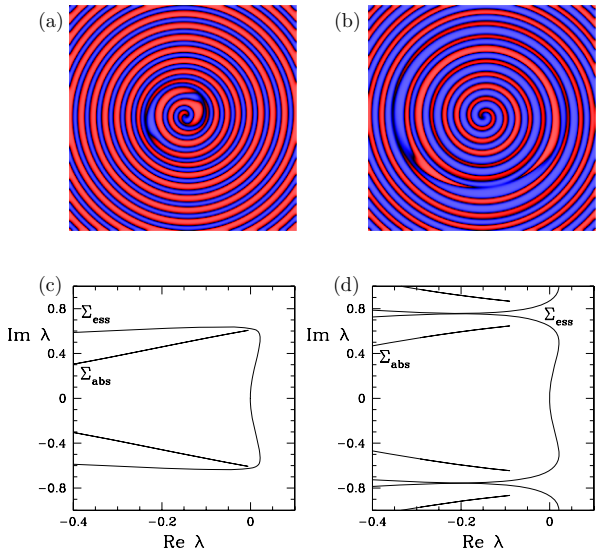


Figure 3. *Two examples of spiral breakup—core breakup* (62427_01.gif [2.6MB]) *on the left and far-field breakup* (62427_02.gif [6.7MB]) *on the right. The top plots show the u chemical field at about the time of breakup in numerical simulations in square geometries with homogeneous Neumann boundary conditions. Domains are of length* 160 *on a side. The bottom shows the absolute and essential spectra obtained by Sandstede and Scheel for the parameter values used in the simulations. Note that these spectra repeat periodically in the imaginary direction, but this can be seen only in* (d)*. Model parameters are (left)* $a = 0.75$*,* $b = 0.0006$*, and* $\epsilon = 0.0741$*; (right)* $a = 0.84$*,* $b = -0.045$*, and* $\epsilon = 0.075$*.*

place in the outer regions of the spiral and is called far-field breakup [4, 18, 22, 26, 29].

The case of far-field breakup, Figure 3(right), has been the subject of several past studies [1, 2, 4, 9, 18, 22, 25, 26, 29]. The breakup can be mostly understood from analysis and simulations of one-dimensional systems. While many of these studies are based on the complex Ginzburg–Landau equation, results appear to be similar for the case of reaction-diffusion equations [2, 4, 26]. The typical scenario is that as a parameter is varied, the spiral first becomes convectively unstable. In a bounded domain the onset of convective instability does not generally lead to breakup, because unstable modes typically propagate away from the core and are not reflected at the boundary. As the parameter is varied further, the spiral becomes absolutely unstable. Only at the absolute-instability threshold will instability surely occur in a bounded domain. Absolute instability corresponds to a growing "global mode" [25, 26], which here means an eigenfunction on the bounded domain whose eigenvalue has positive real part.

In the analysis of Sandstede and Scheel, convective instability is signaled by the crossing of the essential spectrum into the right-half plane [21, 22]. Figure 3(d) shows that the spiral is convectively unstable. However, the spiral was already convectively unstable prior to the breakup seen in Figure 3, and this is not the cause of breakup. The breakup is caused by an eigenvalue with positive real part and the corresponding global mode on the finite (bounded) domain. In principle such an eigenvalue could be associated with the absolute spectrum or it

could be a point eigenvalue. In Figure 3(d) the absolute spectrum is away from the imaginary axis and thus is not expected to play a direct role in the far-field breakup. Thus we expect there to be at least one positive point eigenvalue not contained in Figure 3(d).

It is worth being clear about potentially confusing terminology. Absolute instability is not associated only with the absolute spectrum. The union of the absolute spectrum and the point spectrum determines absolute stability. If part of the absolute spectrum lies in the right-half plane, then the spiral will necessarily be absolutely unstable. However, the converse is not true, since absolute instability can arise due to point eigenvalues, even if the absolute spectrum lies entirely in the left-half plane.

We should warn the reader that simulations at the stated parameters in Figure 3(b) are very sensitive to numerical resolution. As we shall see, this is because the particular parameters considered by Sandstede and Scheel are extremely close to a transition between far-field and core breakup.

The case of core breakup, Figure 3(left), has not been extensively analyzed, in large part because one cannot expect to capture much of the spiral core structure in one-dimensional studies. (See, however, [2].) Sandstede and Scheel [22] have raised the possibility that core breakup may be due to the absolute spectrum. In Figure 3(c) it can be seen that the absolute spectrum is very near the imaginary axis, although it is entirely in the left-half plane. (The argument of Sandstede and Scheel is not simply that the absolute spectrum is close to the real axis, but the details are not important here.) The other possibility for core breakup is that the instability is again due to point eigenvalues. The essential spectrum in 3(c) extends into the right-half plane, and so the spiral is convectively unstable.

Computing the eigenvalue spectra of spiral waves on large domains has thus become important. First and foremost, it is desirable to test the absolute spectra of Sandstede and Scheel in at least a few cases. The primary issue is whether or not eigenvalues tend to the absolute spectrum for typical domains sizes used in studies of spiral waves, e.g., domains such as those in Figure 3. The theory is still developing, and we would like to know whether absolute spectra in fact have any implications for domains of reasonable size. The other important issue which computations can address is the abundance and importance of point eigenvalues not predicted by the absolute spectrum. For example, it is desirable to know how many point eigenvalues are present within the region shown in Figure 3(c)–(d), how many of these eigenvalues have positive real part, and whether or not these are associated with breakup. For these reasons we have undertaken the large-scale eigenvalue computation reported here.

Throughout this paper we shall use *point eigenvalue* to mean those eigenvalues that remain isolated as the domain radius becomes large, as contrasted with the eigenvalues associated with the absolute spectrum that approach one another as the radius becomes large. To simplify discussion we shall use *positive eigenvalue* to mean an eigenvalue, or a complex-conjugate pair of eigenvalues, with positive real part. Similarly we shall use *positive eigenfunction* to mean an eigenfunction whose corresponding eigenvalue has positive real part.

2. Model and methods.

2.1. Model. We will consider a standard two-component reaction-diffusion model [5] given by the equations

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(u,v), \tag{2.1a}$$

$$\frac{\partial v}{\partial t} = \delta\nabla^2 v + g(u,v), \tag{2.1b}$$

where

$$f(u,v) = \frac{1}{\epsilon}u(1-u)\left(u - \frac{v+b}{a}\right). \tag{2.2}$$

There is freedom in the choice of $g(u,v)$, and our methods will not depend on this choice. However, the results we report will be for the g proposed by Bär and Eiswirth [3] and used by Sandstede and Scheel [22], namely,

$$g(u,v) = \begin{cases} -v, & 0 \le u < 1/3, \\ 1 - 6.75u(u-1)^2 - v, & 1/3 \le u \le 1, \\ 1 - v, & 1 < u. \end{cases} \tag{2.3}$$

The equations are posed on a disk of radius R and with homogeneous Neumann boundary conditions at $r = R$:

$$\frac{\partial u}{\partial r}(R,\theta) = \frac{\partial v}{\partial r}(R,\theta) = 0,$$

where r, θ are standard polar coordinates. For chemically reacting systems these are the most natural boundary conditions, as they correspond to zero chemical flux through the boundary of the domain. Other boundary conditions could give different spiral solutions and linear stability spectra on finite domains, but we do not consider any other boundary conditions here.

The parameters of the model are kinetics parameters a, b, and ϵ and the diffusion coefficient δ. If $b > 0$, the equations model an excitable medium. In this case the homogeneous state with $u = v = 0$ everywhere is linearly stable, and finite amplitude perturbations are required to initiate waves. The perturbation threshold is set by b/a. For $b < 0$ the equations model an oscillatory medium. In both cases ϵ controls the time-scale ratio between the u- and v-equations. We consider $\epsilon \ll 1$ corresponding to a fast time-scale for u relative to v. We shall report results only for the case $\delta = 0$; $\delta = 1$ is the other case commonly considered. As stated at the outset, these equations model generic features of spiral waves in a variety of excitable and oscillatory media.

2.2. Computational tasks. Consider rotating-wave solutions of (2.1) rotating at frequency ω. We use (u^*, v^*) to denote such solutions and refer to them as steady spirals, since these are steady states when viewed in the frame of reference which is rotating with the spiral. Transforming to a system of coordinates corotating at frequency ω, steady spirals obey the equations

$$0 = \nabla^2 u^* + \omega\frac{\partial u^*}{\partial \theta} + f(u^*, v^*), \tag{2.4a}$$

$$0 = \omega\frac{\partial v^*}{\partial \theta} + g(u^*, v^*), \tag{2.4b}$$

subject to homogeneous Neumann boundary conditions. These steady-state equations can be written in the form

$$\mathcal{F}\begin{pmatrix} u^* \\ v^* \end{pmatrix} = 0, \tag{2.5}$$

where $\mathcal{F}$ is the nonlinear operator given by the right-hand side of (2.4).

Next, given a steady spiral, we seek the leading part of its linear stability spectrum. Consider the linearized evolution equations, in the rotating frame, for infinitesimal perturbations (u', v') of the steady solution (u^*, v^*):

$$\frac{\partial u'}{\partial t} = \nabla^2 u' + \omega \frac{\partial u'}{\partial \theta} + f_u(u^*, v^*)u' + f_v(u^*, v^*)v', \tag{2.6a}$$

$$\frac{\partial v'}{\partial t} = \omega \frac{\partial v'}{\partial \theta} + g_u(u^*, v^*)u' + g_v(u^*, v^*)v', \tag{2.6b}$$

where $f_u, \ldots, g_v$ denote the derivatives of the kinetic functions. In this frame of reference the eigenvalue problem is

$$\mathcal{L}\begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \lambda \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix}, \tag{2.7}$$

where $\left(\begin{smallmatrix} \tilde{u} \\ \tilde{v} \end{smallmatrix}\right)$ are eigenfunctions, λ are the corresponding eigenvalues, and $\mathcal{L}$ is

$$\mathcal{L} = \begin{pmatrix} \nabla^2 + \omega\partial_\theta + f_u(u^*, v^*) & f_v(u^*, v^*) \\ g_u(u^*, v^*) & \omega\partial_\theta + g_v(u^*, v^*) \end{pmatrix}. \tag{2.8}$$

Thus our primary numerical tasks are the solution of steady-state problem (2.5) and the determination of the leading eigenvalues of problem (2.7).

In addition, it is necessary to perform a few simulations of the time-dependent equations (2.1), e.g., the simulations shown in Figure 3. In the case of Figure 3 the numerical methods are described fully elsewhere [5, 12] and will not be discussed here.

2.3. Computational methods. Equations (2.5) and (2.7) are common in large-scale numerical bifurcation problems, and the computational methods we employ are more or less standard [11]. For completeness we provide a basic description of our methods and stress a few points concerning implementation which are essential to the efficiency of the computations.

The fields are discretized on a regular polar grid $(r_j, \theta_k) = (j\triangle r, k\triangle\theta)$, where $0 < j \leq N_r$ and $0 \leq k < N_\theta$, plus the center point $(0, 0)$. There are thus $N_r N_\theta + 1$ grid points. The r-derivatives in the differential operators are evaluated using second-order finite differences, taking into account the boundary condition at $r = R$. The θ-derivatives are evaluated spectrally using Fourier transforms. In this way (2.5) and (2.7) become

$$\mathbf{F}(\mathbf{U}^*) = 0, \tag{2.9}$$

$$\mathbf{L}\tilde{\mathbf{U}} = \lambda\tilde{\mathbf{U}}, \tag{2.10}$$

where the $\mathbf{U}$'s are vectors of length $N = 2(N_r N_\theta + 1)$, $\mathbf{F}$ is a nonlinear function, and $\mathbf{L}$ is an $N \times N$ matrix.

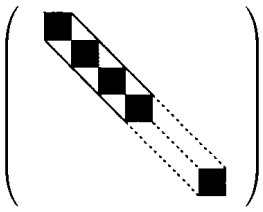


Figure 4. *Banded structure of matrix* **DF** *or* **L**. *Solid black represents full blocks of size* $2N_\theta \times 2N_\theta$ *mainly due to using spectral representation of* θ*-derivatives. The lines are due to the second-order finite-difference representation of* r*-derivatives.*

Newton's method is used to solve steady-state problem (2.9). One iteration of Newton's method is

$$\mathbf{DF}(\mathbf{U}_n)\triangle\mathbf{U}_n = -\mathbf{F}(\mathbf{U}_n), \tag{2.11}$$

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \triangle\mathbf{U}_n, \tag{2.12}$$

where $\mathbf{DF}(\mathbf{U}_n)$ is the linearization of $\mathbf{F}$ about the current iterate $\mathbf{U}_n$. This is the same matrix as $\mathbf{L}$, except that it is evaluated at $\mathbf{U}_n$ rather than at the steady state $\mathbf{U}^*$.

The work of each Newton's iteration is dominated by the work necessary to solve the $N \times N$ linear system of equations (2.11) for the nth correction $\triangle\mathbf{U}_n$. This can be done by a direct method if care is taken to order variables so as to keep the matrix bandwidth of **DF** as small as possible. Let u_{jk} and v_{jk} be values at grid point (r_j, θ_k). Then these are ordered in U_i such that the chemical species changes fastest with index i, followed by the angular index k, followed by the radial index j. This ordering is not that suggested by (2.8). With the ordering we use, the bandwidth is approximately $4N_\theta$. See Figure 4. Even for moderately large discretizations a direct method can be used to solve (2.11). For example, on a grid $N_r \times N_\theta = 600 \times 256$, N is approximately 3×10^5, while the bandwidth is only about 10^3.

The only other issue concerning the steady-state computations is that the frequency ω must be found in addition to fields u^* and v^*. The existence of the additional unknown is consistent with the fact that the solution to (2.4) is not unique due to the rotational symmetry in θ. One more algebraic equation must be added to (2.9) to remove the phase degeneracy and thus give $N + 1$ equations in $N + 1$ unknowns. The constraint we add is simply to fix u as some point in the domain. While the constraint destroys the banded structure of **DF**, a Sherman–Morrison technique [19] is used to find solutions of the augmented linear system using only the banded **DF**.

We now describe our computations of the leading eigenvalue spectrum of **L**. The basis of our approach is to employ a Cayley transformation to transform the eigenvalues we seek (those with largest real part) to dominant eigenvalues (of largest magnitude), and then to find iteratively dominant eigenvalues of the transformed operator. Reference [17] gives a nice review of such methods. Specifically, we consider the matrix **A** defined by the Cayley transform

$$\mathbf{A} = (\xi\mathbf{I} + \mathbf{L})^{-1}(\eta\mathbf{I} + \mathbf{L}), \tag{2.13}$$

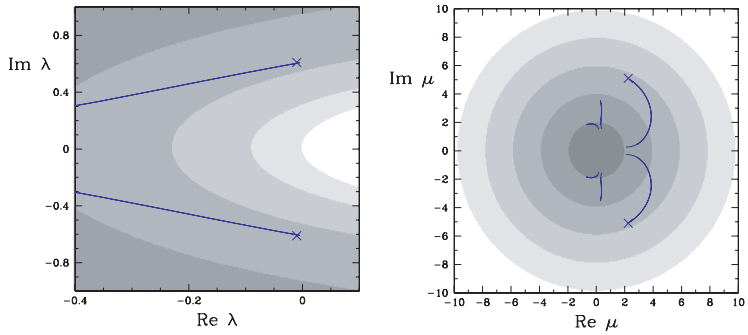


Figure 5. *Plot showing the effect of the Cayley transform on the absolute spectrum in Figure 3(c). Left is the original and right is the transformed spectrum. The right includes two periodic repeats of the absolute spectrum, which are outside of the region shown on the left. Shading indicates magnitude of eigenvalues after the transformation.*

where ξ and η are real parameters and $\mathbf{I}$ is the identity. Letting μ and λ be the eigenvalues of $\mathbf{A}$ and $\mathbf{L}$, respectively, we have the relation

$$\mu = \frac{\eta + \lambda}{\xi + \lambda}. \tag{2.14}$$

The parameters ξ and η can be adjusted so as to map the regions of interest in the λ-plane to large magnitude in the μ-plane. Using the predicted absolute spectra of Sandstede and Scheel, it is easy to find appropriate values of ξ and η. Figure 5 shows the effect of the Cayley transform on the absolute spectrum for one of the cases predicted by Sandstede and Scheel [22] (the other case is similar) for the values of ξ and η used in our computations: $\xi = -0.4$ and $\eta = 4.0$. In the μ-plane we include transforms of two periodic repeats of the absolute spectrum, one corresponding to the larger imaginary part and one to the smaller imaginary part (in the λ-plane). These repeats are outside the region of the λ-plane shown. Most of the eigenvalues of $\mathbf{L}$ lie far to the left in the λ-plane (outside the range of the figure). These are all mapped to near the origin by (2.13).

There is no need to form or store the matrix $\mathbf{A}$ in order to iteratively calculate its eigenvalues. All one needs is the ability to compute $\mathbf{AU}$ for arbitrary $\mathbf{U}$. This is accomplished using the same basic technique as in Newton's method. Letting $\mathbf{U}' = \mathbf{AU}$, we see that $\mathbf{U}'$ obeys

$$(\xi\mathbf{I} + \mathbf{L})\mathbf{U}' = (\eta\mathbf{I} + \mathbf{L})\mathbf{U}. \tag{2.15}$$

However, $(\xi\mathbf{I} + \mathbf{L})$ has the same structure, in particular the same bandwidth, as $\mathbf{L}$, and $(\eta\mathbf{I} + \mathbf{L})$ requires mostly the same computations as evaluating $\mathbf{F}$. Thus we act with $\mathbf{A}$ on $\mathbf{U}$ by computing $(\eta\mathbf{I} + \mathbf{L})\mathbf{U}$ to form a right-hand side and then solving a linear system with matrix $(\xi\mathbf{I} + \mathbf{L})$. Since this is a fixed matrix, for any given $\mathbf{L}$, it is factored only once for all actions of $\mathbf{A}$.

Dominant eigenvalues $\mathbf{A}$ are easily found by subspace iteration [17, 27]. This method guarantees that we can obtain any required number of dominant eigenvalues to arbitrarily high

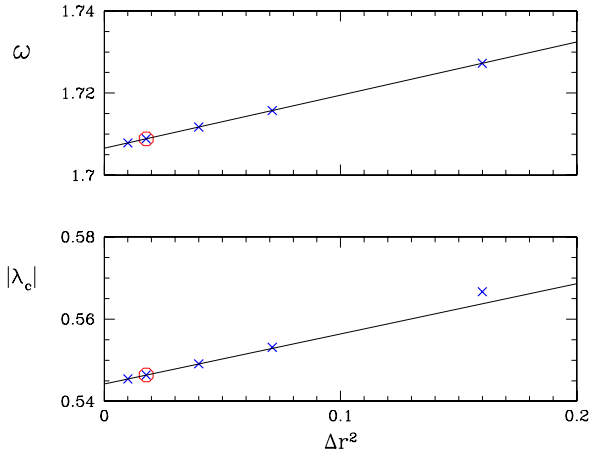


Figure 6. *Graphs illustrating the convergence of steady states (top) and eigenvalues (bottom) as functions of grid resolution $\triangle r^2$, where ω is the spiral frequency and λ_c is a complex eigenvalue. The domain radius is $R = 40$. Crosses represent $N_\theta = 256$. For $\triangle r = 0.1333$ computations have been repeated with $N_\theta = 128$, and results are shown with circles. Parameters are $a = 0.75$, $b = 0.0006$, and $\epsilon = 0.0741$.*

precision. While Arnoldi's method generally converges faster, in practice we find that with this method not all eigenvalues we require are found with high enough precision. While there are methods, such as block Arnoldi, which could probably address this, we have used subspace iteration. From the eigenvalues μ of $\mathbf{A}$ we invert (2.14) to find the required eigenvalues λ.

2.4. Accuracy. We conclude this section by considering the accuracy of our computations and providing details of numerical parameters used in the results reported. The sources of error are the following:

1. discretization error of the steady state problem, i.e., approximation of (2.5) by (2.9);
2. residual error arising from determining the roots of (2.9) to a finite accuracy;
3. discretization error of the eigenvalue problem, i.e., approximation of (2.7) by (2.10);
4. residual error arising from computing eigenvalue/eigenfunction pairs (2.10) to finite accuracy.

The two residual errors are least important. We always iterate sufficiently to reduce these to negligible size. The following hold for all reported results. Solutions $\mathbf{U}^*$ of (2.9) are found such that $\|\mathbf{F}(\mathbf{U}^*)\| < 10^{-8}$. Solutions of (2.10) are found such that $\|\mathbf{L}\tilde{\mathbf{U}} - \lambda\tilde{\mathbf{U}}\| < 10^{-8}$, where $\|\tilde{\mathbf{U}}\| = 1$. The norm is the vector 2-norm. The dimension k of the subspaces used in subspace iteration is $k = 30$ for $R = 20$ and $R = 40$, and $k = 75$ for $R = 80$. In all cases we stop iterations when $\sim 0.7k$ of the eigenvalue-eigenvector pairs have residual less than 10^{-8}. In the case of $R = 80$, we thus obtain 53 pairs with the required residual. We initially start with a subspace generated from k random vectors, but we restart iterations from previous runs when necessary.

The discretization errors are dominated, in both the steady-state and eigenvalue computations, by the second-order finite-derivative approximation to the r-derivatives in the Laplacian operator. This is expected since the θ-derivatives are computed with spectral accuracy. Thus the dominant error in the computations depends on $\triangle r$ in a well-understood way. Figure 6 shows examples of how results from steady-state and eigenvalue computations scale with $\triangle r^2$.

The domain radius is $R = 40$, half the maximum considered in our work. For the steady states we show the frequency ω, and for the eigenvalues we show the magnitude of λ_c, a complex eigenvalue associated with core breakup that will be considered in detail later in the paper. In both cases the second-order convergence is evident. We are interested only in leading eigenvalues (roughly 10^2 out of 10^5), all of which correspond to eigenfunctions with variation on approximately the same spatial scale (roughly the wavelength of the underlying spiral) so that the effects of the finite-difference discretization will be approximately the same for all eigenvalues we report.

Based on these plots, we use $\triangle r = 0.1333$ for all results reported in section 3. At this value of $\triangle r$ we have performed computations at both $N_\theta = 128$ and $N_\theta = 256$. These results show clearly that $N_\theta = 128$ gives sufficient resolution for domain radius $R = 40$. Therefore, for radii up to at least $R = 80$, $N_\theta = 256$ should produce errors smaller than the already small errors due to the radial discretization. In summary, for all results in section 3 we use $\triangle r = 0.1333$ and $N_\theta = 256$.

3. Results.

3.1. Spectra. We begin with results for the eigenvalue spectra. Figure 7 shows leading eigenvalues of $\mathbf{L}$ for the two cases considered by Sandstede and Scheel. The spectrum corresponding to core breakup is at the top, and the spectrum corresponding to far-field breakup is at the bottom. In each figure eigenvalues computed for three domain radii, $R = 20$, $R = 40$, and $R = 80$, are plotted as points with dashed lines connecting eigenvalues associated with the absolute spectrum. For comparison, the absolute and essential spectra obtained by Sandstede and Scheel are shown as solid curves.

Before considering either case in detail, we note that the predominant feature in both is the presence of many eigenvalues which approach the predicted absolute spectra as the domain radius increases. Each doubling of R results in approximately a halving of the distance of eigenvalues to the absolute spectrum, thus supporting $1/R$ convergence to the absolute spectrum. Furthermore the density of eigenvalues approximately doubles with each doubling of R. We return to this while considering each case in more detail. It should be noted that in the far-field case we have not obtained all eigenvalues associated with the periodic repeats of the absolute spectrum for $R = 80$ due to the difficulties of computing these with sufficient accuracy. This results in an abrupt termination of the eigenvalue branches for $R = 80$ at the top and bottom of Figure 7(b).

Consider first the spectrum in Figure 7(a) corresponding to core breakup. Within the region of the complex plane shown, there are three point eigenvalues. All other eigenvalues are associated with the absolute spectrum. Specifically, we find three eigenvalues which are insensitive to the domain radius and which are separated from the absolute spectrum. Of these, one is the zero eigenvalue arising due to rotational symmetry. There are three points indistinguishable from zero in Figure 7(a) corresponding to the three domain radii studied. The other two point eigenvalues are a complex-conjugate pair at approximately $0.050 \pm 0.543i$. As we shall show in section 3.3, these eigenvalues are associated with core breakup. We will denote them by λ_c. (These are the eigenvalues considered in the convergence study in Figure 6.) All other eigenvalues vary with domain radius and approach the absolute spectrum as the radius becomes large.

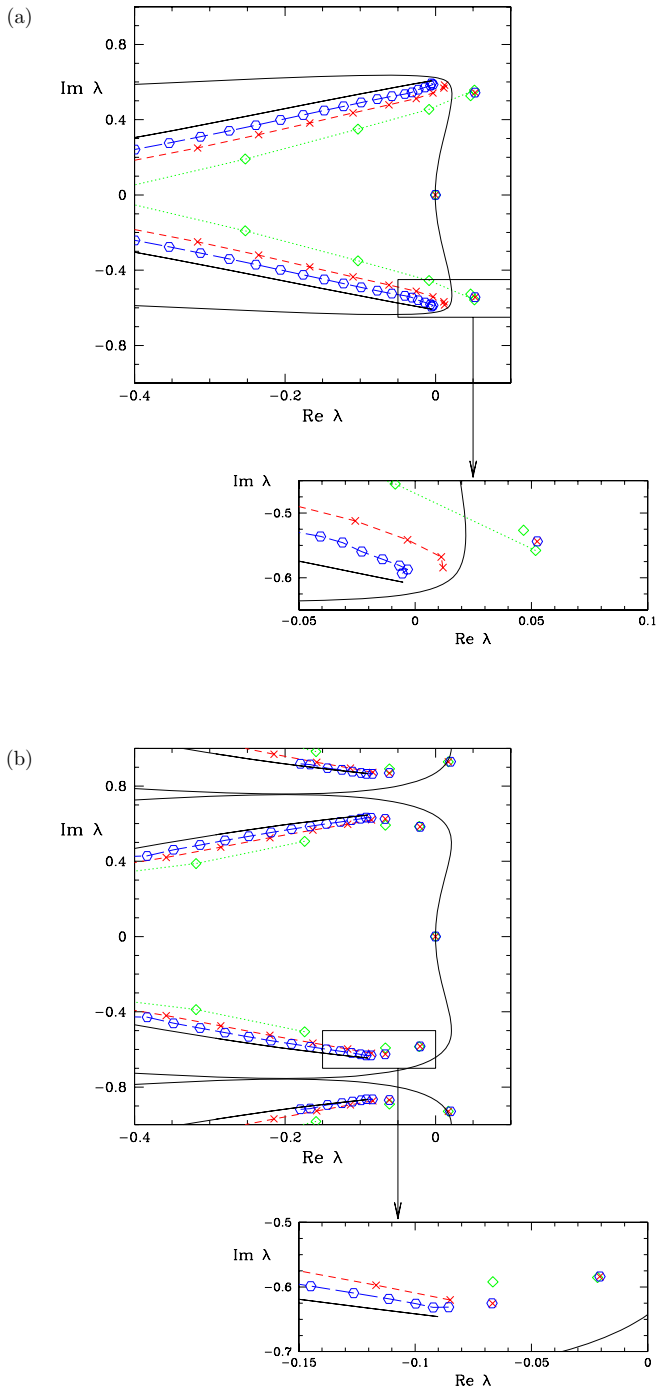


Figure 7. *Eigenvalue spectra. Shown are spectra corresponding to* (a) *core breakup and* (b) *far-field breakup. In each case eigenvalues are shown for three domain radii:* $R = 20$ *(green diamonds),* $R = 40$ *(red crosses), and* $R = 80$ *(blue circles). For each radius, the eigenvalues associated with the absolute spectrum are connected with lines. Point eigenvalue are not. Parameters are as follows:* (a) $a = 0.75$, $b = 0.0006$, $\epsilon = 0.0741$, $\omega = 1.71$*;* (b) $a = 0.84$, $b = -0.045$, $\epsilon = 0.0751$, $\omega = 1.50$.

The enlargement in Figure 7(a) clarifies the situation around the complex point eigenvalues. Even on the enlarged scale the point eigenvalues for $R = 40$ and $R = 80$ coincide. At $R = 20$, however, there are two nearby eigenvalues. The lower one corresponds to the absolute spectrum (indicated by the connecting dashed line) because it moves, as R is increased, toward the absolute spectrum. The other eigenvalue converges, as R is increased, to the point eigenvalue. Note that while the edge of the eigenvalue branch associated with the absolute spectrum is near the point eigenvalue at $R = 20$, the branch does not approach the point eigenvalue as the domain becomes small. It is nevertheless interesting that the point eigenvalue is near the edge of the absolute spectrum. We find this throughout and return to it in the conclusion.

As already noted, the distance of eigenvalues to the absolute spectrum is approximately proportional to $1/R$, and the density of eigenvalues is approximately proportional to R. Because we are not able to extend the computations significantly beyond the radius $R = 80$ (already quite large), there is not enough data to draw strong conclusions about these scalings. In particular it is not clear from the data whether or not the scaling in the vicinity of the edge of the absolute spectrum is different from those elsewhere. Near the edge of the spectrum the eigenvalues are more dense and closer to the absolute spectrum than elsewhere. More importantly we observe a curving and perhaps folding at the edge of the eigenvalue branch as the radius becomes large. This would again suggest a different scaling at the spectrum's edge, but the numerical results are inconclusive.

We have focused our study on the eigenvalues within the region shown in Figure 7(a), but we have computed some eigenvalues outside of this region. In particular our iterative technique frequently finds eigenvalues associated with the periodic repetition of the absolute spectrum in the complex plane. We have not attempted to resolve the details of the other eigenvalue branches. Also there is a complex-conjugate pair of point eigenvalues near $0 \pm i\omega$ due to approximate translational symmetry.

Now consider the spectrum corresponding to far-field breakup. In Figure 7(b) we find four complex-conjugate eigenvalue pairs that we can clearly classify as point eigenvalues. One of these pairs has small positive real part, and hence the spiral wave is absolutely linearly unstable; see section 3.3. Again we observe that the point eigenvalues, except for the zero eigenvalue, appear near the edge of the absolute spectrum.

We observe approximately $1/R$ convergence of eigenvalues to the absolute spectrum. The only apparent deviation is again at the edge of the spectrum. In this case we do not observe any curving of the eigenvalue branch as seen in the enlargement in Figure 7(a); however, we do find that the right-most point of the computed branch does not appear to approach the edge of the predicted absolute spectrum. One possibility is that this last eigenvalue is in fact a point eigenvalue very close to the edge of the absolute spectrum.

3.2. Eigenfunctions. In Figures 8 and 9 we plot eigenfunctions for a representative selection of eigenvalues. All eigenfunctions have been obtained on a domain with $R = 80$, the largest we consider. Each eigenfunction is shown in two formats. In the left column eigenfunctions are visualized on the computational domain. Specifically the $\tilde{u}$-field of the real part of the eigenfunction is plotted, with black used where the field values are near zero. The imaginary parts of the eigenfunctions are not fundamentally different. In the right column

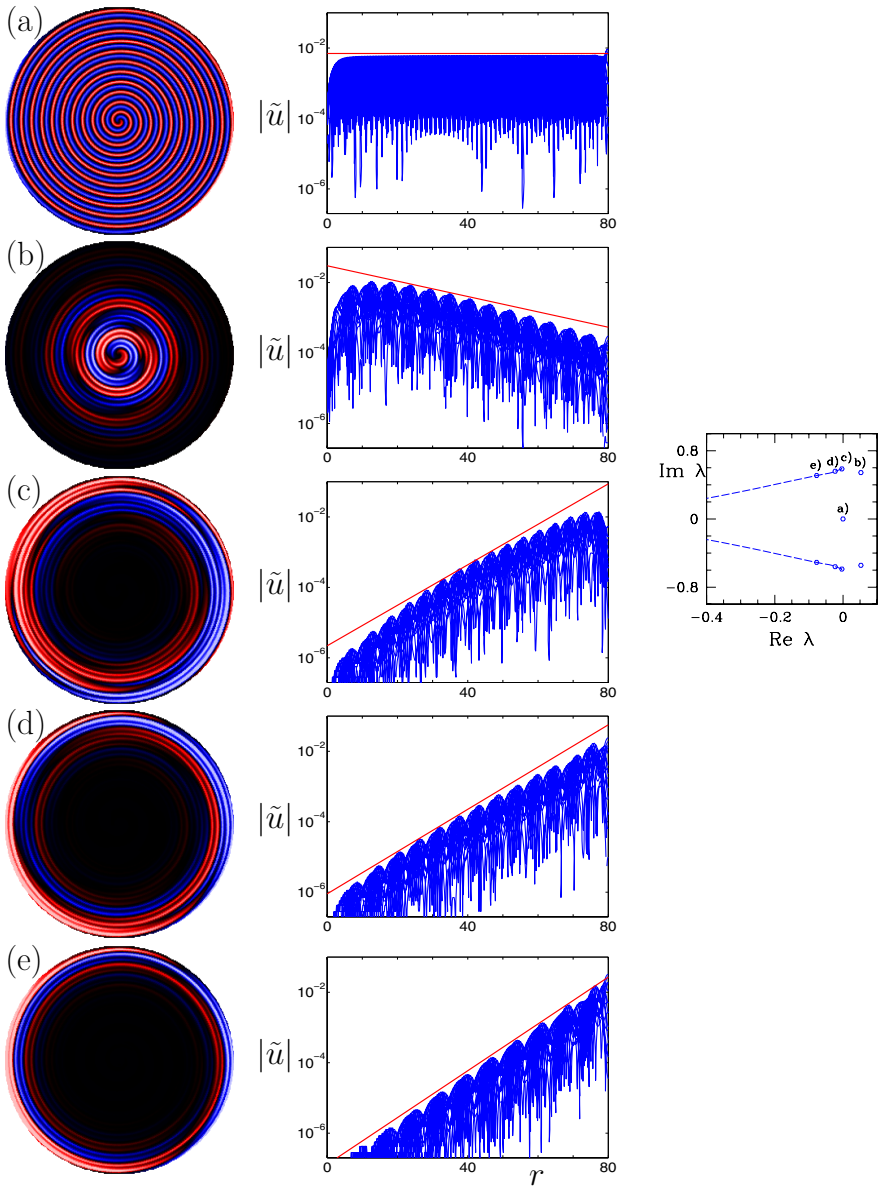


Figure 8. *Representative eigenfunctions for parameters corresponding to core breakup. The left column shows the real part of each eigenfunction. The $\tilde{u}$-field is plotted, with black used where the field is near zero. The right column shows the r-dependence of $|\tilde{u}|$, with predicted growth/decay rates also shown with lines (see text). Plot at the far right is a guide to the corresponding eigenvalues.* (a) *Zero (rotational) eigenvalue,* (b) *positive eigenvalue λ_c, and* (c), (d), (e) *three representative eigenvalues associated with the absolute spectrum. Parameters as in Figure* 7(a).

the modulus of each eigenfunction is shown as a function of radius. Specifically, we generate 16 radial sections $|\tilde{u}(r, \theta_s)|$, where $\theta_s = s\pi/8$ and $s = 0, 1, \ldots, 15$, and plot all 16 sections simultaneously as a function of r. The envelope of these sections gives a simple representation of the modulus of the eigenfunction as a function of r.

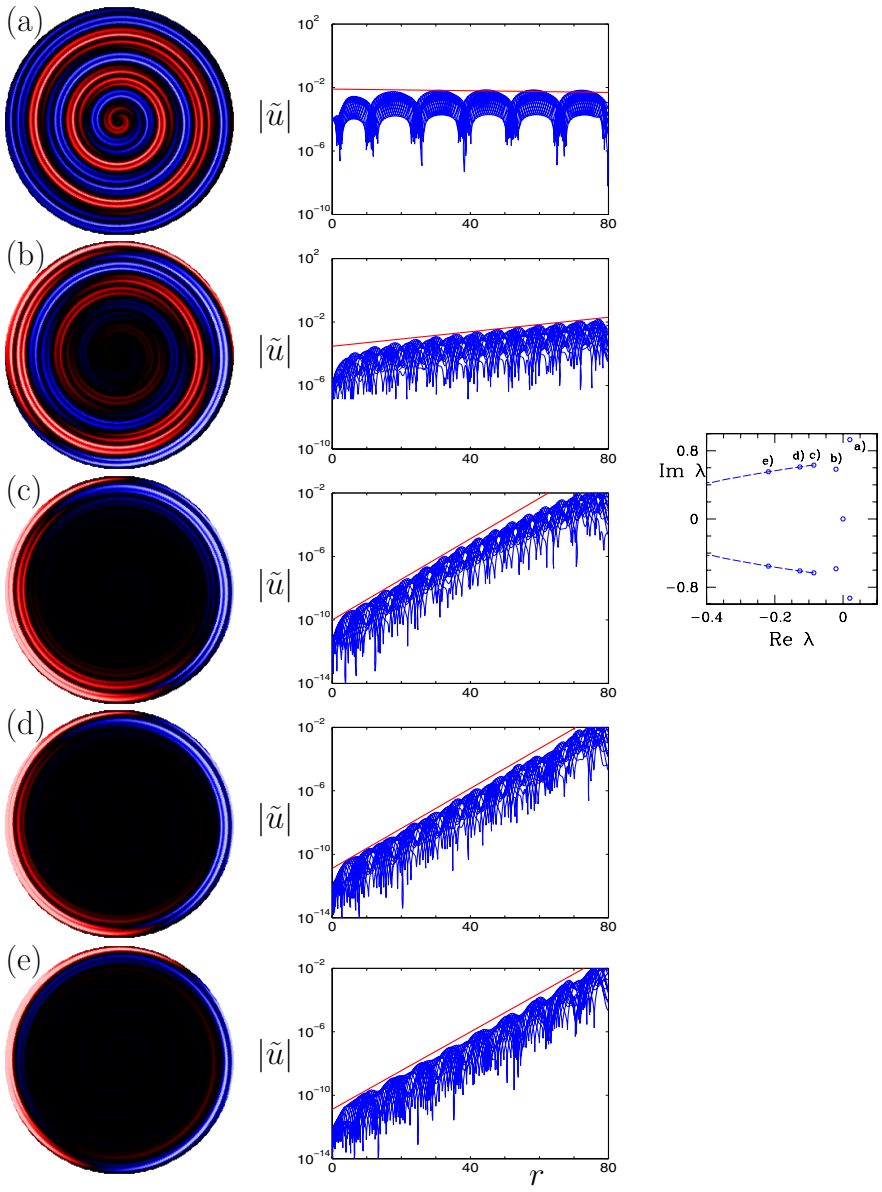


Figure 9. *Representative eigenfunctions for parameters corresponding to far-field breakup. Same format is used as in Figure 8. The rotational eigenfunction is not shown. Two eigenfunctions corresponding to point eigenvalues are shown with* (a) *slightly positive eigenvalue and* (b) *slightly negative eigenvalue.* (c), (d), (e) *are three representative eigenvalues associated with the absolute spectrum. Parameters as in Figure 7*(b).

The technique used by Sandstede and Scheel [23] to obtain absolute and essential spectra also predicts large-r behavior of eigenfunctions. The main prediction is that if an eigenvalue is to the left of the essential spectrum, then the corresponding eigenfunctions will be exponentially growing at large r, whereas if an eigenvalue is to the right of the essential spectrum, then the corresponding eigenfunctions will be exponentially decaying at large r. In addition to

the general prediction, the numerical technique employed by Sandstede and Scheel to obtain spectra for specific problems also provides the growth/decay rates for eigenfunctions. These rates [20] are indicated by the (red) lines in the right column of Figures 8 and 9. For eigenfunctions corresponding to the absolute spectrum, the predicted exponential growth rates have been taken from the point on the absolute spectrum nearest to the computed eigenvalue. Only the slope of the lines is relevant. The intercept is chosen for ease of comparison with the eigenfunctions.

Consider first the eigenfunctions in Figure 8 corresponding to the case of core breakup. The top eigenfunction is the zero mode due to rotational symmetry. This eigenfunction is given by the θ-derivative of the underlying spiral wave and hence closely resembles the spiral. The eigenfunction neither grows nor decays at large r.

Figure 8(b) shows the eigenfunction corresponding to the positive complex-conjugate point eigenvalues λ_c. Since the eigenvalues are to the right of the essential spectrum, the eigenfunction decays with r. While there is generally good agreement between the computed eigenfunction and prediction, there is some deviation from prediction that is more pronounced at larger r.

Figures 8(c)–(e) show three eigenfunctions associated with the absolute spectrum. The agreement between the computed eigenfunctions and predictions is extremely good away from the edge of the absolute spectrum. Near the edge the agreement is less good. In particular, eigenfunctions are not pure exponential, even at large r, and the computed eigenfunctions systematically grow more slowly than prediction. While not shown, we find that the eigenfunctions computed on smaller domains show even slower growth as a function of r. This would suggest that the deviation shown in Figure 8(c) is due to finite domain size.

Figure 9 shows eigenfunctions corresponding to parameters for which far-field breakup occurs. We plot eigenfunctions corresponding to two of the complex-conjugate point eigenvalues and show three representative eigenfunctions associated with the absolute spectrum. The eigenfunction corresponding to the zero eigenvalue is similar to Figure 8(a) and is not shown.

The eigenfunctions corresponding to the point eigenvalues agree very well with prediction. The growth rate of the positive eigenfunction in Figure 9(a) is very small since the corresponding eigenvalue is almost exactly on the essential spectrum (Figure 7(b)). This is a fortuitous situation which illustrates nicely that the essential spectrum delimits the crossover from growth to decay of eigenfunctions. While quantitatively the agreement is very good, there is a qualitative disagreement between the computed eigenfunction and prediction. Our computed eigenfunction is growing with r, indicating that the eigenvalue is actually slightly to the left of the essential spectrum, whereas in Figure 7(b) the eigenvalue is slightly to the right of the essential spectrum, and the predicted exponent is slightly negative. Quantitatively the difference is very small and is likely due to a small numerical difference, e.g., a difference between the value of ω found in our computations and that used by Sandstede and Scheel. The closeness of this eigenvalue to the essential spectrum is just by chance. If parameters are changed, the eigenvalue moves away from the essential spectrum. It is because of this closeness of the eigenvalue to the essential spectrum that numerical simulations at these parameter values are so sensitive to numerical resolution (as noted in section 1).

The eigenfunction in Figure 9(b) is exponentially growing since the corresponding eigenvalue is to the left of the essential spectrum. There are no observable deviations from pure

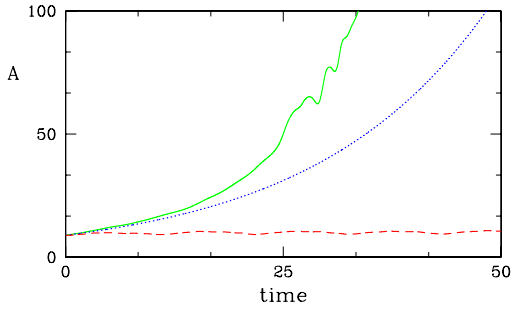


Figure 10. *Two time series showing the evolution starting from different perturbations of the steady spiral with parameters leading to core breakup. In one case (solid green) the perturbation is the eigenfunction corresponding to positive eigenvalue λ_c ((b) in Figure 8). In the other (dashed red) the perturbation is the eigenfunction corresponding to right-most eigenvalue associated with the absolute spectrum ((c) in Figure 8). The dotted blue curve shows exponential growth at rate given by λ_c. Parameters are as in Figure 7(a).*

exponential growth at large r, and the agreement with prediction is very good.

The three eigenfunctions associated with the absolute spectrum show the same trend as in Figure 8. The agreement between the computed eigenfunctions and predictions is extremely good away from the edge of the absolute spectrum, while near the edge eigenfunctions are not pure exponential and systematically grow more slowly than prediction. This case is even more striking than in Figure 8 for the following reasons. The growth rates in Figure 8 are much larger than in Figure 9. (Note the scale change.) The numerical values representing the eigenfunctions span a larger range, and yet the computed eigenfunctions away from the edge agree very well with predictions. There is every reason to believe that these eigenfunctions are numerically well resolved within the finite domain. Unlike the case in Figure 8(c), here the eigenfunction closest to the edge of the spectrum, Figure 9(c), deviates from exponential growth only at large r. There is a clear range r, up to approximately $r = 40$, where the eigenfunction agrees well with the predicted exponential growth. This strongly suggests that the lack of agreement is due to finite-size effects. It is nevertheless interesting that these are more pronounced near the edge of the spectrum.

3.3. Implications for breakup. We now return to the issue of spiral breakup discussed at the outset (Figure 3). We begin with the case of core breakup. Recall that while this was treated by Sandstede and Scheel [22, 23], they were not able to draw definite conclusions about the role of the absolute and essential spectrum in core breakup. It is already clear from the spectra in Figure 7 that the steady spiral is linearly unstable due to the presence of positive point eigenvalues λ_c. Here we present additional nonlinear simulations of the breakup.

Figure 10 shows the time evolution from two different initial conditions, each composed of the steady spiral plus a small amount of one of the computed eigenfunctions. The amplitude plotted is defined as $A = \min_\theta \|U^* - R_\theta U\|$, where R_θ is a rotation by angle θ. The norm is the vector 2-norm. Essentially A is the norm of the difference between the u-field of the steady spiral, U^*, and the u-field of the nonlinear solution, U. The minimization over rotation is included to take into account the small drift of the nonlinear solution relative to the steady spiral.

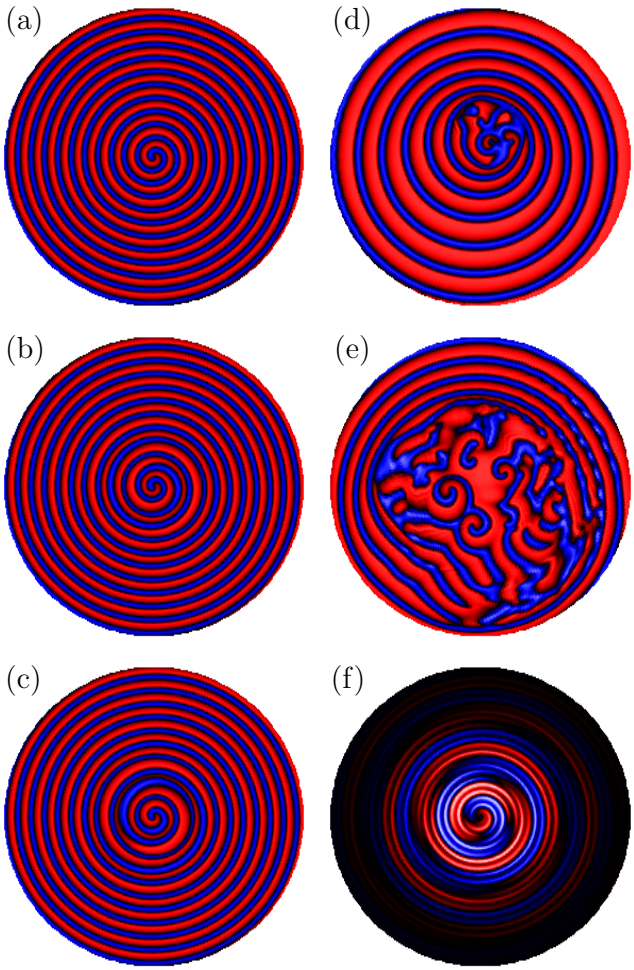


Figure 11. *Snapshots of evolution from the perturbed steady spiral leading to far-field breakup (solid green curve in Figure 10). The u-field is shown with $u \simeq 0$ blue and $u \simeq 1$ red.* (a) $t = 0$, (b) $t = 20$ *($\simeq 5$ periods),* (c) $t = 25$, (d) $t = 85$ *($\simeq 23$ periods),* (e) $t = 120$ *($\simeq 33$ periods),* (f) *eigenfunction. Parameters are as in Figure 7*(a).

Consider the evolution starting from the initial condition formed from the positive eigenfunction corresponding to λ_c. Accompanying visualizations are presented in Figure 11. The dynamics is initially linear, obeying the exponential growth dictated by the real part of λ_c. After a short time the growth becomes nonlinear, and almost immediately core breakup occurs [Figure 11(c); time 25]. Beyond this time the amplitude A loses most of its meaning. Visualizations at much later times are shown. One of the more striking aspects of the breakup is that it occurs at $r \simeq 20$, not at the center of the spiral. This radius is near where the unstable eigenfunction has maximal magnitude. Visually one sees the similarity between the nonlinear breakup and the unstable eigenfunction in Figure 11.

The initial nonlinear growth in Figure 10 is faster than linear. This means that, at lowest

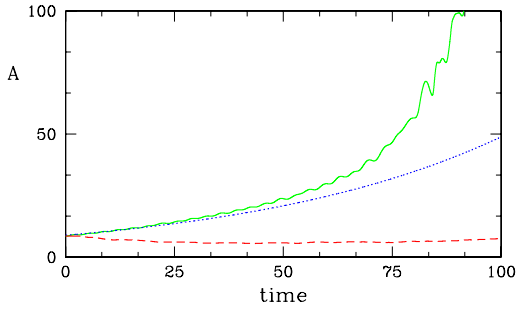


Figure 12. *Two time series showing the evolution starting from different perturbations of the steady spiral with parameters leading to far-field breakup. In one case (solid green) the perturbation is the eigenfunction corresponding to a positive eigenvalue (Figure 9(a)). In the other (dashed red) the perturbation is the eigenfunction corresponding to the weakly stable point eigenvalue (Figure 9(b)). The dotted blue curve shows exponential growth at a rate given by the positive eigenvalue. Parameters are as in Figure 7(b).*

order, the effect of nonlinearity on the instability is destabilizing. Such behavior occurs, for example, sufficiently close to a subcritical bifurcation; see, e.g., [13]. This nonlinear destabilization is consistent with the fact that small positive eigenvalues lead to the dramatic breakup of the spiral wave. If nonlinearity were stabilizing, one would expect the linear instability to saturate in a state resembling a superposition of the original spiral and a small amount of the unstable eigenmode (as occurs, for example, in spiral meandering [6, 7, 8]). We note that in systems such as this one the behavior can change very rapidly with parameters following a bifurcation [25, 26], and hence we are not able to conclude that the nonlinear growth follows from a subcritical bifurcation, only that at these parameter values it is destabilizing.

For completeness we have also computed the nonlinear evolution from an initial condition formed from the eigenfunction corresponding to the right-most eigenvalue of the absolute spectrum (Figure 8(c)). Figure 10 shows the initial dynamics from this simulation. Not surprisingly A does not change much on the scale of Figure 10 because the associated eigenvalue is very near zero. The simulation eventually leads to core breakup if run long enough. However, this is simply because the steady spiral is linearly unstable. When breakup does eventually occur, it follows the same route (e.g., same exponential growth) as for the initial condition based on the positive eigenfunction.

The conclusion is that, in this case, core breakup is due to nonlinear effects following from linear instability due to a complex-conjugate pair of point eigenvalues. The absolute spectrum plays no direct role in the spiral breakup.

Next we briefly consider far-field breakup. We have directly computed the eigenfunction associated with absolute instability causing far-field breakup, Figure 9(a). The leading eigenfunction shows exactly the long wavelength modulation of the steady spiral expected for this instability [1, 4, 9, 18, 26, 29]. Figures 12 and 13 show the dynamics following from the steady spiral perturbed by the unstable eigenfunction. The dynamics is initially that of exponential growth with the expected growth rate. The growth becomes nonlinear, and long-wavelength modulation of the spiral becomes visible (time 80 in Figure 13). Shortly thereafter the spiral breaks near the domain boundary. At these parameter values the eigenfunction grows

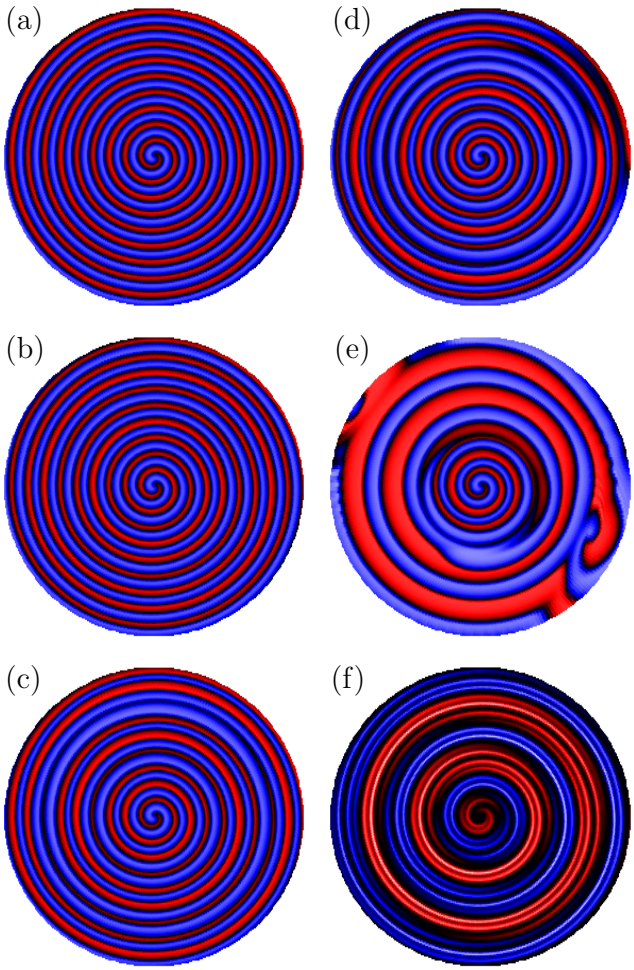


Figure 13. *Snapshots of evolution from the perturbed steady spiral leading to far-field breakup (solid green curve in Figure 12). The u-field is shown.* (a) $t = 0$, (b) $t = 50$ *(*$\simeq 12$ *periods),* (c) $t = 80$ *(*$\simeq 19$ *periods),* (d) $t = 85$, (e) $t = 120$ *(*$\simeq 29$ *periods),* (f) *eigenfunction. Parameters are as in Figure 7*(b).

weakly with radius, as seen in Figure 9(a), and for this reason one would expect a preference for breakup near the domain boundary. However, in this case the qualitative character of the eigenfunction depends sensitively on parameters, and for slightly different parameters the eigenfunction may decay weakly with the radius.

The far-field case is similar to the core breakup case in most other respects. The nonlinear growth in Figure 12 is faster than linear. No other eigenvalues appear to play a direct role in the far-field breakup. Figure 12 shows the evolution from an initial condition formed from the eigenfunction corresponding to the complex-conjugate point eigenvalues near the imaginary axis (Figure 9(b)).

4. Conclusions. In this paper we have examined in detail the linear stability spectra and associated eigenfunctions for spiral waves in large domains. Everywhere, except possibly near the edges of the absolute spectra, numerically computed eigenvalues and eigenfunctions agree extremely well with the results of Sandstede and Scheel. Our results answer the question posed at the outset. Absolute spectra can be relevant and predictive for typical domain sizes used in studies of spiral waves. Even in domains containing only a few spiral wavelengths (the case $R = 20$) eigenvalues show signs of the absolute spectrum—they lie along curves located roughly in the correct part of the complex plane. For domains containing five spiral wavelengths or more ($R \gtrsim 40$) eigenvalues lie quite close to the absolute spectra. Of course these results are for the particular equations and parameters studied here, and absolute spectra will not necessarily be such good predictors for domains of these sizes in other systems. Nevertheless, in at least two cases, one with excitable dynamics and one with oscillatory dynamics, absolute spectra are predictive.

The computed eigenvalues support convergence to the absolute spectrum inversely with the domain radius, at least away from the edge of the absolute spectrum. In most cases the eigenfunctions agree with the exponential forms deduced by Sandstede and Scheel. This is even the case for point eigenvalues not associated with the absolute spectrum. Near the edges of the absolute spectrum, however, eigenfunctions do not exhibit exponential growth at large radius, even in the largest domains we have considered. While results suggest that this is due to finite-size effects, more work is necessary to understand the behavior of eigenvalues and eigenfunctions near the edges of the spectrum.

We have computed the positive point eigenvalues giving rise to both core and far-field breakup of spiral waves. The essential difference between the two cases is the form of the eigenfunctions. For core breakup the eigenfunction has a maximum not far from the core of the spiral and decays at large radius. For far-field breakup the eigenfunction grows with radius. Moreover, the far-field eigenfunction shows long-wavelength modulation known to precede far-field breakup. Nonlinearity also plays a role in breakup, and we have presented nonlinear simulations showing the destabilizing effects of nonlinearity in each case.

The most intriguing aspect of this work is that all point eigenvalues we have found appear near edges of the absolute spectra. This may be a coincidence, but it would not seem so from Figure 7. We leave this as an open problem.

Acknowledgments. We are very grateful to Bjorn Sandstede and Arnd Scheel for helpful discussions and for kindly providing data used for many of the comparisons presented.

REFERENCES

[1] I. Aranson, L. Aranson, L. Kramer, and A. Weber, *Stability limits of spirals and traveling waves in nonequilibrium media*, Phys. Rev. A, 46 (1992), pp. R2992–R2995.

[2] M. Bär and L. Brusch, *Breakup of spiral waves caused by radial dynamics: Eckhaus and finite wave-number instabilities*, New J. Phys., 6 (2004), pp. 1–22.

[3] M. Bär and M. Eiswirth, *Turbulence due to spiral breakup in a continuous excitable medium*, Phys. Rev. E, 48 (1993), pp. R1635–R1637.

[4] M. Bär and M. Or-Guil, *Alternative scenarios of spiral breakup in a reaction-diffusion model with excitable and oscillatory dynamics*, Phys. Rev. Lett., 82 (1999), pp. 1160–1163.

[5] D. Barkley, *A model for fast computer-simulation of waves in excitable media*, Phys. D, 49 (1991), pp. 61–70.
[6] D. Barkley, *Linear-stability analysis of rotating spiral waves in excitable media*, Phys. Rev. Lett., 68 (1992), pp. 2090–2093.
[7] D. Barkley, *Euclidean symmetry and the dynamics of rotating spiral waves*, Phys. Rev. Lett., 72 (1994), pp. 164–167.
[8] D. Barkley, *Spiral meandering*, in Chemical Waves and Patterns, R. Kapral and K. Showalter, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995, pp. 163–190.
[9] L. Brusch, A. Torcini, and M. Bär, *Doppler effect of nonlinear waves and superspirals in oscillatory media*, Phys. Rev. Lett., 91 (2003), paper 108302.
[10] M. C. Cross and P. C. Hohenberg, *Pattern-formation outside of equilibrium*, Rev. Mod. Phys., 65 (1993), pp. 851–1112.
[11] E. Doedel and L. S. Tuckerman, eds., *Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems*, IMA Vol. Math. Appl. 119, Springer, New York, 2000.
[12] M. Dowle, R. M. Mantel, and D. Barkley, *Fast simulations of waves in three-dimensional excitable media*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 7 (1997), pp. 2529–2546.
[13] R. D. Henderson and D. Barkley, *Secondary instability in the wake of a circular cylinder*, Phys. Fluids, 8 (1996), pp. 1683–1685.
[14] R. Kapral and K. Showalter, eds., *Chemical Waves and Patterns*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.
[15] A. Karma, *Spiral breakup in model equations of action potential propagation in cardiac tissue*, Phys. Rev. Lett., 71 (1993), p. 1103–1106.
[16] J. Keener and J. Sneyd, *Mathematical Physiology*, Springer, New York, 1998.
[17] K. Meerbergen and D. Roose, *Matrix transformations for computing rightmost eigenvalues of large sparse non-symmetric eigenvalue problems*, IMA J. Numer. Anal., 16 (1996), pp. 297–346.
[18] Q. Ouyang and J. M. Flesselles, *Transition from spirals to defect turbulence driven by a convective instability*, Nature, 379 (1996), pp. 143–146.
[19] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, UK, 1992.
[20] B. Sandstede, *private communication*, 2004.
[21] B. Sandstede and A. Scheel, *Absolute and convective instabilities of waves on unbounded and large bounded domains*, Phys. D, 145 (2000), pp. 233–277.
[22] B. Sandstede and A. Scheel, *Absolute versus convective instability of spiral waves*, Phys. Rev. E, 62 (2000), pp. 7708–7714.
[23] B. Sandstede and A. Scheel, *Superspiral structures of meandering and drifting spiral waves*, Phys. Rev. Lett., 86 (2001), pp. 171–174.
[24] B. Sandstede and A. Scheel, *Stability, and Instability, of Spiral Waves*, manuscript.
[25] S. Tobias, M. Proctor, and E. Knobloch, *Convective and absolute instabilities of fluid flows in finite geometry*, Phys. D, 113 (1998), pp. 43–72.
[26] S. M. Tobias and E. Knobloch, *Breakup of spiral waves into chemical turbulence*, Phys. Rev. Lett., 80 (1998), pp. 4811–4814.
[27] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
[28] A. T. Winfree, *When Time Breaks Down*, Princeton University Press, Princeton, NJ, 1987.
[29] L. Zhou and Q. Ouyang, *Experimental studies on long-wavelength instability and spiral breakup in a reaction-diffusion system*, Phys. Rev. Lett., 85 (2000), pp. 1650–1653.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 2, pp. 179–204

Entrainment and Chaos in a Pulse-Driven Hodgkin–Huxley Oscillator*

Kevin K. Lin†

Abstract. The Hodgkin–Huxley model describes action potential generation in certain types of neurons and is a standard model for conductance-based, excitable cells. Following the early work of Winfree and Best, this paper explores the response of a spontaneously spiking Hodgkin–Huxley neuron model to a periodic pulsatile drive. The response as a function of drive period and amplitude is systematically characterized. A wide range of qualitatively distinct responses are found, including entrainment to the input pulse train and persistent chaos. These observations are consistent with a theory of kicked oscillators developed by Q. Wang and L.-S. Young. In addition to general features predicted by Wang–Young theory, it is found that most combinations of drive period and amplitude lead to entrainment instead of chaos. This preference for entrainment over chaos is explained by the structure of the Hodgkin–Huxley phase-resetting curve.

Key words. entrainment and phase-locking, strange attractors, Hodgkin–Huxley model, driven nonlinear oscillators

AMS subject classifications. 37N25, 65P99

DOI. 10.1137/050631677

1. Introduction.

The Hodgkin–Huxley model describes action potential generation in certain types of neurons and is a standard model for conductance-based, excitable cells [5, 18, 20]. There is an extensive literature on the response of the Hodgkin–Huxley model to different types of inputs [1, 2, 11, 14, 15, 16, 17, 19, 24, 25], and understanding how single neurons respond to external forcing continues to be relevant for the study of information transmission in neural systems [21, 23]. Because neurons typically communicate via pulsatile synaptic events, it is natural to investigate the response of the Hodgkin–Huxley model to pulsatile inputs. Early studies by Best [3] and Winfree [38] examine the response of a Hodgkin–Huxley model to periodic impulse trains, characterizing in detail the structure of phase singularities and the transition from degree 1 to degree 0 phase resetting. However, their work does not systematically address the asymptotic dynamical behavior as a function of drive period and amplitude.[1]

This paper studies a spontaneously spiking (i.e., oscillatory) Hodgkin–Huxley neuron model driven by periodic, pulsatile input of fixed amplitude and period, and systematically classifies the response as a function of drive period and amplitude. It is found that the following hold:

*Received by the editors May 16, 2005; accepted for publication (in revised form) by D. Terman January 19, 2006; published electronically April 21, 2006. This work is partially supported by an NSF Postdoctoral Fellowship.
http://www.siam.org/journals/siads/5-2/63167.html
†Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012 (klin@cims.nyu.edu).
[1]Takabe, Aihara, and Matsumoto [32] appear to have carried out such a systematic study. However, I was able to locate only an abstract.

1. In response to periodic pulsatile forcing of fixed amplitude A and period T, a spontaneously spiking Hodgkin–Huxley system can exhibit a wide range of distinct behaviors depending on A and T:
   (a) *Entrainment.* The driven system is stably periodic, and its period is a rational multiple of the drive period T.
   (b) *Transient chaos.* The system experiences a transient period of exponential instability before entraining to the input. This transient chaos is caused by a Smale horseshoe [13].
   (c) *Chaos.* The system becomes fully chaotic: it possesses a positive Lyapunov exponent and a mixing attractor. (See [39] for a review of these concepts.)

   The response of the pulse-driven neuron is approximately T_0-periodic in the drive period T, where T_0 is the intrinsic period of the unforced Hodgkin–Huxley oscillator. For example, if the pulse-driven oscillator is chaotic for some drive amplitude A and drive period T, then it is likely to be chaotic when driven by a pulse train of amplitude A with period near $T + T_0$.
2. The scenarios enumerated above are *prevalent* in the sense that they correspond to positive-area subsets of the drive period–drive amplitude space. Prevalence, together with the approximate periodicity stated above, implies that each scenario occurs with positive "probability." (See the discussion of Figure 3 in section 3 for the precise meaning of probability in this context.) The range of responses and their prevalence are consistent with a theory of nonlinear oscillators developed recently by Wang and Young [33, 34, 35, 36].
3. While chaotic behavior is readily observable, most combinations of drive period and drive amplitude lead to entrainment instead of chaos. This preference for entrainment can be explained by the structure of the phase-resetting curve (see section 4) of the Hodgkin–Huxley system.

This paper relies heavily on numerical computation, the conceptual framework provided by Winfree's theory of biological rhythms [38], and the work of Wang and Young on nonlinear oscillators [34, 35]. Phase-resetting curves, introduced by Winfree, play a particularly important role here. The phase-resetting curve of a nonlinear oscillator is an interval map which captures the asymptotic response of a nonlinear oscillator to a single, pulsatile perturbation. Because they are one-dimensional objects, phase-resetting curves are often easier to understand than the nonlinear oscillators they represent. They are frequently used to infer stable dynamical behavior like phase locking. Wang–Young theory provides a mathematical framework for using phase-resetting curves to infer the existence and prevalence of chaotic behavior. Rather than numerically verify the hypotheses of their theorems, we have opted to examine the consequences of the theory directly, relying on a combination of numerical simulation and geometric reasoning to characterize the specific response of the Hodgkin–Huxley model to a periodic pulsatile drive.

For the sake of clarity, parameters are selected to ensure that the Hodgkin–Huxley system possesses a unique limit cycle and no other attracting invariant set. This corresponds to a repeatedly spiking neuron with an unstable rest state. While the scenarios stated above should still hold when the limit cycle coexists with other stable invariant sets, this choice simplifies the interpretation of numerical simulations. Otherwise, a trajectory may jump out of the

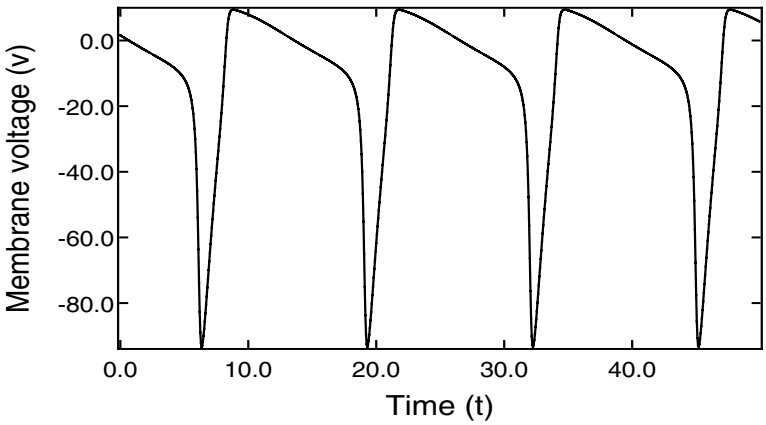


Figure 1. *The time course for the Hodgkin–Huxley equations at the parameter values (2.5). The rapid "spike" followed by a long "recovery" period is typical of the Hodgkin–Huxley equations. Note that this paper uses the original definition of membrane voltage used by Hodgkin and Huxley, so that action potentials are downward (rather than upward) spikes.*

basin of the limit cycle, which obscures the mechanism described by Wang–Young theory. This last scenario has been investigated thoroughly by Winfree [38] and Best [3].

The rest of this paper is organized as follows. Section 2 briefly reviews the unforced Hodgkin–Huxley equations and their properties. Main numerical results are summarized in section 3 and discussed in section 4. Section 5 discusses further numerical results, addressing some issues raised in sections 3 and 4. Section 6 discusses possible extensions and generalizations.

2. Brief review of the Hodgkin–Huxley model. The Hodgkin–Huxley equations are a system of nonlinear ordinary differential equations[2] that describe the way neurons generate spatially and temporally localized electrical pulses [5, 18, 20]. These electrical pulses, called *action potentials*, are the primary way in which neurons transmit information. Action potentials are triggered by sufficiently large membrane voltages, which can be set up by the influx of ions into the cell. A neuron is said to *fire* or *spike* when it generates an action potential (Figure 1). The Hodgkin–Huxley model describes action potential generation in terms of the membrane voltage and dimensionless *gating variables* which quantify the effective permeability (or *conductance*) of the membrane to various types of ions.

The original Hodgkin–Huxley equations model action potential generation in the squid giant axon. This giant axon contains two types of membrane ion channels. One type of channel is specific to potassium ions, the other to sodium ions. The state variables of the model are the membrane voltage v, the activation n of the potassium channels, and the

[2]This paper does not treat the Hodgkin–Huxley PDEs: spatial dependence is not relevant here.

activation m and inactivation h of the sodium channels. The equations are [18]

$$\begin{aligned} \dot{v} &= C^{-1}\left[-I - \bar{g}_{\mathrm{K}} n^4 (v - v_{\mathrm{K}}) - \bar{g}_{\mathrm{Na}} m^3 h (v - v_{\mathrm{Na}}) - \bar{g}_{\mathrm{leak}} (v - v_{\mathrm{leak}})\right], \\ \dot{m} &= \alpha_{\mathrm{m}}(v)(1-m) - \beta_{\mathrm{m}}(v) m, \\ \dot{n} &= \alpha_{\mathrm{n}}(v)(1-n) - \beta_{\mathrm{n}}(v) n, \\ \dot{h} &= \alpha_{\mathrm{h}}(v)(1-h) - \beta_{\mathrm{h}}(v) h, \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} \alpha_{\mathrm{m}}(v) &= \Psi\left(\tfrac{v+25}{10}\right), & \beta_{\mathrm{m}}(v) &= 4\exp\left(\tfrac{v}{18}\right), \\ \alpha_{\mathrm{n}}(v) &= 0.1\Psi\left(\tfrac{v+10}{10}\right), & \beta_{\mathrm{n}}(v) &= 0.125\exp\left(\tfrac{v}{80}\right), \\ \alpha_{\mathrm{h}}(v) &= 0.07\exp\left(\tfrac{v}{20}\right), & \beta_{\mathrm{h}}(v) &= \tfrac{1}{1+\exp\left(\frac{v+30}{10}\right)}, \\ \Psi(v) &= \tfrac{v}{\exp(v)-1}. \end{aligned} \tag{2.2}$$

Each ion channel consists of independent identical subunits, which must all open to allow ions to pass through. The gating variables m, n, and h take values in $(0, 1)$ and represent the fraction of subunits which are open. The term n^4 enters into the potassium conductance because potassium channels consist of four identical subunits; analogous structures account for the m^3h term in the sodium conductance [5]. The gating variable equations are master equations for continuous-time Markov chains with voltage-dependent transition rates α and β; the Markov chains describe the opening and closing of the corresponding channel subunits. The $\dot{v}$ equation is Kirchoff's current law. Action potentials are *downward* voltage spikes and a positive I corresponds to an *inflow* of positively-charged ions. The voltage convention here is that of [18] and opposite contemporary usage: the membrane voltage v is defined by

$$v = \text{voltage outside} - \text{voltage inside}.$$

Action potentials are generally initiated by perturbations to the membrane voltage. Such perturbations may be caused, for instance, by the flow of ions across the cell membrane. Because neurons transmit signals through spatially and temporally localized pulses, it is natural to model stimuli as impulses [31]. The simplest type of repetitive, pulsatile stimulus to a neuron is a periodic impulse train. This means replacing the $\dot{v}$ equation above by

$$\begin{aligned} \dot{v} = C^{-1}&\left[-I - \bar{g}_{\mathrm{K}} n^4 (v - v_{\mathrm{K}}) - \bar{g}_{\mathrm{Na}} m^3 h (v - v_{\mathrm{Na}}) - \bar{g}_{\mathrm{leak}} (v - v_{\mathrm{leak}})\right] \\ &+ A \sum_{k \in \mathbb{Z}} G(t - kT), \end{aligned} \tag{2.3}$$

where G is a "bump" function such that $\int G(t)\, dt = 1$. For simplicity, most of this paper uses the choice $G(t) = \delta(t)$; section 5.2 discusses the response of the Hodgkin–Huxley system to a pulsatile drive with

$$G(t) = \begin{cases} 1/t_0, & 0 \le t \le t_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

Mathematically, one can also choose to perturb the gating variables, but such perturbations are not entirely natural and are not considered here.

This paper uses the original Hodgkin–Huxley parameters [18]:

$$\begin{aligned} v_{\mathrm{Na}} &= -115 \text{ mV}, & \bar{g}_{\mathrm{Na}} &= 120 \text{ m}\Omega^{-1}/\text{cm}^2, \\ v_{\mathrm{K}} &= +12 \text{ mV}, & \bar{g}_{\mathrm{K}} &= 36 \text{ m}\Omega^{-1}/\text{cm}^2, \\ v_{\mathrm{leak}} &= -10.613 \text{ mV}, & \bar{g}_{\mathrm{leak}} &= 0.3 \text{ m}\Omega^{-1}/\text{cm}^2, \\ C &= 1 \ \mu\text{F}/\text{cm}^2. \end{aligned} \tag{2.5}$$

Time is measured in milliseconds and current density in $\mu\text{A}/\text{cm}^2$.

Figure 2 shows a bifurcation diagram for the unforced Hodgkin–Huxley equations. When $I \leq 6$ (not shown in figure), in particular when $I = 0$, the neuron maintains a stable rest state. This corresponds to a branch of stable fixed points continuing beyond the left edge of the diagram. A sufficiently large value of I causes a neuron to fire repeatedly, which corresponds to the creation of a limit cycle through a saddle-node bifurcation of periodic orbits. Further increasing I destabilizes the rest state through a subcritical Hopf bifurcation.

In this paper, the injected current is always set to a value near $I \approx 14$, corresponding to a steady ionic current which destabilizes the rest state. The phenomena studied here are insensitive to the precise value of I, as long as it ensures the existence of a stable limit cycle and an unstable fixed point. As explained in the introduction, these properties simplify the interpretation of numerical simulations. For this choice of I, the Jacobian of the Hodgkin–Huxley vector field (see (2.1)) at the unstable fixed point has two real eigenvalues $\{-4.97815, -0.146991\}$ in the left half plane, and a complex conjugate pair $0.0763367 \pm 0.61866i$ in the right half plane. The fixed point thus has two-dimensional stable and unstable manifolds. The Lyapunov exponents associated with the limit cycle are 0, ≈ -0.20, ≈ -2.0, and ≈ -8.3. Its period is $T_0 \approx 12.944$.

3. Main numerical results.

Lyapunov exponents provide a convenient way to characterize the asymptotic dynamics of (2.3). Let $\phi_t : \mathbb{R}^4 \to \mathbb{R}^4$ denote the flow map generated by the unforced Hodgkin–Huxley equations, T the drive period, and A the drive amplitude. The Poincaré map

$$F_T(v, m, n, h) = \phi_T(v + A, m, n, h) \tag{3.1}$$

takes a Hodgkin–Huxley state vector (v, m, n, h), applies a pulse of amplitude A to the membrane voltage, then evolves it for time T. Iterating the map F_T thus gives a stroboscopic record of the state of our pulse-driven Hodgkin–Huxley system before the arrival of each pulse. The long-term dynamical behavior of the pulse-driven Hodgkin–Huxley oscillator can be deduced from the asymptotic dynamics of F_T, which is characterized by its (largest) Lyapunov exponent $\lambda_{\max}$ [13]:

$$\begin{aligned} \lambda_{\max} < 0 &\iff F_T \text{ has sinks} &&\iff \text{kicked flow is entrained to input,} \\ \lambda_{\max} = 0 &\iff F_T \text{ is quasi-periodic} &&\iff \text{kicked flow drifts relative to input,} \\ \lambda_{\max} > 0 &\iff F_T \text{ is chaotic} &&\iff \text{kicked flow is chaotic.} \end{aligned}$$

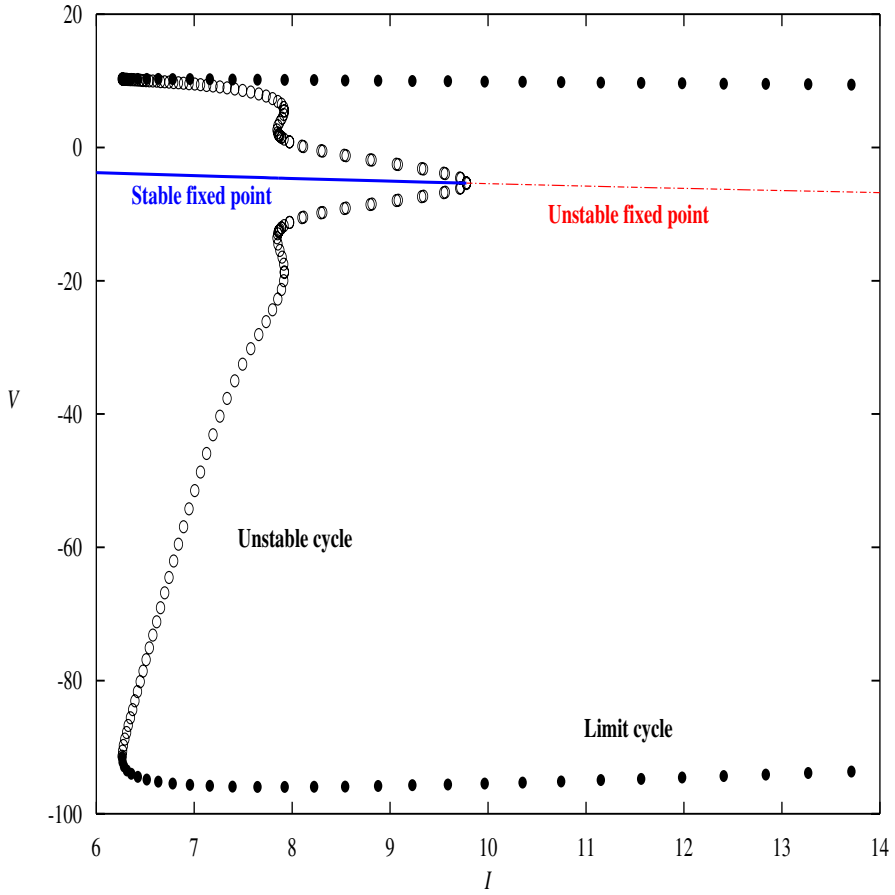


Figure 2. *The bifurcation diagram for the Hodgkin–Huxley equations as the injected current I is varied. The line in the middle marks the v coordinate of the rest state. The solid blue part is stable, while the dashed red part is unstable. Solid black dots near the top and the bottom of the figure are the maximum and minimum v values of limit cycles. Empty black circles are the maximum and minimum v values of unstable cycles. The fixed point undergoes a subcritical Hopf bifurcation as I increases. This diagram is computed using XPPAUT* [7].

(Sinks refer to stable fixed points and stable periodic orbits.) Note that, of the scenarios given in the introduction, transient chaos alone does not appear in this list: Lyapunov exponents, being long-time average quantities, cannot detect transient chaos.

Figure 3 shows the Lyapunov exponents of F_T as a function of T/T_0, where T_0 is the period of the Hodgkin–Huxley limit cycle. Different colors correspond to different values of A. The periodicity of the response as a function of T is apparent. Because the response type as a function of T is approximately identical over each period $[nT_0, (n+1)T_0]$, it makes sense to speak of the probability that a randomly chosen drive period T will elicit a particular asymptotic behavior, for example chaos. More precisely, periodicity ensures that the fraction p_n of drive periods T in $[0, nT_0]$ for which $\lambda_{\max} > 0$ converges to a well-defined limit as $n \to \infty$. Similar statements hold for $\lambda_{\max} < 0$ and $\lambda_{\max} = 0$.

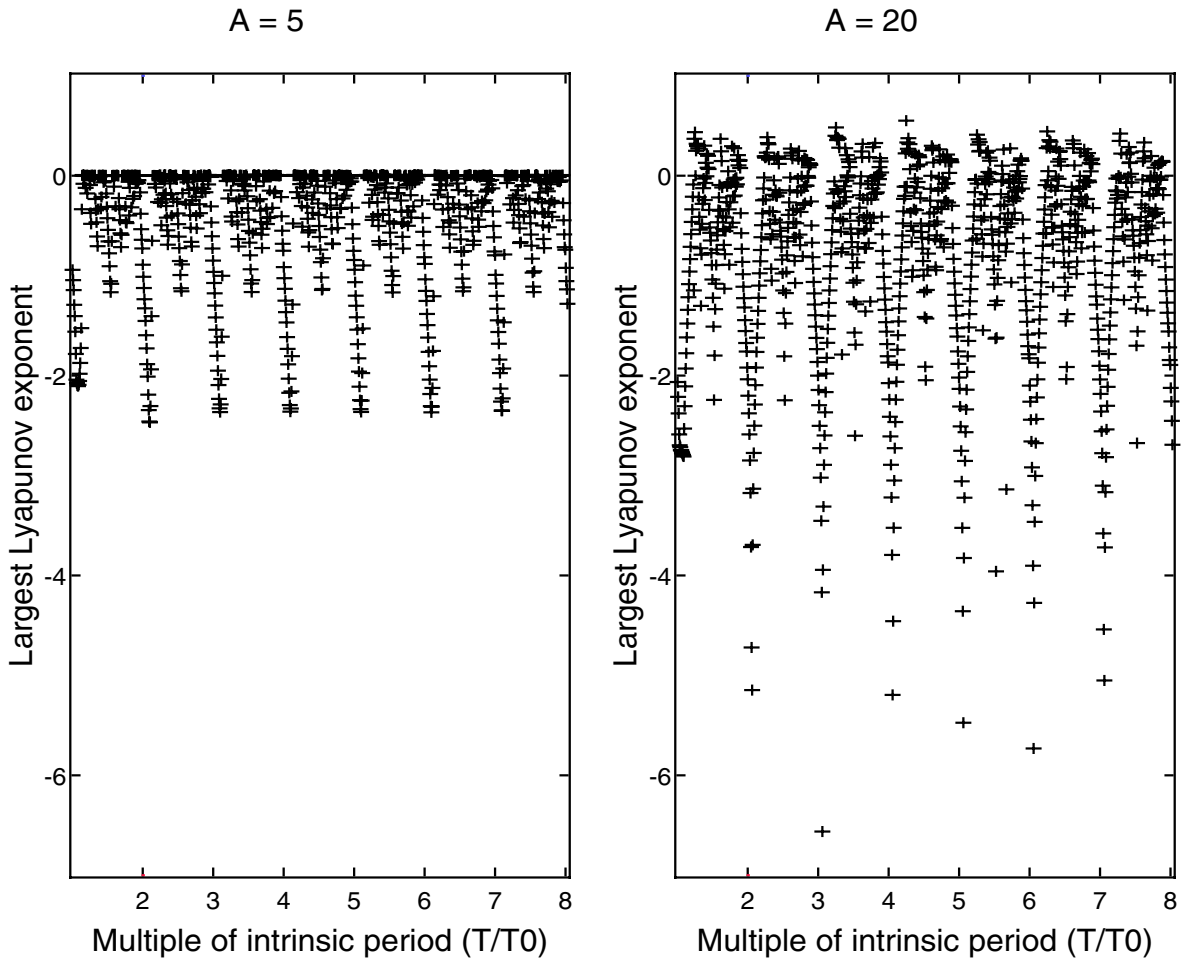


Figure 3. *Asymptotic properties of the pulse-driven flow are described by the dynamics of the time-T map F_T (see (3.1)) and its largest Lyapunov exponent $\lambda_{\max}$. Entrainment corresponds to $\lambda_{\max} < 0$, and chaos corresponds to $\lambda_{\max} > 0$. This figure shows $\lambda_{\max}$ as a function of the drive period T, with T ranging from $T_0 \approx 13$ (the intrinsic period of the unforced Hodgkin–Huxley system; see section 2) to $8 \cdot T_0 \approx 101$.* Left: *Kick amplitude is $A = 5$.* Right: *Kick amplitude is $A = 40$. Note* (i) $\lambda_{\max}(T + T_0) \approx \lambda_{\max}(T)$; (ii) *presence of both positive and negative exponents for strong kicks (right), and only zero and negative exponents for weak kicks (left); and* (iii) *the presence of more negative exponents than positive ones. See section 4 for a discussion. Lyapunov exponents are estimated by iterating F_T for* 1000 *steps and tracking the rate of growth of a tangent vector.*

Figure 4 shows these probabilities as functions of A. At $A = 10$, the probability of obtaining a positive exponent is roughly 20%, and the probability of obtaining a negative exponent is roughly 70%. Thus, if one were to pick T randomly out of an interval $[NT_0, (N+1)T_0]$ for large fixed integer N, the probability that $\lambda_{\max}(F_T) > 0$ would be about 20%. Figure 4 shows that when A is small, the most likely type of behavior is rotation-like behavior. This possibility becomes less likely as A increases. At the same time, sinks and chaos both become more likely, with sinks dominating the scene. One feature of Figure 4 specific to the pulse-driven Hodgkin–Huxley flow is that when A is large, the system prefers entrainment

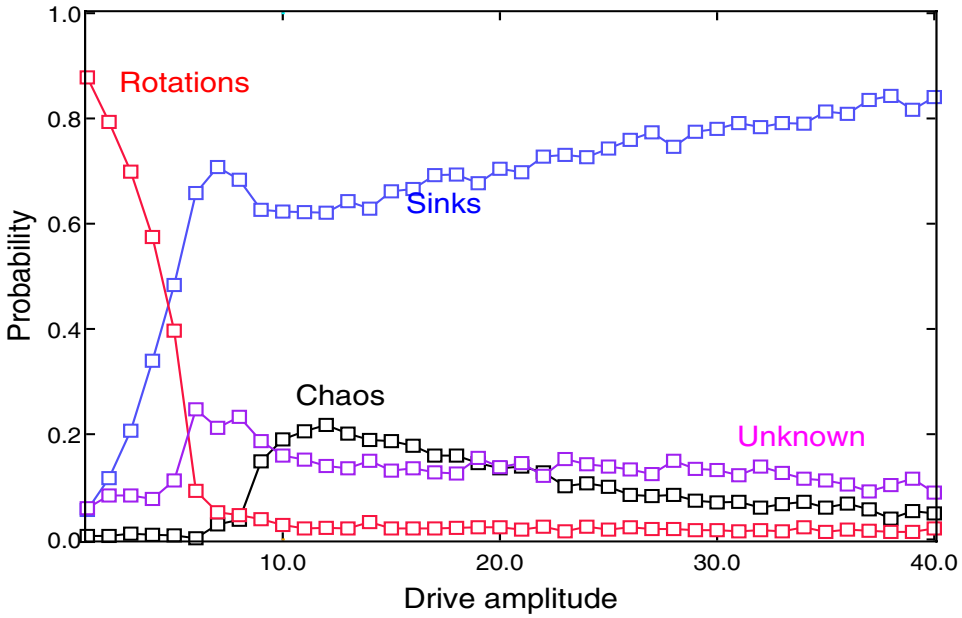


Figure 4. *The probability of different response types, as a function of the drive amplitude A. For each drive amplitude A, the fraction of drive periods $T \in [T_0, 8T_0]$ for which $\lambda_{\max}(F_T) > 0$, etc., is computed by sampling from a uniform grid in $[T_0, 8T_0]$. It is natural to equate these fractions with probabilities because the Lyapunov exponents are roughly periodic functions of T (and become more so as $T \to \infty$), as shown in Figure 3 and explained in section 4.* Empirical definitions: *Let $\hat{\lambda}$ denote the estimated Lyapunov exponent and ϵ the estimated standard error. Then "chaos" is defined as $\hat{\lambda} > 3\epsilon$, "entrainment" $\hat{\lambda} < -3\epsilon$, and "rotation" $|\hat{\lambda}| < \epsilon/3$.*

over chaos, in the sense that entrainment has higher probability. This preference is more pronounced as A increases. Note that in computing Lyapunov exponents numerically, we only have access to finite time information. In principle, this means that it is virtually impossible to distinguish persistent chaotic behavior from transient chaos caused by a "large" horseshoe (but see section 5.1).

In all numerical simulations shown in this paper, (2.1) is integrated using an adaptive integrator of Runge–Kutta–Fehlberg type, with an error tolerance of 10^{-6} (in the sup norm) [30]. The largest Lyapunov exponent $\lambda_{\max}$ of F_T is computed in a straightforward manner, by choosing a nonzero unit vector $w \in \mathbb{R}^4$ and estimating the rate of growth of $||(DF_T)^n w||$. The matrix-vector product $(DF_T)^n w$ is easily computed via the variational equations $\dot{x} = H(x)$, $\dot{w} = DH(x) \cdot w$ for the Hodgkin–Huxley vector field H. (DH is the Jacobian matrix of H; see [10] for details.)

4. Discussion.

4.1. Response to a single pulse: Phase-resetting curves. This section reviews phase-resetting curves. See Winfree [38], Glass and Mackey [11], and Brown, Moehlis, and Holmes [4] for more details and applications, and Guckenheimer and Holmes [13] for background

information on dynamical systems theory. See [7, 8, 9, 37] for further discussions of phase-resetting curves.

Let $\phi_t : \mathbb{R}^n \to \mathbb{R}^n$ be a flow generated by a smooth vector field with a hyperbolic limit cycle γ. Such a limit cycle represents a stable nonlinear oscillator. The basin of attraction of γ is denoted $B(\gamma)$. The hyperbolicity of γ guarantees that points in $B(\gamma)$ converge to γ exponentially fast. (It is convenient to use γ to refer to both the trajectory $\gamma : \mathbb{R} \to \mathbb{R}^n$ and the invariant point set it defines.) An impulsive perturbation ("kick") to the nonlinear oscillator can be defined by specifying a kick amplitude A and a kick direction $\hat{K} : \mathbb{R}^n \to \mathbb{R}^n$ and defining a family of *kick maps*

$$K_A(x) = x + A \cdot \hat{K}(x), \tag{4.1}$$

so that kicks send each point $x \in \mathbb{R}^n$ to $K_A(x)$. For what follows, K_A should be smooth and satisfy $K_A(B(\gamma)) \subset B(\gamma)$.

The Hodgkin–Huxley system with the value of I given in section 2 is a nonlinear oscillator whose basin $B(\gamma)$ is an open subset of $\mathbb{R}^4$. The kick map corresponding to an instantaneous voltage spike is simply $K_A(v, m, n, h) = (v + A, m, n, h)$. As in section 3, it is convenient to introduce the time-T map

$$F_T = \phi_T \circ K_A, \tag{4.2}$$

where $\circ$ denotes function composition. Iterating F_T gives a stroboscopic record of the system state before the arrival of each kick, and thus describes the long-time dynamics of the flow ϕ_t under repeated, T-periodic kicks.

Because the phase dimension n may be large, the dynamics of $F_T : \mathbb{R}^n \to \mathbb{R}^n$ may be difficult to analyze. Winfree observed that every point near the limit cycle γ must converge to γ as $t \to \infty$, and so the flow near γ is dominated by the rotational motion along γ. Thus, one can reduce the dimension of the phase space from n to 1, at least heuristically. To do this, first define the phase function $\theta : \gamma \to [0, T_0)$ by fixing a reference point $x_0 \in \gamma$ and requiring that $\phi_{\theta(x)}(x_0) = x$ for all $x \in \gamma$. By construction, θ satisfies $\frac{d}{dt}(\theta(\gamma(t))) = 1$, $0 \le t < T_0$. The function θ can be extended to a function $\theta : B(\gamma) \to [0, T_0)$ by projecting along strong-stable manifolds:[3] if y is a point in the basin of γ, then $\theta(y)$ is defined to be $\theta(x)$, where x is the unique point such that $y \in W_{\mathrm{ss}}(x)$. This definition of phase preserves the property that $\frac{d}{dt}(\theta(\phi_t(x))) = 1$.

Consider the limit [12]

$$\bar{F}_T = \lim_{n \to +\infty} F_{T+nT_0}. \tag{4.4}$$

The map $\bar{F}_T$ is well defined on the basin of γ and retracts the basin onto γ; i.e., $\bar{F}_T(x) \in \gamma$ for all $x \in B(\gamma)$. Thus, $\bar{F}_T$ induces an interval map $f_T : [0, T_0) \to [0, T_0)$ which, given the current

[3] The *strong-stable manifold* $W_{\mathrm{ss}}(x)$ of $x \in \gamma$ is the set

$$W_{\mathrm{ss}}(x) = \left\{ y \in \mathbb{R}^n : \lim_{n \in \mathbb{Z}, n \to +\infty} \phi_{nT_0}(y) \to x \right\}. \tag{4.3}$$

When the vector field generating ϕ_t is smooth, the strong-stable manifolds are (locally) smooth submanifolds of $\mathbb{R}^n$. The strong-stable linear subspace $E_{\mathrm{ss}}(x)$ is the tangent space of $W_{\mathrm{ss}}(x)$ at x. See [12, 13].

phase of the system, yields the new phase after kicking and evolving the system for time T. That is, $f_T(\theta(x)) = \theta(F_T(x))$ for all $x \in B(\gamma)$.

The map f_T is the *phase-resetting curve*[4] or, more precisely, the finite phase-resetting curve (*infinitesimal phase-resetting curves* [4, 7] are not needed here). By construction, it has the property that

$$f_{T+\delta}(t) = f_T(t) + \delta \pmod{T_0}. \tag{4.5}$$

Thus, the family of maps $\{f_T\}$ is periodic in T.

Periodicity in drive period T. The approximate periodicity of $\lambda_{\max}(F_T)$ seen in Figure 3 is easy to understand heuristically: kicking the oscillator every T seconds and kicking it every $T+T_0$ seconds should yield the same asymptotic response because the oscillator simply traverses γ at frequency $1/T_0$ between kicks. One can restate this using phase-resetting curves: if the drive period T is sufficiently large and $\theta(x_0) = t_0$, then the f_T-orbit $(t_0, f_T(t_0), f_T^2(t_0), \dots)$ should closely follow the phases $(\theta(x_0), \theta(F_T(x_0)), \theta(F_T^2(x_0)), \dots)$ of the corresponding F_T-orbit. Since $f_{T+T_0} = f_T$, this suggests that $\lambda_{\max}(F_{T+T_0}) \approx \lambda_{\max}(F_T)$.

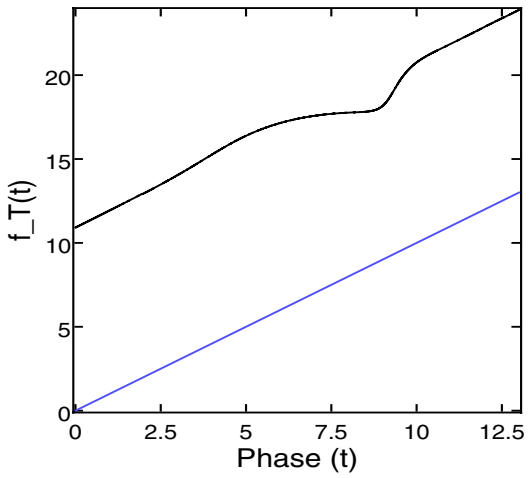
Preference for entrainment. Figure 5 shows phase-resetting curves for the Hodgkin–Huxley equations for various values of drive period T and drive amplitude A. For sufficiently small values of A, the phase-resetting curves are circle diffeomorphisms: either there are sinks (i.e., stable fixed points or stable periodic orbits), or the map is conjugate to a rotation on a circle and the response of the kicked oscillator drifts relative to the periodic drive. As A increases, the graph of f_T rather quickly folds over and acquires critical points. A striking feature of the graphs in Figure 5 is the "plateau," a phase interval over which f_T varies very slowly. Another striking feature is the "kink" around $\theta \approx 9.8$. These features are discussed in more depth in section 5. For now, notice that the plateau provides a simple mechanism for creating sinks: changing the kick period T shifts the graph of f_T vertically. Whenever the graph intersects the diagonal with a derivative $|f_T'| < 1$, then a stable fixed point is created.

This mechanism can be used to verify the results of Figure 4: compute the graph of the first return map of f_T to an interval around the plateau, then shift the graph vertically using a number of different values of T and estimate the fraction of T's for which f_T has a stable fixed point (see Figure 6). Table 1 shows the results. For $A = 10$, the 58% probability of sinks corresponds fairly closely with Figure 4. It is unclear whether the ambiguous exponents in Figure 4 really represent positive or negative Lyapunov exponents. If a significant fraction of the ambiguous exponents are really negative, then they must come from small sinks.

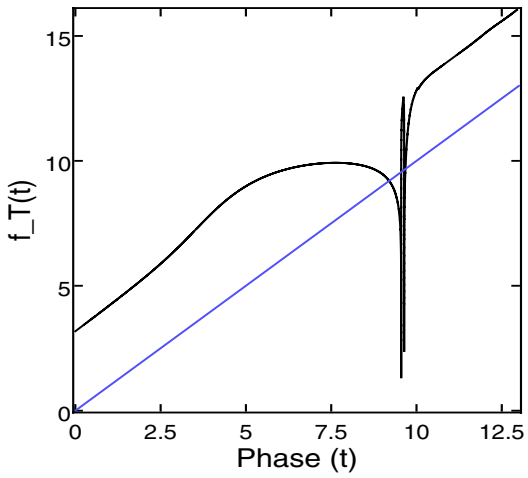
Note on numerics. Phase-resetting curves are computed here using a variation of the Ermentrout–Kopell adjoint method [4, 8]. The method is described in the appendix. A systematic comparison of this method to existing methods for computing phase-resetting curves is beyond the scope of the present paper and will be presented elsewhere.

4.2. Response to repeated pulses: Wang–Young theory. Phase-resetting curves provide simple, intuitive explanations for many dynamical properties of pulse-driven nonlinear oscillators. For our spiking Hodgkin–Huxley oscillator, explicitly computed phase-resetting

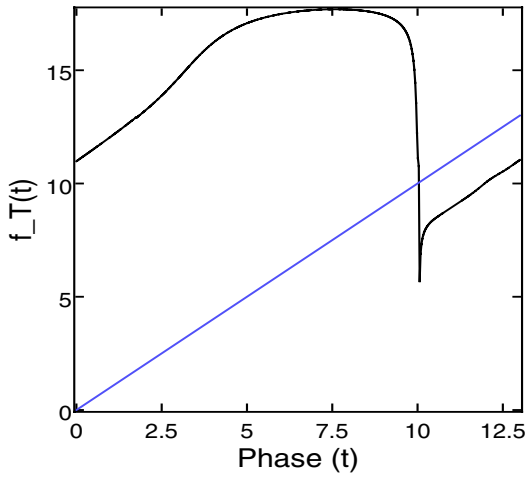
[4] Wang and Young refer to phase-resetting curves as *singular limits*. Phase-resetting curves are also sometimes called *phase transition curves* [11].



(a) $A = 5$



(b) $A = 10$



(c) $A = 20$

Figure 5. *The graph of the lift $\tilde{f}_T$ of f_T, i.e., the unique continuous map $\mathbb{R} \to \mathbb{R}$ such that $\tilde{f}_T = f_T$ on $[0, T_0)$ and $\tilde{f}_T(t + T_0) = \tilde{f}_T(t) \pmod{T_0}$, for the pulse-driven Hodgkin–Huxley equations. Drive amplitudes are given under each panel. The precise value of the drive period T is not so important; varying T shifts the graph vertically (see (4.5)). Note that f_T has winding number 1 for $A = 5$ and $A = 10$, and has winding number 0 for $A = 20$. The numerical data suggests that the degree changes around $A \approx 13.589$; the precise geometric mechanism is not clear. Note that phase ranges from 0 to the intrinsic period $T_0 \approx 12.9$ of the Hodgkin–Huxley limit cycle.*

curves show why our pulse-driven neuron prefers entrainment over chaos. In order to infer asymptotic behavior, there needs to be a correspondence between the orbits of f_T and F_T, and the phase of $x \in B(\gamma)$ generally does not determine the phase of $F_T(x)$: it may only

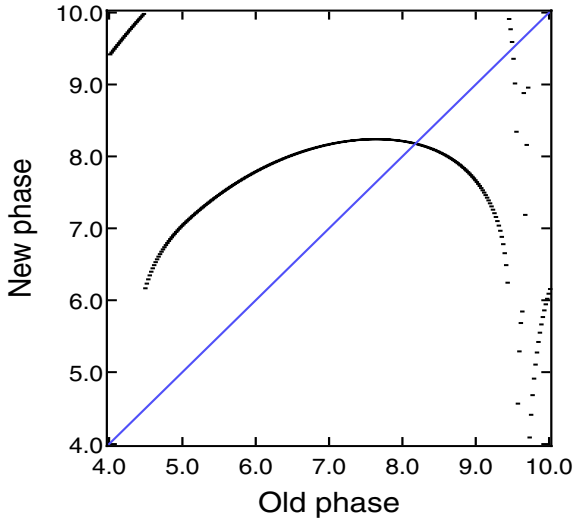


Figure 6. *The first return map R_{f_T} to the interval $[4, 10]$ (chosen to enclose the plateau), for $A = 10$ and $T = 17.6$. The blue line marks the diagonal.*

Table 1

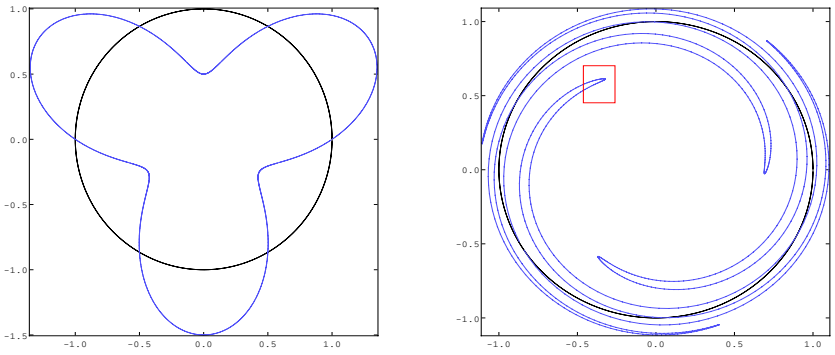
Estimates of the probability of obtaining sinks near the plateau, as a function of A. The data for this table is computed by trying about 40 values of T for each choice of A and examining the graph of the first return map to the interval $[4, 10]$ (chosen to coincide with the "plateau") and its intersection(s) with the diagonal.

Drive amplitude A	Prob. of sink near plateau	Prob. of $\lambda_{\max} < 0$
5	41%	48%
10	58%	62%
20	68%	70%
30	76%	78%

do so approximately for a finite number of iterates. When $\lambda_{\max}(f_T) < 0$, this is enough to show that f_T orbits indeed approximate the phases of F_T. Inferring chaotic behavior for F_T from f_T is far more difficult. Wang–Young theory provides a mathematical framework for inferring chaotic behavior using phase-resetting curves and, in addition, explains why chaotic phenomena (and all the other scenarios) are prevalent.

Shear is an important ingredient of Wang–Young theory. Let γ be the limit cycle which represents the unforced nonlinear oscillator. Near γ, the dynamics follows the periodic rotational motion on γ. Shear refers to the presence of an angular velocity *gradient* around γ: the stronger the shear, the sharper the angular velocity changes at γ. In two dimensions, this means the flow runs much faster on one side of γ than on the other; in the presence of strong shear, strong stable manifolds tend to become more nearly tangent to γ.

Shear and its interaction with kicks are illustrated in a simple model in Figure 7. In the presence of strong shear, most ways of kicking the oscillator which take advantage of shear (e.g., kicks which do *not* take $x \in \gamma$ too close to the strong-stable manifold $W_{ss}(x)$) will cause segments of the limit cycle to stretch and fold as they fall back toward γ. The phase space



(a) (b)

Figure 7. *A simple model (see (4.6)) which illustrates the effect of shear. In* (a)*, a large number of initial conditions are placed around a limit cycle (the black circle), and a radial kick is applied to each point. The blue curve shows the resulting positions of each initial condition. In* (b)*, the kicked points are allowed to flow. The red box contains one of the turning points (see section 4.2).*

stretching caused by shear manifests itself in phase intervals over which $|f_T'| > 1$. This expansion is conducive to chaotic behavior. However, shear also creates regions of contraction around "turning points," an example of which is highlighted in Figure 7. Such turning points correspond to critical points on the phase-resetting curve and can easily counteract the expansion needed for a positive Lyapunov exponent. The competition between expansion and contraction is the main source of difficulty in proving $\lambda_{\max}(F_T) > 0$.

To infer the existence of parameters for which the time-T map F_T is fully chaotic, Wang and Young use results from their previous work on strange attractors with one expanding direction and a roughly toroidal geometry [33, 36]. Their main result gives conditions under which there must be T for which $\lambda_{\max}(F_T) > 0$. Furthermore, one can find such "chaotic parameters" near "nice" values of T for which f_T has a positive Lyapunov exponent. Applying the general theory to kicked oscillators requires checking certain geometric conditions. This has been done for a few concrete classes of models [27, 34, 35]. To illustrate the consequences of the theorems, consider the simple mechanical system [34]

$$\ddot{\theta}(t) + \lambda\dot{\theta}(t) = \mu + A \cdot \hat{K}(\theta(t)) \sum_{n\in\mathbb{Z}} \delta(t - nT). \tag{4.6}$$

This model was first studied by Zaslavsky, who discovered that this simple system can exhibit fully chaotic behavior [40]. For (4.6), it can be shown that the full range of scenarios enumerated in the introduction take place and that they are all prevalent. More precisely, Wang and Young prove (see Theorems 1–3 in [34]) that the following hold:

1. *Invariant curve and weak kicks.* When the drive amplitude A is sufficiently small (which is equivalent to having a large enough contraction rate λ), there exists a simple closed curve $\tilde{\gamma}$ to which all orbits of F_T converge and which is invariant under F_T. Moreover, we have the following dichotomy:

(a) *Quasi-periodic attractors.* There exists a set of T of positive Lebesgue measure for which F_T is topologically conjugate to an irrational rotation. In this case, F_T is uniquely ergodic on $\tilde{\gamma}$.
(b) *Gradient-like dynamics.* There exists an open set of T such that F_T has a finite number of periodic sinks and saddles on $\tilde{\gamma}$, and every orbit converges to one of these periodic orbits.

2. *Gradient-like dynamics without an invariant curve.* As A increases (or λ decreases), the invariant curve $\tilde{\gamma}$ breaks up. Nevertheless, there continues to be an invariant set (no longer a simple closed curve) on which gradient-like dynamics persists.
3. *Transient chaos.* For even larger A or smaller λ, Smale horseshoes (see [13]) will form. Horseshoes can coexist with sinks and saddles, creating transient chaos.
4. *Chaos.* In the presence of sufficiently strong shear, there exists a positive measure set of drive periods T for which F_T is fully chaotic in the sense that it possesses (i) a strange attractor with a positive Lyapunov exponent, (ii) at least one and at most finitely many ergodic SRB measures[5] with no zero Lyapunov exponents, (iii) a central limit theorem, and (iv) exponential decay of correlations if a power F_T^N is mixing for some SRB measure ν.

Note that this list of (fairly well understood) scenarios may not be exhaustive. Other scenarios or combinations of scenarios are not excluded by the theory. Also, the kicks in (4.6) are purely radial. This is not strictly necessary; any kick map which takes advantage of shear will do. See [34] for precise conditions and proofs.

5. Further results.

5.1. More on the Hodgkin–Huxley phase-resetting curve.

Plateau. The plateau in the phase-resetting curve for our pulse-driven Hodgkin–Huxley model (see Figure 5) corresponds to a segment $\bar{\gamma}$ of the limit cycle γ, which becomes nearly parallel to a strong-stable manifold after receiving a kick. This can be seen by examining the factors which contribute to the derivative f_T' and which can potentially cause f_T' to become small over a relatively large phase interval. This can be checked by writing f_T as a composition of other functions and differentiating.

Let $\gamma : \mathbb{R} \to \mathbb{R}^4$ denote the limit cycle trajectory. If we choose $\gamma(0)$ so that $\theta(\gamma(0)) = 0$, then $\theta(\gamma(t)) = t$ for all $t \in [0, T_0)$, and

$$\begin{aligned} f_T &= \theta \circ F_T \circ \gamma \\ &= \theta \circ \phi_T \circ K_A \circ \gamma. \end{aligned}$$

Changing T does not affect f_T', so we can set $T = 0$. Let $f = f_0$. Then $f = \theta \circ K_A \circ \gamma$, and the chain rule gives

$$f' = (D\theta \circ K_A \circ \gamma) \cdot (DK_A \circ \gamma) \cdot \dot{\gamma}. \tag{5.1}$$

[5]SRB measures are natural invariant measures for dissipative dynamical systems. They characterize the asymptotic behavior of a Lebesgue-positive measure set of initial conditions and have a number of nice mathematical properties. See Young [39] for an introduction.

However, $K_A(v,m,n,h) = (v + A, m, n, h)$, and thus its Jacobian DK_A is the identity matrix, and for all $t \in [0, T_0)$,

$$\begin{aligned} f'(t) &= D\theta(K_A(\gamma(t))) \cdot \dot\gamma(t) \\ &= |D\theta(K_A(\gamma(t)))| \cdot |\dot\gamma(t)| \cdot \cos(\text{angle}(D\theta(K_A(\gamma(t))), \dot\gamma(t))) \\ &= |D\theta(K_A(\gamma(t)))| \cdot |\dot\gamma(t)| \cdot \sin(\measuredangle(t)), \end{aligned} \tag{5.2}$$

where $\measuredangle(t)$ is the angle between $\dot\gamma(t)$ and the strong-stable manifold at $K_A(\gamma(t))$. The last step uses the fact that the phase function $\theta : B(\gamma) \to [0, T_0)$ is constant on strong-stable manifolds (see section 4). This implies that the gradient $D\theta(x)$ is everywhere orthogonal to $W_{ss}(x^*)$, where x^* is the unique point in γ having the same phase as x. The factors in (5.2) thus have simple, geometric meaning: $\gamma(t)$ is a point on the limit cycle γ, and $|\dot\gamma(t)|$ is the speed of the limit cycle at that point; $\measuredangle(t)$ is the angle between $\dot\gamma(t)$ and the strong-stable manifold at $K_A(\gamma(t))$; and $|D\theta|$ measures the rate at which the phase is changing at $K_A(\gamma(t))$.

Figure 8 shows f' alongside the three factors in (5.2). The figure shows that the plateau, where f'_T becomes nearly 0 over a long phase interval, coincides with the near-vanishing of $\measuredangle(t)$. The other factors of f' stay nearly constant over this interval. Thus, there is a segment $\bar\gamma$ of the limit cycle γ, corresponding to the phase interval where $\measuredangle(t)$ is small, such that $K_A(\bar\gamma)$ is nearly tangent to a strong-stable manifold. That the segment $\bar\gamma$, which may be small as a subset of $\mathbb{R}^4$, corresponds to a large phase interval is due to the relatively slow speed of the limit cycle near $\bar\gamma$.

What this argument does not explain is the robustness of this tangency (equivalently, the robustness of the plateau) as the drive amplitude A increases (see Figure 5). This requires a detailed analysis of the geometry (in $\mathbb{R}^4$!) of the strong-stable manifolds (see section 6).

Numerical computation of $D\theta$. Figure 8 requires the numerical computation of the gradient $D\theta(x)$ for $x \in B(\gamma)$. This can be done as follows.

Fix $x \in B(\gamma)$ and consider $\phi^n_{T_0}(x)$. Clearly, the limit $\lim_{n\to\infty} \phi^n_{T_0}(x) = x^*$ exists and has the property that $x^* \in \gamma$, $\theta(x^*) = \theta(x)$, and $x \in W_{ss}(x^*)$. Set $\pi_{ss}(x) = x^*$. Then π_{ss} projects $B(\gamma)$ onto γ and is the identity map on γ. Furthermore, the nullspace of the Jacobian matrix $D\pi_{ss}(x)$ of π_{ss} is the tangent to $W_{ss}(\pi_{ss}(x))$ at x, by construction.

To compute $D\theta(x)$, the foregoing discussion suggests that we compute $D\phi^n_{T_0}(x)$ for some large finite n. For any finite n, the singular values of $D\phi^n_{T_0}(x)$ consist of a dominant singular value σ_1 and three nearly zero singular values σ_2, σ_3, and σ_4. The $\sigma_i \to 0$ as $n \to \infty$ for $i = 2, 3, 4$. Denote the left and right singular vectors associated with σ_1 by u and v. It is easy to check that the right eigenvector v is orthogonal to the null space of $D\pi_{ss}(x)$ and hence tangent to $D\theta$.

This computation requires a relatively accurate estimate of the intrinsic period T_0 of the limit cycle γ, without which the computation would not converge. This paper adopts the following strategy: instead of estimating T_0 just once and reusing its value, solve the system of 24 equations

$$\dot x_1 = H(x_1), \quad \dot x_2 = H(x_2), \quad \dot J = DH(x_2) \cdot J \tag{5.3}$$

with initial conditions

$$x_1(0) \in \gamma, \quad x_2(0) = x, \quad J(0) = \mathrm{Id}_{4\times 4}, \tag{5.4}$$

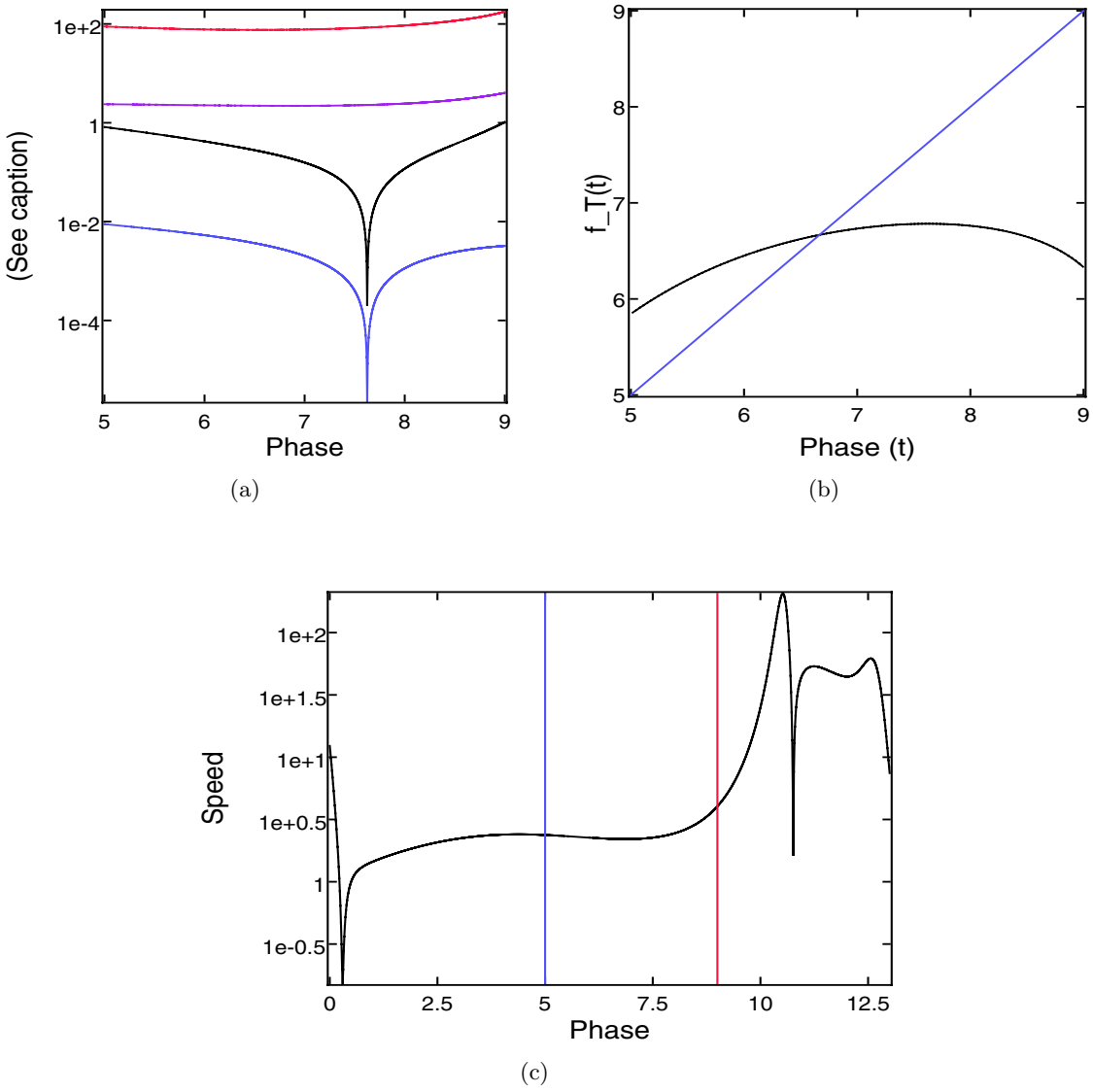
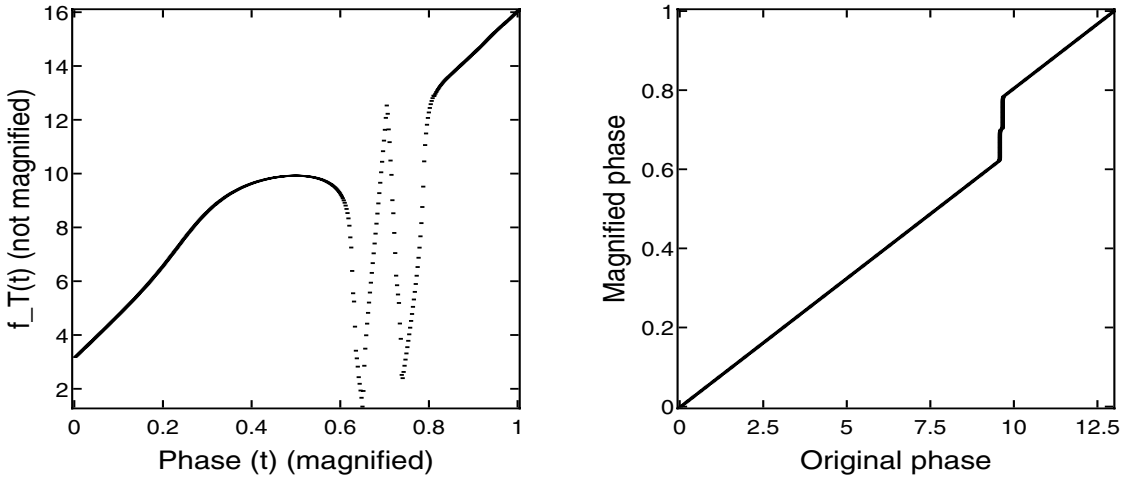


Figure 8. *The origin of the plateau: in* (a)*, the curves are* (i) black: $|f'(t)|$*;* (ii) blue: $|\sin(\angle(t))|$*;* (iii) red: $|D\theta(K_A(t))|$*;* (iv) purple: $|\dot{\gamma}(t)|$*. In* (b)*, the graph of f_T near the plateau is shown for reference. Here, $A = 10$.* (c) *The speed of the Hodgkin–Huxley flow along γ. The vertical lines mark the interval $[5, 9]$, which is part of the plateau.*

where H is the Hodgkin–Huxley flow field and x is the point at which we would like to evaluate $D\theta$. Note that $x_1, x_2 \in \mathbb{R}^4$ and $J \in \mathbb{R}^{4\times 4}$. The solution of these equations then gives $x_2(t) = \phi_t(x)$ and $J(t) = D\phi_t(x)$. The reference trajectory x_1 is used only to count the number of periods which have elapsed, and the trajectory (x_2, J) is used to compute $D\pi_{\text{ss}}(x)$. This procedure works fairly well in practice.



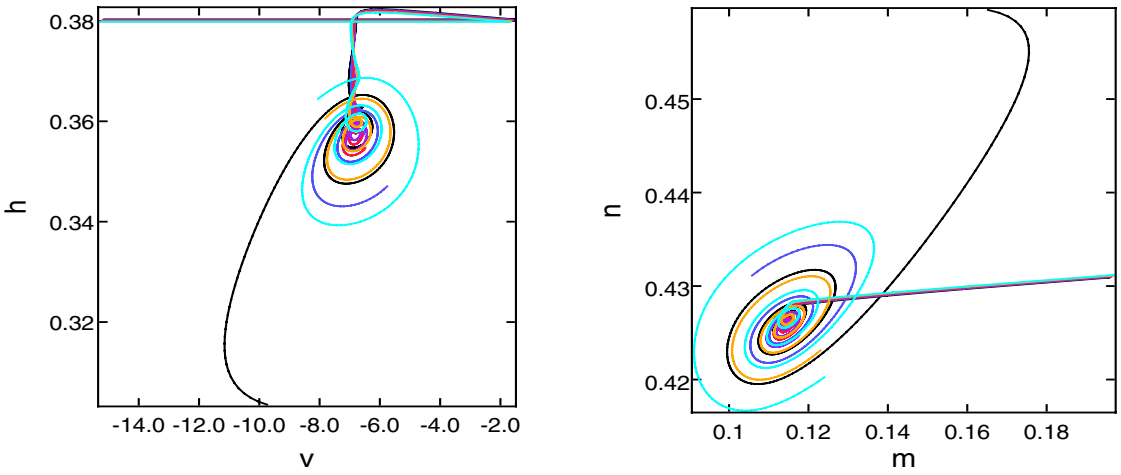
(a) (b)

Figure 9. (a) *The graph of* $f_T \circ g^{-1}$ *with drive amplitude* $A = 10$*, with the abscissa shown in a new coordinate system* $\theta' = g(\theta)$ *to magnify the region around the "kink." No interpolation is done in this figure: only actually computed points are shown.* (b) *The graph of the coordinate transformation* g*. The map* g *is generated automatically by the simple adaptive algorithm described in the appendix.*

Kink. It is natural to ask whether f_T (see Figure 5), for $A = 10$ or $A = 20$, is discontinuous around the kink. A discontinuity indicates that there are points in a neighborhood of γ which can be kicked outside the basin of γ. This is not likely the case: Figure 9 shows a magnified view of the phase-resetting curve near the kink; the graph does not include any numerical interpolation. The figure is obtained by fixing a small parameter $\delta > 0$ and adaptively refining the grid $\{\theta_n\}$ on which the phase-resetting curve is evaluated until $|f_T(\theta_{n+1}) - f_T(\theta_n)| \le \delta$. In Figure 9, δ is set to 0.1. The adaptive procedure (see the appendix) continues to converge for smaller values of δ.

Figure 10 suggests an explanation for the kink: that it is likely caused by a segment of γ being kicked near the stable manifold of the unstable fixed point. This would cause the segment to wind around the stable manifold and eventually spiral away from the fixed point. (Recall that the two unstable eigenvalues of the fixed point form a complex conjugate pair.) In the process the kicked segment spreads apart, and its subsets pick up different amounts of time delays. However, because the Hodgkin–Huxley phase space is four-dimensional, Figure 10 cannot give a reliable picture of the dynamics: projecting onto two dimensions loses too much information.

The scenario sketched above predicts that there exist a critical kick amplitude A_{crit} at which $K_A(\gamma)$ intersects the stable manifold of the fixed point. (There may be more than one intersection, and more than one value of A which cause intersections.) As $A \to A_{\text{crit}}$, the phase-resetting curve should start winding around S^1 more and more. This can be numerically tested: an estimate of A_{crit} is computed using the Nelder–Meade algorithm [30] to minimize the closest distance of a trajectory to the fixed point. This yields a critical value $A_{\text{crit}} \approx 13.58953\ldots$.



(a) (b)

Figure 10. *A suggestive picture: in* (a)*, a segment of γ, starting in the upper left corner of the picture, is kicked straight across to the upper right corner. It then follows the flow toward the fixed point for some time before spiraling away. The overall direction of motion is top to bottom.* (b) *Another view of the approach to the fixed point. The overall direction of motion here is right to left. The kick amplitude is $A = 13.589$.*

When $A = A_{\text{crit}}$ exactly, f_T should wind around infinitely many times and possess a singularity near the location(s) of intersection. For $A \neq A_{\text{crit}}$, f_T remains smooth, but as $A \to A_{\text{crit}}$, f_T should develop a singularity and blow up. See Figure 11.

Horseshoes and transient chaos. Wang–Young theory also guarantees the existence of T's for which F_T exhibits transient chaos; i.e., F_T possesses a Smale horseshoe [13] together with a sink. The coexistence of a horseshoe with a sink has the following effect on the dynamics: almost every F_T-orbit would eventually fall into a sink, but an orbit which wanders near a horseshoe would dance around unpredictably for a finite number of iterations. Two nearby orbits which enter the vicinity of a horseshoe can emerge widely separated and fall into the sink out of phase (unless the sink happens to be a fixed point). In terms of time series data, this kind of behavior can be recognized by looking at *pairs* of trajectories and finding that they chaotically "flutter" about before settling down into a steady periodic motion, likely out of phase.

In contrast to entrainment and chaos, transient chaotic behavior is difficult to observe in the pulse-driven Hodgkin–Huxley system. This is because the most likely place to find a horseshoe is near the kink, where the expansion is so strong that most trajectories escape very quickly. Nevertheless, it is possible to find indirect evidence for horseshoes in the pulse-driven Hodgkin–Huxley model. To do so, one looks for an interval $I \subset [0, T_0)$ such that $f_T(I)$ gets mapped completely across I at least twice. It is easy, for example, to find a "small" horseshoe around the kink in the phase-resetting curve; see Figure 12. The phase interval I tells us the rough location of a horseshoe for F_T.

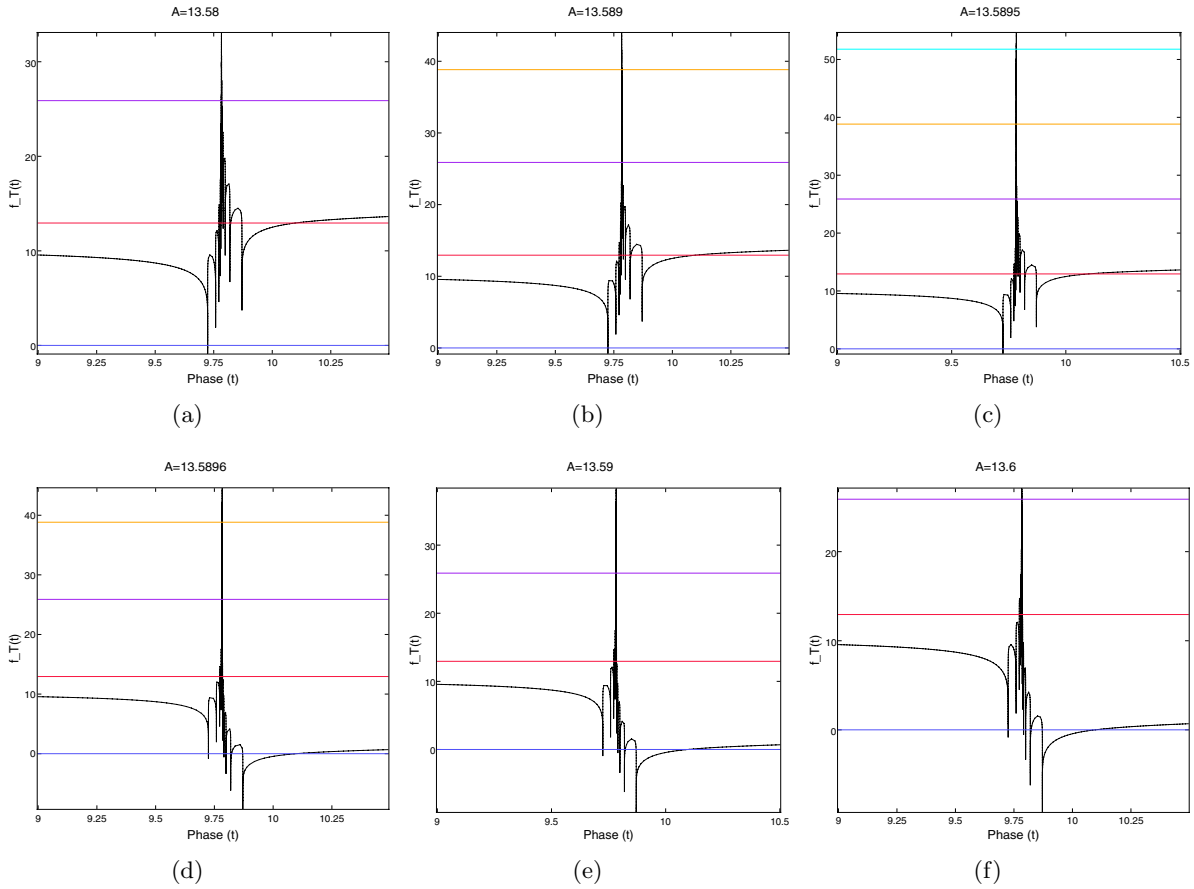


Figure 11. *Graphs of f_T for kick amplitudes A approaching A_{crit} from below* (a)–(c) *and above* (d)–(f). *The horizontal lines mark integral multiples of T_0.*

To go from such an interval I to a horseshoe for the full map F_T, it is necessary to (i) blow up the corresponding segment of γ to form an open set $U \subset \mathbb{R}^4$ such that $F_T(U)$ intersects U at least twice, and such that the intersection stretches all the way across U in the unstable direction (along γ); and (ii) find invariant cones [13]. This can be done in a straightforward manner and is not discussed further here.

5.2. Miscellany.

Decay of correlations. Wang–Young theory predicts that when the dynamics of a pulse-driven oscillator is chaotic and there is a unique SRB measure, time correlations (more precisely time autocovariance functions) decay exponentially fast in time. Figure 13 shows the time autocovariance function

$$C_{vv}(n) = \int (v \circ F_T^n) \cdot v \, d\mu - \left(\int v \, d\mu \right)^2 \tag{5.5}$$

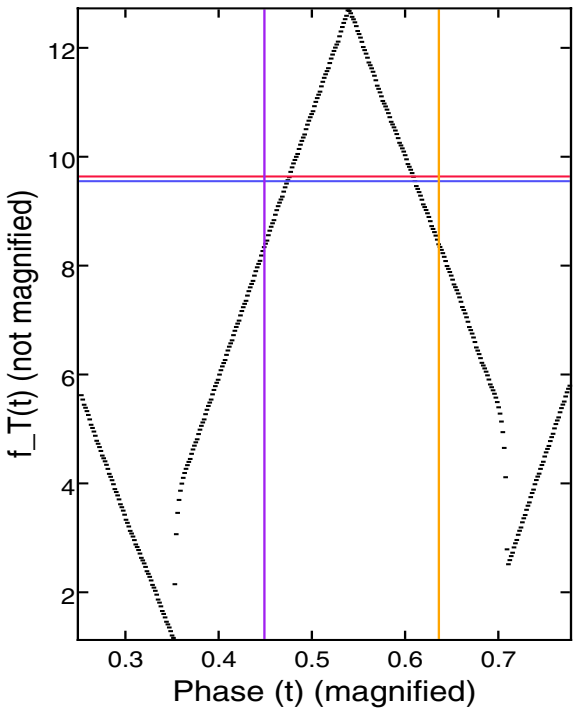
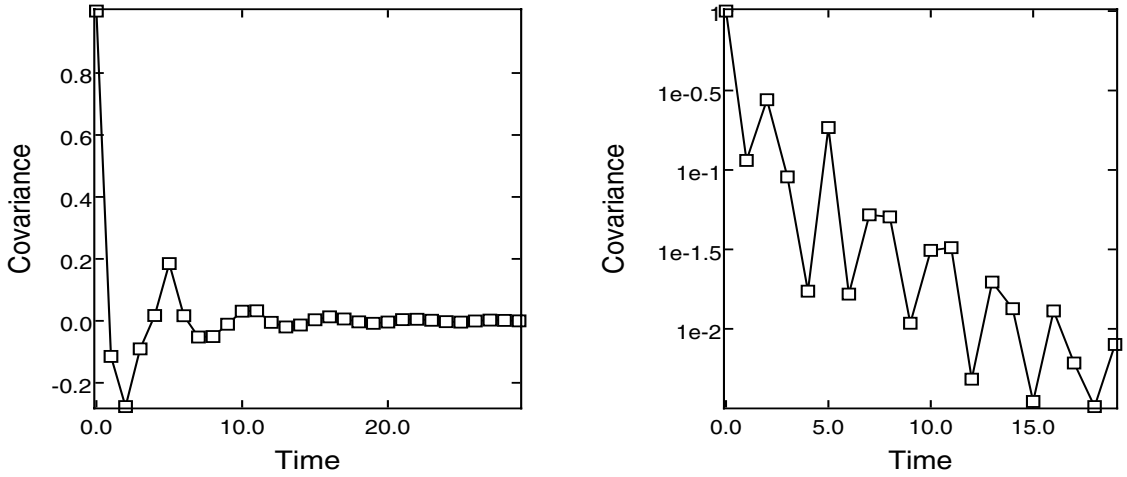


Figure 12. *The phase-resetting map and an interval I (marked by the straight lines) which maps across itself twice. The abscissa (but not the ordinate) is shown in transformed coordinates. Here, $A = 10$ and $T = 81$. This is a "small" horseshoe: because the derivative of f_T is so large there (on the order of $10^3 \sim 10^4$), most numerically computed orbits escape the horseshoe after a few iterations.*

for the voltage variable v (μ is an ergodic invariant measure). While it clearly decays as $n \to \infty$ and thus provides evidence that the invariant measure is mixing, the data is not sufficient to confirm that the decay is exponential.

Response to finite-duration pulses. A natural variation on the numerical experiments of previous sections is to replace instantaneous impulses with finite-duration pulses. Heuristically, if the pulse duration t_0 is less than the fastest of the intrinsic time scales of γ, the resulting response should be essentially the same as the response to instantaneous impulses. With $I \approx 14$, these time scales are 12.944 (= the period), 5.1, 0.50, and 0.12 (corresponding to the negative Lyapunov exponents).

Figure 14 summarizes the numerical results for $t_0 = 0.05$ (shorter than all time scales), 0.3 (shorter than all but one time scale), 2.75 (shorter than all but two fastest time scales), and 9.0 (very slow, not really pulsatile in any sense of the word). These graphs should be compared to Figure 4. The pulse amplitude is adjusted so that the *total* amount of charge delivered is the same as for an impulse of amplitude A. Interestingly enough, the behaviors seen earlier are quite robust and disappear only when $t_0 = 9$. These results suggest that the contracting directions do not mix very much over γ, and only the slowest contracting time scale participates in the production of chaotic behavior.



(a) Linear (b) Semilog

Figure 13. *Normalized autocorrelation function $C_{vv}(n)$ for the v variable of the time-T map F_T. This graph indicates that the dynamics is mixing, but the data is insufficient to confirm that the system is exponentially mixing.*

6. Outlook. The results reported here show that the pulse-driven Hodgkin–Huxley model (2.3) responds to low-frequency (relative to the intrinsic period T_0 of the spiking neuron) periodic impulse forcing in a wide range of ways. Depending on the drive period and drive amplitude, the response can range from entrainment to fully chaotic behavior. This is consistent with the predictions of Wang–Young theory. Furthermore, as shown in section 4, it is possible to explain some phenomena specific to our pulse-driven Hodgkin–Huxley oscillator in terms of special features of the phase-resetting curve and provide a partial understanding of the source of these features.

Some interesting directions for future work include the following.

Random kick times. The shape of the Hodgkin–Huxley phase-resetting curve suggests that if one were to drive a Hodgkin–Huxley neuron using a pulse train with *random* kick times, the resulting random dynamical system could have a negative Lyapunov exponent. This is because the phase-resetting curve moves up and down from kick to kick, and for any kick time distribution which is sufficiently uniform (e.g., an exponential distribution), the probability that the plateau intersects the diagonal is high. The size of the plateau suggests that over many iterates, contraction may dominate expansion, leading to a negative Lyapunov exponent. A negative Lyapunov exponent implies that two Hodgkin–Huxley neurons, when driven by a *common* pulse train with random kick times, will synchronize. That is, the plateau provides a way to create a "random fixed point" [22]. These predictions are consistent with preliminary numerical results and with a perturbation theory developed by Nakao et al. for randomly kicked oscillators in the limit of weak kicks [26]. The heuristic geometric argument sketched above may lead to an extension of their result to the regime of strong kicks.

This synchronization mechanism has also been studied numerically by Doi in the context of

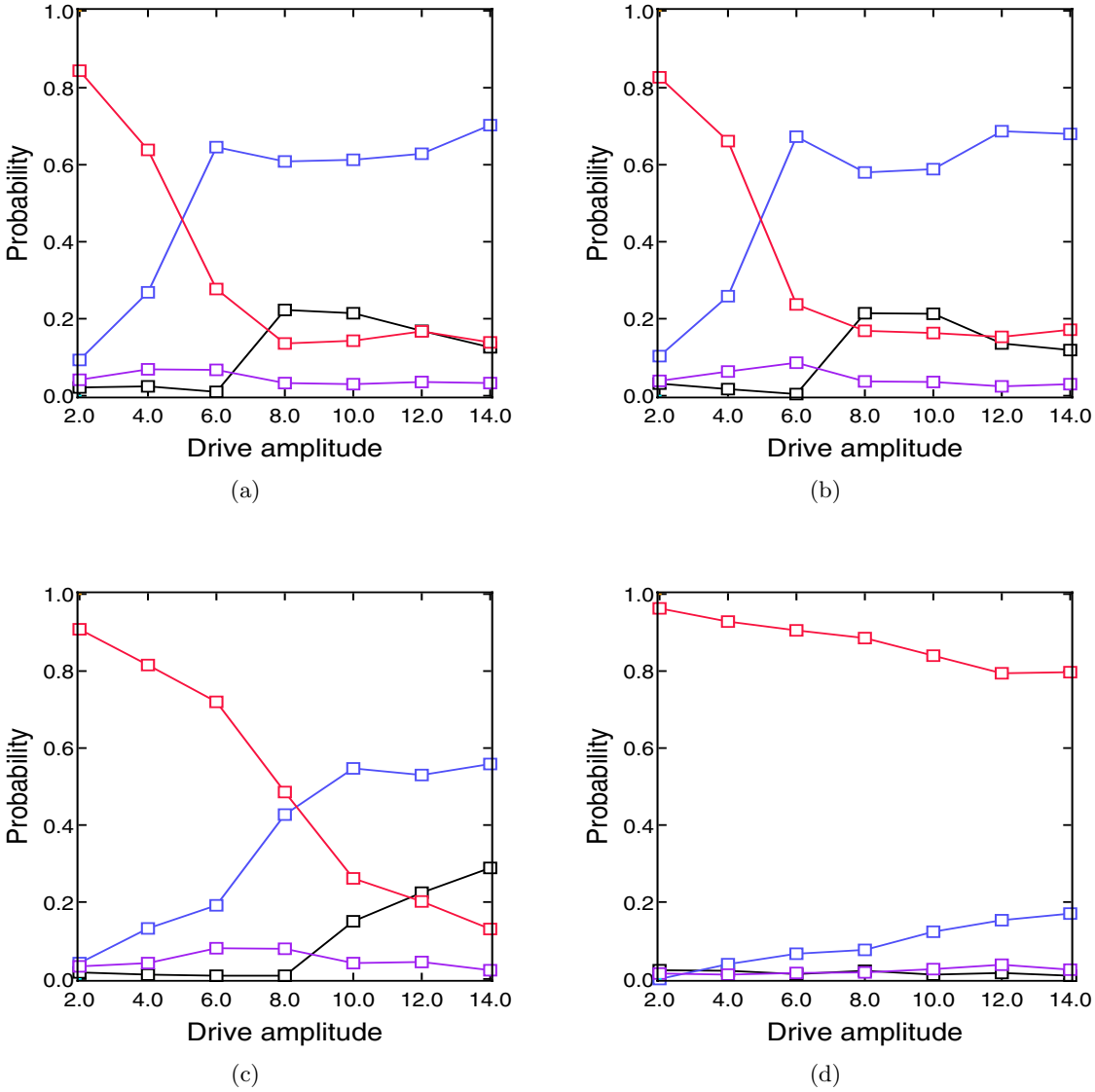


Figure 14. *Distribution of Lyapunov exponents for $I \approx 14$ with finite-duration pulse of duration* (a) $t_0 = 0.05$, (b) $t_0 = 0.3$, (c) $t_0 = 2.75$, *and* (d) $t_0 = 9.0$. *See Figure 4 caption for details. Recall that the Lyapunov exponents of the unperturbed limit cycle γ are -0.196, -2.01, and -8.31, corresponding to relaxation times of* 5.1, 0.50, *and* 0.12.

a simple piecewise linear map [6]. In addition, there is an extensive literature on noise-induced synchrony in neural models, including white-noise-driven Hodgkin–Huxley equations [29, 41]. Models exhibiting noise-induced synchrony provide a concrete framework for exploring neural reliability [21, 23].

Robustness of the phase-resetting curve. How robust are the features (plateau, kink) of the

Hodgkin–Huxley phase-resetting curve under perturbations to parameters? How robust is the geometry of the near-tangency of kicked segments and strong-stable manifolds responsible for forming the plateau? As the range of phenomena predicted by Wang–Young theory may be present in two-dimensional models like the Morris–Lecar or FitzHugh–Nagumo, these models may provide a good starting point for exploring these questions in lower-dimensional settings.

Appendix. Computation of phase-resetting curves. All numerical calculations of phase-resetting curves reported in this paper use the algorithm described here. It is closely related to an algorithm due to Ermentrout and Kopell [8]. It is presented here for the sake of completeness; a systematic comparison with existing methods will appear elsewhere. The algorithm computes the phase-resetting curve for *finite-size perturbations* but can be adapted to compute phase-resetting curves for infinitesimal perturbations. See Appendix A in [4] and [28, 37] for more general discussions of phase-resetting curves, and Ermentrout, Pascal, and Gutkin [9] for a discussion of computing phase-resetting curves experimentally.

The basic idea is to numerically compute the strong-stable linear subspaces along the limit cycle. Then, using these linear subspaces as approximations to strong-stable submanifolds, project a kicked point down to the limit cycle, where the phase can be estimated. The algorithm really computes strong-stable linear subspaces along γ and uses these linear subspaces to approximate the phase-resetting curve.

Some preliminaries: the limiting map $\bar{F}_T = \lim_{n\to\infty} F_{T+nT_0}$ can be characterized abstractly by the equation

$$\bar{F}_T = \pi_{ss} \circ \phi_T \circ K_A = \phi_T \circ \pi_{ss} \circ K_A, \tag{A.1}$$

where $\pi_{ss}(x) = y$ if and only if $x \in W_{ss}(y)$, i.e., $\pi_{ss} : B(\gamma) \to \gamma$ maps the basin of γ onto the limit cycle γ along strong-stable manifolds. The abstract notations (and notions) have their uses: the projection π_{ss} encapsulates the properties of the strong-stable manifolds; e.g., by definition, the strong-stable manifold $W_{ss}(\pi_{ss}(x))$ passes through x for any $x \in B(\gamma)$, and the nullspace of the Jacobian matrix $D\pi_{ss}(x)$ is precisely the tangent space of the strong-stable manifold $W_{ss}(\pi_{ss}(x))$ at x.[6]

Algorithm (phase resetting curves via stable subspaces).

1. Estimate the period T_0 of the limit cycle γ by numerically solving the unforced equations starting with a point on or near γ.
2. Discretize γ by subdividing the time interval $[0, T_0)$ into N intervals and computing the corresponding points $x_i \in \gamma$. Fix an arbitrary reference point x_0 on γ so that each point on γ can be assigned a unique phase $\theta \in [0, T_0)$.
3. For each point x_i computed in the previous step, compute the Jacobian $DH(x_i)$ of the Hodgkin–Huxley flow field H at that point.

[6]Note that the commutation relation $\phi_t \circ \pi_{ss} = \pi_{ss} \circ \phi_t$ expresses the invariance of the strong-stable foliation under ϕ_t. The map K_A, in general, has nothing to do with the flow and does not commute with the other maps.

4. Using the results of the previous two steps, solve

$$\begin{aligned} \dot{x} &= -H(x), \\ \dot{\xi} &= \eta - \langle \eta, \xi \rangle \xi, \\ \eta &= DH(x)^T \xi, \end{aligned} \tag{A.2}$$

using the grid points $\{x_i\}$ computed in the previous steps. The $\dot{x}$ part of the equation above is clearly numerically unstable, but that is not a problem because we already have a numerical representation of γ.
The equations above are a variant of the usual method for computing Lyapunov exponents [13]. They preserve the length of $\xi(t)$, although in practice it is necessary to rescale $\xi(t)$ to ensure that this constraint is maintained. As $t \to \infty$, $\xi(t)$ becomes orthogonal to the strong-stable linear subspace $E_{\text{ss}}(x(t))$ of γ. The subspace $E_{\text{ss}}(x(t))$ is tangent to the strong-stable manifold $W_{\text{ss}}(x(t))$ at $x(t)$.[7]
5. Using (A.1) in combination with the linear subspaces computed in the previous step, we can now approximate the phase-resetting curve. Start with a point $x \in \gamma$ and compute $\Phi_t(K_A(x))$ for increasing t. Let $t_0 > 0$ be the minimum positive time at which (i) $\Phi_{t_0}(K_A(x))$ has returned to a small, fixed neighborhood of γ (in this paper this is chosen to be a neighborhood of distance 10^{-4} around γ), and (ii) $\Phi_{t_0}(K_A(x))$ lies within one of the precomputed linear subspaces $E_{\text{ss}}(x_*)$ for some point x_*. Let θ_* denote the phase of the point x_*. Then the new phase of the system is $(T + \theta_* - t_0) \pmod{T_0}$.
6. Proceed to the next grid point and repeat.
7. When the derivative of the phase-resetting curve becomes large or infinite, it may be necessary to adaptively generate the grid points on which the curve is evaluated. Generally speaking, the grid $\{x_i\}$ constructed in step 2 need not equal the grid $\{x_i'\}$ on which the phase-resetting curve is evaluated. In particular, the grid $\{x_i'\}$ can be adaptively chosen to ensure that $\left|\hat{f}_a(x_{i+1}') - \hat{f}_a(x_i')\right| \leq \epsilon$, where $\hat{f}_a$ denotes the computed phase-resetting curve and ϵ is a fixed number, in this paper usually 0.1. This adaptive mechanism provides a way to detect discontinuities in f_T.

Acknowledgments. It is a pleasure to thank Lai-Sang Young, Eric Shea-Brown, Adi Rangan, Charlie Peskin, John Rinzel, and Louis Tao for many helpful and stimulating conversations. I am also grateful to the referees for their help in improving the exposition and for pointing out relevant references.

[7] These equations can be generalized to the following:

$$\begin{aligned} \dot{x} &= -H(x), \\ \dot{\xi}_i &= \eta_i - \langle \eta_i, \xi_i \rangle \xi_i - \textstyle\sum_{j<i} \left(\langle \xi_i, \dot{\xi}_j \rangle + \langle \eta_i, \xi_j \rangle \right) \xi_j, \\ \eta_i &= DH(x)^T \xi_i. \end{aligned} \tag{A.3}$$

If the vectors (ξ_i) form an orthonormal basis at $t = 0$, then the equations will guarantee that $(\xi_i(t))$ are orthonormal for all $t > 0$. Again, it will be necessary to perform Gram–Schmidt orthogonalizations periodically to maintain this constraint. The vector $\xi_1(t)$, as before, converges to a vector orthogonal to $E_{\text{ss}}(x(t))$. Thus $(\xi_2(t), \xi_3(t), \xi_4(t))$ span $E_{\text{ss}}(x(t))$. Similarly, the vectors $(\xi_3(t), \xi_4(t))$ span the subspace consisting of the two fastest contracting directions, and $(\xi_4(t))$ spans the fastest contracting direction.

REFERENCES

[1] K. Aihara and G. Matsumoto, *Chaotic oscillations and bifurcations in squid giant axons*, in Chaos, Nonlinear Science: Theory and Applications, Manchester University Press, Manchester, UK, 1986.

[2] J. C. Alexander, E. J. Doedel, and H. G. Othmer, *On the resonance structure in a forced excitable system*, SIAM J. Appl. Math., 50 (1990), pp. 1373–1418.

[3] E. N. Best, *Null space in the Hodgkin-Huxley equations: A critical test*, Biophys. J., 27 (1979), pp. 87–104.

[4] E. Brown, J. Moehlis, and P. Holmes, *On the phase reduction and response dynamics of neural oscillator populations*, Neural Comput., 16 (2004), pp. 673–715.

[5] J. Cronin, *Mathematical Aspects of Hodgkin-Huxley Neural Theory*, Cambridge University Press, Cambridge, UK, 1987.

[6] S. Doi, *A chaotic map with a flat segment can produce a noise-induced order*, J. Statist. Phys., 55 (1989), pp. 941–964.

[7] G. B. Ermentrout, *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*, Software Environ. Tools 14, SIAM, Philadelphia, 2002.

[8] G. B. Ermentrout and N. Kopell, *Multiple pulse interactions and averaging in systems of coupled neural oscillators*, J. Math. Biol., 29 (1991), pp. 195–217.

[9] G. B. Ermentrout, M. Pascal, and B. Gutkin, *The effects of spike frequency adaptation and negative feedback on the synchronization of neural oscillators*, Neural Comput., 13 (2001), pp. 1285–1310.

[10] K. Geist, U. Parlitz, and W. Lauterborn, *Comparison of different methods for computing Lyapunov exponents*, Progr. Theoret. Phys., 83 (1990), pp. 875–893.

[11] L. Glass and M. C. Mackey, *From Clocks to Chaos: The Rhythms of Life*, Princeton University Press, Princeton, NJ, 1988.

[12] J. Guckenheimer, *Isochrons and phaseless sets*, J. Theoret. Biol., 1 (1974), pp. 259–273.

[13] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.

[14] R. Guttman, L. Feldman, and E. Jakobsson, *Frequency entrainment of squid axon membrane*, J. Membrane Biol., 56 (1980), pp. 9–18.

[15] H. Hayashi, S. Ishizuka, and K. Hirakawa, *Chaotic response of the pacemaker neuron*, J. Phys. Soc. Japan, 54 (1985), pp. 2337–2346.

[16] H. Hayashi, S. Ishizuka, M. Ohta, and K. Hirakawa, *Chaotic behavior in the* onchidium *giant neuron under sinusoidal stimulation*, Phys. Lett. A, 88 (1982), pp. 435–438.

[17] H. Hayashi, M. Nakao, and K. Hirakawa, *Entrained, harmonic, quasiperiodic and chaotic responses of the self-sustained oscillation of* nitella *to sinusoidal stimulation*, J. Phys. Soc. Japan, 52 (1983), pp. 344–351.

[18] A. L. Hodgkin and A. F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, J. Physiol., 117 (1952), pp. 500–544.

[19] A. V. Holden, *The response of excitable membrane models to a cyclic input*, Biology and Cybernetics, 21 (1975), pp. 1–7.

[20] D. H. Hubel, *Eye, Brain, and Vision*, Scientific American Library, W. H. Freeman, New York, 1988.

[21] J. D. Hunter, J. G. Milton, P. J. Thomas, and J. D. Cowan, *Resonance effect for neural spike time reliability*, J. Neurophysiol., 80 (1998), pp. 1427–1438.

[22] F. Ledrappier and L.-S. Young, *Entropy formula for random transformations*, Probab. Theory Related Fields, 80 (1988), pp. 217–240.

[23] Z. F. Mainen and T. J. Sejnowski, *Reliability of spike timing in neocortical neurons*, Science, 268 (1995), pp. 1503–1506.

[24] G. Matsumoto, K. Aihara, Y. Hanyu, N. Takahashi, S. Yoshizawa, and J. ichi Nagumo, *Chaos and phase locking in normal squid axons*, Phys. Lett. A, 123 (1987), pp. 162–166.

[25] G. Matsumoto, N. Takahashi, and Y. Hanyu, *Chaos, phase locking and bifurcation in normal squid axons*, in Chaos in Biological Systems, Plenum Press, New York, 1987, pp. 143–156.

[26] H. Nakao, K. suke Arai, K. Nagai, Y. Tsubo, and Y. Kuramoto, *Synchrony of limit-cycle oscillators induced by random external impulses*, Phys. Rev. E (3), 72 (2005), paper 026220.

[27] A. Oksasoglu and Q. Wang, *Strange attractors in periodically-kicked Chua's circuit*, Internat. J. Bifur.

Chaos Appl. Sci. Engrg., 16 (2005), pp. 83–98.
[28] S. A. Oprisan and C. C. Canavier, *The influence of limit cycle topology on the phase resetting curve*, Neural Comput., 14 (2002), pp. 1027–1057.
[29] K. Pakdaman and S. Tanabe, *Random dynamics of the Hodgkin-Huxley neuron model*, Phys. Rev. E (3), 64 (2001), paper 050902.
[30] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, UK, 1992.
[31] F. Rieke, *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA, 1997.
[32] T. Takabe, K. Aihara, and G. Matsumoto, *Response characteristics of the Hodgkin-Huxley equations to pulse-train stimulation*, Transactions of IEICE Japan (Japanese edition), J71-A (1988), pp. 744–750.
[33] Q. Wang and L.-S. Young, *Strange attractors with one direction of instability*, Comm. Math. Phys., 218 (2001), pp. 1–97.
[34] Q. Wang and L.-S. Young, *From invariant curves to strange attractors*, Comm. Math. Phys., 225 (2002), pp. 275–304.
[35] Q. Wang and L.-S. Young, *Strange attractors in periodically-kicked limit cycles and Hopf bifurcations*, Comm. Math. Phys., 240 (2003), pp. 509–529.
[36] Q. Wang and L.-S. Young, *Toward a theory of rank-1 attractors*, Ann. Math., to appear.
[37] T. L. Williams and G. Bowtell, *The calculation of frequency-shift functions for chains of coupled oscillators, with application to a network model of the lamprey locomotor pattern generator*, J. Comput. Neurosci., 4 (1997), pp. 47–55.
[38] A. Winfree, *The Geometry of Biological Time*, 2nd ed., Springer-Verlag, New York, 2000.
[39] L.-S. Young, *Ergodic theory of differentiable dynamical systems*, in Real and Complex Dynamics, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 464, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995, pp. 293–336.
[40] G. Zaslavsky, *The simplest case of a strange attractor*, Phys. Lett. A, 69 (1978), pp. 145–147.
[41] C. Zhou and J. Kurths, *Noise-induced synchronization and coherence resonance of a Hodgkin-Huxley model of thermally sensitive neurons*, Chaos, 13 (2003), pp. 401–409.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 2, pp. 205–251

Nilpotent Hopf Bifurcations in Coupled Cell Systems*

Toby Elmhirst† and Martin Golubitsky‡

Abstract. Network architecture can lead to robust synchrony in coupled systems and, surprisingly, to codimension one bifurcations from synchronous equilibria at which the associated Jacobian is nilpotent. We prove three theorems concerning nilpotent Hopf bifurcations from synchronous equilibria to periodic solutions, where the critical eigenvalues have algebraic multiplicity two and geometric multiplicity one, and discuss these results in the context of three different networks in which the bifurcations occur generically. Phenomena stemming from these bifurcations include multiple periodic solutions, solutions that grow at a rate faster than the standard $\lambda^{\frac{1}{2}}$, and solutions that grow slower than the standard $\lambda^{\frac{1}{2}}$. These different bifurcations depend on the network architecture and, in particular, on the flow-invariant subspaces that are forced to exist by the architecture.

Key words. Hopf bifurcation, coupled cells, nonsemisimple normal form

AMS subject classifications. 34C23, 34C25, 37G05

DOI. 10.1137/050635559

1. Introduction.

Stewart, Golubitsky, Pivato, and Török (see [9, 5]) formalized the concept of a *coupled cell network*, where a *cell* is a system of ordinary differential equations (ODEs) and a *coupled cell system* consists of cells whose equations are coupled. These researchers defined the architecture of coupled cell networks and developed a theory that shows how network architecture leads to synchrony. The architecture of a coupled cell network is a graph that indicates which cells have the same phase space, which cells are coupled to which, and which couplings are the same. Coupled cell systems with a given architecture are called *admissible*. In this theory, local network symmetries (which form a groupoid; see [9] for details) generalize symmetry as a way of organizing network dynamics, and synchrony-breaking bifurcations replace symmetry-breaking bifurcations as a basic way in which transitions to complicated dynamics occur.

This paper is concerned with *homogeneous networks*. These are networks in which there is only one type of cell and one type of coupling. In particular, the differential equations defining the time evolution of each cell in any admissible system are identical. Thus these networks have the property that the diagonal subspace Δ, formed by setting the coordinates in all cells equal, is flow-invariant for all admissible coupled cell systems. It is therefore expected that branches of synchronous equilibria can exist in Δ and that synchrony-breaking bifurcations from these equilibria (bifurcations in which critical eigenvectors of the Jacobian J at the equilibrium are not in Δ) can occur naturally as one parameter in the differential equations is varied.

*Received by the editors August 2, 2005; accepted for publication (in revised form) by T. Kaper January 24, 2006; published electronically April 21, 2006. This work was supported in part by NSF grant DMS-0244529.
http://www.siam.org/journals/siads/5-2/63555.html
†Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2 (toby@math.uh.edu).
‡Department of Mathematics, University of Houston, Houston, TX 77204-3008 (mg@math.uh.edu).

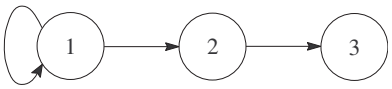


Figure 1. *The three-cell feed-forward chain.*

Golubitsky, Nicol, and Stewart [2] observed that a certain three-cell feed-forward network (see Figure 1) has codimension one synchrony-breaking bifurcations where J restricted to the center subspace is nonsemisimple. The corresponding Hopf bifurcation, which we call a *nilpotent Hopf bifurcation*, leads to periodic solutions whose amplitudes grow with the surprising growth rate of 1/6, rather than the expected growth rate from Hopf bifurcation of 1/2. Leite [6] and Leite and Golubitsky [7] showed that there are 34 different types of homogeneous three-cell networks where the number of inputs to each cell is either one or two, and that several of these, in addition to the feed-forward network, lead to nilpotent bifurcations.

In this paper we develop an approach to nilpotent Hopf bifurcation theory which enables us to complete the work in [2] by showing that the 1/6 power growth rate is generic in the codimension one nilpotent Hopf bifurcations of the feed-forward network, and to classify the periodic solutions that can emanate from codimension one nilpotent Hopf bifurcations in certain other homogeneous cell networks.

Nilpotent (or 1:1 resonant nonsemisimple) Hopf bifurcations have been considered previously in a generic setting in [10, 1, 8]. Without the coupled cell framework, such bifurcations occur in codimension three. However, the structure imposed on admissible vector fields at the linear level by certain network architectures implies that nilpotent Hopf bifurcations can occur in these systems at codimension one. Moreover, these same architectures can also put restrictions on the higher order terms of admissible vector fields, which force surprising branching of the solutions.

When investigating structured systems, one fundamental question is "How does the architecture of the system affect the dynamics," and already we see unexpected, complex behavior in simple-looking systems. In addition to aiding our understanding of natural systems, we also expect applications of a more synthetic nature. In particular, we believe that the "amplification" seen in the 1/6 growth rate in the network of Figure 1 could have interesting engineering consequences.

We begin with a brief review of ordinary Hopf bifurcation and a summary of our main results.

The standard Hopf theorems. Hopf bifurcation occurs at an equilibrium x_0 and at a parameter value λ_0 of

$$\dot{x} = F(x, \lambda), \qquad x \in \mathbf{R}^n, \lambda \in \mathbf{R}, \tag{1.1}$$

when the linearization of $(\mathrm{d}F)_{x_0,\lambda_0}$ has a pair of purely imaginary eigenvalues. Generically, the critical eigenvalues are simple, and no other eigenvalues lie on the imaginary axis. Under these assumptions we may assume, after a change of coordinates and a rescaling of time, that $x_0 = 0$, $\lambda_0 = 0$, the critical eigenvalues of $(\mathrm{d}F)_{0,0}$ are $\pm i$, and locally $F(0, \lambda) = 0$.

Let $\sigma(\lambda) + i\omega(\lambda)$, where $\sigma(0) = 0$ and $\omega(0) = 1$, be the eigenvalue extension in $(\mathrm{d}F)_{0,\lambda}$ of

i. Assume that the *eigenvalue crossing condition* holds, that is,

$$\sigma'(0) \neq 0. \tag{1.2}$$

The two standard Hopf bifurcation theorems [4] then state the following:

1. There is a unique branch of small amplitude periodic solutions to (1.1) with period near 2π.
2. Generically, the amplitude of the periodic solutions on this branch grows at a rate of order $\lambda^{1/2}$; that is, the bifurcation is of pitchfork type.

The main results. We present three theorems concerning nilpotent Hopf bifurcations in coupled cell systems. Specifically, in this paper we shall use the term *nilpotent Hopf bifurcation* to indicate that there are critical eigenvalues $\pm\omega i$ of $(\mathrm{d}F)_{0,0}$ at the equilibrium $(0,0)$, where $\omega > 0$, that are each double, but with only one (complex conjugate) pair of corresponding eigenvectors. Typically we will also rescale time so that $\omega = 1$. We note that network architectures can easily be found that lead to codimension one bifurcations in admissible vector fields in which the critical eigenvalues have algebraic multiplicity greater than two and geometric multiplicity one. For example, the n-cell feed-forward chain, obtained by attaching additional cells to the end of the network in Figure 1, has eigenvalues of algebraic multiplicity $n-1$ and geometric multiplicity one.

We show below that nilpotent Hopf bifurcations can occur generically in codimension one bifurcations from a synchronous equilibrium in coupled cell networks. This point was noted previously in [2, 6]. Our focus here is on the nonlinear theory, in which we show that network architecture can lead generically to multiple periodic solutions whose amplitude growth rate is greater than, equal to, or less than $1/2$.

This variety in nilpotent Hopf bifurcations is due to the type of nonlinear degeneracies forced by different network architectures on their admissible vector fields. In our approach we study classes of networks whose architectures force, in codimension one, a particular type of nonlinear degeneracy in the Liapunov–Schmidt reduced equation. Given this assumption on architecture, we classify the branches of periodic solutions that occur generically in codimension one bifurcations. Each network architecture can, in principle, lead to different codimension one bifurcations, just as each symmetry group can lead to different equivariant bifurcations.

We illustrate our results by discussing the following three specific networks:

(a) the three-cell feed-forward network in Figure 1, whose nilpotent Hopf bifurcation generically leads to two branches of periodic solutions with amplitude growth at rates of $1/6$ and $1/2$. The existence of these solutions in a restricted class of coupled cell systems is shown in [2].

(b) the three-cell network in Figure 2, whose nilpotent Hopf bifurcation generically leads either to two or four branches of periodic solutions with amplitude growth at the standard rate of $\lambda^{1/2}$.

(c) the five-cell network in Figure 3, whose nilpotent Hopf bifurcation generically leads to two branches of periodic solutions with amplitude growth at rate λ.

Coupled cell networks and nilpotent normal forms. A general theory for the differential equations associated with coupled cell networks is outlined in [9, 5]. In particular, an algo-

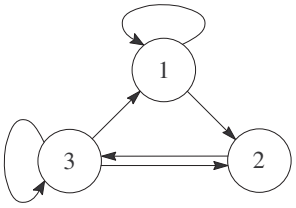


Figure 2. *A three-cell network with nilpotent linear part.*

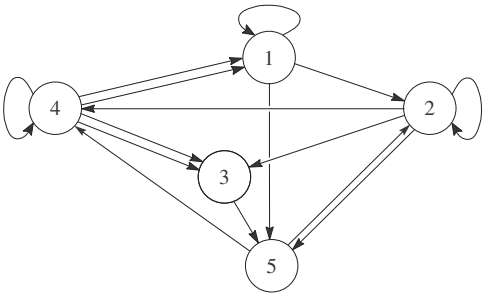


Figure 3. *A five-cell network with nilpotent linear part.*

rithmic way of identifying a class of systems of differential equations with a directed graph is given. The identification is reasonably intuitive, so we do not describe the general setup here. Rather, we just list the results for the three networks of Figures 1–3.

(a) Following [2], the coupled cell systems corresponding to the three-cell feed-forward network in Figure 1 have the form

$$\begin{aligned} \dot{x}_1 &= f(x_1, x_1), \\ \dot{x}_2 &= f(x_2, x_1), \\ \dot{x}_3 &= f(x_3, x_2), \end{aligned} \tag{1.3}$$

where $x_1, x_2, x_3 \in \mathbf{R}^k$ and $f : \mathbf{R}^k \times \mathbf{R}^k \to \mathbf{R}^k$ is arbitrary. Note that the synchrony subspace $x_1 = x_2 = x_3$ is flow-invariant for such systems, and the existence of a synchronous equilibrium (satisfying $f(a, a) = 0$) is to be expected. Without loss of generality we may assume that the synchronous equilibrium is at the origin. The Jacobian of (1.3) at the origin has the form

$$J = \begin{pmatrix} A+B & 0 & 0 \\ B & A & 0 \\ 0 & B & A \end{pmatrix},$$

where $A = f_1(\mathbf{0})$ is the linearized internal dynamics and $B = f_2(\mathbf{0})$ is the linearized coupling. The $3k$ eigenvalues and eigenvectors of J are

Eigenvector	Eigenvalues	Algebraic multiplicity	Geometric multiplicity
$(0, 0, u)^t$	A	2	1
$(v, v, v)^t$	$A + B$	1	1

where u is an eigenvector of A and v is an eigenvector of $A+B$. It follows that when $k \geq 2$, (1.3) can have a codimension one nilpotent Hopf bifurcation.

(b) The coupled cell systems corresponding to the three-cell network in Figure 2 have the form

$$\begin{aligned} \dot{x}_1 &= f(x_1, \overline{x_1, x_3}), \\ \dot{x}_2 &= f(x_2, \overline{x_1, x_3}), \\ \dot{x}_3 &= f(x_3, \overline{x_2, x_3}), \end{aligned} \tag{1.4}$$

where $x_1, x_2, x_3 \in \mathbf{R}^k$ and the overbar indicates that $f : \mathbf{R}^k \times \mathbf{R}^{2k} \to \mathbf{R}^k$ satisfies $f(a,b,c) = f(a,c,b)$. The synchrony subspace $x_1 = x_2 = x_3$ is still flow-invariant for such systems, and the existence of a synchronous equilibrium, which without loss of generality we may assume is at the origin, is to be expected. The Jacobian of (1.4) at the origin has the form

$$J = \begin{pmatrix} A+B & 0 & B \\ B & A & B \\ 0 & B & A+B \end{pmatrix}, \tag{1.5}$$

where $A = f_1(\mathbf{0})$ is the linearized internal dynamics and $B = f_2(\mathbf{0}) = f_3(\mathbf{0})$ is the linearized coupling. The $3k$ eigenvalues and eigenvectors of J are

Eigenvector	Eigenvalues	Algebraic multiplicity	Geometric multiplicity
$(u,u,-u)^t$	A	2	1
$(v,v,v)^t$	$A+2B$	1	1

where u is an eigenvector of A and v is an eigenvector of $A+2B$. Thus, if $k \geq 2$, codimension one nilpotent Hopf bifurcations occur in this network as well.

(c) The coupled cell systems corresponding to the five-cell network in Figure 3 have the form

$$\begin{aligned} \dot{x}_1 &= f(x_1, \overline{x_1, x_4, x_4}), \\ \dot{x}_2 &= f(x_2, \overline{x_1, x_2, x_5}), \\ \dot{x}_3 &= f(x_3, \overline{x_2, x_4, x_4}), \\ \dot{x}_4 &= f(x_4, \overline{x_2, x_4, x_5}), \\ \dot{x}_5 &= f(x_5, \overline{x_1, x_2, x_3}), \end{aligned} \tag{1.6}$$

where $x_j \in \mathbf{R}^k$, $f : \mathbf{R}^k \times \mathbf{R}^{3k} \to \mathbf{R}^k$, and the overbar indicates that $f(a,b,c,d)$ is invariant under permutation of b,c,d. The synchrony subspace $x_1 = x_2 = x_3 = x_4 = x_5$ is flow-invariant, and a synchronous equilibrium, which again we may assume is at the origin, is therefore to be expected. The Jacobian of (1.6) at the origin is

$$J = \begin{pmatrix} A+B & 0 & 0 & 2B & 0 \\ B & A+B & 0 & 0 & B \\ 0 & B & A & 2B & 0 \\ 0 & B & 0 & A+B & B \\ B & B & B & 0 & A \end{pmatrix},$$

where $A = f_1(\mathbf{0})$ is the linearized internal dynamics and $B = f_2(\mathbf{0}) = f_3(\mathbf{0}) = f_4(\mathbf{0})$ is the linearized coupling. The $5k$ eigenvalues of J are

Eigenvalues	Algebraic multiplicity	Geometric multiplicity
$A \pm iB$	2	1
$A + 3B$	1	1

Assuming that (1.6) depends on a parameter λ, we can arrange for a codimension one nilpotent Hopf bifurcation at $\lambda = 0$ by taking $k = 1$, $A(\lambda) = \lambda$, and $B(\lambda) \equiv -1$.

Review of the Liapunov–Schmidt reduction proof of Hopf bifurcation. We use the standard procedure of Liapunov–Schmidt reduction for finding periodic solutions through Hopf bifurcation (see [3]), but nilpotence dramatically changes this analysis.

Since periodic solutions to (1.1) will not in general have period 2π, rescale time in the usual way by letting $s = (1 + \tau)t$ so that (1.1) becomes

$$(1+\tau)\frac{\mathrm{d}z}{\mathrm{d}s} = F(z,\lambda). \tag{1.7}$$

Fixing the period allows us to define the operator $\Phi : \mathcal{C}^1_{2\pi} \times \mathbf{R} \times \mathbf{R} \longrightarrow \mathcal{C}_{2\pi}$ by

$$\Phi(x,\lambda,\tau) = (1+\tau)\frac{\mathrm{d}x}{\mathrm{d}s} - F(x,\lambda), \tag{1.8}$$

where $\mathcal{C}_{2\pi}$ and $\mathcal{C}^1_{2\pi}$ are respectively the Banach spaces of continuous and continuously differentiable 2π-periodic functions $x : \mathbf{S}^1 \longrightarrow \mathbf{R}^n$. Note that

(a) the solutions to $\Phi(x,\lambda,\tau) = 0$ correspond to near 2π-periodic solutions of (1.1),
(b) $\Phi(0,\lambda,\tau) \equiv 0$ since $F(0,\lambda) \equiv 0$,
(c) Φ is $\mathbf{S}^1$-equivariant, where $\theta \in \mathbf{S}^1$ acts on $u \in \mathcal{C}_{2\pi}$ by

$$(\theta \cdot u)(s) = u(s - \theta).$$

In standard Hopf bifurcation, the kernel and cokernel of the Frechet derivative $(\mathrm{d}\Phi)_{\mathbf{0}}$ are two-dimensional, and these spaces may be identified with $\mathbf{C}$. Liapunov–Schmidt reduction implies the existence of a mapping $\phi : \mathbf{C}\times\mathbf{R}\times\mathbf{R} \to \mathbf{C}$, whose zeros near the origin parameterize the small amplitude periodic solutions of $\Phi = 0$. Moreover, this reduction can be performed to preserve symmetry; that is, we can assume that

$$\phi(e^{i\theta}z,\lambda,\tau) = e^{i\theta}\phi(z,\lambda,\tau). \tag{1.9}$$

It follows that

$$\phi(z,\lambda,\tau) = p(|z|^2,\lambda,\tau)z + q(|z|^2,\lambda,\tau)iz, \tag{1.10}$$

where p, q are real-valued smooth functions satisfying $p(\mathbf{0}) = q(\mathbf{0}) = 0$. Using (1.9), we need only look for solutions where $z = x \in \mathbf{R}$. Hence solutions to $\phi = 0$ are of two types: $x = 0$ (the trivial equilibrium) and solutions to the system $p = q = 0$ (the desired small amplitude periodic solutions).

In standard Hopf bifurcation, a calculation shows that $q_\tau(\mathbf{0}) = -1$. Hence, the equation $q = 0$ can be solved by the implicit function theorem for $\tau = \tau(x^2, \lambda)$, where $\tau(\mathbf{0}) = 0$, and small amplitude periodic solutions to (1.1) are found by solving

$$r(x^2, \lambda) \equiv p(x^2, \lambda, \tau(x^2, \lambda)) = 0. \tag{1.11}$$

Another (more complicated) calculation shows that

$$r_\lambda(\mathbf{0}) = \sigma'(0).$$

It follows from the eigenvalue crossing condition that $r = 0$ can be solved by another application of the implicit function theorem for $\lambda = \lambda(x^2)$, where $\lambda(0) = 0$, and the first Hopf theorem is proved. Setting $u = x^2$, the second Hopf theorem (the square root growth of amplitude) is proved by making the genericity assumption $r_u(\mathbf{0}) \neq 0$. The calculation of $r_u(\mathbf{0})$ in terms of (1.1) is the most difficult of the calculations.

Statements of the main theorems. In a nilpotent Hopf bifurcation the kernel and cokernel of the Frechet derivative $(\mathrm{d}\Phi)_{\mathbf{0}}$ are still two-dimensional—just like ordinary Hopf bifurcation. It follows that the Liapunov–Schmidt reduced equation $\phi = 0$ has the same $\mathbf{S}^1$-equivariance as in standard Hopf bifurcation and hence has the form of (1.10).

We show in section 2.2 that a nilpotent Hopf bifurcation leads to the following result.

Proposition 1.1. *Let p and q be as in (1.10). Then*

$$p_\tau(\mathbf{0}) = 0, \qquad q_\tau(\mathbf{0}) = 0, \tag{1.12}$$

$$p_\lambda(\mathbf{0}) = 0, \qquad q_\lambda(\mathbf{0}) = 0. \tag{1.13}$$

It follows from Proposition 1.1 that we cannot employ the implicit function theorem in the same way as it is used in the standard Hopf theorem; higher derivatives are necessary to understand the bifurcation. Note that Proposition 1.1 does not require any assumptions about network architecture. Indeed, it follows simply from the fact that $(\mathrm{d}F)_{\mathbf{0}}$ is nilpotent.

All is not lost, however. In section 2.3 we assume that F is a homogeneous coupled cell system with nilpotent linearization, and we obtain the following generalization of the eigenvalue crossing condition (see [10] for a version of this proposition for generic vector fields).

Proposition 1.2. *If F is a homogeneous coupled cell system, then*

$$p_{\tau\tau}(\mathbf{0}) = -2, \qquad q_{\tau\tau}(\mathbf{0}) = 0, \tag{1.14}$$

$$p_{\lambda\lambda}(\mathbf{0}) = 2(\sigma'(0)^2 - \omega'(0)^2), \qquad q_{\lambda\lambda}(\mathbf{0}) = -4\sigma'(0)\omega'(0), \tag{1.15}$$

$$p_{\lambda\tau}(\mathbf{0}) = 2\omega'(0), \qquad q_{\lambda\tau}(\mathbf{0}) = 2\sigma'(0). \tag{1.16}$$

These explicit calculations enable us to proceed with the calculation of solutions to $p = q = 0$. The three theorems mentioned previously can now be stated (in reverse order) and the proofs sketched.

We begin by stating the following theorem, which is proved in section 3.

Theorem 1.3. *Assume*

$$q_u(\mathbf{0}) \neq 0 \tag{1.17}$$

and the eigenvalue crossing condition (1.2). *Then there exist two branches of near* 2π*-periodic solutions, each one growing at order* λ. *Moreover, one branch is subcritical, and the other is supercritical.*

This theorem is proved by a straightforward application of the implicit function theorem. We also show that the network in Figure 3 generically satisfies (1.17).

Observe that Theorem 1.3 refers to the class of networks whose architecture does not force restrictions on the nonlinear terms of the reduced equation. The linear structure alone forces a different branching pattern than for generic Hopf bifurcation. It does not follow, however, that this situation is somehow the "generic" case for coupled cell networks. Indeed, there are network architectures, such as the feed-forward network, that do force $p_u = q_u = 0$.

In section 4 we prove the next theorem.

Theorem 1.4. *Assume*

$$p_u(\mathbf{0}) = q_u(\mathbf{0}) = 0, \tag{1.18}$$

$$\begin{aligned} &p_{uu}(\mathbf{0}) \neq -\tfrac{1}{2}p_{u\tau}(\mathbf{0})^2, && p_{u\lambda}(\mathbf{0}) \neq -\omega'(0)p_{u\tau}(\mathbf{0}),\\ &q_{uu}(\mathbf{0}) \neq -p_{u\tau}(\mathbf{0})q_{u\tau}(\mathbf{0}), && q_{u\tau}(\mathbf{0}) \neq 0,\\ &q_{u\lambda}(\mathbf{0}) \neq -\big(\sigma'(0)p_{u\tau}(\mathbf{0}) + \omega'(0)q_{u\tau}(\mathbf{0})\big), \end{aligned} \tag{1.19}$$

$$p_{uu}(\mathbf{0}) \neq \frac{q_{uu}(\mathbf{0})}{q_{u\tau}(\mathbf{0})}\left(p_{u\tau}(\mathbf{0}) + \frac{q_{uu}(\mathbf{0})}{2q_{u\tau}(\mathbf{0})}\right), \tag{1.20}$$

and the eigenvalue crossing condition (1.2). *Then there exist either two or four branches of near* 2π*-periodic solutions to* (1.1), *with the number of branches depending on* F, *and each branch grows as* $\lambda^{\frac{1}{2}}$.

The constraint (1.18) means that the implicit function theorem cannot be used to solve $p = q = 0$, so the proof of Theorem 1.4 takes a form different from that for Theorem 1.3. We show that there are either two or four branches of solutions to $p = q = 0$ (depending on the uu, $u\tau$, and $u\lambda$ derivatives of p and q). Each of these solution branches is defined by $\lambda = O(u)$, and the growth rate of the amplitude is the standard 1/2 power. The proof of this theorem, given in section 4, is based on showing that generically solution branches to $p = q = 0$ are determined at quadratic order in u, λ, and τ. Then in section 4.3 we show that the network in Figure 2 generically satisfies (1.18) and (1.19).

In section 5 we prove a third result.

Theorem 1.5. *Generic nilpotent Hopf bifurcation in the feed-forward chain yields two branches of near* 2π*-periodic solutions, the amplitude of one growing as* $\lambda^{\frac{1}{2}}$ *and the amplitude of the other growing as* $\lambda^{\frac{1}{6}}$.

We show, using Poincaré–Birkhoff normal form techniques, that the feed-forward network has two branches of solutions: one with growth rate 1/2 and the other with growth rate 1/6. This step builds on the results in [2]. Then we show that the existence of these branches of solutions implies that

$$p_u(\mathbf{0}) = p_{uu}(\mathbf{0}) = p_{uuu}(\mathbf{0}) = q_u(\mathbf{0}) = q_{uu}(\mathbf{0}) = q_{uuu}(\mathbf{0}) = 0. \tag{1.21}$$

Furthermore, using the Liapunov–Schmidt reduction, we show that

$$p_{uuuu}(\mathbf{0}) \neq 0 \quad \text{and} \quad q_{uuuu}(\mathbf{0}) \neq 0. \tag{1.22}$$

It follows that there are no other branches of solutions to $p = q = 0$, or else these fourth derivatives would also vanish.

In all three theorems and their corresponding examples, flow-invariant subspaces play a vital role since they force the derivatives of p and q to vanish in (1.18) and (1.21). The five-cell network of Figure 3 has no nontrivial invariant subspaces, and thus the derivatives are not constrained. On the other hand, the feed-forward chain in Figure 1 and the three-cell network of Figure 2 both possess a synchrony subspace $S = \{(u, u, v) : u, v \in \mathbf{R}^k\}$. Furthermore, at nilpotent Hopf bifurcations both networks satisfy the following:

(1.23) The network has a flow-invariant subspace, S, that contains the critical eigenspace but does not contain the corresponding generalized eigenspace.

An immediate consequence of (1.23) is the following.

Proposition 1.6. *Suppose that a coupled cell system satisfies* (1.23). *Then at a nilpotent Hopf bifurcation there exists a branch of solutions that grows at* $O(\lambda^{\frac{1}{2}})$, *and* (1.18) *holds.*

Proof. Note that (1.23) implies that we can restrict the vector field to S, and because S does not contain the generalized eigenvectors we can apply the standard Hopf theorem to deduce that there is a standard Hopf bifurcation in S. Thus there is at least one branch that grows as $O(\lambda^{\frac{1}{2}})$. This branch can be parameterized by u, so that

$$p(u, \lambda(u), \tau(u)) \equiv 0 \quad \text{and} \quad q(u, \lambda(u), \tau(u)) \equiv 0.$$

Differentiating both expressions with respect to u and evaluating at the origin yields

$$p_u(\mathbf{0}) + p_\lambda(\mathbf{0})\lambda_u(0) + p_\tau(\mathbf{0})\tau_u(0) = 0 \quad \text{and} \quad q_u(\mathbf{0}) + q_\lambda(\mathbf{0})\lambda_u(0) + q_\tau(\mathbf{0})\tau_u(0) = 0.$$

So by Proposition 1.1, $p_u(\mathbf{0}) = q_u(\mathbf{0}) = 0$. ∎

Observe that the feed-forward chain has an additional flow-invariant subspace $\hat{S} = \{(0, 0, v) : v \in \mathbf{R}^k\}$, and it turns out that $\hat{S}$ also satisfies (1.23). However, $\hat{S}$ is not a synchrony subspace since it is not a polydiagonal, and this implies stronger consequences than in Proposition 1.6. It is this that underlies the additional degeneracies in (1.21).

Theorems 1.3, 1.4, and 1.5 are concerned only with the existence of solutions branches, and we do not consider the stability of the solutions or other dynamical features of the systems. We would expect such an analysis to turn up some interesting features.

In section 6 we give three more examples of three-cell networks that can undergo nilpotent Hopf bifurcation in codimension one. One of these falls into the same category as the network in Figure 2, in that it has two or four branches of solutions given by Theorem 1.4. The other two are similar to the feed-forward chain of Figure 1, in that they have two branches of solutions, one growing at $O(\lambda^{\frac{1}{2}})$ and the other at $O(\lambda^{\frac{1}{6}})$.

Finally, we have placed the (lengthy) expressions for the derivatives of p and q into the appendix, so as not to distract from the flow of the calculations.

2. Reduction with nilpotent normal form. In this section we derive information about the λ and τ derivatives of the Liapunov–Schmidt reduced mapping. Throughout we make the standard assumptions that there exists a trivial (synchronous) equilibrium $F(0,\lambda)=0$, that $J(\lambda)=(\mathrm{d}F)_{0,\lambda}$ has a complex conjugate pair of eigenvalues $\sigma(\lambda)\pm i\omega(\lambda)$ with $\sigma(0)=0$ and $\omega(0)=1$, and that the eigenvalue crosses the imaginary axis with nonzero speed, $\sigma'(0)\neq 0$. In addition we assume that the critical eigenvalues of $(\mathrm{d}F)_{0,0}$ have algebraic multiplicity 2 and geometric multiplicity 1. We also assume throughout that F is a coupled cell system, since nilpotent Hopf bifurcations are nongeneric in systems without a coupled cell structure.

In section 2.1 we set up the generalities of the Liapunov–Schmidt reduction for a nilpotent Hopf bifurcation and obtain a reduced bifurcation problem of the form (1.10). Then in section 2.2 we prove Proposition 1.1. Finally, in section 2.3 we make the additional assumption that F is a *homogeneous* coupled cell system and prove Proposition 1.2.

2.1. The Liapunov–Schmidt reduction. We study the Liapunov–Schmidt reduction of (1.8) onto the kernel of the linearization of Φ at the origin. The linearization of Φ at the origin is given by

$$\mathcal{L}u \equiv (\mathrm{d}\Phi)_{\mathbf{0}}u = \frac{\mathrm{d}u}{\mathrm{d}s} - Ju, \tag{2.1}$$

where $J=(\mathrm{d}F)_{\mathbf{0}}$. Then $\mathcal{K}=\ker\mathcal{L}$ consists of the 2π-periodic solutions to the linear system

$$\frac{\mathrm{d}u}{\mathrm{d}s} = Ju. \tag{2.2}$$

The Liapunov–Schmidt reduction requires an invariant splitting,

$$\mathcal{C}^1_{2\pi} = \mathcal{K}\oplus\mathcal{M},$$
$$\mathcal{C}_{2\pi} = \mathcal{N}\oplus\mathcal{R},$$

where $\mathcal{R}=\operatorname{range}\mathcal{L}$, such that $\mathcal{L}|_{\mathcal{M}}:\mathcal{M}\to\mathcal{R}$ is invertible. As we now show, we can take $\mathcal{M}=\mathcal{K}^{\perp}$ and $\mathcal{N}=\mathcal{K}^*$, where $\mathcal{K}^*$ is the kernel of

$$\mathcal{L}^*u = -\frac{\mathrm{d}u}{\mathrm{d}s} - J^tu, \tag{2.3}$$

which is the adjoint of $\mathcal{L}$ with respect to the inner product

$$\langle u,v\rangle = \frac{1}{2\pi}\int_0^{2\pi}\overline{v(s)^t}u(s)\,\mathrm{d}s. \tag{2.4}$$

Proposition 2.1. *Assume that J has eigenvalues $\pm i$ with algebraic multiplicity 2 and geometric multiplicity 1 and no other eigenvalues on the imaginary axis. Then*

1. $\dim\mathcal{K}=\dim\mathcal{K}^*=2$;
2. *there are bases $\{v_1,v_2\}$ for $\mathcal{K}$ and $\{v_1^*,v_2^*\}$ for $\mathcal{K}^*$ such that $\mathcal{K}$ and $\mathcal{K}^*$ can be identified with $\mathbf{R}^2$ so that $\mathbf{S}^1$ acts on both spaces by counterclockwise rotation;*
3. *there are invariant splittings of $C^1_{2\pi}$ and $\mathcal{C}_{2\pi}$:*

$$\text{(a)}\quad \mathcal{C}_{2\pi}=\mathcal{K}^*\oplus\mathcal{R},$$
$$\text{(b)}\quad \mathcal{C}^1_{2\pi}=\mathcal{K}\oplus\mathcal{K}^{\perp}. \tag{2.5}$$

Proof. We show that $\mathcal{K}$ and $\mathcal{K}^*$ are two-dimensional by constructing bases. Note that if $n > 4$, then J has eigenvalues off the imaginary axis, so solutions not lying in the space spanned by the eigenvectors of $\pm i$ will not be periodic.

Let $c \in \mathbf{C}^n$ be an eigenvector of J with eigenvalue i. Then setting

$$v_1(s) = \mathrm{Re}\{e^{is}c\}, \qquad v_2(s) = \mathrm{Im}\{e^{is}c\} \tag{2.6}$$

forms a basis for $\mathcal{K}$, and in particular, $\dim \mathcal{K} = 2$.

Since J^t has the same eigenvalues as J, and in particular has double eigenvalues $\pm i$ and no other eigenvalues on the imaginary axis, we can construct a basis for $\mathcal{K}^*$ in a similar way. Let $d \in \mathbf{C}^n$ be an eigenvector $J^t d = -id$. Then

$$v_1^*(s) = \mathrm{Re}\{e^{is}d\} \quad \text{and} \quad v_2^*(s) = \mathrm{Im}\{e^{is}d\} \tag{2.7}$$

forms a basis for $\mathcal{K}^*$, and $\dim \mathcal{K}^* = 2$.

We can identify $\mathcal{K}$ and $\mathcal{K}^*$ with $\mathbf{R}^2$ via the mappings

$$(x, y) \mapsto xv_1 + yv_2 \quad \text{and} \quad (x, y) \mapsto xv_1^* + yv_2^*. \tag{2.8}$$

Observe that $\mathbf{S}^1$ acts on $\mathcal{K}$ as

$$\theta \cdot v_1(s) = v_1(s - \theta) = \mathrm{Re}\{e^{-i\theta}e^{is}c\} = \cos\theta\, v_1(s) + \sin\theta\, v_2(s),$$

and similarly

$$\theta \cdot v_2(s) = -\sin\theta\, v_1(s) + \cos\theta\, v_2(s).$$

Therefore, the action of $\mathbf{S}^1$ on $\mathcal{K}$, coordinatized by (2.8), is given by

$$\theta \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2.9}$$

That is, $\theta \in \mathbf{S}^1$ acts by counterclockwise rotation through θ. By the same argument applied to the identification given for $\mathcal{K}^*$ in (2.8), $\mathbf{S}^1$ also acts on $\mathcal{K}^*$ by counterclockwise rotation as in (2.9).

The invariant splittings given in (2.5a,b) follow from the fact that $\mathcal{L}$ is Fredholm of index zero. Specifically, the Fredholm alternative states that

$$\mathcal{R}^{\perp} = \mathcal{K}^*, \tag{2.10}$$

which gives the decomposition in (2.5a). ■

An important point of departure of this case from the standard Hopf bifurcation is that here we have

$$\mathcal{K}^* \perp \mathcal{K}, \tag{2.11}$$

as is shown in Lemma 2.2 below. This is essentially the reason why the first λ and τ derivatives of p and q vanish in Proposition 1.1. The derivatives in Proposition 1.2 come from the fact

that the generalized eigenvectors of J with eigenvalues $\pm i$ are not orthogonal to $\mathcal{K}^*$. Notice also that $\mathcal{K} \perp \mathcal{K}^* = \mathcal{R}^\perp$ implies that $\mathcal{K} \subset \mathcal{R}$.

Let b be a generalized eigenvector of J such that

$$Jb = ib + c, \tag{2.12}$$

and choose b so that

$$b^t \bar{c} = 0.$$

Then define

$$u_1 = \mathrm{Re}\{e^{is} b\} \quad \text{and} \quad u_2 = \mathrm{Im}\{e^{is} b\} \tag{2.13}$$

by analogy with v_j and v_j^*. Note that

$$\mathcal{L} u_j = -v_j.$$

The following lemma summarizes the relations between b, c, and d, and can be contrasted with [3, Chapter VIII, Lemma 2.4].

Lemma 2.2. *Let F be any vector field such that $(\mathrm{d}F)_{0,0}$ is nilpotent, and let b, c, and d be defined as above. Then*

$$c^t \bar{d} = 0 \tag{2.14}$$

and the eigenvector d can be scaled so that

$$b^t \bar{d} = 2. \tag{2.15}$$

Proof. Observe that

$$ib^t \bar{d} = b^t [J^t \bar{d}] = [Jb]^t \bar{d} = (ib + c)^t \bar{d} = ib^t \bar{d} + c^t \bar{d}.$$

Therefore, $c^t \bar{d} = 0$.

Similarly, let v be any eigenvector of J with eigenvalue μ. Then

$$\mu v^t \bar{d} = [Jv]^t \bar{d} = v^t J^t \bar{d} = iv^t \bar{d}. \tag{2.16}$$

Thus every eigenvector of J with eigenvalue different from i is orthogonal to $\bar{d}$. But $c^t \bar{d} = 0$, so $b^t \bar{d}$ must be nonzero or $\bar{d}$ is orthogonal to every eigenvector of J, which is a contradiction. Therefore $b^t \bar{d} \neq 0$, and we can scale d so that $b^t \bar{d} = 2$. ∎

Before continuing with the reduction, we give the following useful formulas concerning v_j, v_j^*, and u_j.

Lemma 2.3.

$$\frac{\mathrm{d}v_1}{\mathrm{d}s} = -v_2, \qquad \frac{\mathrm{d}v_1}{\mathrm{d}s} = -v_2, \qquad \frac{\mathrm{d}u_1}{\mathrm{d}s} = -u_2, \qquad \frac{\mathrm{d}u_2}{\mathrm{d}s} = u_1, \tag{2.17}$$

$$\langle v_j^*, v_k \rangle = 0, \qquad \langle v_j^*, u_k \rangle = \delta_{jk}, \tag{2.18}$$

where $\langle \cdot, \cdot \rangle$ *is the inner product defined in* (2.4).

Proof. These formulas are straightforward calculations from the definitions of v_j, v_j^*, and u_j in (2.6), (2.7), and (2.13). We give two examples of these calculations; the others follow very similar lines.

Observe that (2.6) implies

$$\frac{\mathrm{d}v_1}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}\mathrm{Re}\{e^{is}c\} = \mathrm{Re}\{ie^{is}c\} = -\mathrm{Im}\{e^{is}c\} = -v_2.$$

Similar calculations yield the other derivatives in (2.17).

We verify only that $\langle v_j^*, u_k \rangle = \delta_{jk}$. Since

$$\begin{aligned} v_1^* &= \mathrm{Re}\{e^{is}d\} = \tfrac{1}{2}(e^{is}d + e^{-is}\bar{d}), & v_2^* &= \mathrm{Im}\{e^{is}d\} = -\tfrac{i}{2}(e^{is}d - e^{-is}\bar{d}), \\ u_1 &= \mathrm{Re}\{e^{is}d\} = \tfrac{1}{2}(e^{is}b + e^{-is}\bar{b}), & u_2 &= \mathrm{Im}\{e^{is}d\} = -\tfrac{i}{2}(e^{is}b - e^{-is}\bar{b}), \end{aligned}$$

we can write

$$v_j^* = \frac{(-i)^{j-1}}{2}(e^{is}d + (-1)^{j-1}e^{-is}\bar{d}) \quad \text{and} \quad u_k = \frac{(-i)^{k-1}}{2}(e^{is}b + (-1)^{k-1}e^{-is}\bar{b}).$$

Then using (2.4) and the fact that $b^t\bar{d} = 2$ by (2.15),

$$\begin{aligned} \langle v_j^*, u_k \rangle &= \frac{i^{k-1}(-i)^{j-1}}{8\pi}\int_0^{2\pi}(e^{-is}\bar{b} + (-1)^{k-1}e^{is}b)^t(e^{is}d + (-1)^{j-1}e^{-is}\bar{d})\mathrm{d}s \\ &= \frac{i^{k-1}(-i)^{j-1}}{8\pi}\int_0^{2\pi}\big((-1)^{k-1}e^{2is}b^td + (-1)^{j+k}b^t\bar{d} \\ &\qquad + \bar{b}^td + (-1)^{j-1}e^{-2is}\bar{b}^t\bar{d}^t\big)\mathrm{d}s \\ &= \frac{1}{2}i^{k-1}(-i)^{j-1}(1 + (-1)^{j+k}) \\ &= \delta_{jk}. \qquad \blacksquare \end{aligned}$$

To continue the reduction, let $E : \mathcal{C}_{2\pi} \to \mathcal{R}$ be the projection with $\ker E = \mathcal{K}^*$, and write $x \in \mathcal{C}^1_{2\pi}$ as $x = v + w$, where $v \in \mathcal{K}$ and $w \in \mathcal{K}^\perp$. Then $\Phi(v + w, \lambda, \tau) = 0$ if and only if

$$\begin{aligned} &\text{(a)} \quad E\Phi(v + w, \lambda, \tau) = 0, \\ &\text{(b)} \quad (I - E)\Phi(v + w, \lambda, \tau) = 0, \end{aligned} \tag{2.19}$$

where $I - E$ is the complementary projection of $\mathcal{C}_{2\pi}$ onto $\mathcal{K}^*$ with kernel $\mathcal{R}$. The differential of $E\Phi(v + w, \lambda, \tau)$ with respect to w at the origin is just $\mathcal{L}|_{\mathcal{K}^\perp}$, and $\mathcal{L}|_{\mathcal{K}^\perp} : \mathcal{K}^\perp \to \mathcal{R}$ is invertible because $\mathcal{L}$ is Fredholm of index zero. Thus the implicit function theorem implies that (2.19a) can be solved for $w = W(v, \lambda, \tau)$, where $W : \mathcal{K} \times \mathbf{R} \times \mathbf{R} \to \mathcal{K}^\perp$ is such that $W(\mathbf{0}) = 0$ and

$$E\Phi(xv_0 + W(xv_0, \lambda, \tau), \lambda, \tau) \equiv 0 \tag{2.20}$$

for any $v_0 \in \mathcal{K}$ and $x \in \mathbf{R}$. Solving (2.19b) is therefore equivalent to solving $\Phi(u, \lambda, \tau) = 0$, and hence, to finding the near 2π-periodic solutions to (1.1).

The following lemma greatly simplifies a number of formulas later on.

Lemma 2.4.

$$W_x(\mathbf{0}) = W_\alpha(\mathbf{0}) = W_{\alpha\beta}(\mathbf{0}) = 0, \tag{2.21}$$

where α and β are placeholders for λ and τ.

Proof. Differentiating (2.20) with respect to x and evaluating all derivatives at the origin yields

$$\mathrm{d}\Phi(v_0 + W_x(\mathbf{0})) = 0$$

and therefore

$$0 = \mathcal{L}v_0 + \mathcal{L}W_x(\mathbf{0}) = \mathcal{L}W_x(\mathbf{0}).$$

However, $W_x(\mathbf{0}) \in \mathcal{K}^\perp$, and thus $W_x(\mathbf{0}) = 0$.

Observe that setting $v = 0$, $w = W(0,\lambda,\tau)$ in (2.19a) yields

$$E\left((1+\tau)\frac{\mathrm{d}}{\mathrm{d}s}W(0,\lambda,\tau) - F(W(0,\lambda,\tau),\lambda)\right) \equiv 0.$$

This is solved for $W(0,\lambda,\tau) \equiv 0$ since $F(0,\lambda) \equiv 0$, and this must be the only solution since the implicit function theorem guarantees that $W(0,\lambda,\tau)$ is the unique solution to (2.19a). This implies that all λ and τ derivatives of W evaluated at the origin vanish. ∎

Define the reduced mapping $\phi : \mathcal{K} \times \mathbf{R} \times \mathbf{R} \to \mathcal{K}^*$ by

$$\phi(v,\lambda,\tau) = (I - E)\Phi(v + W(v,\lambda,\tau),\lambda,\tau). \tag{2.22}$$

Then for $j = 1, 2$

$$\phi_j(x,y,\lambda,\tau) = \left\langle v_j^*, \phi(xv_1 + yv_2,\lambda,\tau)\right\rangle.$$

Since $\mathcal{K}$ and $\mathcal{K}^*$ are invariant subspaces under the action of $\mathbf{S}^1$ as specified in (2.9), it follows from Proposition 3.3 of Chapter VII in [3] that ϕ commutes with this action of $\mathbf{S}^1$ on $\mathcal{K}$ and $\mathcal{K}^*$. Therefore, by Proposition 2.3 of Chapter VIII in [3], the reduced mapping has the form in (1.10), where x and y come from the identification of $\mathcal{K}$ with $\mathbf{C}$ given by (2.8).

Solutions to (1.7) that are 2π-periodic correspond to solutions to $\phi(x,y,\lambda,\tau) = 0$, and by (1.10) these solutions are given by $x = y = 0$ or $p = q = 0$. The former correspond to the trivial steady-state solutions $z = 0$ to (1.7), whereas the latter (with $x^2 + y^2 > 0$) correspond to nonconstant 2π-periodic solutions to (1.7).

Because ϕ is $\mathbf{S}^1$-equivariant, solutions to $\phi(x,y,\lambda,\tau) = 0$ come in group orbits, and we can therefore rotate the plane so that $y = 0$ and

$$x \geq 0. \tag{2.23}$$

In particular,

$$\phi_1(x,0,\lambda,\tau) = \langle v_1^*, \phi(xv_1,\lambda,\tau)\rangle = p(x^2,\lambda,\tau)x,$$
$$\phi_2(x,0,\lambda,\tau) = \langle v_2^*, \phi(xv_1,\lambda,\tau)\rangle = q(x^2,\lambda,\tau)x.$$

2.2. Linear terms in the reduced equation. We now consider the λ and τ derivatives of p and q. The first x derivatives of ϕ,

$$\frac{\partial \phi_1}{\partial x}(\mathbf{0}) = p(\mathbf{0}) \quad \text{and} \quad \frac{\partial \phi_2}{\partial x}(\mathbf{0}) = q(\mathbf{0}),$$

are both zero because linear terms vanish in the Liapunov–Schmidt reduction. Indeed, suppose that linear terms remained in $\phi = (I - E)\Phi$. Then we could have used the implicit function theorem to solve for these terms as we did for (2.19a). Since we could not do this, there can be no linear terms in ϕ. The second x derivatives,

$$\frac{\partial^2 \phi_1}{\partial x^2}(\mathbf{0}) = p_x(\mathbf{0}) \quad \text{and} \quad \frac{\partial^2 \phi_2}{\partial x^2}(\mathbf{0}) = q_x(\mathbf{0}),$$

are also clearly zero because p and q are quadratic in x.

Proof of Proposition 1.1. For any parameter α, the general formula for $\phi_{\alpha x}$ is

$$\phi_{\alpha x} = (I - E)\left(\mathrm{d}\Phi_\alpha(v_1 + W_x) + \mathrm{d}\Phi(W_{\alpha x}) + \mathrm{d}^2\Phi(v_1 + W_x, W_\alpha)\right), \tag{2.24}$$

from (A.5), but we can simplify this by using (2.21). With these results, and bearing in mind that

(a) $\mathrm{d}^2\Phi(\cdot,\cdot)$ is bilinear, so any terms of the form $\mathrm{d}\Phi_\alpha(\cdot, 0)$ vanish, and

(b) $\operatorname{range} \mathrm{d}\Phi = \ker(I - E)$, so any terms in the range of $\mathrm{d}\Phi$ also vanish,

formula (2.24) becomes

$$\phi_{\alpha x} = (I - E)\left(\mathrm{d}\Phi_\alpha(v_1)\right). \tag{2.25}$$

To verify (1.12), (2.25) implies

$$\phi_{\tau x} = (I - E)\left(\mathrm{d}\Phi_\tau(v_1)\right).$$

Observe that

$$\mathrm{d}\Phi_\tau(v_1) = \frac{\mathrm{d}v_1}{\mathrm{d}s} = -v_2,$$

since $\mathrm{d}\Phi_\tau = \frac{\mathrm{d}}{\mathrm{d}s}$ and by (2.17). Since $v_2 \in \mathcal{K} \subset \mathcal{R} = \ker(I - E)$, it follows that $\phi_{\tau x} = 0$ and therefore that $p_\tau(0,0,0) = q_\tau(0,0,0) = 0$. This verifies (1.12).

To verify (1.13), (2.25) implies

$$\phi_{\lambda x} = (I - E)\left(\mathrm{d}\Phi_\lambda(v_1)\right).$$

Observe that

$$\Phi_\lambda(x, \lambda, \tau) = -F_\lambda(x, \lambda) \tag{2.26}$$

and that $F(x, \lambda)$ can be written as

$$F(x, \lambda) = J(\lambda)x + h.o.t.$$

Thus $F_\lambda(x,\lambda) = J'(\lambda)x + h.o.t.$, and therefore, evaluating at the origin,

$$\mathrm{d}\Phi_\lambda = -\mathrm{d}F_\lambda(x,\lambda)|_{\mathbf{0}} = -J'(0). \tag{2.27}$$

In (2.28) in Lemma 2.5, below, we prove that

$$J'(0)c = (\sigma'(0) + i\omega'(0))c - (J - iI_n)c'(0),$$

from which it follows that

$$\begin{aligned} J'(0)v_1 &= \mathrm{Re}\{e^{is}J'(0)c\} \\ &= \sigma'(0)\mathrm{Re}\{e^{is}c\} + \omega'(0)\mathrm{Re}\{ie^{is}c\} - \mathrm{Re}\{(J - iI_n)c'(0)\} \\ &= \sigma'(0)v_1 - \omega'(0)v_2 - \mathrm{Re}\{e^{is}(J - iI_n)c'(0)\}. \end{aligned}$$

So, recalling from (2.18) that $\langle v_i^*, v_j\rangle = 0$,

$$\begin{aligned} \left\langle v_j^*, \mathrm{d}\Phi_\lambda(v_1)\right\rangle &= \left\langle v_j^*, \sigma'(0)v_1 - \omega'(0)v_2\right\rangle - \left\langle v_j^*, \mathrm{Re}\{e^{is}(J - iI_n)c'(0)\}\right\rangle \\ &= \left\langle v_j^*, \mathrm{Re}\{e^{is}(J - iI_n)c'(0)\}\right\rangle. \end{aligned}$$

Observe that

$$[(J - iI_n)c'(0)]^t\bar{d} = c'(0)^t(J - iI_n)^t\bar{d} = 0,$$

since $\bar{d}$ is an eigenvector of J^t with eigenvalue i. Therefore,

$$\begin{aligned} \left\langle v_1^*, \mathrm{Re}\{e^{is}(J - iI_n)c'(0)\}\right\rangle &= \frac{1}{2}\mathrm{Re}\{[(J - iI_n)c'(0)]^t\bar{d}\} = 0, \\ \left\langle v_2^*, \mathrm{Re}\{e^{is}(J - iI_n)c'(0)\}\right\rangle &= -\frac{1}{2}\mathrm{Im}\{[(J - iI_n)c'(0)]^t\bar{d}\} = 0, \end{aligned}$$

which verifies (1.13). ■

The following calculations are needed.

Lemma 2.5. *Let $\mu(\lambda) = \sigma(\lambda) + i\omega(\lambda)$ be the eigenvalue of $J(\lambda)$ such that $\mu(0) = i$ with eigenvector $c(\lambda)$ such that $c(0) = c$. Let $b(\lambda)$ be the corresponding generalized eigenvector such that $J(\lambda)b(\lambda) = \mu(\lambda)b(\lambda) + c(\lambda)$. Then*

$$J'(0)c = \mu'(0)c - (J - iI_n)c'(0), \tag{2.28}$$

$$J''(0)c = \mu''(0)c - 2(J'(0) - \mu'(0)I_n)c'(0) - (J - iI_n)c''(0), \tag{2.29}$$

$$J'(0)b = \mu'(0)b - (J - iI_n)b'(0) + c'(0). \tag{2.30}$$

Proof. Since $c(\lambda)$ is an eigenvector of $J(\lambda)$ with eigenvalue $\mu(\lambda)$,

$$J(\lambda)c(\lambda) = \mu(\lambda)c(\lambda). \tag{2.31}$$

Differentiating (2.31) with respect to λ and evaluating at $\lambda = 0$ gives

$$J'(0)c(0) + Jc'(0) = \mu'(0)c(0) + \mu(0)c'(0) = \mu'(0)c(0) + ic'(0),$$

since $\mu(0) = i$, and this rearranges to give (2.28).

Differentiating (2.31) twice with respect to λ, we obtain

$$J''(0)c(0) + 2J'(0)c'(0) + Jc''(0) = \mu''(0)c(0) + 2\mu'(0)c'(0) + ic''(0),$$

which rearranges to give (2.29).

Similarly, for the generalized eigenvector we have

$$J(\lambda)b(\lambda) = \mu(\lambda)b(\lambda) + c(\lambda),$$

and differentiating this with respect to λ and evaluating at the origin, we obtain

$$J'(0)b(0) + Jb'(0) = \mu'(0)b(0) + \mu(0)b'(0) + c'(0),$$

which rearranges to give (2.30). ■

2.3. Quadratic terms in the reduced equation. To prove Proposition 1.2 we require Lemmas 2.6 and 2.7, which we prove at the end of the section.

Lemma 2.6. *For any vector field with nilpotent linearization*

$$W_{\tau x} = -u_2, \tag{2.32}$$

$$W_{\lambda x} = -\sigma'(0)u_1 + \omega'(0)u_2 + \mathrm{Re}\{e^{is}(c'(0) - \zeta c)\}, \tag{2.33}$$

where

$$\zeta = \frac{1}{\|c\|^2} c'(0)^t \bar{c} \tag{2.34}$$

so that ζc is the projection of $c'(0)$ onto the critical eigenspace of J.

Lemma 2.7. *Suppose that F is a coupled cell system such that $(\mathrm{d}F)_{\mathbf{0}}$ is nilpotent and that*

(a) *each cell in the network has identical linearized internal dynamics,*

(b) *the linearized coupling between any two cells takes the form $mB(\lambda)$, where $B(\lambda)$ is a $k \times k$ matrix and $m \in \mathbf{R}$.*

Then

$$c^{(l)}(0)^t \bar{d} = 0 \qquad \text{for all } l \geq 0, \tag{2.35}$$

$$[J^{(k)}(0)c^{(l)}(0)]^t \bar{d} = 0 \qquad \text{for all } k, l \geq 0, \tag{2.36}$$

where

$$J^{(k)}(0) = \left.\frac{\partial^k J(\lambda)}{\partial \lambda^k}\right|_{\lambda=0} \quad \text{and} \quad c^{(l)}(0) = \left.\frac{\partial^l c(\lambda)}{\partial \lambda^l}\right|_{\lambda=0}.$$

Remark 2.8. Assumptions (a) and (b) in Lemma 2.7 are instant if the network is homogeneous.

Proof of Proposition 1.2. The general formula for $\phi_{\alpha\beta x}$ is given by (A.6), but the same arguments that we used to derive (2.25) can be used to obtain

$$\phi_{\alpha\beta x} = (I - E)\big(\mathrm{d}\Phi_{\alpha\beta}(v_1) + \mathrm{d}\Phi_{\alpha}(W_{\beta x}) + \mathrm{d}\Phi_{\beta}(W_{\alpha x})\big). \tag{2.37}$$

To verify (1.14), we have from (2.37)

$$\phi_{\tau\tau x} = (I-E)\big(\mathrm{d}\Phi_{\tau\tau}(v_1) + 2\mathrm{d}\Phi_\tau(W_{\tau x})\big) = (I-E)\left(2\frac{\mathrm{d}}{\mathrm{d}s}W_{\tau x}\right),$$

where the second equality follows because $\mathrm{d}\Phi_\tau = \frac{\mathrm{d}}{\mathrm{d}s}$ and $\Phi_{\tau\tau} = 0$. By the formula for $W_{\tau x}$ in (2.32) and the derivative given in (2.17),

$$\frac{\mathrm{d}}{\mathrm{d}s}W_{\tau x} = -\frac{\mathrm{d}}{\mathrm{d}s}u_2 = -u_1.$$

So, by (2.18), the jth component of $\phi_{\tau\tau x}$ is

$$\phi_{\tau\tau x,j} = -2\left\langle v_j^*, u_1\right\rangle = -2\delta_{j1}, \qquad j = 1, 2.$$

For the $\lambda\lambda$ derivative in (1.15), we have from (2.37)

$$\phi_{\lambda\lambda x} = (I-E)\big(\mathrm{d}\Phi_{\lambda\lambda}(v_1) + 2\mathrm{d}\Phi_\lambda(W_{\lambda x})\big). \tag{2.38}$$

Consider the first term $(I-E)(\mathrm{d}\Phi_{\lambda\lambda}(v_1))$, and note that (2.27) can be extended to the second λ derivative, to yield

$$\mathrm{d}\Phi_{\lambda\lambda}(v_1) = -\mathrm{d}F_{\lambda\lambda}(v_1) = -J''(0)v_1 = -\mathrm{Re}\{e^{is}J''(0)c\}.$$

Thus, since $[J''(0)c]^t\bar{d} = 0$ by (2.36), we have

$$\phi_{\lambda\lambda x,j} = \left\langle v_j^*, \mathrm{d}\Phi_{\lambda\lambda}(v_1)\right\rangle = -\left\langle v_j^*, \mathrm{Re}\{e^{is}J''(0)c\}\right\rangle = 0.$$

Therefore by (2.27), equation (2.38) becomes

$$\phi_{\lambda\lambda x} = 2(I-E)\big(\mathrm{d}\Phi_\lambda(W_{\lambda x}\big) = -2(I-E)\big(J'(0)W_{\lambda x}\big),$$

and by formula (2.33) for $W_{\lambda x}$

$$J'(0)W_{\lambda x} = -\sigma'(0)\mathrm{Re}\{e^{is}J'(0)b\} + \omega'(0)\mathrm{Im}\{e^{is}J'(0)b\} + \mathrm{Re}\{e^{is}J'(0)(c'(0)-\zeta c)\}.$$

Using (2.30) for $J'(0)b$ and rearranging, this becomes

$$J'(0)W_{\lambda x} = -\mathrm{Re}\big\{e^{is}\big(\mu'(0)^2 b - \mu'(0)(J-iI_n)b'(0) + \mu'(0)c'(0) - J'(0)(c'(0)-\zeta c)\big)\big\},$$

so that

$$\begin{aligned}\phi_{\lambda\lambda x,1} &= -2\left\langle v_1^*, J'(0)W_{\lambda x}\right\rangle \\ &= \mathrm{Re}\big\{\mu'(0)^2 b^t\bar{d} - \mu'(0)[(J-iI_n)b'(0)]^t\bar{d} \\ &\qquad + \mu'(0)c'(0)^t\bar{d} - [J'(0)c'(0)]^t\bar{d} + [\zeta J'(0)c]^t\bar{d}\big\},\end{aligned} \tag{2.39}$$

with $\phi_{\lambda\lambda x,2}$ being minus the imaginary part of the same expression.

Observe that the last three terms in (2.39) vanish by Lemma 2.7 and that the second term vanishes because

$$[(J - iI_n)b'(0)]^t\bar{d} = b'(0)^t[(J - iI_n)^t\bar{d}] = 0,$$

since $\bar{d}$ is an eigenvector of J^t with eigenvalue i. Hence,

$$\begin{aligned}\phi_{\lambda\lambda x,1} &= -2\left\langle v_1^*, J'(0)W_{\lambda x}\right\rangle = \mathrm{Re}\{\mu'(0)^2 b^t\bar{d}\},\\ \phi_{\lambda\lambda x,2} &= -2\left\langle v_2^*, J'(0)W_{\lambda x}\right\rangle = -\mathrm{Im}\{\mu'(0)^2 b^t\bar{d}\},\end{aligned}$$

which, given that $b^t\bar{d} = 2$ by (2.15), proves (1.15).

Turning now to (1.16), we have from (2.37)

$$\phi_{\lambda\tau x} = (I - E)\big(\mathrm{d}\Phi_{\lambda\tau}(v_1) + \mathrm{d}\Phi_\lambda(W_{\tau x}) + \mathrm{d}\Phi_\tau(W_{\lambda x})\big). \tag{2.40}$$

First note that $\mathrm{d}\Phi_{\lambda\tau} = 0$. For the second term in (2.40), (2.27) and (2.32) imply that

$$\mathrm{d}\Phi_\lambda(W_{\tau x}) = J'(0)u_2 = \mathrm{Im}(e^{is}J'(0)b).$$

Expanding $J'(0)b$ by (2.30), we obtain

$$\begin{aligned}\mathrm{d}\Phi_\lambda(W_{\tau x}) &= \mathrm{Im}\{e^{is}(\mu'(0)b - (J - iI_n)b'(0) + c'(0))\}\\ &= \omega'(0)u_1 + \sigma'(0)u_2 - \mathrm{Im}\{e^{is}[(J - iI_n)b'(0) + c'(0)]\}.\end{aligned} \tag{2.41}$$

For the third term in (2.40), we have from (2.33) and the fact that $\mathrm{d}\Phi_\tau = \frac{\mathrm{d}}{\mathrm{d}s}$ that

$$\begin{aligned}\mathrm{d}\Phi_\tau(W_{\lambda x}) &= -\sigma'(0)\dot{u}_1 + \omega'(0)\dot{u}_2 + \mathrm{Re}\{ie^{is}(c'(0) - \zeta c)\}\\ &= \omega'(0)u_1 + \sigma'(0)u_2 - \mathrm{Im}\{e^{is}(c'(0) - \zeta c)\}.\end{aligned}$$

Putting this together with (2.41) and taking inner products yields

$$\begin{aligned}\left\langle v_j^*, \mathrm{d}\Phi_\lambda(W_{\tau x}) + \mathrm{d}\Phi_\tau(W_{\lambda x})\right\rangle &= \langle v_j^*, 2\omega'(0)u_1 + 2\sigma'(0)u_2\\ &\qquad - \mathrm{Im}\{e^{is}[(J - iI_n)b'(0) + c'(0) + c'(0) - \zeta c]\}\rangle\\ &= 2\omega'(0)\left\langle v_j^*, u_1\right\rangle + 2\sigma'(0)\left\langle v_j^*, u_2\right\rangle,\end{aligned}$$

where the second line follows because $c'(0)^t\bar{d} = 0 = c^t\bar{d}$ and $[(J - iI_n)b'(0)]^t\bar{d} = b'(0)^t(J - iI_n)^t\bar{d} = 0$. The formulas in (1.16) then follow from Lemma 2.3. ■

Proof of Lemma 2.6. To show (2.32), we have from formula (A.14) and from the facts that $W_x = W_\tau = 0$, $\mathrm{d}\Phi_\tau = \frac{\mathrm{d}}{\mathrm{d}s}$, and $\mathcal{L}u_2 = -v_2$ that

$$W_{\tau x} = -\mathcal{L}^{-1}E(\mathrm{d}\Phi_\tau(v_1)) = -\mathcal{L}^{-1}E\left(\frac{\mathrm{d}v_1}{\mathrm{d}s}\right) = \mathcal{L}^{-1}E(v_2) = -u_2,$$

as required.

To prove (2.33), we use (A.14) and (2.27) to obtain

$$W_{\lambda x} = -\mathcal{L}^{-1}E(\mathrm{d}\Phi_\lambda(v_1)) = \mathcal{L}^{-1}E(J'(0)v_1).$$

So, using formula (2.28) for $J'(0)v_1 = \mathrm{Re}\{e^{is}J'(0)c\}$, $W_{\lambda x}$ is the solution to the differential equation

$$\mathcal{L}W_{\lambda x} = \sigma'(0)v_1 - \omega'(0)v_2 - \mathrm{Re}\{e^{is}(J - iI_n)c'(0)\}. \tag{2.42}$$

Write $c'(0) = c_0 + c_1$, where $c_0 \in \ker(J - iI_n)$ and $c_1 \in \ker(J - iI_n)^{\perp}$. Then

$$c_0 = \frac{c'(0)^t \bar{c}}{\|c\|^2} c$$

and

$$\mathrm{Re}\{e^{is}(J - iI_n)c'(0)\} = \mathrm{Re}\left\{e^{is}(J - iI_n)\left(c'(0) - \frac{c'(0)^t\bar{c}}{\|c\|}c\right)\right\}.$$

With this substitution it is straightforward to check that (2.42) is solved for $W_{\lambda x}$ as in (2.33). ■

Proof of Lemma 2.7. For any two matrices M and N we define $[N]_M$ to be the matrix obtained by replacing every entry m_{ij} in M with the block $m_{ij}N$. Using this notation, the Jacobian of a coupled cell system satisfying (a) and (b) has the form

$$\begin{aligned} J(\lambda) &= \begin{pmatrix} A(\lambda) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A(\lambda) \end{pmatrix} + \begin{pmatrix} m_{11}B(\lambda) & \cdots & m_{1n}B(\lambda) \\ \vdots & \ddots & \vdots \\ m_{n1}B(\lambda) & \cdots & m_{nn}B(\lambda) \end{pmatrix} \\ &= [A(\lambda)]_{I_n} + [B(\lambda)]_M, \end{aligned} \tag{2.43}$$

where $A(\lambda)$ is the linearized internal dynamics, $B(\lambda)$ is the linearized coupling, and $m_{ij} \in \mathbf{R}$.

As shown in [6], the eigenvalues of a homogeneous network are the eigenvalues of the $k \times k$ matrices $A(\lambda) + \mu_j B(\lambda)$, where $\mu_1, \ldots, \mu_n$ are the eigenvalues of M. Fix μ_c such that $A(0) + \mu_c B(0)$ has a critical eigenvalue, and observe that μ_c is independent of λ. Note that μ_c has algebraic multiplicity 2 and geometric multiplicity 1. Then $A(0)^t + \mu_c B(0)^t$ also has a critical eigenvalue.

The eigenvectors $u(\lambda)$ of $J(\lambda)$ have the form

$$u(\lambda) = \begin{pmatrix} v_1 w(\lambda) \\ \vdots \\ v_n w(\lambda) \end{pmatrix} = [w(\lambda)]_v,$$

where $w(\lambda) \in \mathbf{C}^k$ is an eigenvector of $A(\lambda) + \mu_j B(\lambda)$ and $v \in \mathbf{C}^n$ is an eigenvector of M. Note that v does not depend on λ since the matrix M does not depend on λ. Thus we can write

$$c(\lambda) = [\beta(\lambda)]_\alpha \quad \text{and} \quad d(\lambda) = [\delta(\lambda)]_\gamma,$$

where α and γ are respectively the appropriate eigenvectors of M and M^t with eigenvalues μ_c, and $\beta(\lambda)$ and $\delta(\lambda)$ are respectively the eigenvectors of $A(\lambda) + \mu_c B(\lambda)$ and $A(\lambda)^t + \mu_c B(\lambda)^t$ with eigenvalues $\sigma(\lambda) + i\omega(\lambda)$. Then

$$c(\lambda)^t \bar{d}(\lambda) = ([\beta(\lambda)]_\alpha)^t [\bar{\delta}(\lambda)]_{\bar{\gamma}} = (\alpha^t \bar{\gamma})(\beta(\lambda)^t \bar{\delta}(\lambda)). \tag{2.44}$$

Observe that $\beta(0)^t\bar{\delta}(0) \neq 0$ since $\beta(0)$ and $\bar{\delta}(0)$ are respectively the eigenvectors of $A(0) + \mu_c B(0)$ and $(A(0) + \mu_c B(0))^t$ with the same simple eigenvalue i, and recall from Lemma 2.2 that $c^t\bar{d} = 0$ since $(\mathrm{d}F)_{\mathbf{0}}$ is nilpotent. Then (2.44) implies $\alpha^t\bar{\gamma} = 0$.

Hence,

$$c^{(l)}(0)^t\bar{d}(0) = ([\beta^{(l)}(0)]_\alpha)^t[\bar{\delta}(0)]_{\bar{\gamma}} = (\alpha^t\bar{\gamma})(\beta^{(l)}(0)^t\bar{\delta}(0)) = 0,$$

which proves (2.35). To show (2.36), use (2.43) to calculate

$$\begin{aligned}[J^{(k)}(0)c^{(l)}(0)]^t\bar{d}(0) &= \left(J^{(k)}(0)[\beta^{(l)}(0)]_\alpha\right)^t[\bar{\delta}(0)]_{\bar{\gamma}} \\ &= (\alpha^t\bar{\gamma})\left((A^{(k)}(0) + \mu_c B^{(k)}(0))\beta^{(l)}(0)\right)^t\bar{\delta}(0) \\ &= 0\end{aligned}$$

since $\alpha^t\bar{\gamma} = 0$. ∎

3. Hopf bifurcation with linear u terms. In this section we consider nilpotent Hopf bifurcations in coupled cell systems that satisfy (1.17). Note that we are making two assumptions in (1.17); that the network architecture does not force $q_u(\mathbf{0})$ to vanish, and that the bifurcation is generic for that network. Theorem 1.3 is proved below, and the corresponding bifurcation diagram is shown in Figure 4. We then show that there exists a vector field on the five-cell network of Figure 3 such that $q_u(\mathbf{0}) \neq 0$. Hence, by Theorem 1.3, this system has two branches of solutions that grow linearly with λ.

Proof of Theorem 1.3. By Propositions 1.1 and 1.2, we can write

$$p(u, \lambda, \tau) = p_u(\mathbf{0})u - \tau^2 + (\sigma'(0)^2 - \omega'(0)^2)\lambda^2 + 2\omega'(0)\lambda\tau + \cdots, \tag{3.1}$$

$$q(u, \lambda, \tau) = q_u(\mathbf{0})u - 2\sigma'(0)\omega'(0)\lambda^2 + 2\sigma'(0)\lambda\tau + \cdots, \tag{3.2}$$

and we require solutions to $p = q = 0$. Assuming (1.17) and applying the implicit function theorem, we can solve $q = 0$ near the origin for

$$u = \frac{2\sigma'(0)\lambda}{q_u(\mathbf{0})}(\omega'(0)\lambda - \tau) + \cdots.$$

Substituting this into (3.1) and setting $p = 0$ yields the following equation in τ and λ:

$$\tau^2 + 2\left(\frac{p_u(\mathbf{0})}{q_u(\mathbf{0})}\sigma'(0) - \omega'(0)\right)\lambda\tau + \left(\omega'(0)^2 - \frac{2p_u(\mathbf{0})}{p_u(\mathbf{0})}\sigma'(0)\omega'(0) - \sigma'(0)^2\right)\lambda^2 + \cdots = 0,$$

where quadratic terms are solved for

$$\tau = \left[\omega'(0) - \frac{\sigma'(0)}{q_u(\mathbf{0})}\left(p_u(\mathbf{0}) \pm \sqrt{p_u(\mathbf{0})^2 + q_u(\mathbf{0})^2}\right)\right]\lambda. \tag{3.3}$$

Since these solutions are real and distinct we can apply the recognition problem for simple bifurcation using τ as the state variable and λ as the bifurcation parameter to prove that higher order terms are unimportant. See [3, Proposition 9.3, p. 95].

Substitute (3.3) into $q = 0$, for q as in (3.2), and rearrange to obtain

$$u = \frac{2\sigma'(0)^2}{q_u(\mathbf{0})^2}\left(p_u(\mathbf{0}) \pm \sqrt{p_u(\mathbf{0})^2 + q_u(\mathbf{0})^2}\right)\lambda^2.$$

Note that only the + sign is relevant, since we require $u = x^2 > 0$. Thus, since $u = x^2$ grows as $O(\lambda^2)$, x grows as $O(\lambda)$, and we obtain the bifurcation diagram in Figure 4. ∎

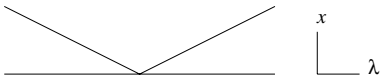


Figure 4. *Bifurcation diagram for nilpotent Hopf bifurcation with* $q_u(\mathbf{0}) \neq 0$.

The five-cell network of Figure 3. To illustrate the results of this section we consider the five-cell system shown in Figure 3 and defined by (1.6) with the vector field defined by

$$f(x_i; x_j, x_k, x_l) = \lambda x_i - x_j - x_k - x_l - x_i^3. \tag{3.4}$$

We compute $p_u(\mathbf{0})$ and $q_u(\mathbf{0})$ for this system and show that $p_u(\mathbf{0}) \neq 0 \neq q_u(\mathbf{0})$.

Recall that

$$p_u(\mathbf{0}) = \frac{1}{2}p_{xx}(\mathbf{0}) = \frac{1}{2}\frac{\partial^3\phi_1}{\partial x^3}(\mathbf{0}) \quad \text{and} \quad q_u(\mathbf{0}) = \frac{1}{2}q_{xx}(\mathbf{0}) = \frac{1}{2}\frac{\partial^3\phi_2}{\partial x^3}(\mathbf{0}).$$

By (A.1) we have

$$\phi_{xxx} = (I - E)(\mathrm{d}^3\Phi(v_1, v_1, v_1) + 3\mathrm{d}^2\Phi(v_1, W_{xx})) \tag{3.5}$$

since $\mathrm{d}\Phi(W_{xxx}) \in \mathcal{R} = \ker(I - E)$.

As a reminder of the higher order differentials of Φ, let $u_1, \ldots, u_m$ be any functions in $\mathcal{C}_{2\pi}$, and let $u_{i,j}$ denote the jth component of u_i. By definition, evaluating all derivatives at the origin, we have

$$\begin{aligned}(\mathrm{d}^m\Phi)_{0,0,0}(u_1, \ldots, u_m) &= \frac{\partial}{\partial t_1} \cdots \frac{\partial}{\partial t_m}\Phi(t_1u_1 + \cdots + t_mu_m, 0, 0)\Big|_{t=0} \\ &= -\sum_{i_1,\ldots,i_m=1}^{n} \frac{\partial^m F}{\partial x_{i_1} \ldots \partial x_{i_m}}\Big|_{0,0} u_{1,i_1} \cdots u_{m,i_m} \\ &= -(\mathrm{d}^m F)_{0,0,0}(u_1, \ldots, u_m).\end{aligned}$$

See [3, pp. 31–32]. Letting $v_{1,j}$ denote the jth component of v_1 gives

$$\begin{aligned}\mathrm{d}^3\Phi(v_1, v_1, v_1) &= -\sum_{i,j,k=1}^{n} \frac{\partial^3 F}{\partial x_i \partial x_j \partial x_k}\Big|_0 v_{1,i}v_{1,j}v_{1,k} \\ &= -\frac{1}{4}\mathrm{Re}\{e^{3is}\mathrm{d}^3F(c, c, c) + 3e^{is}\mathrm{d}^3F(c, c, \bar{c})\}\end{aligned} \tag{3.6}$$

since

$$\begin{aligned}v_{1,i}v_{1,j}v_{1,k} &= \frac{1}{8}(e^{is}c_i + e^{-is}\bar{c}_i)(e^{is}c_j + e^{-is}\bar{c}_j)(e^{is}c_k + e^{-is}\bar{c}_k) \\ &= \frac{1}{4}\mathrm{Re}\{e^{3is}c_ic_jc_k + e^{is}(c_ic_j\bar{c}_k + c_i\bar{c}_jc_k + \bar{c}_ic_jc_k)\}.\end{aligned}$$

Therefore, since second derivatives of F in (3.4) vanish, formula (3.5) becomes

$$\phi_{xxx} = (I-E)\mathrm{d}^3\Phi(v_1,v_1,v_1) = -\frac{1}{4}(I-E)\mathrm{Re}\big\{e^{3is}\mathrm{d}^3F(c,c,c) + 3e^{is}\mathrm{d}^3F(c,c,\bar{c})\big\}.$$

Thus the components of ϕ_{xxx} on $\mathcal{K}^*$ are

$$p_u(\mathbf{0}) = -\frac{1}{8}\left\langle v_1^*, \mathrm{Re}\{e^{3is}\mathrm{d}^3F(c,c,c) + e^{is}3\mathrm{d}^3F(c,c,\bar{c})\}\right\rangle = -\frac{3}{16}\mathrm{Re}\{\mathrm{d}^3F(c,c,\bar{c})^t\bar{d}\},$$
$$q_u(\mathbf{0}) = -\frac{1}{8}\left\langle v_2^*, \mathrm{Re}\{e^{3is}\mathrm{d}^3F(c,c,c) + e^{is}3\mathrm{d}^3F(c,c,\bar{c})\}\right\rangle = \frac{3}{16}\mathrm{Im}\{\mathrm{d}^3F(c,c,\bar{c})^t\bar{d}\}.$$

To compute $\mathrm{d}^3F(c,c,\bar{c})$, rewrite the equations in the form $\dot{x}_i = f_i(x)$ for $i = 1,\dots,5$ and observe that

$$\frac{\partial^3 f_i}{\partial x_i^3} = -6 \quad \text{for} \quad i = 1,\dots,5 \qquad \text{and} \qquad \frac{\partial^3 f_i}{\partial x_j^3} = 0 \quad \text{if} \quad i \neq j.$$

Hence

$$\mathrm{d}^3F(c,c,\bar{c}) = \sum_{i,j,k=1}^{5} \left.\frac{\partial^3 F}{\partial x_i^3}\right|_0 c_i c_j \bar{c}_k = -6\begin{pmatrix} c_1c_1\bar{c}_1 \\ c_2c_2\bar{c}_2 \\ c_3c_3\bar{c}_3 \\ c_4c_4\bar{c}_4 \\ c_5c_5\bar{c}_5 \end{pmatrix}.$$

The critical eigenspace is spanned by the real and imaginary parts of

$$c = (2, -2+2i, -4i, -1-i, 2)^t,$$

and therefore

$$\mathrm{d}^3F(c,c,\bar{c}) = -12(4, -8+8i, -32i, -1-i, 4)^t.$$

The generalized eigenvector orthogonal to c is

$$b = \frac{1}{17}(-45-27i, 4-18i, 14-12i, -8+36i, 57+7i),$$

and the eigenvector of J^t with eigenvalue $-i$ is

$$d = \frac{1}{20}(-3-i, 1-3i, 1-3i, -2+6i, 3+i)^t,$$

which is chosen so that $b^t\bar{d} = 2$, as in (2.15). Hence

$$\mathrm{d}^3F(c,c,\bar{c})^t\bar{d} = -36+24i,$$

and therefore, since $p_u(\mathbf{0}) = \frac{1}{2}p_{xx}(\mathbf{0})$ and $q_u(\mathbf{0}) = \frac{1}{2}q_{xx}(\mathbf{0})$,

$$p_u(\mathbf{0}) = -\frac{3}{16}\mathrm{Re}\{-36+24i\} = \frac{27}{4} \quad \text{and} \quad q_u(\mathbf{0}) = \frac{3}{16}\mathrm{Im}\{-36+24i\} = \frac{9}{2}.$$

Since $q_u(\mathbf{0}) \neq 0$, Theorem 1.3 guarantees linear growth near to a bifurcation.

4. Hopf bifurcation with quadratic u terms. In this section we consider nilpotent Hopf bifurcations in networks that satisfy (1.18) and (1.19). In section 4.1 we prove Theorem 1.4. Then in section 4.2 we discuss the implications of (1.23) for the Liapunov–Schmidt reduction and derive expressions for the partial derivatives of the reduced mapping. Finally, in section 4.3 we show that the network in Figure 2 satisfies (1.23) and, using the formulas derived in section 4.2, that admissible vector fields can be chosen to give either two or four branches, as stated in Theorem 1.4, with any combination of super- and subcritical branches.

The number of branches is determined by the number of points in the intersection of two quadratics in the (u,τ)-plane. These quadratics are not arbitrary because, as we show in Lemma 4.2, the null intersection is not possible. However, in the two or four branch cases any configuration of super- and subcritical branches may be obtained by admissible vector fields for the network in Figure 2.

4.1. Proof of Theorem 1.4. Proposition 1.2 and (1.18) imply that p and q can be written as

$$\begin{aligned} p(u,\lambda,\tau) &= \tfrac{1}{2}p_{uu}(\mathbf{0})u^2 + p_{u\tau}(\mathbf{0})u\tau + p_{u\lambda}(\mathbf{0})u\lambda - \tau^2 \\ &\quad + (\sigma'(0)^2 - \omega'(0)^2)\lambda^2 + 2\omega'(0)\lambda\tau + \cdots, \\ q(u,\lambda,\tau) &= \tfrac{1}{2}q_{uu}(\mathbf{0})u^2 + q_{u\tau}(\mathbf{0})u\tau + q_{u\lambda}(\mathbf{0})u\lambda \\ &\quad - 2\sigma'(0)\omega'(0)\lambda^2 + 2\sigma'(0)\lambda\tau + \cdots. \end{aligned} \tag{4.1}$$

We seek solutions to $p(u,\lambda,\tau) = q(u,\lambda,\tau) = 0$.

The first step in finding solutions to (4.1) is to introduce changes of coordinates that simplify the equations but do not affect the branching behavior at the bifurcation.

Lemma 4.1. *The mapping (p,q) in (4.1) is strongly equivalent to*

$$\begin{aligned} p(u,\lambda,\tau) &= \alpha u^2 + \gamma u\lambda - \tau^2 + \sigma'(0)^2\lambda^2 + \cdots, \\ q(u,\lambda,\tau) &= au^2 + bu\tau + cu\lambda + 2\sigma'(0)\lambda\tau + \cdots, \end{aligned} \tag{4.2}$$

where

$$\begin{aligned} \alpha &= \frac{1}{4}\big(2p_{uu}(\mathbf{0}) + p_{u\tau}(\mathbf{0})^2\big), & a &= \frac{1}{2}\big(q_{uu}(\mathbf{0}) + p_{u\tau}(\mathbf{0})q_{u\tau}(\mathbf{0})\big), \\ & & b &= q_{u\tau}(\mathbf{0}), \\ \gamma &= p_{u\lambda}(\mathbf{0}) + \omega'(0)p_{u\tau}(\mathbf{0}), & c &= q_{u\lambda}(\mathbf{0}) + \sigma'(0)p_{u\tau}(\mathbf{0}) + \omega'(0)q_{u\tau}(\mathbf{0}). \end{aligned} \tag{4.3}$$

Furthermore, the nondegeneracy conditions in (1.19) imply that all coefficients in (4.2) are nonzero, and condition (1.20) implies that when $\lambda = 0$ the only solution to $p = q = 0$ near the origin is the origin itself.

We assume that the coefficients in (4.2) are independent and arbitrary. In section 4.3 we use the results derived in section 4.2 to show that this assumption is valid for the network shown in Figure 2.

The second step in finding solutions to (4.1) is to truncate p and q at quadratic order and to introduce similarity variables

$$\hat{u} = \lambda u \quad \text{and} \quad \hat{\tau} = \tau\lambda. \tag{4.4}$$

We then prove that generically there are two or four branches of solutions to $p = q = 0$ in two stages. Consider the equations

$$\begin{aligned} \hat{p}(\hat{u}, \hat{\tau}) &\equiv \alpha\hat{u}^2 + \gamma\hat{u} - \hat{\tau}^2 + \sigma'(0)^2 \quad = 0, \\ \hat{q}(\hat{u}, \hat{\tau}) &\equiv a\hat{u}^2 + b\hat{u}\hat{\tau} + c\hat{u} + 2\sigma'(0)\tau = 0. \end{aligned} \tag{4.5}$$

Note that solutions to (4.5) correspond to lines of solutions (parametrized by λ) in the zeros for the quadratic truncations of p and q in (4.2). We prove the following.

Lemma 4.2. *Assume* (1.19). *Then, generically, solutions to* (4.5) *consist of either two or four points, with the precise number depending on the coefficients in* (4.3).

Recall that the only solutions to (4.2) of interest are those with $u = x^2 > 0$, since we require real solutions for x. However, any solution $(\hat{u}_0, \hat{\tau}_0)$ to (4.5) corresponds to a ray of solutions to the quadratic truncations of p and q. If $\hat{u}_0 > 0$, the ray consists of points $u_0 = \hat{u}_0\lambda$ and $\tau_0 = \hat{\tau}_0\lambda$, where $\lambda \geq 0$. On the other hand, if $\hat{u}_0 < 0$, the ray consists of (u_0, τ_0), where $\lambda < 0$. Each ray of solutions in (u, τ) space corresponds to a parabola of solutions in (x, τ, λ) space, where the solutions are supercritical if $\hat{u}_0 > 0$ and subcritical if $\hat{u}_0 < 0$.

Finally, we use hyperbolicity to justify truncating (4.2) at quadratic order.

Lemma 4.3. *Generically in the coefficients* (4.3), *all solutions to* (4.5) *are hyperbolic.*

Lemma 4.3 implies that higher order perturbations of the truncated equations merely move the branches of solutions in (u, λ, τ) space and do not affect the existence of the branches. The proof of Theorem 1.4 follows from Lemmas 4.1, 4.2, and 4.3.

Figures 5, 6, and 7 show that two or four branches with all choices of super- and subcritical branches are indeed possible, given that the coefficients in (4.2) are arbitrary. Recall that negative $\hat{u}$ solutions to (4.5) correspond to subcritical branches of solutions to $p = q = 0$ and that positive solutions correspond to supercritical branches.

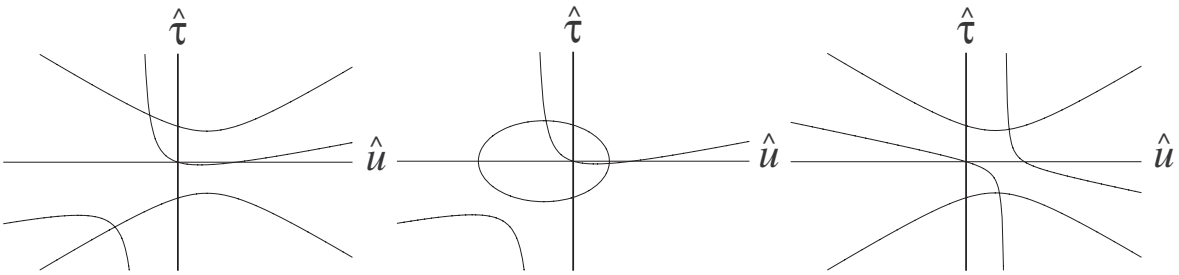


Figure 5. *Two solutions to* (4.5) *with* $\gamma = -1$, $a = -1$, $c = 1$, $\sigma'(0) = 1$. *Left: Two solutions with* $\hat{u} < 0$; $\alpha = 1$, $b = 3$. *Center: One solution with* $\hat{u} < 0$ *and one with* $\hat{u} > 0$; $\alpha = -1$, $b = 3$. *Right: Two solutions with* $\hat{u} > 0$; $\alpha = 1$, $b = -3$.

Proof of Lemma 4.1. For strong equivalence (see [4]) we may make changes of coordinates of the form

$$g(x, \tau, \lambda) = S(x, \tau, \lambda)\phi(X(x, \tau, \lambda), T(x, \tau, \lambda), \lambda).$$

Thus we can transform τ by

$$\tau \mapsto \tau + \frac{p_{u\tau}(\mathbf{0})}{2}u + \omega'(0)\lambda,$$

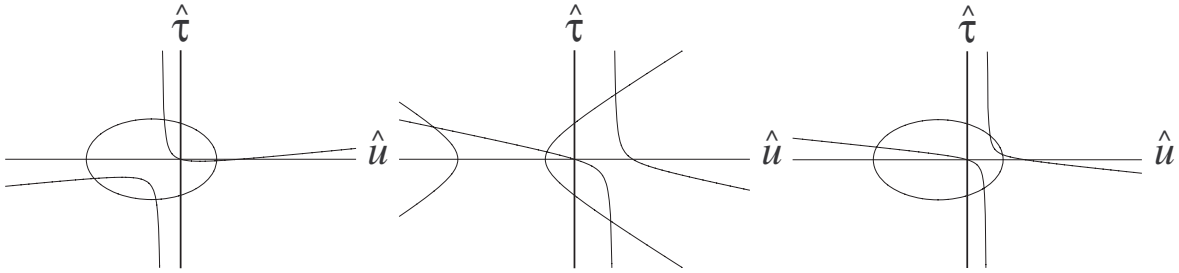


Figure 6. *Four solutions to* (4.5) *with* $a = -1$, $c = 1$, $\sigma'(0) = 1$. *Left: Three solutions with* $\hat{u} < 0$ *and one with* $\hat{u} > 0$; $\alpha = -1$, $\gamma = -1$, $b = 6$. *Center: Two solutions with* $\hat{u} < 0$ *and two with* $\hat{u} > 0$; $\alpha = 1$, $\gamma = 2.5$, $b = -3$. *Right: One solution with* $\hat{u} < 0$ *and three with* $\hat{u} > 0$; $\alpha = -1$, $\gamma = -1$, $b = -6$.

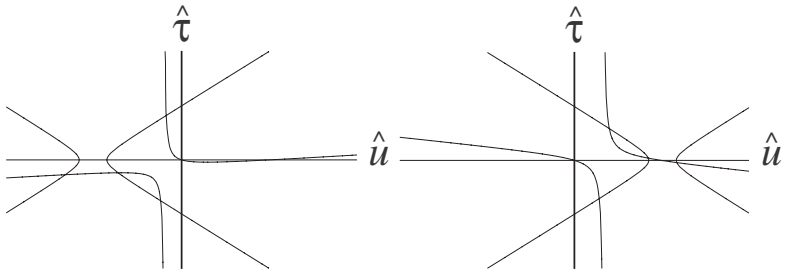


Figure 7. *Four solutions to* (4.5) *with* $\alpha = 1$, $a = -1$, $c = 1$, $\sigma'(0) = 1$. *Left: Four solutions with* $\hat{u} < 0$; $\gamma = 2.025$, $b = 10$. *Right: Four solutions with* $\hat{u} > 0$; $\gamma = -2.025$, $b = -6$.

which yields

$$
\begin{aligned}
p(u,\lambda,\tau) &= \tfrac{1}{4}\big(2p_{uu}(\mathbf{0}) + p_{u\tau}(\mathbf{0})^2\big)u^2 + \big(p_{u\lambda}(\mathbf{0}) + \omega'(0)p_{u\tau}(\mathbf{0})\big)u\lambda \\
&\quad - \tau^2 + \sigma'(0)^2\lambda^2 + \cdots, \\
q(u,\lambda,\tau) &= \tfrac{1}{2}\big(q_{uu}(\mathbf{0}) + p_{u\tau}(\mathbf{0})q_{u\tau}(\mathbf{0})\big)u^2 + q_{u\tau}(\mathbf{0})u\tau \\
&\quad + \big(q_{u\lambda}(\mathbf{0}) + \omega'(0)q_{u\tau}(\mathbf{0}) + \sigma'(0)p_{u\tau}(\mathbf{0})\big)u\lambda + 2\sigma'(0)\lambda\tau + \cdots.
\end{aligned} \tag{4.6}
$$

This gives (4.2) and the coefficients in (4.3). Moreover, (1.19) implies that all coefficients are nonzero.

Note that if $\lambda = 0$, then

$$
p(u,0,\tau) = \alpha u^2 - \tau^2 + \cdots \qquad \text{and} \qquad q(u,0,\tau) = au^2 + bu\tau + \cdots.
$$

Use the implicit function theorem to solve $q = 0$ to quadratic order for $\tau = -\frac{a}{b}u + \cdots$, and substitute into $p = 0$, giving

$$
\left(\alpha - \frac{a^2}{b^2}\right)u^2 + \cdots = 0,
$$

which has nonzero solutions for u only if $\alpha b^2 - a^2 = 0$. By (4.3), $\alpha b^2 - a^2 \neq 0$ is equivalent to (1.20). So the only solution for $\lambda = 0$ is the origin. ■

Proof of Lemma 4.2. We show that (1.19) implies that there always exists at least one solution. Therefore, generically, there must be either two or four solutions.

Observe that $\hat{q} = 0$ can be solved for

$$\hat{\tau} = \hat{\tau}(\hat{u}) = -\frac{a\hat{u}^2 + c\hat{u}}{b\hat{u} + 2\sigma'(0)}. \tag{4.7}$$

Substituting this into $\hat{p}$ in (4.5) yields

$$\hat{p}(\hat{u}, \hat{\tau}(\hat{u})) = \frac{1}{(b\hat{u} + 2\sigma'(0))^2}\left[(\alpha\hat{u}^2 + \gamma\hat{u} + \sigma'(0)^2)(b\hat{u} + 2\sigma'(0))^2 - (a\hat{u}^2 + c\hat{u})^2\right].$$

Thus $\hat{p} = 0$ only when

$$h(\hat{u}) \equiv (\alpha\hat{u}^2 + \gamma\hat{u} + \sigma'(0)^2)(b\hat{u} + 2\sigma'(0))^2 - (a\hat{u}^2 + c\hat{u})^2 = 0. \tag{4.8}$$

Observe that

$$h(0) = 4\sigma'(0)^4 > 0 \quad \text{and} \quad h\left(-\frac{2\sigma'(0)}{b}\right) = -(a\hat{u}^2 + c\hat{u})^2 \leq 0.$$

Hence, by the mean value theorem, there exists a $\hat{u}_0$ between 0 and $-2\sigma'(0)/b$ such that $h(\hat{u}_0) = 0$. Therefore, for $\hat{\tau}_0 = \hat{\tau}(\hat{u}_0)$ as given in (4.7), $\hat{p}(\hat{u}_0, \hat{\tau}_0) = 0$. Since there is at least one solution to (4.5), generically there must be either two or four solutions. ■

Proof of Lemma 4.3. We show that the Jacobian of the mapping

$$R(\hat{u}, \hat{\tau}) = (\hat{p}(\hat{u}, \hat{\tau}), \hat{q}(\hat{u}, \hat{\tau}))$$

is singular at a solution to (4.5) only if $\hat{u}$ satisfies a quartic equation that is different from $h = 0$. Hence, generically, a point $\hat{u}_0$ will not solve both equations simultaneously.

Observe that the Jacobian of R is

$$\mathrm{d}_{\hat{u},\hat{\tau}}R = \begin{pmatrix} \frac{\partial \hat{p}}{\partial \hat{u}} & \frac{\partial \hat{p}}{\partial \hat{\tau}} \\ \frac{\partial \hat{q}}{\partial \hat{u}} & \frac{\partial \hat{q}}{\partial \hat{\tau}} \end{pmatrix} = \begin{pmatrix} 2\alpha\hat{u} + \gamma & -2\hat{\tau} \\ 2a\hat{u} + b\hat{\tau} + c & b\hat{u} + 2\sigma'(0) \end{pmatrix}.$$

Thus $\mathrm{d}R$ is singular only if its determinant

$$(2\alpha\hat{u} + \gamma)(b\hat{u} + 2\sigma'(0)) + 2\hat{\tau}(2a\hat{u} + b\hat{\tau} + c) = 0.$$

Evaluating this on the manifold of solutions $\hat{\tau} = \hat{\tau}(\hat{u})$ to $\hat{q} = 0$ given by (4.7) yields a second quartic equation:

$$2(a\hat{u} + c)\big(b(a\hat{u} + c)\hat{u} - (2a\hat{u} + c)(b\hat{u} + 2\sigma'(0))\big)\hat{u} + (2\alpha\hat{u} + \gamma)(b\hat{u} + 2\sigma'(0))^3 = 0. \tag{4.9}$$

Generically in the coefficients, the two quartics (4.8) and (4.9) will not have any simultaneous roots. Therefore, solutions will generically be hyperbolic. ■

4.2. Liapunov–Schmidt reduction with a flow-invariant synchrony subspace. In this section we discuss the effect that (1.23) has on the derivatives of the reduced mapping ϕ. The existence of a flow-invariant synchrony subspace S forces certain derivatives of the reduced equation ϕ to vanish because trajectories in S are trapped in S for all time. The space of 2π-periodic solutions that lie inside S forms a subspace of $\mathcal{C}_{2\pi}$, and so we define

$$\mathcal{S}_{2\pi} = \{u \in \mathcal{C}_{2\pi} \,:\, u(s) \in S \text{ for all } s\}\,. \tag{4.10}$$

The essential point is that if $u_1, \dots, u_k \in \mathcal{S}_{2\pi}$, then $\mathrm{d}^k\Phi(u_1, \dots, u_k) \in \mathcal{R}$ and thus vanishes under the projection $(I - E)$ onto $\mathcal{K}^*$. This and related results are given by the following lemma.

Lemma 4.4. *Suppose that a coupled cell network has a synchrony subspace S satisfying (1.23). Then*

(a) *if $w \in S$, then*

$$w^t \bar{d} = 0; \tag{4.11}$$

(b) *if $c_1, \dots, c_k \in S$, then*

$$\mathrm{d}^k F(c_1, \dots, c_k) \in S; \tag{4.12}$$

(c) *if $u_1, \dots, u_k \in \mathcal{S}_{2\pi}$, then*

$$\mathrm{d}^k\Phi(u_1, \dots, u_k) \in \mathcal{S}_{2\pi}; \tag{4.13}$$

(d)

$$\mathcal{S}_{2\pi} \subseteq \mathcal{R}. \tag{4.14}$$

Proof. To prove (4.11), note that any synchrony subspace, being invariant under J, must be the direct sum of eigenspaces of J. In the proof of Lemma 2.2 it was shown that the only eigenvector to which $\bar{d}$ is not orthogonal is the generalized eigenvector b, which is not contained in S by hypothesis (1.23). Therefore $\bar{d}$ is orthogonal to every vector in S.

The statement in (4.12) follows simply because S is invariant for F and hence for all differentials of F. Similarly for (4.13), $\mathcal{S}_{2\pi}$ is invariant for Φ and hence for all differentials of Φ.

Finally, to prove (4.14) note that, by the definitions of v_1^* and v_2^* in (2.7), and by (4.11), trajectories in $\mathcal{S}_{2\pi}$ are orthogonal to v_1^* and v_2^*. So $\mathcal{S}_{2\pi} \subseteq (\mathcal{K}^*)^\perp = \mathcal{R}$ by (2.11). ∎

With these results in mind we compute the second partial derivatives of p and q. To simplify these calculations we assume that F is odd, since it turns out that this is sufficient for our purposes in section 4.3. To begin, we require the following lemma.

Lemma 4.5. *Assume that a coupled cell system satisfies (1.23) with flow-invariant subspace S and that F is odd. Then*

$$W_{xx}(\mathbf{0}) = W_{xxxx}(\mathbf{0}) = W_{\tau xx}(\mathbf{0}) = W_{\lambda xx}(\mathbf{0}) = 0 \tag{4.15}$$

and

$$W_{xxx}(\mathbf{0}) = -\frac{1}{4}\mathrm{Re}\{e^{3is}\eta_3 + e^{is}\eta_1\}, \tag{4.16}$$

where η_1 *and* η_3 *are such that*

$$(J - iI_n)\eta_1 = 3\mathrm{d}^3F(c, c, \bar{c}) \quad \textit{and} \quad (J - 3iI_n)\eta_3 = \mathrm{d}^3F(c, c, c). \tag{4.17}$$

Proof. The statements in (4.15) follow immediately from the assumption that F is odd. This assumption also implies that formula (A.11) in the appendix becomes

$$W_{xxx} = -\mathcal{L}^{-1}(\mathrm{d}^3\Phi(v_1, v_1, v_1)).$$

By (3.6) it follows that W_{xxx} is the solution to the differential equation

$$\mathcal{L}W_{xxx} = \frac{1}{4}\mathrm{Re}\big\{e^{3is}\mathrm{d}^3F(c, c, c) + 3e^{is}\mathrm{d}^3F(c, c, \bar{c})\big\}.$$

It is straightforward to verify that this solution is given by (4.16). ∎

Then we have the following.

Proposition 4.6. *Assume the same hypotheses as in Lemma 4.5. Then*

$$p_{uu}(\mathbf{0}) = \frac{5}{16}\mathrm{Re}\{\xi_1^t\bar{d}\}, \qquad q_{uu}(\mathbf{0}) = -\frac{5}{16}\mathrm{Im}\{\xi_1^t\bar{d}\}, \tag{4.18}$$

$$p_{u\tau}(\mathbf{0}) = -\frac{1}{8}\mathrm{Im}\{\xi_2^t\bar{d}\}, \qquad q_{u\tau}(\mathbf{0}) = -\frac{1}{8}\mathrm{Re}\{\xi_2^t\bar{d}\}, \tag{4.19}$$

$$p_{u\lambda}(\mathbf{0}) = \frac{1}{8}\mathrm{Re}\{\xi_3^t\bar{d}\}, \qquad q_{u\lambda}(\mathbf{0}) = -\frac{1}{8}\mathrm{Im}\{\xi_3^t\bar{d}\}, \tag{4.20}$$

where

$$\xi_1 = 2\mathrm{d}^3F(c, \bar{c}, \eta_1) + \mathrm{d}^3F(c, c, \bar{\eta}_1), \tag{4.21}$$

$$\xi_2 = 6\mathrm{d}^3F(c, \bar{c}, b) - 3\mathrm{d}^3F(c, c, \bar{b}) + \eta_1, \tag{4.22}$$

$$\xi_3 = 6\mu'(0)\mathrm{d}^3F(c, \bar{c}, b) + 3\bar{\mu}'(0)\mathrm{d}^3F(c, c, \bar{b}) + J'(0)\eta_1, \tag{4.23}$$

where $\mu'(0) = \sigma'(0) + i\omega'(0)$.

Proof. Applying Lemmas 4.4 and 4.5 to formula (A.2) yields

$$\phi_{xxxxx} = 10(I - E)\mathrm{d}^3\Phi(v_1, v_1, W_{xxx}). \tag{4.24}$$

Using (4.16) and the linearity of $\mathrm{d}^3\Phi$, we compute

$$\begin{aligned}\mathrm{d}^3\Phi(v_1, v_1, W_{xxx}) = \tfrac{1}{16}\mathrm{Re}\big\{&e^{5is}\mathrm{d}^3F(c, c, \eta_3)\\ &+ e^{3is}\big(\mathrm{d}^3F(c, c, \eta_1) + 2\mathrm{d}^3F(c, \bar{c}, \eta_3)\big)\\ &+ e^{is}\big(2\mathrm{d}^3F(c, \bar{c}, \eta_1) + \mathrm{d}^3F(c, c, \bar{\eta}_1) + \mathrm{d}^3F(\bar{c}, \bar{c}, \eta_3)\big)\big\}.\end{aligned} \tag{4.25}$$

Note that, by (4.17), $\eta_3 \in S$ since $\mathrm{d}^3F(c,c,c) \in S$ and S is invariant under $J - 3iI_n$. Therefore $e^{is}\mathrm{d}^3F(\bar{c},\bar{c},\eta_3) \in \mathcal{R}$ by Lemma 4.4. Since only terms involving e^{is} have nonzero projection onto $\mathcal{K}^*$, we obtain

$$\phi_{xxxxx,j} = 10\,\langle v_j^*, \mathrm{d}^3\Phi(v_1,v_1,W_{xxx})\rangle = \frac{5}{8}\langle v_j^*, \mathrm{Re}\{e^{is}\big(2\mathrm{d}^3F(c,\bar{c},\eta_1) + \mathrm{d}^3F(c,c,\bar{\eta}_1)\big)\}\rangle.$$

From this it is straightforward to verify that

$$\phi_{xxxxx,1} = \frac{5}{16}\mathrm{Re}\{\xi_1^t\bar{d}\} \quad \text{and} \quad \phi_{xxxxx,2} = -\frac{5}{16}\mathrm{Im}\{\xi_1^t\bar{d}\},$$

where ξ_1 is as defined in (4.21).

Now consider the formulas in (4.19) and (4.20). Using formula (A.8) and recalling from (2.21) and (4.15) that $W_\tau = W_\lambda = W_{xx} = 0$, we obtain

$$\phi_{\alpha xxx} = (I-E)\big(3\mathrm{d}^3\Phi(v_1,v_1,W_{\alpha x}) + \mathrm{d}^3\Phi_\alpha(v_1,v_1,v_1) + \mathrm{d}\Phi_\alpha(W_{xxx})\big), \tag{4.26}$$

where α is either τ or λ.

Consider the case where $\alpha = \tau$. By the fact that $\mathrm{d}^k\Phi_\tau = \frac{\mathrm{d}}{\mathrm{d}s}$ if $k = 1$ and 0 if $k > 1$, we obtain

$$\mathrm{d}^3\Phi_\tau(v_1,v_1,v_1) = 0 \quad \text{and} \quad \mathrm{d}\Phi_\tau(W_{xxx}) = \frac{1}{4}\mathrm{Im}\{3e^{3is}\eta_3 + e^{is}\eta_1\}. \tag{4.27}$$

Since $W_{\tau x} = -u_2$ by (2.32) we compute

$$\mathrm{d}^3\Phi(v_1,v_1,W_{\tau x}) = \frac{1}{4}\mathrm{Im}\big\{e^{3is}\mathrm{d}^3F(c,c,b) + e^{is}(2\mathrm{d}^3F(c,\bar{c},b) - \mathrm{d}^3F(c,c,\bar{b}))\big\}. \tag{4.28}$$

Substituting (4.27) and (4.28) into (4.26), we obtain

$$\phi_{\tau xxx,j} = \frac{1}{4}\langle v_j^*, \mathrm{Im}\big\{e^{is}\big(6\mathrm{d}^3F(c,\bar{c},b) - 3\mathrm{d}^3F(c,c,\bar{b}) + \eta_1\big)\big\}\rangle$$

and hence

$$\phi_{\tau xxx,1} = -\frac{1}{8}\mathrm{Im}\{\xi_2^t\bar{d}\} \quad \text{and} \quad \phi_{\tau xxx,2} = -\frac{1}{8}\mathrm{Re}\{\xi_2^t\bar{d}\},$$

where ξ_2 is defined in (4.22).

In the case when $\alpha = \lambda$ we have from (4.26)

$$\phi_{\lambda xxx} = (I-E)\big(3\mathrm{d}^3\Phi(v_1,v_1,W_{\lambda x}) + \mathrm{d}^3\Phi_\lambda(v_1,v_1,v_1) + \mathrm{d}\Phi_\lambda(W_{xxx})\big). \tag{4.29}$$

Observe that $\mathrm{d}^3\Phi_\lambda(v_1,v_1,v_1) = -\mathrm{d}^3F_\lambda(v_1,v_1,v_1)$ and note that F_λ is a vector field on the same network as F and therefore has the same flow-invariant subspaces. Hence, by Lemma 4.4, $(I-E)\mathrm{d}^3\Phi(v_1,v_1,v_1) = 0$ since $v_1 \in \mathcal{S}_{2\pi}$. Then using formulas (2.33) for $W_{\lambda x}$ and (4.16) for W_{xxx} we have

$$\begin{aligned}\mathrm{d}^3\Phi(v_1,v_1,W_{\lambda x}) = \frac{1}{4}\mathrm{Re}\big\{&e^{3is}\big(\mu'(0)\mathrm{d}^3F(c,c,b) - \mathrm{d}^3F(c,c,c'(0)-\zeta c)\big)\\ &+ e^{is}\big(2\mu'(0)\mathrm{d}^3F(c,\bar{c},b) + \bar{\mu}'(0)\mathrm{d}^3F(c,c,\bar{b})\\ &\quad - 2\mathrm{d}^3F(c,\bar{c},c'(0)-\zeta c) - \mathrm{d}^3F(c,c,\bar{c}'(0)-\bar{\zeta}\bar{c})\big)\big\},\\ \mathrm{d}\Phi_\lambda(W_{xxx}) = \frac{1}{4}\mathrm{Re}\{&e^{3is}J'(0)\eta_3 + e^{is}J'(0)\eta_1\},\end{aligned}$$

where ζ is given by (2.34) and $c(\lambda)$ is the continuation of the critical eigenvector for $J(\lambda)$ so that $c(0) = c$. Note that $c(\lambda) \in S$ for all λ. Therefore, $c'(0) \in S$, and $\mathrm{d}^3F(c, c, c'(0) - \zeta c) \in S$ by (4.12). Thus we obtain

$$\phi_{\lambda xxx,j} = \frac{1}{4}\left\langle v_j^*, \mathrm{Re}\left\{e^{is}\left(6\mu'(0)\mathrm{d}^3F(c,\bar{c},b) + 3\bar{\mu}'(0)\mathrm{d}^3F(c,c,\bar{b}) + J'(0)\eta_1\right)\right\}\right\rangle.$$

It is now straightforward to verify that

$$\phi_{\lambda xxx,1} = \frac{1}{8}\mathrm{Re}\{\xi_3^t\bar{d}\} \quad \text{and} \quad \phi_{\lambda xxx,2} = -\frac{1}{8}\mathrm{Im}\{\xi_3^t\bar{d}\},$$

where ξ_3 is as in (4.23). ■

4.3. The three-cell network of Figure 2. In this section we show that the network in Figure 2 can have bifurcations with either two or four branches of solutions, depending on the specific vector field, as stated in Theorem 1.4. Recall that $S = \{(u, u, v) : u, v \in \mathbf{R}^n\}$ is a flow-invariant synchrony subspace for the network, and that the critical eigenvector has the form $c = (a, a, -a)$, where $Aa = ia$. The corresponding generalized eigenvector has the form

$$b = \zeta\begin{pmatrix} a \\ -a \\ 0 \end{pmatrix} + \begin{pmatrix} w \\ w \\ -w \end{pmatrix},$$

where ζ and w are chosen so that $(A - iI_k)w = -(\zeta B - I_k)u$. Hence, S contains the critical eigenspace but not the generalized eigenspace, since $(u, -u, 0)$ breaks synchrony. Therefore the network satisfies (1.23), and thus Proposition 1.6 applies and (1.18) holds. In Proposition 4.7 we use the formulas in Proposition 4.6 to show that the second partial derivatives of p and q are independent and arbitrary, and hence that (1.19) and (1.20) also hold generically. Therefore, in this network the coefficients of (4.2) can be varied to obtain either two or four branches of solutions, as stated in Theorem 1.4.

Proposition 4.7. *The partial derivatives $p_{uu}(\mathbf{0})$, $p_{u\lambda}(\mathbf{0})$, $p_{u\tau}(\mathbf{0})$, $q_{uu}(\mathbf{0})$, $q_{u\lambda}(\mathbf{0})$, and $q_{u\tau}(\mathbf{0})$ are arbitrary and can be varied independently for the network defined by (1.4).*

Proof. We show that the real and imaginary parts of $\xi_j^t\bar{d}$ are arbitrary and independent, for ξ_j defined in (4.21) to (4.23). In fact, we can do this in a very restricted setting to simplify the calculations. Showing that the derivatives of p and q are arbitrary and independent under restrictive assumptions clearly implies that the result holds for generic vector fields on this network.

The first of these assumptions is that each cell has two-dimensional internal dynamics so that

$$z_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad z_2 = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}, \quad \text{and} \quad z_3 = \begin{pmatrix} x_5 \\ x_6 \end{pmatrix}.$$

We also assume that

$$A(0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad B(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The matrix $A'(0)$ is left arbitrary.

Let $a = (i, 1)^t$ be the eigenvector of A with eigenvalue i, and set

$$c = (a, a, -a)^t. \tag{4.30}$$

Then since $B = I_2$, the corresponding generalized eigenvector is

$$b = (a, -a, 0)^t. \tag{4.31}$$

Finally, observe that $a = (i, 1)^t$ is also an eigenvector of A^t such that $A^t a = -ia$, so

$$d = (a, -a, 0)^t$$

is an eigenvector of J^t with eigenvalue $-i$.

We also make a number of assumptions about the higher derivatives of F. As in Lemma 4.5 and Proposition 4.6, we assume that F is odd. Since $n = 2$ we can write

$$f(u, \overline{v, w}, \lambda) = \begin{pmatrix} g(u, \overline{v, w}, \lambda) \\ h(u, \overline{v, w}, \lambda) \end{pmatrix},$$

where $g, h : \mathbf{R}^6 \times \mathbf{R} \longrightarrow \mathbf{R}$, with u being the internal variable and v and w being the input variables. Now set all third derivatives equal to zero except for $f_{u_1u_1u_1}$, which we fix at

$$f_{u_1u_1u_1} = \begin{pmatrix} g_{u_1u_1u_1} \\ h_{u_1u_1u_1} \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \tag{4.32}$$

and

$$f_{u_2u_2v_2} = \begin{pmatrix} g_{u_2u_2v_2} \\ h_{u_2u_2v_2} \end{pmatrix} \quad \text{and} \quad f_{v_1v_1w_2} = \begin{pmatrix} g_{v_1v_1w_2} \\ h_{v_1v_1w_2} \end{pmatrix},$$

which we leave arbitrary. Note that $f_{u_2u_2v_2} = f_{u_2u_2w_2}$ and $f_{v_1v_1w_2} = f_{w_1w_1v_2}$ by the invariance of f.

First we compute η_1 and show that $[J'(0)\eta_1]^t \bar{d} \neq 0$. Then it follows that $\xi_3^t \bar{d}$ is arbitrary and independent of $\xi_1^t \bar{d}$ and $\xi_2^t \bar{d}$ due to the occurrence of the term $J'(0)\eta_1$ in (4.23). Finally we show that under the above assumptions about F,

$$\begin{pmatrix} \xi_1^t \bar{d} \\ \xi_2^t \bar{d} \end{pmatrix} = 3 \begin{pmatrix} -6i & 6 & -4i & 4 \\ -i & 1 & 2i & -2 \end{pmatrix} \begin{pmatrix} g_{u_2u_2v_2} \\ h_{u_2u_2v_2} \\ g_{v_1v_1w_2} \\ h_{v_1v_1w_2} \end{pmatrix} + \begin{pmatrix} 72 \\ 24 \end{pmatrix}. \tag{4.33}$$

Because the matrix in (4.33) has full rank, and because $g_{u_2u_2v_2}$, $h_{u_2u_2v_2}$, $g_{v_1v_1w_2}$, and $h_{v_1v_1w_2}$ are arbitrary, the real and imaginary parts of $\xi_1^t \bar{d}$ and $\xi_2^t \bar{d}$ can be manipulated arbitrarily from just these four derivatives of f. Therefore, the coefficients in (4.2) are arbitrary and can be varied independently for this network.

Recall from (4.17) that η_1 is such that $(J - iI_n)\eta_1 = 3\mathrm{d}^3F(c, c, \bar{c})$. Using (4.32), we compute

$$\mathrm{d}^3F(c, c, \bar{c}) = \begin{pmatrix} if_{u_1u_1u_1} \\ if_{u_1u_1u_1} \\ -if_{u_1u_1u_1} \end{pmatrix} = \begin{pmatrix} 2i \\ 0 \\ 2i \\ 0 \\ -2i \\ 0 \end{pmatrix}.$$

Then it is easy to verify that

$$\eta_1 = 3b - \frac{3i}{2}\bar{c} \tag{4.34}$$

by using the definitions of c and b in (4.30) and (4.31). Hence

$$[J'(0)\eta_1]^t\bar{d} = 3[J'(0)b]^t\bar{d} - \frac{3i}{2}[J'(0)\bar{c}]^t\bar{d}.$$

Observe that $J'(0)\bar{c} \in S$ since $\bar{c} \in S$ and S is invariant for $J'(0)$. So by (4.11), $[J'(0)\bar{c}]^t\bar{d} = 0$. Also observe that

$$J'(0)b = \begin{pmatrix} A'(0)a \\ -A'(0)a \\ 0 \end{pmatrix} + \begin{pmatrix} B'(0)a \\ B'(0)a \\ -B'(0)a \end{pmatrix},$$

so that

$$[J'(0)\eta_1]^t\bar{d} = 6[A'(0)a]^t\bar{a}$$

since $(B'(0)a, B'(0)a, -B'(0)a)^t \in S$ is orthogonal to $\bar{d}$. Since the entries in $A'(0)$ are arbitrary, $[J'(0)\eta_1]^t\bar{d}$ is an arbitrary complex number, and therefore $p_{u\lambda}(\mathbf{0})$ and $q_{u\lambda}(\mathbf{0})$ can be varied independently from each other. Furthermore, since only $\xi_3^t\bar{d}$ depends on this term, $p_{u\lambda}(\mathbf{0})$ and $q_{u\lambda}(\mathbf{0})$ are independent of the other derivatives. We will therefore consider only $\xi_1^t\bar{d}$ and $\xi_2^t\bar{d}$ from here on.

By plugging (4.34) into the expressions for ξ_1 and ξ_2 in (4.21) and (4.22), using the linearity of d^3F, and observing that $\bar{c}^t\bar{d} = c^t\bar{d} = 0$, we obtain

$$\begin{aligned} \xi_1^t\bar{d} &= 12\mathrm{d}^3F(c, \bar{c}, b)^t\bar{d} + 6\mathrm{d}^3F(c, c, \bar{b})^t\bar{d}, \\ \xi_2^t\bar{d} &= 6\mathrm{d}^3F(c, \bar{c}, b)^t\bar{d} - 3\mathrm{d}^3F(c, c, \bar{b})^t\bar{d} + 12. \end{aligned} \tag{4.35}$$

Using the definitions of $c = (a, a, -a)$ and $b = (a, -a, 0)$, we obtain by direct calculation

$$\mathrm{d}^3F(c, \bar{c}, b) = \begin{pmatrix} if_{u_1u_1u_1} + f_{v_1v_1w_2} + 2f_{u_2u_2v_2} \\ -if_{u_1u_1u_1} + f_{v_1v_1w_2} + f_{u_2u_2v_2} \\ -3f_{v_1v_1w_2} - f_{u_2u_2v_2} \end{pmatrix},$$

$$\mathrm{d}^3F(c, c, \bar{b}) = \begin{pmatrix} if_{u_1u_1u_1} - f_{v_1v_1w_2} + 2f_{u_2u_2v_2} \\ -if_{u_1u_1u_1} - 3f_{v_1v_1w_2} + f_{u_2u_2v_2} \\ 3f_{v_1v_1w_2} - f_{u_2u_2v_2} \end{pmatrix},$$

and thus we compute (4.35) as

$$\begin{aligned}\xi_1^t \bar{d} &= 6[6if_{u_1u_1u_1} + 3f_{u_2u_2v_2} + 2f_{v_1v_1w_2}]^t \bar{a} \\ &= -18(ig_{u_2u_2v_2} - h_{u_2u_2v_2}) - 12(ig_{v_1v_1w_2} - h_{v_1v_1w_2}) + 72, \\ \xi_2^t \bar{d} &= 3[2if_{u_1u_1u_1} + f_{u_2u_2v_2} - 2f_{v_1v_1w_2}]^t \bar{a} + 12 \\ &= -3(ig_{u_2u_2v_2} - h_{u_2u_2v_2}) + 6(ig_{v_1v_1w_2} - h_{v_1v_1w_2}) + 24.\end{aligned}$$

It is a straightforward matter to show that this can be written as (4.33). ■

5. Hopf bifurcation in the feed-forward chain. The proof of Theorem 1.5 divides into two parts. First, in section 5.1 we prove the following.

Proposition 5.1. *At a nilpotent Hopf bifurcation in the feed-forward chain there exist two branches of near 2π-periodic solutions, one growing as $\lambda^{\frac{1}{2}}$ and the other growing as $\lambda^{\frac{1}{6}}$.*

Then in section 5.2 we prove the following.

Proposition 5.2. *The branches given by Proposition 5.1 are generically the only branches.*

Before we delve into the proofs of Propositions 5.1 and 5.2 we need to consider the various invariant subspaces that play a role later on. Recall that the critical eigenvector of J is $c = (0, 0, a)^t$, where a is an eigenvector of A with eigenvalue i. The corresponding generalized eigenvector is

$$b = (0, \zeta a, w), \tag{5.1}$$

where $\zeta \in \mathbf{C}$ and $w \in \mathbf{C}^k$ are chosen so that $w^t \bar{a} = 0$ and $(A - iI_n)w = -(\zeta B - I_n)a$, and hence $\bar{c}^t b = 0$.

Observe that

$$\begin{aligned} S &= \left\{(u, u, v) \ : \ u, v \in \mathbf{R}^k\right\}, \\ \hat{S} &= \left\{(0, 0, u) \ : \ u \in \mathbf{R}^k\right\} \end{aligned} \tag{5.2}$$

are flow-invariant subspaces for (1.3), which both contain the critical eigenvector c but not the generalized eigenvector given by (5.1). The feed-forward chain thus satisfies (1.23) for both S and $\hat{S}$. Thus Proposition 1.6 implies that there exists a branch of solutions in $\hat{S}$ that grows as $O(\lambda^{\frac{1}{2}})$.

By analogy with (4.10), we define

$$\hat{\mathcal{S}}_{2\pi} = \left\{u \in \mathcal{C}_{2\pi} \ : \ u(s) \in \hat{S} \text{ for all } s\right\}.$$

Recall from Lemma 4.4 that $\mathcal{S}_{2\pi} \subset \mathcal{R}$. Thus we have

$$\hat{\mathcal{S}}_{2\pi} \subset \mathcal{S}_{2\pi} \subset \mathcal{R}. \tag{5.3}$$

Suppose that (1.3) depends on a bifurcation parameter λ and undergoes a nilpotent Hopf bifurcation at $\lambda = 0$. Assume that the eigenvalues of $B = B(0)$ have negative real part and

that z_1 has reached its asymptotic state $z_1 = 0$, so that we can restrict our attention to the subsystem

$$\begin{aligned} \dot{y} &= f(y, 0, \lambda) = g(y, \lambda), \\ \dot{z} &= f(z, y, \lambda) = h(z, y, \lambda) \end{aligned} \tag{5.4}$$

with linearization

$$J(\lambda) = (\mathrm{d}F)_{0,\lambda} = \begin{pmatrix} A(\lambda) & 0 \\ B(\lambda) & A(\lambda) \end{pmatrix}.$$

Let $\mu(\lambda) = \sigma(\lambda) + i\omega(\lambda)$ such that $\sigma(0) = 0$ and $\omega(\lambda) = 1$ be the continuation of the critical eigenvalue of $A(0)$.

Proposition 5.1 is proved in [2, Lemma 6.1] with the assumption that the normal form of the vector field on the center manifold is $\mathbf{S}^1$-equivariant, where $\mathbf{S}^1$ acts as

$$f(e^{i\theta} y, e^{i\theta} z) = e^{i\theta} f(y, z). \tag{5.5}$$

In section 5.1 we show that this assumption is satisfied generically, and so [2, Lemma 6.1] holds in full generality. Specifically, we prove the following.

Proposition 5.3. *Up to third order, the normal form of the subsystem* (5.4) *is*

$$\begin{aligned} \dot{y} &= \mu(\lambda) y + c_3(\lambda) y^2 \bar{y} + O(5), \\ \dot{z} &= \mu(\lambda) z + y + c_3(\lambda) z^2 \bar{z} + \alpha \bar{y} z^2 + \beta y z \bar{z} + O(4), \end{aligned} \tag{5.6}$$

which is $\mathbf{S}^1$*-equivariant under the action given by* (5.5).

Because of the work in [2], Proposition 5.3 suffices to prove Proposition 5.1 for the truncated equations without higher order terms. For the truncated equations the $\lambda^{1/6}$ branch is also shown in [2] to consist of asymptotically stable solutions. Hence, a scaling argument may be used to prove the result for (5.6).

The following two lemmas prove Proposition 5.2 and hence Theorem 1.5.

Lemma 5.4. *At a nilpotent Hopf bifurcation in the feed-forward chain the reduced equation* (1.10) *satisfies* (1.21).

Lemma 5.5. *The feed-forward chain generically satisfies* (1.22).

These lemmas are proved in section 5.2 and use Liapunov–Schmidt reduction instead of the normal form from (5.6). It may be possible to prove Theorem 1.5 using only normal form methods, but this requires proving that (5.6) is equivariant to all orders. While we fully expect this to be the case, a proof remains elusive.

5.1. Proof of Proposition 5.3. By [2, Lemma 6.2] there exists a center manifold $\mathcal{M}$ for (5.4) such that the vector field on $\mathcal{M}$ is in skew-product form. On this center manifold we can therefore change coordinates to put $(\mathrm{d}F)_{0,\lambda}$ in complex Jordan form:

$$(\mathrm{d}F)_{0,\lambda} = \begin{pmatrix} A(\lambda) & 0 \\ I_2 & A(\lambda) \end{pmatrix},$$

where

$$A(\lambda) = |\mu(\lambda)| \begin{pmatrix} e^{i\theta(\lambda)} & 0 \\ 0 & e^{-i\theta(\lambda)} \end{pmatrix}.$$

Thus we consider

$$\begin{aligned} \dot{y} &= |\mu(\lambda)|e^{i\theta(\lambda)}y + G(y, \bar{y}, \lambda), \\ \dot{z} &= |\mu(\lambda)|e^{i\theta(\lambda)}z + y + H(y, \bar{y}, z, \bar{z}, \lambda), \end{aligned} \tag{5.7}$$

where G and H are $O(2)$.

First we make identical changes on y and z,

$$y \mapsto y + \phi(y, \bar{y}) \quad \text{and} \quad z \mapsto z + \phi(z, \bar{z}), \tag{5.8}$$

to put $\dot{y} = g(y, \lambda)$ in standard normal form for Hopf bifurcation. Since the changes are the same on both variables we have

$$\begin{aligned} &\text{(a)} \quad \dot{y} = \mu(\lambda)y + c(|y|^2, \lambda)y, \\ &\text{(b)} \quad \dot{z} = \mu(\lambda)z + c(|z|^2, \lambda)z + y + H_2 + H_3 + \cdots, \end{aligned} \tag{5.9}$$

where H_j is order j in y, $\bar{y}$, z, and $\bar{z}$, and

$$H_j(0, 0, z, \bar{z}, \lambda) = 0 \tag{5.10}$$

since all the terms depending only on z and $\bar{z}$ appear in $c(|z|^2, \lambda)z$.

Next we make changes of the form

$$y \mapsto y \quad \text{and} \quad z \mapsto \phi(y, \bar{y}, z, \bar{z}), \tag{5.11}$$

where ϕ is order 2. Substituting this into (5.9b), we obtain

$$\dot{z} = \mu z + y + \mu\phi - \mu\phi_y y - \bar{\mu}\phi_{\bar{y}}\bar{y} - \mu\phi_z z - \phi_z y - \bar{\mu}\phi_{\bar{z}}\bar{z} - \phi_{\bar{z}}\bar{y} + H_2 + O(3),$$

and so second order terms can be eliminated if we can choose ϕ so that

$$\mu\phi - \mu\phi_y y - \bar{\mu}\phi_{\bar{y}}\bar{y} - \mu\phi_z z - \phi_z y - \bar{\mu}\phi_{\bar{z}}\bar{z} - \phi_{\bar{z}}\bar{y} + H_2 = 0.$$

Let M_k denote the space of order k monomials in y, $\bar{y}$, z, and $\bar{z}$, and define

$$\tilde{M}_k = M_{k-1}y + M_{k-1}\bar{y}.$$

Then

$$\tilde{M}_2 = \text{span}\left\{y^2, y\bar{y}, yz, y\bar{z}, \bar{y}^2, \bar{y}z, \bar{y}\bar{z}\right\}, \tag{5.12}$$

and H_2 is a linear combination of elements of $\tilde{M}_2$ by (5.10).

Define the map $\Psi_2 : \tilde{M}_2 \longrightarrow \tilde{M}_2$ by

$$\Psi_2(\phi) = \mu\phi - \mu\phi_y y - \bar{\mu}\phi_{\bar{y}}\bar{y} - \mu\phi_z z - \phi_z y - \bar{\mu}\phi_{\bar{z}}\bar{z} - \phi_{\bar{z}}\bar{y}.$$

Then we seek solutions to the linear equation

$$\Psi_2(\phi) + H_2 = 0.$$

The action of Ψ_2 on the basis elements in (5.12) yields

$$\begin{aligned} y^2 &\mapsto -\mu y^2, & y\bar{y} &\mapsto -\bar{\mu} y\bar{y}, \\ yz &\mapsto -\mu yz - y^2, & y\bar{z} &\mapsto -\bar{\mu} y\bar{z} - y\bar{y}, \\ \bar{y}^2 &\mapsto (\mu - 2\bar{\mu})\bar{y}^2, & \bar{y}z &\mapsto -\bar{\mu}\bar{y}z - y\bar{y}. \\ \bar{y}\bar{z} &\mapsto (\mu - 2\bar{\mu})\bar{y}\bar{z} - \bar{y}^2, \end{aligned}$$

Thus, with respect to this basis, Ψ_2 can be written as

	y^2	$y\bar{y}$	yz	$y\bar{z}$	$\bar{y}^2$	$\bar{y}z$	$\bar{y}\bar{z}$
y^2	$-\mu$	$\cdot$	-1	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$y\bar{y}$	$\cdot$	$-\bar{\mu}$	$\cdot$	-1	$\cdot$	-1	$\cdot$
yz	$\cdot$	$\cdot$	$-\mu$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$y\bar{z}$	$\cdot$	$\cdot$	$\cdot$	$-\bar{\mu}$	$\cdot$	$\cdot$	$\cdot$
$\bar{y}^2$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\mu - 2\bar{\mu}$	$\cdot$	-1
$\bar{y}z$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$-\bar{\mu}$	$\cdot$
$\bar{y}\bar{z}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\mu - 2\bar{\mu}$

Note that $\mu(0) \neq 0$ and $\mu(0) = -\bar{\mu}(0)$, so for sufficiently small λ, $\mu(\lambda) \neq 0$ and $\mu - 2\bar{\mu} \neq 0$. Therefore, for sufficiently small λ, range $\Psi_2 = \tilde{M}_2$, so all quadratics may be eliminated.

Moving on to cubic terms, we again make changes of the form (5.11), but with ϕ being order 3. Making this substitution into (5.9b), we obtain

$$\dot{z} = \mu z + y + \mu\phi - \mu\phi_y y - \bar{\mu}\phi_{\bar{y}}\bar{y} - \mu\phi_z z - \phi_z y - \bar{\mu}\phi_{\bar{z}}\bar{z} - \phi_{\bar{z}}\bar{y} + c_3 z^2\bar{z} + H_3 + O(4),$$

and so third order terms can be eliminated if we solve the linear equation $\Psi_3(\phi) + H_3 = 0$, where $\Psi_3 : \tilde{M}_3 \longrightarrow \tilde{M}_3$ is defined by

$$\Psi_3(\phi) = \mu\phi - \mu\phi_y y - \bar{\mu}\phi_{\bar{y}}\bar{y} - \mu\phi_z z - \phi_z y - \bar{\mu}\phi_{\bar{z}}\bar{z} - \phi_{\bar{z}}\bar{y}.$$

Observe that

$$\begin{aligned} \tilde{M}_3 = \operatorname{span}\{&y^3, y^2\bar{y}, y\bar{y}^2, y^2 z, y^2\bar{z}, y\bar{y}z, y\bar{y}\bar{z}, yz^2, yz\bar{z}, y\bar{z}^2, \bar{y}^3, \\ &\bar{y}^2 z, \bar{y}^2\bar{z}, \bar{y}z^2, \bar{y}z\bar{z}, \bar{y}\bar{z}^2\}. \end{aligned} \tag{5.13}$$

Table 1

Matrix representation of the mapping $\Psi_3 : \tilde{M}_3 \longrightarrow \tilde{M}_3$ with respect to the basis in (5.13), where $\alpha = -(\mu + \bar{\mu})$ and $\beta = \mu - 3\bar{\mu}$.

	y^3	$y^2\bar{y}$	$y\bar{y}^2$	y^2z	$y^2\bar{z}$	$y\bar{y}z$	$y\bar{y}\bar{z}$	yz^2	$yz\bar{z}$	$y\bar{z}^2$	$\bar{y}^3$	$\bar{y}^2z$	$\bar{y}^2\bar{z}$	$\bar{y}z^2$	$\bar{y}z\bar{z}$	$\bar{y}\bar{z}^2$
y^3	-2μ	·	·	-1	·	·	·	·	·	·	·	·	·	·	·	·
$y^2\bar{y}$	·	α	·	·	-1	-1	·	·	·	·	·	·	·	·	·	·
$y\bar{y}^2$	·	·	$-2\bar{\mu}$	·	·	·	-1	·	·	·	·	-1	·	·	·	·
y^2z	·	·	·	-2μ	·	·	·	-2	·	·	·	·	·	·	·	·
$y^2\bar{z}$	·	·	·	·	α	·	·	·	-1	·	·	·	·	·	·	·
$y\bar{y}z$	·	·	·	·	·	α	·	·	-1	·	·	·	·	-2	·	·
$y\bar{y}\bar{z}$	·	·	·	·	·	·	$-2\bar{\mu}$	·	·	-2	·	·	·	·	-1	·
yz^2	·	·	·	·	·	·	·	-2μ	·	·	·	·	·	·	·	·
$yz\bar{z}$	·	·	·	·	·	·	·	·	α	·	·	·	·	·	·	·
$y\bar{z}^2$	·	·	·	·	·	·	·	·	·	$-2\bar{\mu}$	·	·	·	·	·	·
$\bar{y}^3$	·	·	·	·	·	·	·	·	·	·	β	·	-1	·	·	·
$\bar{y}^2z$	·	·	·	·	·	·	·	·	·	·	·	$-2\bar{\mu}$	·	·	-1	·
$\bar{y}^2\bar{z}$	·	·	·	·	·	·	·	·	·	·	·	·	β	·	·	-2
$\bar{y}z^2$	·	·	·	·	·	·	·	·	·	·	·	·	·	α	·	·
$\bar{y}z\bar{z}$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	$-2\bar{\mu}$	·
$\bar{y}\bar{z}^2$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	β

Then Ψ_3 applied to each of the basis elements in (5.13) yields

$$
\begin{aligned}
y^3 &\mapsto -2\mu y^3, & y^2\bar{y} &\mapsto -(\mu+\bar{\mu})y^2\bar{y},\\
y\bar{y}^2 &\mapsto -2\bar{\mu}y\bar{y}^2, & y^2z &\mapsto -2\mu y^2z - y^3,\\
y^2\bar{z} &\mapsto -(\mu+\bar{\mu})y^2\bar{z} - y^2\bar{y}, & y\bar{y}z &\mapsto -(\mu+\bar{\mu})y\bar{y}z - y^2\bar{y},\\
y\bar{y}\bar{z} &\mapsto -2\bar{\mu}y\bar{y}\bar{z} - y\bar{y}^2, & yz^2 &\mapsto -2\mu yz^2 - 2y^2z,\\
yz\bar{z} &\mapsto -(\mu+\bar{\mu})yz\bar{z} - y^2\bar{z} - y\bar{y}z, & y\bar{z}^2 &\mapsto -2\bar{\mu}y\bar{z}^2 - 2y\bar{y}\bar{z},\\
\bar{y}^3 &\mapsto (\mu-3\bar{\mu})\bar{y}^3, & \bar{y}^2z &\mapsto -2\bar{\mu}\bar{y}^2z - y\bar{y}^2,\\
\bar{y}^2\bar{z} &\mapsto \mu\bar{y}^2\bar{z} - 2\bar{\mu}\bar{y}^2\bar{z} - \bar{y}^3, & \bar{y}z^2 &\mapsto -(\mu+\bar{\mu})\bar{y}z^2 - 2y\bar{y}z,\\
\bar{y}z\bar{z} &\mapsto -2\bar{\mu}\bar{y}z\bar{z} - y\bar{y}\bar{z} - \bar{y}^2z, & \bar{y}\bar{z}^2 &\mapsto -(\mu+\bar{\mu})\bar{y}\bar{z}^2 - 2\bar{y}^2\bar{z}.
\end{aligned}
$$

Thus Ψ_3 can be represented by the matrix shown in Table 1.

Observe that

$$-2\mu(0) = -2i, \quad -2\bar{\mu}(0) = 2i, \quad \mu(0) - 3\bar{\mu}(0) = 4i, \quad \text{and} \quad \mu(0) + \bar{\mu}(0) = 0.$$

Thus for λ sufficiently close to 0

$$-2\mu(\lambda) \neq 0, \quad -2\bar{\mu}(\lambda) \neq 0, \quad \text{and} \quad \mu(\lambda) - 3\bar{\mu}(\lambda) \neq 0.$$

It is straightforward to check that

$$\ker \Psi|_{\lambda=0} = \left\{\bar{y}z^2, yz\bar{z}\right\}.$$

Thus for λ sufficiently small, third order terms other than $\bar{y}z^2$ or $yz\bar{z}$ can be eliminated. Therefore the normal form up to third order is as in (5.6).

It is straightforward to verify that $h(y,z) = \alpha\bar{y}z^2 + \beta yz\bar{z}$ is $\mathbf{S}^1$-equivariant under the action of (5.5). Thus it follows that (5.6) is also $\mathbf{S}^1$-equivariant. ■

5.2. Proof of Lemmas 5.4 and 5.5.

Proof of Lemma 5.4. Observe that Proposition 1.6 implies that $p_u(\mathbf{0}) = q_u(\mathbf{0}) = 0$ since the feed-forward chain satisfies (1.23) for S and $\hat{S}$. Then by Propositions 1.1 and 1.2 the general form of the Liapunov–Schmidt reduced equation is

$$\begin{aligned} 0 &= p(u,\lambda,\tau) = u\hat{p}(u,\lambda,\tau) - \tau^2 + \lambda^2 + O(|\tau,\lambda|^3), \\ 0 &= q(u,\lambda,\tau) = u\hat{q}(u,\lambda,\tau) + 2\tau\lambda + O(|\tau,\lambda|^3). \end{aligned} \tag{5.14}$$

Consider the branch of solutions on which $u = x^2$ grows at $O(\lambda^{\frac{1}{3}})$, and introduce a scaling parameter s such that $\lambda = s^3$. Then since τ scales linearly with λ we have

$$u = sv(s) \quad \text{and} \quad \tau = s^3\tilde{\tau}(s),$$

where $v(0) \neq 0$ and $\tilde{\tau}(0) \neq 0$. Then (5.14) becomes

$$\begin{aligned} 0 &= sv\hat{p}(sv, s^3, s^3\tilde{\tau}) - s^6\tilde{\tau}^2 + s^6 + O(s^9), \\ 0 &= sv\hat{q}(sv, s^3, s^3\tilde{\tau}) + 2s^6\tilde{\tau} + O(s^9). \end{aligned} \tag{5.15}$$

Expanding $\hat{p}$ and $\hat{q}$ in powers of s, we obtain

$$\begin{aligned} \hat{p}(sv, s^3, s^3\tilde{\tau}) &= s\hat{p}_u v + s^2\hat{p}_{uu}v + O(s^3), \\ \hat{q}(sv, s^3, s^3\tilde{\tau}) &= s\hat{q}_u v + s^2\hat{q}_{uu}v + O(s^3), \end{aligned}$$

and so (5.15) becomes

$$\begin{aligned} 0 &= s^2\hat{p}_u v^2 + s^3\hat{p}_{uu}v^2 + O(s^4), \\ 0 &= s^2\hat{q}_u v^2 + s^3\hat{q}_{uu}v^2 + O(s^4). \end{aligned}$$

Equating powers of s, we obtain

$$\hat{p}_u = \hat{p}_{uu} = \hat{q}_u = \hat{q}_{uu} = 0,$$

which implies the result, by definition of $\hat{p}$ and $\hat{q}$ in (5.14). ■

Proof of Lemma 5.5. Since the feed-forward chain satisfies (1.23) for S and $\hat{S}$, Lemma 4.4 applies. However, the skew-product form of the feed-forward chain and the fact that $\hat{S}$ is not polydiagonal lead to the following stronger form of Lemma 4.4.

Lemma 5.6. *If one of the arguments $c_1, \ldots, c_m$ lies in $\hat{S}$, then*

$$\mathrm{d}^m F(c_1, \ldots, c_m) \in \hat{S}. \tag{5.16}$$

If one of the arguments $u_1, \ldots, u_m$ lies in $\hat{\mathcal{S}}_{2\pi}$, then

$$\mathrm{d}^m \Phi(u_1, \ldots, u_m) \in \hat{\mathcal{S}}_{2\pi}.$$

Proof. Suppose that one of the arguments $c_1, \dots, c_m$ lies in $\hat{S}$. Since $\mathrm{d}^m F$ is symmetric in $c_1, \dots, c_m$ we can assume without loss of generality that this is the first component. Then $c_{1,j} = 0$ if $j \leq 2k$, so

$$\mathrm{d}^m F(c_1, \dots, c_m) = \sum_{\substack{i_1 = 2k+1, \dots, 3k \\ i_2, \dots, i_m = 1, \dots, 3k}} \left. \frac{\partial^m F}{\partial x_{i_1} \cdots \partial x_{i_m}} \right|_0 c_{1,i_1} \cdots c_{m,i_m}, \tag{5.17}$$

since all terms with $i_1 \leq 2k$ vanish.

Observe that for $i_1 = 2k+1, \dots, 3k$,

$$\left. \frac{\partial^m F}{\partial x_{i_1} \cdots \partial x_{i_m}} \right|_0 = \begin{pmatrix} \frac{\partial^m f(z_1, z_1)}{\partial x_{i_1} \cdots \partial x_{i_m}} \\ \frac{\partial^m f(z_2, z_1)}{\partial x_{i_1} \cdots \partial x_{i_m}} \\ \frac{\partial^m f(z_3, z_2)}{\partial x_{i_1} \cdots \partial x_{i_m}} \end{pmatrix}_0 = \begin{pmatrix} 0 \\ 0 \\ \frac{\partial^m f(z_3, z_2)}{\partial x_{i_1} \cdots \partial x_{i_m}} \end{pmatrix}_0,$$

and so (5.17) becomes

$$\mathrm{d}^m F(c_1, \dots, c_m) = \sum_{\substack{i_1 = 2k+1, \dots, 3k \\ i_2, \dots, i_m = 1, \dots, 2k}} \begin{pmatrix} 0 \\ 0 \\ \frac{\partial^m f(z_3, z_2)}{\partial x_{i_1} \cdots \partial x_{i_m}} \end{pmatrix}_0 c_{1,i_1} \cdots c_{m,i_m},$$

which lies in $\hat{S}$ by the definition in (5.2).

Similarly, suppose that $u_1 \in \hat{\mathcal{S}}_{2\pi}$. Then each u_1 is a linear combination of terms of the form $e^{lis} c_1$ for some $c_1 \in \hat{S}$, $l \in \mathbf{Z}$. So $\mathrm{d}^m \Phi(u_1, \dots, u_m)$ is a linear combination of terms of the form

$$e^{lis} \mathrm{d}^m F(c_1, \dots, c_m),$$

which lie in $\hat{\mathcal{S}}_{2\pi}$ since $\mathrm{d}^m F(c_1, \dots, c_m) \in \hat{S}$ by (5.16). ∎

In line with previous calculations, we assume that F is odd: if the result is true in this restricted case, then it will certainly be true generically. To simplify notation, let $\phi_{(k)}$ and $W_{(k)}$ denote the kth x derivatives of ϕ and W. Then using the formula for $\phi_{(9)}$ given in (A.4) along with Lemmas 4.5 and 5.6 and the fact that $\mathrm{d}\Phi(W_{(9)}) \in \mathcal{R}$ by definition, we obtain

$$\phi_{(9)} = 280(I - E)\mathrm{d}^3 \Phi(W_{(3)}, W_{(3)}, W_{(3)}). \tag{5.18}$$

We claim that

$$\left\langle v_j^*, \mathrm{d}^3 \Phi(W_{(3)}, W_{(3)}, W_{(3)}) \right\rangle = \left\langle v_j^*, \mathrm{Re}\{e^{is} \mathrm{d}^3 F(\eta_1, \eta_1, \bar{\eta}_1)\} \right\rangle. \tag{5.19}$$

To see this, observe that $c = (0, 0, a) \in \hat{S}$, and thus $\mathrm{d}^3 F(c, c, c)$ and $\mathrm{d}^3 F(c, c, \bar{c})$ both lie in $\hat{S}$ by (5.16). Now consider the occurrence of η_3 in the formula for $W_{(3)}$ in Lemma 4.5 and observe that $\eta_3 = (J - 3iI_n)^{-1} \mathrm{d}^3 F(c, c, c)$ also lies in $\hat{S}$ because $\hat{S}$ is invariant for $(J - 3iI_n)^{-1}$. Thus any terms in the expansion of $\mathrm{d}^3 \Phi(W_{(3)}, W_{(3)}, W_{(3)})$ of the form $e^{\pm mis} \mathrm{d}^3 F(\eta_3, \cdot, \cdot)$ will lie in $\hat{\mathcal{S}}_{2\pi}$

by Lemma 5.6 and vanish in the projection onto $\mathcal{K}^*$ by (5.3). Furthermore, $e^{\pm 3is}\mathrm{d}^3F(\eta_1,\eta_1,\eta_1)$ also vanishes in the projection, and we are left with (5.19). Therefore, if we show that

$$\mathrm{d}^3F(\eta_1,\eta_1,\bar{\eta}_1)^t\bar{d}\neq 0, \tag{5.20}$$

then the result will follow.

Recall from (4.17) that

$$(J-iI_n)\eta_1 = 3\mathrm{d}^3F(c,c,\bar{c}).$$

Since $\mathrm{d}^3F(c,c,\bar{c})\in\hat{S}$, and since there are no other constraints on $\mathrm{d}^3F(c,c,\bar{c})$, the projection onto the critical eigenspace E_i will generically be nonzero. Note also that $(J-iI_n)$ is not invertible. The kernel of $(J-iI_n)$ is E_i, and the preimage of E_i under $(J-iI_n)$ is the generalized eigenspace G_i. Hence, the projection of η_1 onto G_i will also be generically nonzero in order to pick up the component of $\mathrm{d}^3F(c,c,\bar{c})$ in E_i. Thus we can write

$$\eta_1 = \alpha b + w,$$

where $\alpha\in\mathbf{C}$ and $w\in\hat{S}-E_i$. Therefore $\eta_1\notin S$ since $b\notin S$.

Thus, using the linearity of d^3F, we have

$$\mathrm{d}^3F(\eta_1,\eta_1,\bar{\eta}_1)^t\bar{d} = \alpha^3\mathrm{d}^3F(b,b,\bar{b})^t\bar{d}$$

since any terms $\mathrm{d}^3F(w,\cdot,\cdot)$ lie in $\hat{S}$ by (5.16) and are therefore orthogonal to d by (4.11). Generically $\alpha^3\mathrm{d}^3F(b,b,\bar{b})^t\bar{d}\neq 0$, because $b^t\bar{d}\neq 0$ by (2.15). This proves (5.20), and hence Proposition 5.5. ■

6. Further examples of nilpotent Hopf bifurcation. In this section we consider three additional examples of three-cell networks, shown in Figures 8, 9, and 10, that can have nilpotent Hopf bifurcations.

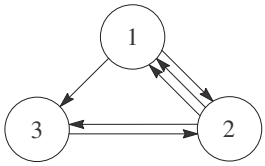


Figure 8. *Another three-cell network with two or four branches at a nilpotent Hopf bifurcation.*

6.1. Another network with multiple $O(\lambda^{\frac{1}{2}})$ branches. The network in Figure 8 is defined by

$$\begin{aligned}\dot{x}_1 &= f(x_1,\overline{x_2,x_2}),\\ \dot{x}_2 &= f(x_2,\overline{x_1,x_3}),\\ \dot{x}_2 &= f(x_3,\overline{x_1,x_2})\end{aligned} \tag{6.1}$$

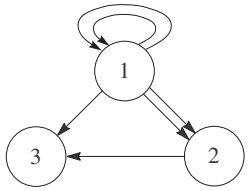


Figure 9. *A feed-forward-like three-cell network.*

and has Jacobian

$$J = \begin{pmatrix} A & 2B & 0 \\ B & A & B \\ B & B & A \end{pmatrix}.$$

The $3k$ eigenvalues and eigenvectors of J are

Eigenvector	Eigenvalues	Algebraic multiplicity	Geometric multiplicity
$(-2u, u, u)^t$	$A - B$	2	1
$(v, v, v)^t$	$A + 2B$	1	1

where u is an eigenvector of $A - B$ and v is an eigenvector of $A + 2B$. It follows that when $k \geq 2$, (6.1) can have a codimension one nilpotent Hopf bifurcation if $A - B$ has a purely imaginary pair of eigenvalues.

Suppose that a is the critical eigenvector of $A - B$. Then the critical eigenvector of J and the corresponding generalized eigenvector are

$$c = \begin{pmatrix} -2a \\ a \\ a \end{pmatrix} \quad \text{and} \quad b = \zeta \begin{pmatrix} 2a \\ -7a \\ 11a \end{pmatrix} + \begin{pmatrix} -2w \\ w \\ w \end{pmatrix},$$

where $\zeta \in \mathbf{C}$ and $w \in \mathbf{C}^k$ are chosen so that $\bar{w}^t a = 0$ and $(A - B - iI_k)w = -(6\zeta B - I_k)a$.

Observe that $S = \{(u, v, v) : u, v \in \mathbf{R}^k\}$ is a synchrony subspace for this network and that S contains the critical eigenspace but not the generalized eigenspace. Hence this network satisfies (1.23). Therefore, Proposition 1.6 implies that there exists a branch of solutions that grows at $O(\lambda^{\frac{1}{2}})$, and that (1.18) holds. Thus, in the absence of any further constraints that force $p_{uu} = q_{uu} = 0$, it follows from Theorem 1.4 that there exist two or four solutions, each growing at $O(\lambda^{\frac{1}{2}})$. We do not verify the absence of such constraints here, but we note that the absence of any other flow-invariant subspaces suggests, by analogy with the networks studied previously, that the second derivatives of p and q are indeed unconstrained.

6.2. Two networks with branches that grow at $O(\lambda^{\frac{1}{6}})$. In addition to the three-cell feed-forward chain of section 5, there are two other three-cell networks, shown in Figures 9 and 10, that can have branches of solutions that grow at $O(\lambda^{\frac{1}{6}})$.

The network in Figure 9 is defined by

$$\begin{aligned} \dot{x}_1 &= f(x_1, \overline{x_1, x_1}), \\ \dot{x}_2 &= f(x_2, \overline{x_1, x_1}), \\ \dot{x}_3 &= f(x_3, \overline{x_1, x_2}) \end{aligned} \tag{6.2}$$

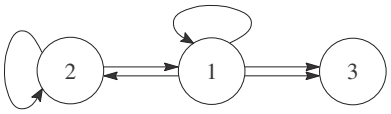


Figure 10. *Another feed-forward-like three-cell network.*

and has Jacobian

$$J = \begin{pmatrix} A+2B & 0 & 0 \\ 2B & A & 0 \\ B & B & A \end{pmatrix}.$$

The network in Figure 10 is defined by

$$\begin{aligned} \dot{x}_1 &= f(x_1, \overline{x_1, x_2}), \\ \dot{x}_2 &= f(x_2, \overline{x_1, x_2}), \\ \dot{x}_3 &= f(x_3, \overline{x_1, x_1}) \end{aligned} \tag{6.3}$$

and has Jacobian

$$J = \begin{pmatrix} A+B & B & 0 \\ B & A+B & 0 \\ 2B & 0 & A \end{pmatrix}.$$

In both cases, the $3k$ eigenvalues and eigenvectors of J are

Eigenvector	Eigenvalues	Algebraic multiplicity	Geometric multiplicity
$(0,0,u)^t$	A	2	1
$(v,v,v)^t$	$A+2B$	1	1

where u is an eigenvector of A and v is an eigenvector of $A+2B$. It follows that when $k \geq 2$, (6.2) and (6.3) can have a codimension one nilpotent Hopf bifurcation if A has a purely imaginary pair of eigenvalues.

Consider first the network shown in Figure 9 and defined by (6.2). Suppose that a is the critical eigenvector of A. Then the critical eigenvector of J and the corresponding generalized eigenvector are

$$c = \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 \\ \zeta a \\ w \end{pmatrix},$$

where $\zeta \in \mathbf{C}$ and $w \in \mathbf{C}^k$ are chosen so that $\bar{w}^t a = 0$ and $(A - iI_k)w = -(\zeta B - I_k)a$.

Observe that $S = \{(u,u,v) : u,v \in \mathbf{R}^k\}$ and $\hat{S} = \{(0,0,v) : v \in \mathbf{R}^k\}$ are invariant subspaces for this network, which both contain the critical eigenspace but not the generalized eigenspace. Hence this network satisfies (1.23), and Proposition 1.6 implies that there exists a branch of solutions that grows at $O(\lambda^{\frac{1}{2}})$ and that (1.18) holds. This branch is obtained by

restricting the system to $\hat{S}$ so that $x_1 = x_2 = 0$ and observing that cell 3 undergoes a standard Hopf bifurcation.

This branch of solutions is unstable because the origin is unstable for cell 2. However, the same argument used in [2, Lemma 6.1] can be employed to show that there exists an additional branch of solutions that grows at $O(\lambda^{\frac{1}{6}})$. Suppose that the eigenvalues of B are negative so that the origin in the first cell is stable for $\dot{x}_1 = f(x_1, \overline{x_1, x_1}, \lambda)$ if λ is sufficiently small. Thus we may assume that $x_1 = 0$.

$$\begin{aligned}\dot{x}_2 &= f(x_2, \overline{0, 0}, \lambda) = g(x_2, \lambda),\\ \dot{x}_3 &= f(x_3, \overline{0, x_2}, \lambda) = h(x_3, x_2, \lambda),\end{aligned}$$

which is precisely the form of the reduced feedforward network in (5.4). Thus the same $\mathbf{S}^1$-equivariant normal form can be obtained as in section 5.1, and hence the arguments of [2, Lemma 6.1] are applicable.

Now consider the network shown in Figure 10 and defined by (6.3). Suppose that a is the critical eigenvector of A. Then the critical eigenvector of J and the corresponding generalized eigenvector are

$$c = \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} \zeta a \\ -\zeta a \\ w \end{pmatrix},$$

where $\zeta \in \mathbf{C}$ and $w \in \mathbf{C}^k$ are chosen so that $\bar{w}^t a = 0$ and $(A - iI_k)w = -(2\zeta B - I_k)a$.

Again, observe that $S = \{(u, u, v) : u, v \in \mathbf{R}^k\}$ and $\hat{S} = \{(0, 0, v) : v \in \mathbf{R}^k\}$ are invariant subspaces that satisfy (1.23), so that Proposition 1.6 implies a branch of solutions that grows at $O(\lambda^{\frac{1}{2}})$ and that (1.18) holds.

Observe that cells 1 and 2 form a $\mathbb{Z}_2$-equivariant subsystem that is not influenced by cell 3. A synchrony-breaking Hopf bifurcation in this subsystem yields a branch of periodic solutions that grows as $\lambda^{\frac{1}{2}}$ and satisfies

$$x_1(t) = x_2\left(t + \tfrac{1}{2}\right).$$

Thus the bifurcation in cell 3 is forced by $x_1(t)$ in exactly the same way as it is forced by cell 2 in the feed-forward chain. Assuming that the normal form on the center manifold is $\mathbf{S}^1$-equivariant under the action in (5.5), it follows from the proof of [2, Lemma 6.1] that there exists a branch that grows at $O(\lambda^{\frac{1}{6}})$.

Appendix. Derivatives of the reduced mapping. The following is a collection of all the derivatives of the reduced mapping ϕ. These are derived by definition; see [3, pp. 31–33]. For higher derivatives we use the notation $\phi_{(k)}$ and $W_{(k)}$ to signify the kth x derivative of ϕ and W. In order to keep these formulas as readable as possible we have used the fact that $W_x(0, 0, 0) = 0$, but it should be remembered that in deriving $\phi_{(k+1)}$ from $\phi_{(k)}$, an argument of v_1 should be read as $v_1 + W_x$.

$$\phi_{xxx} = (I - E)(\mathrm{d}^3\Phi(v_1, v_1, v_1) + 3\mathrm{d}^2\Phi(v_1, W_{xx}) + \mathrm{d}\Phi(W_{xxx})) \tag{A.1}$$

$$\begin{aligned}\phi_{(5)} = (I-E)\big(&\mathrm{d}^5\Phi(v_1,v_1,v_1,v_1,v_1) + 10\mathrm{d}^4\Phi(v_1,v_1,v_1,W_{xx})\\ &+ 15\mathrm{d}^3\Phi(v_1,W_{xx},W_{xx}) + 10\mathrm{d}^3\Phi(v_1,v_1,W_{xxx})\\ &+ 10\mathrm{d}^2\Phi(W_{xx},W_{xxx}) + 5\mathrm{d}^2\Phi(v_1,W_{(4)}) + \mathrm{d}\Phi(W_{(5)})\big)\end{aligned} \tag{A.2}$$

$$\begin{aligned}\phi_{(7)} = (I-E)\big(&\mathrm{d}^7\Phi(v_1,v_1,v_1,v_1,v_1,v_1,v_1) + 21\mathrm{d}^6\Phi(v_1,v_1,v_1,v_1,v_1,W_{xx})\\ &+ 105\mathrm{d}^5\Phi(v_1,v_1,v_1,W_{xx},W_{xx}) + 35\mathrm{d}^5\Phi(v_1,v_1,v_1,v_1,W_{xxx})\\ &+ 105\mathrm{d}^4\Phi(v_1,W_{xx},W_{xx},W_{xx}) + 210\mathrm{d}^4\Phi(v_1,v_1,W_{xx},W_{xxx})\\ &+ 35\mathrm{d}^4\Phi(v_1,v_1,v_1,W_{(4)}) + 105\mathrm{d}^3\Phi(W_{xx},W_{xx},W_{xxx})\\ &+ 70\mathrm{d}^3\Phi(v_1,W_{xxx},W_{xxx}) + 105\mathrm{d}^3\Phi(v_1,W_{xx},W_{(4)})\\ &+ 21\mathrm{d}^3\Phi(v_1,v_1,W_{(5)}) + 35\mathrm{d}^2\Phi(W_{xxx},W_{(4)})\\ &+ 21\mathrm{d}^2\Phi(W_{xx},W_{(5)}) + 7\mathrm{d}^2\Phi(v_1,W_{(6)}) + \mathrm{d}\Phi(W_{(7)})\big)\end{aligned} \tag{A.3}$$

$$\begin{aligned}\phi_{(9)} = (I-E)\big(&\mathrm{d}^9\Phi(v_1,v_1,v_1,v_1,v_1,v_1,v_1,v_1,v_1) + 36\mathrm{d}^8\Phi(v_1,v_1,v_1,v_1,v_1,v_1,v_1,W_{xx})\\ &+ 378\mathrm{d}^7\Phi(v_1,v_1,v_1,v_1,v_1,W_{xx},W_{xx}) + 84\mathrm{d}^7\Phi(v_1,v_1,v_1,v_1,v_1,v_1,W_{xxx})\\ &+ 1260\mathrm{d}^6\Phi(v_1,v_1,v_1,W_{xx},W_{xx},W_{xx}) + 1260\mathrm{d}^6\Phi(v_1,v_1,v_1,v_1,W_{xx},W_{xxx})\\ &+ 126\mathrm{d}^6\Phi(v_1,v_1,v_1,v_1,v_1,W_{(4)}) + 945\mathrm{d}^5\Phi(v_1,W_{xx},W_{xx},W_{xx},W_{xx})\\ &+ 3780\mathrm{d}^5\Phi(v_1,v_1,W_{xx},W_{xx},W_{xxx}) + 840\mathrm{d}^5\Phi(v_1,v_1,v_1,W_{xxx},W_{xxx})\\ &+ 1260\mathrm{d}^5\Phi(v_1,v_1,v_1,W_{xx},W_{(4)}) + 126\mathrm{d}^5\Phi(v_1,v_1,v_1,v_1,W_{(5)})\\ &+ 1260\mathrm{d}^4\Phi(W_{xx},W_{xx},W_{xx},W_{xxx}) + 2520\mathrm{d}^4\Phi(v_1,W_{xx},W_{xxx},W_{xxx})\\ &+ 1890\mathrm{d}^4\Phi(v_1,W_{xx},W_{xx},W_{(4)}) + 1260\mathrm{d}^4\Phi(v_1,v_1,W_{xxx},W_{(4)})\\ &+ 756\mathrm{d}^4\Phi(v_1,v_1,W_{xx},W_{(5)}) + 84\mathrm{d}^4\Phi(v_1,v_1,v_1,W_{(6)})\\ &+ 280\mathrm{d}^3\Phi(W_{xxx},W_{xxx},W_{xxx}) + 1260\mathrm{d}^3\Phi(W_{xx},W_{xxx},W_{(4)})\\ &+ 378\mathrm{d}^3\Phi(W_{xx},W_{xx},W_{(5)}) + 315\mathrm{d}^3\Phi(v_1,W_{(4)},W_{(4)})\\ &+ 504\mathrm{d}^3\Phi(v_1,W_{xxx},W_{(5)}) + 252\mathrm{d}^3\Phi(v_1,W_{xx},W_{(6)})\\ &+ 36\mathrm{d}^3\Phi(v_1,v_1,W_{(7)}) + 126\mathrm{d}^2\Phi(W_{(4)},W_{(5)}) + 84\mathrm{d}^2\Phi(W_{xxx},W_{(6)})\\ &+ 36\mathrm{d}^2\Phi(W_{xx},W_{(7)}) + 9\mathrm{d}^2\Phi(v_1,W_{(8)}) + \mathrm{d}\Phi(W_{(9)})\big)\end{aligned} \tag{A.4}$$

The following are differentials involving parameters α and β:

$$\phi_{\alpha x} = (I-E)\left(\mathrm{d}\Phi_\alpha(v_1) + \mathrm{d}\Phi(W_{\alpha x}) + \mathrm{d}^2\Phi(v_1,W_\alpha)\right), \tag{A.5}$$

$$\begin{aligned}\phi_{\alpha\beta x} = (I-E)\big(&\mathrm{d}\Phi_{\alpha\beta}(v_1) + \mathrm{d}\Phi_\alpha(W_{\beta x}) + \mathrm{d}\Phi_\beta(W_{\alpha x}) + \mathrm{d}\Phi(W_{\alpha\beta x})\\ &+ \mathrm{d}^2\Phi_\alpha(v_1,W_\beta) + \mathrm{d}^2\Phi_\beta(v_1,W_\alpha) + \mathrm{d}^2\Phi(W_{\alpha x},W_\beta)\\ &+ \mathrm{d}^2\Phi(W_{\beta x},W_\alpha) + \mathrm{d}^2\Phi(v_1,W_{\alpha\beta}) + \mathrm{d}^3\Phi(v_1,W_\alpha,W_\beta)\big),\end{aligned} \tag{A.6}$$

$$\begin{aligned}\phi_{\alpha xx} = (I-E)\big(&\mathrm{d}^3\Phi(v_1,v_1,W_\alpha) + 2\mathrm{d}^2\Phi(v_1,W_{\alpha x}) + \mathrm{d}^2\Phi(W_{xx},W_\alpha)\\ &+ \mathrm{d}\Phi(W_{\alpha xx}) + \mathrm{d}^2\Phi_\alpha(v_1,v_1) + \mathrm{d}\Phi_\alpha(W_{xx})\big),\end{aligned} \tag{A.7}$$

$$\begin{aligned}\phi_{\alpha xxx} = (I-E)\big(&\mathrm{d}^4\Phi(v_1,v_1,v_1,W_\alpha) + 3\mathrm{d}^3\Phi(v_1,v_1,W_{\alpha x})\\ &+ 3\mathrm{d}^3\Phi(v_1,W_{xx},W_\alpha) + 3\mathrm{d}^2\Phi(v_1,W_{\alpha xx})\\ &+ 3\mathrm{d}^2\Phi(W_{xx},W_{\alpha x}) + \mathrm{d}^2\Phi(W_{xxx},W_\alpha) + \mathrm{d}\Phi(W_{\alpha xxx})\\ &+ \mathrm{d}^3\Phi_\alpha(v_1,v_1,v_1) + 3\mathrm{d}^2\Phi_\alpha(v_1,W_{xx}) + \mathrm{d}\Phi_\alpha(W_{xxx})\big).\end{aligned} \tag{A.8}$$

The following formulas for the $W_{(k)}$ are obtained by differentiating

$$E\Phi(xv_1 + W(xv_1,\lambda,\tau),\lambda,\tau) \equiv 0 \tag{A.9}$$

k times with respect to x. This yields an expression of the form

$$E(\cdots) + E\mathrm{d}\Phi(W_{(k)}) = 0,$$

which can be rearranged to give

$$\mathrm{d}\Phi(W_{(k)}) = -E(\cdots)$$

since $\mathrm{d}\Phi(W_{(k)}) \in \mathcal{R}$ on which E acts as the identity, and hence

$$W_{(k)} = -\mathcal{L}^{-1}E(\cdots).$$

In this way, we obtain

$$W_{xx} = -\mathcal{L}^{-1}E\big(\mathrm{d}^2\Phi(v_1,v_1)\big), \tag{A.10}$$

$$W_{xxx} = -\mathcal{L}^{-1}E\big(\mathrm{d}^3\Phi(v_1,v_1,v_1) + 3\mathrm{d}^2\Phi(v_1,W_{xx})\big), \tag{A.11}$$

$$\begin{aligned}W_{xxxx} = -\mathcal{L}^{-1}E\big(&\mathrm{d}^4\Phi(v_1,v_1,v_1,v_1) + 6\mathrm{d}^3\Phi(v_1,v_1,W_{xx})\\ &+ 3\mathrm{d}^2\Phi(W_{xx},W_{xx}) + 4\mathrm{d}^2\Phi(v_1,W_{xxx})\big).\end{aligned} \tag{A.12}$$

Similarly, we obtain the following expressions for $W_{\alpha x}$ by differentiating (A.9) with respect to α and x to obtain

$$E\left(\mathrm{d}^2\Phi(v_1+W_x,W_\alpha) + \mathrm{d}\Phi(W_{\alpha x}) + \mathrm{d}\Phi_\alpha(v_1+W_x)\right) = 0, \tag{A.13}$$

which rearranges to give

$$W_{\alpha x} = -\mathcal{L}^{-1}E\left(\mathrm{d}^2\Phi(v_1,W_\alpha) + \mathrm{d}\Phi_\alpha(v_1)\right). \tag{A.14}$$

By further differentiation of (A.13) with respect to x and rearranging, we obtain

$$\begin{aligned}W_{\alpha xx} = -\mathcal{L}^{-1}E\big(&\mathrm{d}^3\Phi(v_1,v_1,W_\alpha) + \mathrm{d}^2\Phi_\alpha(v_1,v_1)\\ &+ \mathrm{d}^2\Phi(W_{xx},W_\alpha) + 2\mathrm{d}^2\Phi(v_1,W_{\alpha x}) + \mathrm{d}\Phi_\alpha(W_{xx})\big),\end{aligned} \tag{A.15}$$

$$\begin{aligned}W_{\alpha xxx} = -\mathcal{L}^{-1}E\big(&\mathrm{d}^4\Phi(v_1,v_1,v_1,W_\alpha) + 3\mathrm{d}^3\Phi(v_1,v_1,W_{\alpha x})\\ &+ 3\mathrm{d}^3\Phi(v_1,W_{xx},W_\alpha) + \mathrm{d}^3\Phi_\alpha(v_1,v_1,v_1)\\ &+ 3\mathrm{d}^2\Phi(W_{xx},W_{\alpha x}) + \mathrm{d}^2\Phi(W_{xxx},W_\alpha) + 3\mathrm{d}^2\Phi(v_1,W_{\alpha xx})\\ &+ 3\mathrm{d}^2\Phi_\alpha(v_1,W_{xx}) + \mathrm{d}^2\Phi_\alpha(W_{xxx})\big).\end{aligned} \tag{A.16}$$

Acknowledgments. We thank Yunjiao Wang and Maria Leite for helpful conversations.

REFERENCES

[1] S. van Gils, M. Krupa, and W. Langford, *Hopf bifurcation with non-semisimple* 1:1 *resonance*, Nonlinearity, 3 (1990), pp. 825–850.

[2] M. Golubitsky, M. Nicol, and I. Stewart, *Some curious phenomena in coupled cell networks*, J. Nonlinear Sci., 14 (2004), pp. 119–236.

[3] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. I, Appl. Math. Sci. 51, Springer-Verlag, New York, 1984.

[4] M. Golubitsky, I. N. Stewart, and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. II, Appl. Math. Sci. 69, Springer-Verlag, New York, 1988.

[5] M. Golubitsky, I. Stewart, and A. Török, *Patterns of synchrony in coupled cell networks with multiple arrows*, SIAM J. Appl. Dyn. Syst., 4 (2005), pp. 78–100.

[6] M. Leite, *Homogeneous Three-Cell Networks*, Ph.D. thesis, Department of Mathematics, University of Houston, Houston, TX, 2005.

[7] M. Leite and M. Golubitsky, *Homogeneous three-cell networks*, Networks, submitted.

[8] N. S. Namachchivaya, M. Doyle, W. Langford, and N. Evans, *Normal form for generalized Hopf bifurcation with non-semisimple* 1:1 *resonance*, Z. Angew. Math. Phys., 45 (1994), pp. 312–335.

[9] I. Stewart, M. Golubitsky, and M. Pivato, *Symmetry groupoids and patterns of synchrony in coupled cell networks*, SIAM J. Appl. Dyn. Syst., 2 (2003), pp. 609–646.

[10] A. Vanderbauwhede, *Hopf bifurcation at non-semisimple eigenvalues*, in Multiparameter Bifurcation Theory, M. Golubitsky and J. Guckenheimer, eds., Contemp. Math. 56, AMS, Providence, RI, 1986, pp. 343–353.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 2, pp. 252–279

Binary Asteroid Observation Orbits from a Global Dynamical Perspective*

F. Gabern†‡, W. S. Koon†, J. E. Marsden†, and D. J. Scheeres§

Abstract. We study spacecraft motion near a binary asteroid by means of theoretical and computational tools from geometric mechanics and dynamical systems. We model the system assuming that one of the asteroids is a rigid body (ellipsoid) and the other a sphere. In particular, we are interested in finding periodic and quasi-periodic orbits for the spacecraft near the asteroid pair that are suitable for observations and measurements. First, using reduction theory, we study the full two body problem (gravitational interaction between the ellipsoid and the sphere) and use the energy-momentum method to prove nonlinear stability of certain relative equilibria. This study allows us to construct the restricted full three-body problem (RF3BP) for the spacecraft motion around the binary, assuming that the asteroid pair is in relative equilibrium. Then, we compute the modified Lagrangian fixed points and study their spectral stability. The fixed points of the restricted three-body problem are modified in the RF3BP because one of the primaries is a rigid body and not a point mass. A systematic study depending on the parameters of the problem is performed in an effort to understand the rigid body effects on the Lagrangian stability regions. Finally, using frequency analysis, we study the global dynamics near these modified Lagrangian points. From this global picture, we are able to identify (almost-) invariant tori in the stability region near the modified Lagrangian points. Quasi-periodic trajectories on these invariant tori are potentially convenient places to park the spacecraft while it is observing the asteroid pair.

Key words. asteroid pairs, spacecraft dynamics, stability, frequency analysis, invariant tori, quasi-periodic motion

AMS subject classifications. 37N05, 70F07, 70F15

DOI. 10.1137/050641843

1. Introduction.

The dynamics of asteroid pairs has recently become a topic of interest, specially since the first binary asteroid, Ida-Dactyl, was discovered by the Galileo spacecraft in 1993. It is currently estimated that up to 20% of near-earth asteroids (NEA) are binaries [26], and other examples have been found among the asteroid main-belt, in the Trojan swarms and as transneptunian objects [33]. The problem of two rigid bodies orbiting around each other

*Received by the editors October 4, 2005; accepted for publication (in revised form) by T. Kaper February 14, 2006; published electronically June 2, 2006. This work was partly supported by the California Institute of Technology President's Fund, NSF-ITR grant ACI-0204932, NSF grant DMS-0505711, and by ICB, the Institute for Collaborative Biotechnologies, through ARO grant DAAD19-03-D-0004.

http://www.siam.org/journals/siads/5-2/64184.html

†Control and Dynamical Systems, California Institute of Technology, 07-81, Pasadena, CA 91125 (gabern@cds.caltech.edu, koon@cds.caltech.edu, marsden@cds.caltech.edu). The first author was supported by the 2004 Fulbright—Generalitat de Catalunya postdoctoral program, the MCyT/FEDER grant BFM2003-07521-C02-01, and the CIRIT grant 2001SGR–70.

‡Current address: Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain (frederic.gabern@upc.edu).

§Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109-2140 (scheeres@umich.edu). This author acknowledges support from the Jet Propulsion Laboratory/California Institute of Technology through a grant from the Director's Research Discretionary Fund.

attracted by their mutual gravitational forces is challenging from the geometric mechanics point of view [5]. Moreover, a very interesting problem with applications to astrodynamics is the description of the dynamics of a massless particle (e.g., a spacecraft) moving under the influence of the binary [41, 11]. Indeed, binaries can be used as real-life laboratories to test rigid-body gravitational dynamics. As such, these theoretical studies may be relevant for a possible future mission to binaries.

The objective of this paper is to construct and study a model for the motion of a satellite orbiting a binary asteroid. To do this, we first have to develop a model for the asteroid pair itself. A binary asteroid provides a canonical model for general full body problems (FBPs); see [24, 39]. FBPs are concerned with the dynamical interaction between distributed bodies of finite mass. In particular, the full two-body problem (F2BP) considers the dynamics of two spatially extended bodies that interact via their mutual gravitational fields. In this paper, we consider the "sphere restriction" of the F2BP [20]. That is, it is assumed that one of the rigid bodies is spherically symmetric and thus can be considered as a point mass. As for the other rigid body, we assume that it is a triaxial ellipsoid. The study of this simple model of an asteroid pair will give some hints on the dynamics of spacecrafts about binaries, and it can be generalized further using, for instance, more complicated potentials [47].

One of our first goals is to find stable relative equilibria of this F2BP with the property that we can later build models for the satellite motion around the pair. Moreover, as we look first for simple relative equilibria, we also assume that the two bodies are restricted to moving in a plane. This makes the reduction process and the equilibria characterization much simpler, but not too simple, as the coupling between rotational and translational motion is still there. For a probe sent to a binary asteroid, it is plausible to assume that the underlying F2BP is in some type of relative equilibrium. As the study of more complex models will be important in the future, we have to first understand the simplified cases. Thus, we devote the first part of the paper to studying the F2BP by means of reduction methods [28], to identifying its relative equilibria, and to proving nonlinear stability for some cases with the aid of the energy-momentum method [44].

Following this, we choose a particular stable relative equilibria of the F2BP, which corresponds to a periodic orbit in the original system, to construct a model for the satellite orbiting the binary. This model is a restricted problem of three bodies, but one of the primaries is a rigid body. In the literature this type of model has been called the restricted full three-body problem or RF3BP [41].

As is well known [45], the restricted three-body problem (RTBP) has five equilibrium points, and two of them form an equilateral triangle with the primaries. These equilateral equilibrium points, also called Lagrangian points, may persist when one of the primaries is not a point mass but a distributed body [2, 41, 12]. The position and stability properties of these points are, of course, affected by the perturbation and are thus modified. In the second part of the paper, we study how these equilibrium points are modified depending on the variation of the parameters of the problem.

In some previous works [11, 12], we have studied this problem using normal form techniques near the triangular points of some particular RF3BP, which are simpler than the ones presented here. Even though these tools give quite satisfactory results for a range of parameters, the zone around the fixed points where the dynamics can be described by the normal

form is not very big. In this paper, we extend the description of this dynamics to larger regions of the phase space.

To achieve this goal, in the last part of the paper we apply a dynamical systems–based tool, namely frequency map analysis [21, 37, 16], to study the global dynamics around the Lagrangian points. We are able to identify relatively large neighborhoods of these equilibrium points in phase space at which the trajectories are stable for a long time. As far as we know, this is the first time that this powerful numerical tool (frequency analysis) has been applied to orbit mechanics about asteroids.

Previously, a wavelet-based frequency analysis [46] has been used to study the transport in the Sun-Jupiter RTBP. The advantage of this method appears in problems where the frequencies vary with time, for example when there is relatively "fast" transport from one region of phase space to another. In our problem, we are interested in the dynamics close to an elliptic point, and the results given by the wavelet method should be similar to those presented here. In particular, we look for tori that are "invariant enough," i.e., tori where we can place the spacecraft for a long enough time (the meaning of this will be clarified later on). In this case, if there is transport between nearby tori, the transport should be slow.

The paper is organized as follows. In section 2, we develop and study a model of the asteroid pair. The relative equilibria of the reduced binary system are characterized, and their stability is studied by means of the energy-momentum method. In section 3, we construct the model for the satellite orbiting the asteroid pair, i.e., the RF3BP, and we find the new coordinates of the modified Lagrangian points of the RTBP and study its spectral stability. In section 4, we study the global dynamics near the Lagrangian points of the RF3BP by means of frequency analysis methods. This global study allows us to find (almost-) invariant tori and trajectories very close to quasi-periodic, which are pretty suitable for the satellite. Finally, in section 5, our conclusions and future directions are presented. For the convenience of the reader, we add in the appendix a brief review on Abelian reduction theory, which is used in section 2 to study the F2BP.

2. Stability of the asteroid pair.

2.1. Reduced model for the binary. To model the asteroid pair, we consider the mechanical system of a triaxial ellipsoid and a sphere that interact via the mutual gravitational potential and are allowed to move in a plane. In an inertial reference frame, the kinetic energy of the system is

$$K = \frac{1}{2}m\|\dot{\mathbf{q}}\|^2 + \frac{1}{2}M\|\dot{\mathbf{Q}}\|^2 + \frac{1}{2}I_{zz}\dot{\varphi}^2,$$

where $\mathbf{q}$ and $\mathbf{Q}$ are the positions of the sphere's center and the barycenter of the ellipsoid and m and M are, respectively, the masses of the sphere and the ellipsoid. I_{zz} is the component orthogonal to the plane of motion of the inertia tensor of the ellipsoid, and the angle φ is as shown in Figure 1.

This system is invariant under translations. Thus, defining the relative position of the bodies as $\mathbf{r} = \mathbf{q} - \mathbf{Q}$ and taking as unit of mass the reduced mass, i.e., $mM/(m + M) = 1$,

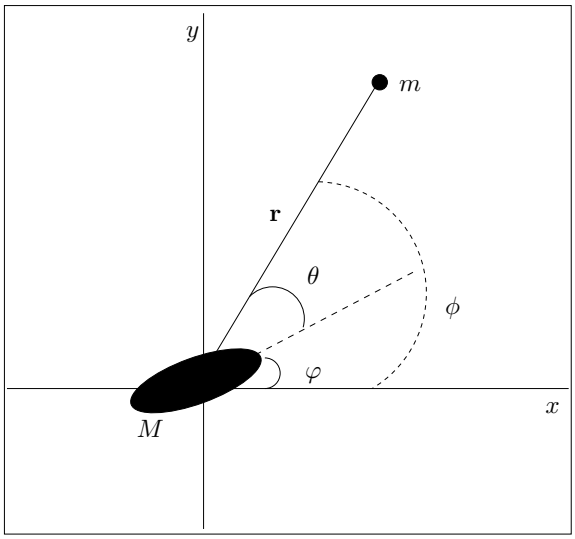


Figure 1. *Gravitational interaction of a rigid body and a sphere in the plane.*

the system is described by the Lagrangian functional

$$L(\mathbf{r}, \varphi, \dot{\mathbf{r}}, \dot{\varphi}) = \frac{1}{2}\|\dot{\mathbf{r}}\|^2 + \frac{1}{2} I_{zz}\dot{\varphi}^2 - V(\mathbf{r}, \varphi), \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidean norm and $V(\cdot)$ is the gravitational potential of the system. The Legendre transform, $(\mathbf{p}, p_\varphi) = (\dot{\mathbf{r}}, I_{zz}\dot{\varphi})$, gives the Hamiltonian formulation of the problem:

$$H = \frac{1}{2}\|\mathbf{p}\|^2 + \frac{1}{2I_{zz}} p_\varphi^2 + V(\mathbf{r}, \varphi). \tag{2.2}$$

We assume that the axes of the ellipsoid, (α, β, γ), are ordered as $0 < \gamma \leq \beta \leq \alpha = 1$ and that the "γ axis" is orthogonal to the plane of motion. Thus, the longest axis of the ellipsoid is taken as the unit of length. The unit of time is taken such that $GmM = 1$. The mass parameter of the system will be denoted $\nu = m/(m + M)$, and then the moment of inertia of the ellipsoid is $I_{zz} = (1 + \beta^2)/(5\nu)$. To write the mutual gravitational potential function of the ellipsoid and sphere, we use Ivory's theorem [25, 38]:

$$V(\mathbf{r}, \varphi) = V(r, \theta) = -\frac{3}{4}\int_{\lambda(r,\theta)}^{+\infty} \Phi(r, \theta; u)\frac{du}{\Delta(u)}, \tag{2.3}$$

where $r = \|\mathbf{r}\|$, $\theta = \phi - \varphi$,

$$\Phi(r, \theta; u) = 1 - \frac{r^2\cos^2\theta}{1 + u} - \frac{r^2\sin^2\theta}{\beta^2 + u},$$

$\Delta(u) = \sqrt{(1 + u)(\beta^2 + u)(\gamma^2 + u)}$, and $\lambda(r, \theta) > 0$ is the largest root of $\Phi(r, \theta; \lambda(r, \theta)) = 0$.

This system still has an overall rotational symmetry; that is, it is invariant under rotations in the plane of motion. In other words, the action of the symmetry group $SO(2)$ leaves the system invariant. Hence, to reduce the dimensionality of the problem, we apply the abelian reduction reviewed in the appendix.

For this purpose, by introducing polar coordinates and relative angles (see [12]), we write the Hamiltonian function in a much more convenient way:

$$H = \frac{1}{2}p_r^2 + \left(\frac{1}{2r^2} + \frac{1}{2I_{zz}}\right)p_\theta^2 + \frac{1}{2I_{zz}}p_\varphi^2 - \frac{1}{I_{zz}}p_\theta p_\varphi + V(r,\theta), \tag{2.4}$$

where $p_\theta = r^2\dot{\theta} + r^2\dot{\varphi}$ and $p_\varphi = r^2\dot{\theta} + (r^2 + I_{zz})\dot{\varphi}$. Notice that φ is a cyclic variable of the Hamiltonian (2.4), and therefore its conjugate momentum p_φ is conserved (Noether's theorem).

We then apply cotangent bundle reduction (see section A.2): The *momentum map* is given by $\mathbf{J}(r,\theta,\varphi,p_r,p_\theta,p_\varphi) = p_\varphi$ and corresponds to the total angular momentum of the system in the new coordinates. The *locked inertia tensor* (instantaneous inertia tensor when the relative motion of the two-body system is locked) is $\mathbb{I}(r,\theta,\varphi) = r^2 + I_{zz}$. The *mechanical connection* is the 1-form given by $\mathcal{A}(r,\theta,\varphi) = \frac{r^2}{r^2+I_{zz}}d\theta + d\varphi$. For a fixed angular momentum $p_\varphi = \mu$, we finally perform the shift from $\mathbf{J}^{-1}(\mu)$ to $\mathbf{J}^{-1}(0)$ as

$$\tilde{p}_r = p_r, \qquad \tilde{p}_\theta = p_\theta - \frac{\mu r^2}{r^2 + I_{zz}}, \qquad \tilde{p}_\varphi = 0.$$

The reduced Hamiltonian in $\mathbf{J}^{-1}(0)/S^1$ has only two degrees of freedom,

$$H_\mu(r,\theta) = \frac{1}{2}\tilde{p}_r^2 + \frac{1}{2}\left(\frac{1}{r^2} + \frac{1}{I_{zz}}\right)\tilde{p}_\theta^2 + V_\mu(r,\theta), \tag{2.5}$$

and $V_\mu(r,\theta)$ is the amended potential,

$$V_\mu(r,\theta) = \frac{\mu^2}{2(r^2 + I_{zz})} - \frac{3}{4}\int_{\lambda(r,\theta)}^{+\infty}\Phi(r,\theta;u)\frac{du}{\Delta(u)}, \tag{2.6}$$

where $\Phi(r,\theta;u)$, $\Delta(u)$, and $\lambda(r,\theta) > 0$ are as defined before, and $\mu \in \mathbb{R}$ is the total angular momentum (fixed). The reduced symplectic form is noncanonical and given by

$$\omega_\mu = dr \wedge d\tilde{p}_r + d\theta \wedge d\tilde{p}_\theta - \frac{2\mu I_{zz} r}{(r^2 + I_{zz})^2}dr \wedge d\theta. \tag{2.7}$$

The equations of motion in the reduced space can be easily derived from $\mathbf{i}_{(\dot{x}^k\partial_{x^k} + \dot{y}_k\partial_{y_k})}\omega_\mu = dH_\mu$, where (x,y) denote the configuration-momenta conjugate pair and $\mathbf{i}_X\Omega$ is the interior product (or contraction) of the vector field X and the 1-form Ω:

$$\dot{r} = \tilde{p}_r, \qquad \dot{\tilde{p}}_r = \frac{\tilde{p}_\theta^2}{r^3} - \frac{\partial V_\mu(r,\theta)}{\partial r} + \frac{2\mu I_{zz} r}{(r^2 + I_{zz})^2}\left(\frac{1}{r^2} + \frac{1}{I_{zz}}\right)\tilde{p}_\theta,$$

$$\dot{\theta} = \left(\frac{1}{r^2} + \frac{1}{I_{zz}}\right)\tilde{p}_\theta, \qquad \dot{\tilde{p}}_\theta = -\frac{\partial V_\mu(r,\theta)}{\partial \theta} - \frac{2\mu I_{zz} r}{(r^2 + I_{zz})^2}\tilde{p}_r.$$

2.2. Relative equilibria. We are now interested in finding the stable (we will make precise later on in which sense) fixed points of the reduced equations. The relative equilibria (or fixed points of the reduced equations) satisfy [27]:

$$\tilde{p}_r = \tilde{p}_\theta = \frac{\partial V_\mu}{\partial r} = \frac{\partial V_\mu}{\partial \theta} = 0.$$

The first two equations give

$$\begin{aligned} p_r &= 0, \\ p_\theta &= \frac{\mu r^2}{r^2 + I_{zz}}, \end{aligned}$$

and the last two give

$$\frac{3}{2} r \left(\cos^2\theta R_\alpha + \sin^2\theta R_\beta\right) - \frac{\mu^2 r}{(r^2 + I_{zz})^2} = 0, \tag{2.8}$$

$$-\frac{3}{4} r^2 \sin 2\theta \left(R_\alpha - R_\beta\right) = 0, \tag{2.9}$$

where R_α and R_β denote the elliptic integrals

$$R_\alpha = \int_{\lambda(r,\theta)}^{+\infty} \frac{du}{(1+u)\Delta(u)}, \qquad R_\beta = \int_{\lambda(r,\theta)}^{+\infty} \frac{du}{(\beta^2+u)\Delta(u)}.$$

2.2.1. Spheroid. First, let us look at the case where the in-plane axes of the ellipsoid are equal. In this case, $0 < \gamma < \beta = \alpha = 1$, and thus $R_\alpha = R_\beta$. Then, (2.9) is satisfied automatically $\forall \theta \in \mathbb{T}$.

From (2.8) and assuming $r > 1$,

$$\frac{\mu^2}{(r^2 + I_{zz})^2} = \frac{3}{2} R_\alpha,$$

where in this case

$$R_\alpha = \int_{r^2-1}^{+\infty} \frac{du}{(1+u)^2\sqrt{\gamma^2+u}}.$$

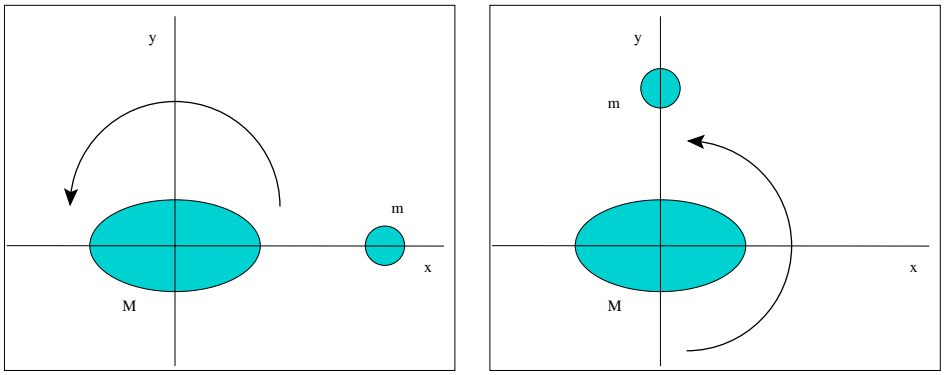
In relative equilibria, the distance between the two bodies is constant ($r \equiv$ constant). We can thus define the following parameter:

$$\omega \equiv \frac{\mu}{r^2 + I_{zz}}. \tag{2.10}$$

The relative equilibria can be seen in the unreduced space as uniformly rotating systems. Then, ω is precisely the angular velocity of this rotating system.

In this degenerate case, the elliptic integral R_α is trivially integrable, and we obtain the following condition for relative equilibria: Given $r > 1$ and $\gamma < 1$,

$$\omega^2 = 3\pi - \frac{3\sqrt{r^2 - 1 + \gamma^2}}{4(1-\gamma^2)r^2} - \frac{3}{2(1-\gamma^2)^{3/2}} \arctan\sqrt{\frac{r^2 - 1 + \gamma^2}{1-\gamma^2}}. \tag{2.11}$$



(a) (b)

Figure 2. *Two types of relative equilibria for the planar ellipsoid–sphere Full 2-Body Problem.* (a) *Long-Axis Equilibrium (LAE).* (b) *Short-Axis Equilibrium (SAE).*

Note: For fixed $\gamma < 1$ and $r > 1$, we have a degenerate "circle" of relative equilibria ($\forall \theta \in \mathbb{T}$) with the binary rotating (in an inertial frame) at a constant angular velocity ω satisfying (2.11). Also, for fixed mass parameter ν, we can compute $I_{zz} = 2/(5\nu)$ and then the angular momentum μ from (2.10).

2.2.2. Triaxial ellipsoid. Now, we focus on the triaxial ellipsoid, where $0 < \gamma \leq \beta < \alpha = 1$. Here, $R_\alpha \neq R_\beta$, and (2.9) yields $\sin 2\theta = 0$, which means that relative equilibria satisfy

$$\theta = k\frac{\pi}{2} \quad \text{for any } k \in \mathbb{Z}.$$

Geometrically, we can distinguish between two types of solutions [40]:

1. LAE (long-axis equilibria) when $\theta = 0$ or π (see Figure 2(a)), and
2. SAE (short-axis equilibria) when $\theta = \pm\frac{\pi}{2}$ (see Figure 2(b)).

Again, defining the angular velocity of the system as in (2.10), we obtain the following relations between ω and r:

1. LAE:

$$\omega^2 = \frac{3}{2}\int_{r^2-1}^{+\infty} \frac{du}{(1+u)\Delta(u)}, \tag{2.12}$$

2. SAE:

$$\omega^2 = \frac{3}{2}\int_{r^2-\beta^2}^{+\infty} \frac{du}{(\beta^2+u)\Delta(u)}, \tag{2.13}$$

where recall that $\Delta(u) = \sqrt{(1+u)(\beta^2+u)(\gamma^2+u)}$. Here, for fixed $0 < \gamma \leq \beta < 1$ and $R \geq 1$, there is a fixed point with coordinates $(r, \theta) = (R, k\pi/2)$ for an angular velocity ω satisfying (2.12) for $k \bmod 2 = 0$, or (2.13) for $k \bmod 2 = 1$. Moreover, given the mass parameter, ν, one can compute the ellipsoid's moment of inertia $I_{zz} = (1+\beta^2)/(5\nu)$ and, using (2.10), the angular momentum of the system μ.

2.3. Stability of relative equilibria: Energy-momentum method. We apply the energy-momentum method of Simo, Lewis, and Marsden [44] to study the stability of the relative equilibria that we have just found. According to this method (see also [31, 3]), to carry out the stability analysis of the LAE and SAE, we must compute $\delta^2 V_\mu$ on the subspace orthogonal to the group G_μ-orbit (in this case, $G_\mu = SO(2) = S^1$).

From (2.6), it is easy to see that

$$\delta^2 V_\mu = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

for $\theta = k\frac{\pi}{2}$ (LAE and SAE) and where $A, B \in \mathbb{R}$ and their values depend, of course, on the particular relative equilibrium.

In the spheroid case, $\alpha = \beta$, and thus

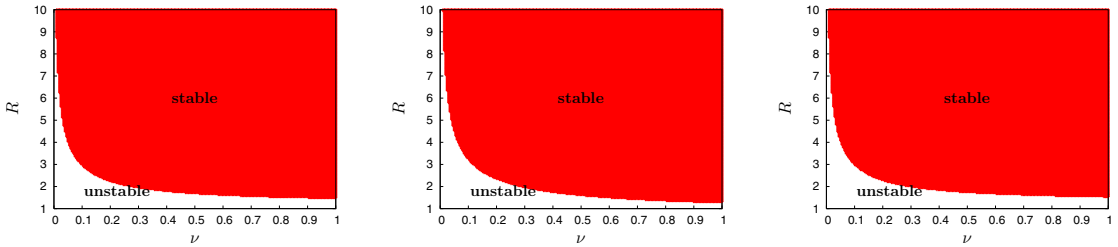
$$A = \frac{4r^2\omega^2}{r^2 + I_{zz}} - \frac{3}{r^2\sqrt{r^2+\gamma^2-1}},$$

where ω satisfies (2.11) and $B = 0$. This is due to the fact that the relative equilibria are degenerate. Therefore, the energy-momentum method is not conclusive for the "circle" of relative equilibria in the spheroid-sphere particular case.

1. *Stability of LAEs.* In this case, $\theta = 0$ or π and thus

$$B = -\frac{3}{2}r^2\left(R_\alpha - R_\beta\right),$$
$$A = \frac{4r^2\omega^2}{r^2 + I_{zz}} - \frac{3}{r\sqrt{(r^2+\beta^2-1)(r^2+\gamma^2-1)}},$$

where ω is given by (2.12). As $\alpha > \beta$, $R_\alpha < R_\beta$ and thus $B > 0$. Then,
(a) if $A > 0$, the LAE is (linearly and nonlinearly) stable;
(b) if $A < 0$, the LAE is unstable (the index is odd [31, 35]).
In Figure 3, we show some ranges for the parameter values that give stability of the LAE.



(a) (b) (c)

Figure 3. LAE. *Nonlinear stability (red zone) and instability (white zone) of the long-axis relative equilibria of the binary with respect to the mass parameter ν and the distance between the two bodies R for some values of the ellipsoid axes:* (a) $\beta = \gamma = 0.5$, (b) $\beta = \gamma = 0.8$, (c) $\beta = 0.5$ *and* $\gamma = 0.25$.

2. *Stability of SAEs.* In this case, $\theta = \pm\frac{\pi}{2}$ and

$$B = \frac{3}{2} r^2 \left(R_\alpha - R_\beta\right),$$
$$A = \frac{4r^2\omega^2}{r^2 + I_{zz}} - \frac{3}{r\sqrt{(r^2+1-\beta^2)(r^2+\gamma^2-\beta^2)}},$$

where ω is given by (2.13). As $\alpha > \beta$, $R_\alpha < R_\beta$ and thus $B < 0$. Then,

(a) if $A > 0$, the SAE is unstable (the index is odd [31, 35]);

(b) if $A < 0$, the stability of the SAE requires further analysis. In this case, we study their spectral stability by computing numerically the eigenvalues of the linearized vector field at the fixed points of the reduced system.

In Figure 4, we show ranges for the parameter values for which the SAE are spectrally stable, unstable, or complex unstable.

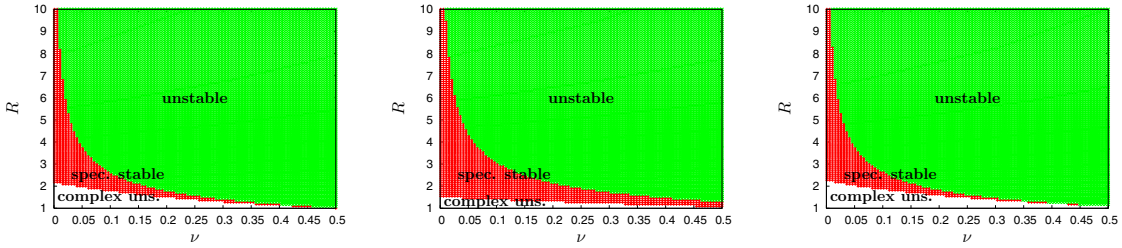


Figure 4. SAE. *Unstable zone (green), complex unstable zone (white), and spectrally stable zone (red) of the short-axis relative equilibria of the binary with respect to the mass parameter ν and the distance between the two bodies R for some values of the ellipsoid axes:* (a) $\beta = \gamma = 0.5$, (b) $\beta = \gamma = 0.8$, (c) $\beta = 0.5$ *and* $\gamma = 0.25$.

3. RF3BP.

We focus now on a satellite influenced by the gravitational potential of the binary. To model the motion of this satellite, we assume that the binary is in a relative equilibrium, and we use the study performed in the last section. As the relative equilibria found in section 2.2 are periodic orbits for the unreduced system, we will write the equations of motion for the satellite in a rotating reference in which the asteroid pair is fixed.

We consider binaries such that the rigid body (ellipsoid) is "big" and the spherical body is "small." Hence, we assume that $\nu \ll 1$. This situation is thought to be the most common in the main belt [33]. Moreover, as for $\nu \ll 1$ and moderate R values the SAE are spectrally stable, we will also assume that the binary is moving in this particular solution of the F2BP. For a study of the motion of a spacecraft near a binary in LAE, see [41].

We will now derive the equations of motion for a satellite orbiting a binary in SAE. We start by assuming that the barycenter of the system is at the origin. As in an inertial reference frame, the SAE is a uniformly rotating motion [12], and we write the equations for the spacecraft in a rotating frame for which the two massive bodies, ellipsoid and sphere, are fixed. More concretely, we choose a reference system where the centers of mass of the ellipsoid and sphere are at the configuration points $(-\nu R, 0, 0)$ and $((1-\nu)R, 0, 0)$, respectively; see Figure 5. In this case, note that the longest axis of the ellipsoid is parallel to the y axis.

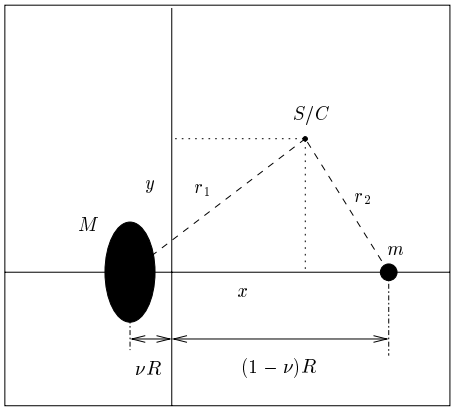


Figure 5. *Model for the spacecraft motion around the binary asteroid. The asteroid pair is assumed to be in SAE, and the equations of motion for the spacecraft are written in a rotating frame.*

Even though the uniformly rotating motion of the binary is in the xy-plane, we will consider that the satellite is allowed to move in the entire three-dimensional xyz-configuration space. The equations of motion for the spacecraft in this situation can be constructed similarly to the RTBP equations [45]. They allow a Hamiltonian formulation and can be obtained from the following Hamiltonian function:

$$\begin{aligned} H(p_x,p_y,p_z,x,y,z) &= \frac{1}{2}\left(p_x^2+p_y^2+p_z^2\right)+\omega\left(yp_x-xp_y\right)+V(x,y,z),\\ V(x,y,z) &= -\frac{3(1-\nu)}{4}\int_{\lambda(x,y,z)}^{+\infty}\Phi(x,y,z;u)\frac{du}{\Delta(u)}-\frac{\nu}{r_2}, \end{aligned} \tag{3.1}$$

where (x,y,z) is the position of the spacecraft in the rotating reference frame, (p_x,p_y,p_z) are the conjugate momenta, and ω is the angular velocity of the rotating system (2.13). We define R to be the distance between the ellipsoid and sphere barycenters. Then, $r_2^2 = (x-(1-\nu)R)^2+y^2+z^2$. The gravitational potential coming from the ellipsoid is computed as before but considering that now the reference frame is tilted by 90°. Then, we compute $V(x,y,z)$ using

$$\Phi(x,y,z;u) = 1-\frac{(x+\nu R)^2}{\beta^2+u}-\frac{y^2}{1+u}-\frac{z^2}{\gamma^2+u}$$

and $\Delta(u) = \sqrt{(1+u)(\beta^2+u)(\gamma^2+u)}$. Finally, $\lambda(x,y,z) > 0$ is defined implicitly as the largest positive root of $\Phi(x,y,z;\lambda(x,y,z)) = 0$.

In this case, the symplectic form of the system is canonical, and the equations of motion for the spacecraft are very easy to derive. From now on, we will denote this problem the

RF3BP, and the differential equations are

$$\begin{aligned}
\dot{x} &= p_x + \omega y, & \dot{p_x} &= \omega p_y - (1-\nu)(x+\nu R)R_\beta - \frac{\nu(x-(1-\nu)R)}{r_2^3},\\
\dot{y} &= p_y - \omega x, & \dot{p_y} &= -\omega p_x - (1-\nu) y R_\alpha - \frac{\nu y}{r_2^3},\\
\dot{z} &= p_z, & \dot{p_z} &= -(1-\nu) z R_\gamma - \frac{\nu z}{r_2^3},
\end{aligned}$$

where

$$\begin{aligned}
R_\alpha &= \frac{3}{2}\int_{\lambda(x,y,z)}^{+\infty} \frac{du}{(1+u)\Delta(u)},\\
R_\beta &= \frac{3}{2}\int_{\lambda(x,y,z)}^{+\infty} \frac{du}{(\beta^2+u)\Delta(u)},\\
R_\gamma &= \frac{3}{2}\int_{\lambda(x,y,z)}^{+\infty} \frac{du}{(\gamma^2+u)\Delta(u)}.
\end{aligned}$$

To compute these elliptic integrals, we use Carlson's elliptic integral of the second kind [4]:

$$\begin{aligned}
R_\alpha(x,y,z) &= R_D(\beta^2+\lambda(x,y,z), \gamma^2+\lambda(x,y,z), 1+\lambda(x,y,z)),\\
R_\beta(x,y,z) &= R_D(1+\lambda(x,y,z), \gamma^2+\lambda(x,y,z), \beta^2+\lambda(x,y,z)),\\
R_\gamma(x,y,z) &= R_D(1+\lambda(x,y,z), \beta^2+\lambda(x,y,z), \gamma^2+\lambda(x,y,z)),
\end{aligned}$$

where $R_D(x,y,z) = \frac{3}{2}\int_0^{+\infty}(t+x)^{-1/2}(t+y)^{-1/2}(t+z)^{-3/2}\,dt$.

3.1. Substitutes of the Lagrangian points. As was mentioned before, it is well known [45] that the RTBP has five equilibrium points. Three of them lie on the x axis and are called collinear or Eulerian points, and two of them, the triangular or Lagrangian points (also known as L_4 and L_5), form an equilateral triangle with the primaries. Here, we are interested in investigating how the rigid-body (ellipsoid) affects the position and stability of these triangular points. See Figure 6.

The fixed points of the RF3BP satisfy the following set of equations:

$$\begin{aligned}
p_x &= -\omega y, & \omega^2 x &= (1-\nu)(x+\nu R)R_\beta + \frac{\nu(x-(1-\nu)R)}{r_2^3},\\
p_y &= \omega x, & \omega^2 y &= (1-\nu) y R_\alpha + \frac{\nu y}{r_2^3},\\
p_z &= 0, & 0 &= (1-\nu) z R_\gamma + \frac{\nu z}{r_2^3}.
\end{aligned}$$

There are three solutions satisfying $y = z = 0$ that correspond to the "perturbed" collinear points. These equilibria are unstable [2] and therefore not suitable for constructing orbits in which to park a spacecraft. A priori, one could think that the collinear points of the RF3BP can have applications similar to the RTBP [29], such as the *Genesis Discovery Mission*, and

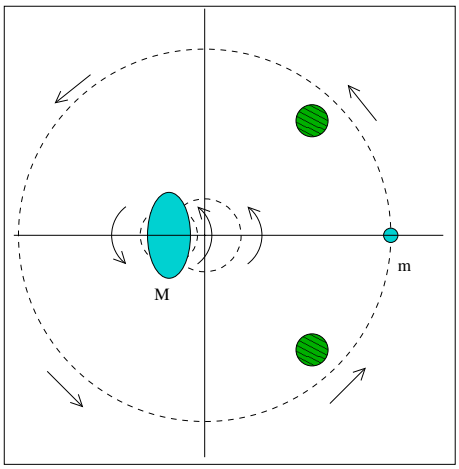


Figure 6. *Schematic picture of the stability zones near the Lagrangian points of the RF3BP.*

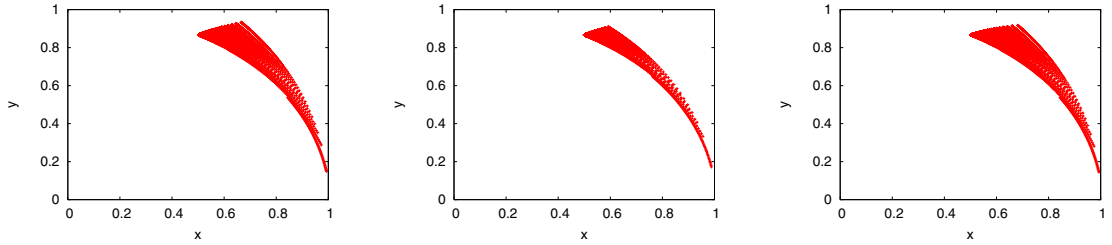
that it can be cheap to park a spacecraft near them by using the so-called center manifold [18, 15]. But this is not true in the current study. The difference is that in the binary asteroid problem the time scale is much shorter. This makes things such that the correction maneuvers, necessary to compensate the normal instability of the orbits, have to be performed too often to be feasible in practice.

We look for solutions such that $z = 0$ and $y \neq 0$. If there were no rigid-body effects, these fixed points would correspond to the RTBP L_4 and L_5. They are thus the substitutes of the Lagrangian points and satisfy

$$\omega^2 x = (1-\nu)(x+\nu R)R_\beta + \frac{\nu(x-(1-\nu)R)}{r_2^3},$$
$$\omega^2 = (1-\nu)R_\alpha + \frac{\nu}{r_2^3}.$$

We have numerically solved these equations for certain values of the parameters β, γ, ν, and R by means of Newton's method. In Figure 7, we plot the (x, y)-projection of these fixed points for $y > 0$ (there is a symmetric solution at $-y$) after shifting and rescaling in such a way that the ellipsoid is centered at $(0, 0)$ and the sphere at $(1, 0)$. After fixing β and γ, we vary the mass parameter ν from 10^{-3} to 0.5 and the distance between primaries from $R = 1$ to $R = 10$. Every red dot in Figure 7 corresponds to the substitute of the L_4 fixed point for a particular set of parameter values. More concretely, the red point at the bottom corner of every figure corresponds to $\nu = 10^{-3}$ and $R = 1$; the point at the top-left corner to $\nu = 0.5$ and $R = 1$; and, the point at the top-right corner to $\nu = 0.5$ and $R = 10$. See caption for more details.

Recall that, in these units, the Lagrangian point (L_4) is at the position $(x, y) = (0.5, \sqrt{3}/2)$. From the pictures, we can see that for small ν (big rigid-body) and short R (the two bodies are close) the position of the fixed point deviates a lot from the L_4 position. For larger ν values and moderate distances between the primaries, the fixed point is closer to the RTBP triangular region.



(a) (b) (c)

Figure 7. *Actual positions* $(x, y, 0)$ *of the substitutes of the Lagrangian point* L_4 *in the* xy*-plane when the mass parameter is varied between* 10^{-3} *and* 0.5 *(right to left) and the distance* R *increases from* 1 *to* 10 *(bottom to top). The positions have been shifted and rescaled in such a way that the ellipsoid is centered at* $(0, 0, 0)$ *and the sphere at* $(1, 0, 0)$*.* (a) $\beta = \gamma = 0.5$*,* (b) $\beta = \gamma = 0.8$*,* (c) $\beta = 0.5$ *and* $\gamma = 0.25$*.*

3.2. Spectral stability of the equilibrium points. We focus now on the study of the stability of the substitutes of the Lagrangian points in the RF3BP. In this section, we consider their spectral stability, i.e., stability w.r.t. eigenvalues of the linearized vector field. In section 4, we will study the nonlinear stability by means of a numerical method.

We begin with the linearization of the system (3.1) at the equilibrium point that substitutes L_4. We then compute numerically its eigenvalues. Similarly at the RTBP, this study shows that for "small" mass parameter values ($\nu \ll 1$), the fixed points are spectrally stable and, for larger ν, they are complex unstable, although in certain cases the critical value is larger than the Routh mass [41].

In Figure 8, we show some examples of the spectral stability behavior of the perturbed triangular fixed points of the RF3BP for different types of ellipsoids. We also superimpose on these pictures the stability zone of the SAE of the corresponding binary system.

In the next section, to perform a numerical global study of the stability region around the perturbed Lagrangian points in the RF3BP, we will be interested in the cases for which the underlying binary system is in a stable SAE and for which the corresponding triangular fixed point is spectrally stable. Thus, we will look at the intersection of the green and red zones in Figure 8.

4. Global dynamics around the Lagrangian points. In this section, the refined Fourier analysis method (see [16]) is used to study the dynamics around the tadpole region of the Lagrangian points. First, we obtain a global dynamical picture of the neighborhood of the fixed points. Later we identify trajectories that are very close to quasi-periodic motion and place the spacecraft in them, to simulate its dynamics while it is in a position in which to make observations of the binary system.

As the computations involved in this section are very intensive, we do not intend to perform a systematic study of the dynamics in terms of the parameters (which we believe was approached in the last sections), but we will choose a particular set of parameter values close to an actual asteroid pair. We thus start by fixing the four parameters of the RF3BP in the following manner: $\beta = 0.7576$, $\gamma = 0.6314$, $\nu = 1.16 \times 10^{-3}$, and $R = 5.873$. These particular values approximate the ones of the binary *Kalliope-Linus* system, placed in the main-belt (see [17]). Even though observations suggest that this asteroid pair is not in SAE,

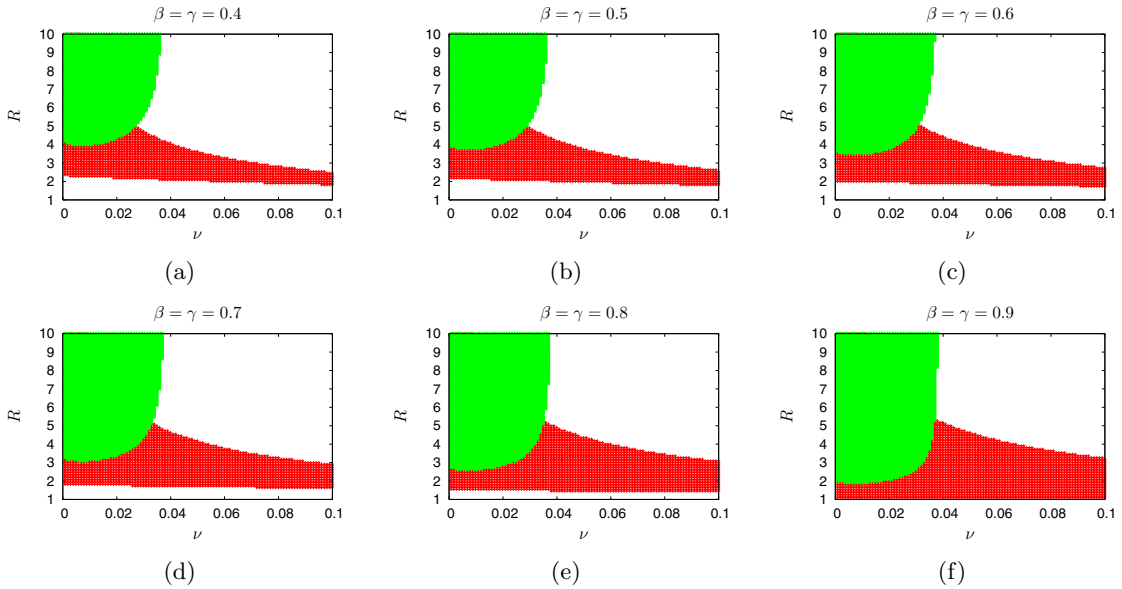


Figure 8. *Green: spectrally stable zones for the perturbed triangular points in the RF3BP. Red: spectrally stable zones for the SAE of the corresponding asteroid pair. The eigenvalues are computed for fixed β and γ, for a mass parameter $\nu \in [0, 0.1]$, and for a distance between the primaries $R \in [1, 10]$. The fixed values of the ellipsoid's axes are* (a) $\beta = \gamma = 0.4$, (b) $\beta = \gamma = 0.5$, (c) $\beta = \gamma = 0.6$, (e) $\beta = \gamma = 0.7$, (f) $\beta = \gamma = 0.8$, (g) $\beta = \gamma = 0.9$.

as an illustration of how the method works we will construct the model for the spacecraft assuming that the underlying binary with these parameter values is actually in SAE. Here, we use the construction of the RF3BP as in section 3. In this case, the uniform rotation of the binary in SAE is $\omega = 7.01844077933 \times 10^{-2}$, which corresponds to a complete revolution every 3.6 terrestrial days.

For this particular set of parameter values, the triangular points of the RF3BP are elliptic fixed points. From the eigenvalues of the linearized system, it is possible to obtain the frequencies of the normal oscillators at the equilibrium point:

$$\begin{aligned} \omega_1 &= 1.72741550738 \times 10^{-2}, \\ \omega_2 &= 6.76474915889 \times 10^{-2}, \\ \omega_3 &= 7.05487253096 \times 10^{-2}. \end{aligned}$$

Under generic conditions, the Kolmogorov–Arnold–Moser (KAM) theorem [19, 1, 34] (see [23] or [43] for a survey) ensures the existence of many quasi-periodic trajectories in a small neighborhood of these fixed points. In practice, the domain of applicability of the KAM theorem is much smaller than the actual stability region, which can be extended far from the elliptic point [32, 14, 8, 9].

In this section, we explore numerically this stability region near the Lagrangian points of the RF3BP by means of the frequency analysis method.

4.1. Frequency analysis and global dynamics. Let

$$f(t) = \sum_{k \in \mathbb{Z}^m} a_k e^{i<k,\ \nu>t}, \quad a_k \in \mathbb{C},$$

be a quasi-periodic function for which we know a table of equidistant values in the time span $[0, T]$. The frequency analysis algorithm will provide the values of the frequencies ν_k and the amplitudes $\tilde{a}_k$ of a function $\tilde{f}(t) = \sum \tilde{a}_k e^{i\nu_k t}$ that approximates $f(t)$ in $[0, T]$. The actual method used for approximating the frequencies is the one given by [16]. The procedure consists, basically, of equating the discrete Fourier transforms of the sampled initial data and of the quasi-periodic approximation. For an introduction to the frequency analysis (FA) method, see [21].

With the help of the FA method, we construct pictures of the global phase space dynamics near the Lagrangian fixed points. As the phase space is six-dimensional (the RF3BP is a three degrees of freedom problem), we need to reduce dimensionality by fixing some coordinates to constants. We thus study particular slices of phase space that are relevant for the dynamics.

The practical implementation is as follows. First, we transform the Cartesian coordinates of (3.1), (x, y, z, p_x, p_y, p_z), to spherical coordinates $(\rho, \theta, \lambda, p_\rho, p_\theta, p_\lambda)$ centered at the rigid-body barycenter by means of the canonical change of variables given by the generating function

$$W(p_x, p_y, p_z, \rho, \theta, \lambda) = -\rho \left(p_x \cos\theta \cos\lambda + p_y \sin\theta \cos\lambda + p_z \sin\lambda \right).$$

This change is useful in the visualization of the global dynamics, since it is well known that the stability region near the RTBP Lagrangian points is of the "banana shape" [32, 13, 8, 9]. Thus, we believe that they are a good set of coordinates with which to investigate the neighborhood of the RF3BP triangular points as well.

The spherical coordinates of the upper ($y > 0$) triangular point for the *Kalliope-Linus* system are

$$\begin{aligned} \rho^t &= 1.0012900026, & p_r^t &= 0.0000000000, \\ \theta^t &= 31.207021475^\circ, & p_\theta^t &= 1.0025816692, \\ \lambda^t &= 0.0000000000, & p_\lambda^t &= 0.0000000000. \end{aligned}$$

The zone of interest is a sufficiently large neighborhood of this point in phase space, and one can fix some coordinates to study some particular slices. We zoom into a window in the (ρ, θ)-space enclosing this triangular point by fixing all the momenta equal to the momenta of the fixed point, $(p_\rho, p_\theta, p_\lambda) = (p_\rho^t, p_\theta^t, p_\lambda^t)$, and by choosing different slices for the inclination $\lambda = \Lambda$, where Λ is a constant. In the experiments, we choose $\Lambda = 0^\circ, 1^\circ, 2^\circ, \ldots, 8^\circ$.

Inside the (ρ, θ)-window we pick a regular mesh of 57,600 initial conditions and integrate them in the interval of time $[0, T]$, with $T = 50{,}000$ adimensional units (this corresponds to about 2,000 terrestrial days in the *Kalliope-Linus* system). Then, we take 32,768 sample points for every trajectory and use the refined Fourier analysis [16] of this sample to evaluate the three basic frequencies (recall that the RF3BP is a three degrees of freedom system) of the orbits that we call $\omega_1^{(1)}$, $\omega_2^{(1)}$, and $\omega_3^{(1)}$. Afterwards, we repeat the integration in the interval of time $[T, 2T]$ and recompute the frequencies. In this case, we call them $\omega_1^{(2)}$, $\omega_2^{(2)}$, and $\omega_3^{(2)}$.

Finally, we consider the values $\delta_j = |1-\omega_j^{(2)}/\omega_j^{(1)}|$, $j = 1, 2, 3$, as an estimation of the diffusion (see [37] and [36]) related to the orbit starting at the phase space point $(\rho, \theta, \lambda, p_\rho, p_\theta, p_\lambda)$. We call δ_j the *diffusion index*. The values of δ_j give an estimation of the chaoticity of the particular orbit. That is, if the trajectory associated with an initial condition is quasi-periodic, then δ_j should be zero.

In Figure 9, we show several contour plots of the function $\sigma_j = \log \delta_j$ for $j = 1$ (we obtain similar pictures for the σ_2 and σ_3 cases). Every picture is performed for a fixed initial inclination $\lambda = \Lambda$, from $\Lambda = 0°$ (top-left) to $\Lambda = 8°$ (bottom-right). We plot a color depending on the value of the diffusion index: blue zones ($\delta_1 < 10^{-10}$) correspond to initial conditions whose trajectories are close to quasi-periodic; yellow-to-red zones ($\delta_1 > 10^{-2}$) are related to strongly irregular and escaping motion.

As we already mentioned, the top-left picture in Figure 9 corresponds to zero initial inclination, i.e., $\Lambda = 0°$. In synodic coordinates, this is the xy plane of motion. As this plane is invariant under the dynamics of (3.1), all orbits starting at this particular slice will remain in this plane and thus will be coplanar to the motion of the binary. We note that the stability zone for this particular plane is quite different from the well-known "banana" region of the RTBP [32, 14, 8]. In particular, for a similar mass parameter $\nu \simeq 10^{-3}$, the stability zones near the RTBP L_4 and L_5 are extended to a much larger domain in the xy plane [9].

When one increases the inclination Λ from 0° to 8°, we see that the stability zone shrinks rapidly until it "breaks" at about $\Lambda \sim 6.5°$ and disappears at about $\Lambda \sim 10°$. This is also very different from the RTBP case, since it is well known (see [14, 36]) that the stable zone for the Sun-Jupiter RTBP ($\nu \simeq 10^{-3}$) grows until $\Lambda \sim 17°$ and does not disappear until $\Lambda > 45°$ (recall the existence of the Trojan asteroids, which can be observed at inclinations larger than 40°; see [7, 10]).

4.2. Gallery of quasi-periodic spacecraft trajectories. We now use the global dynamical pictures obtained in the last section to effectively compute (close-to) quasi-periodic orbits near the upper triangular point of the RF3BP. These orbits will lie on near-invariant tori of dimensions 2 and 3.

To construct these tori, we proceed as follows. First, we choose initial conditions from the global dynamical pictures that have a small associated index (we will make precise later how small). A trajectory corresponding to one of these initial conditions will remain very close to an invariant torus with frequencies $\omega_{1,2,3}$ (computed as in section 4.1). Thus, we can numerically integrate these initial conditions for a "long-enough" time and generate the (almost-) invariant torus (it would be invariant if the diffusion index was exactly zero).

In Figures 10 through 15, we show some examples of these (almost)-invariant tori. All of the images have been generated by looking for initial conditions in the global pictures of Figure 9 with a diffusion index δ_1 smaller than 10^{-12}. In particular, we have integrated them for a time span of $T = 450{,}000$ adimensional units, that is, about 50 terrestrial years. We could thus send a spacecraft to one of these tori, and the satellite would orbit in this (almost) quasi-periodic orbit, without any need of extra propulsion, for at least 50 years; this is a period more than enough for a mission of this type.

More concretely, Figure 10 shows the xy-projections of nine planar tori. These tori lie on the zero inclination ($I = 0°$) plane and, as this is an invariant plane for the RF3BP, spacecraft

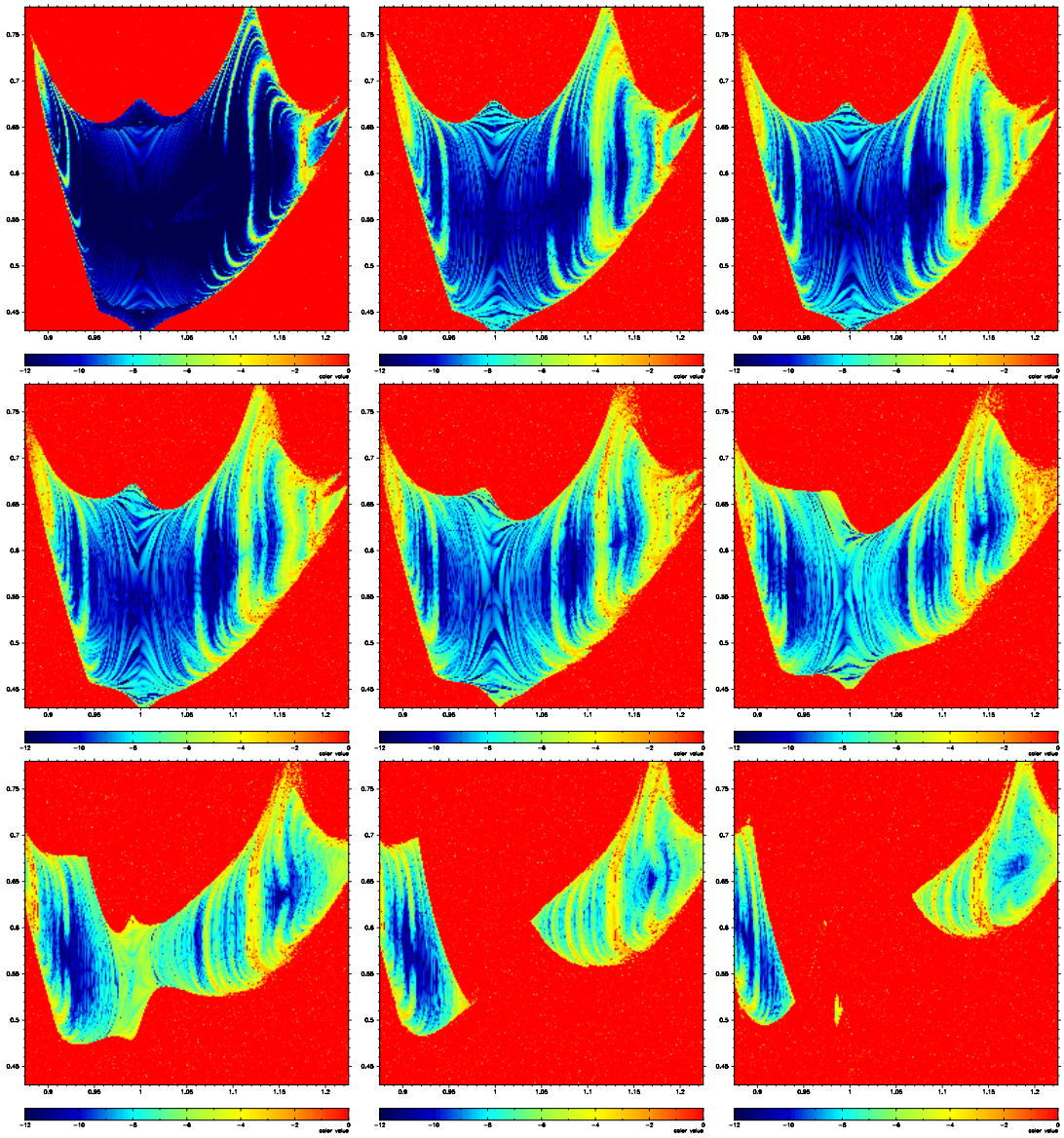


Figure 9. *Global dynamics around the upper triangular point of the* Kalliope–Linus *system. The axes correspond to the* (ρ, θ) *coordinates, where the angle is in radians. The pictures show different slices of phase space for different inclinations* Λ*. From left to right and top to bottom, the portraits correspond to* $\Lambda = 0^\circ$*,* $\Lambda = 1^\circ, \ldots, 8^\circ$*. Blue zones are related to motion close to quasi-periodic, and hence any trajectory starting on them is likely to remain long enough in the neighborhood of the fixed point. Yellow to red zones correspond to initial conditions that escape, go to collision, or have a chaotic behavior in the time-window considered. See text for more details.*

trajectories on these tori will be coplanar with the orbital motion of the binary. In Figures 11 and 12, we plot the xy-, xz-, and $z\dot{z}$-projections of six three-dimensional tori with initial inclination $I = 2^\circ$. In Figure 13, some examples of tori with initial inclination $I = 5^\circ$ are shown. Three examples with initial inclination $I = 6^\circ$, $I = 7^\circ$, and $I = 8^\circ$ are displayed in

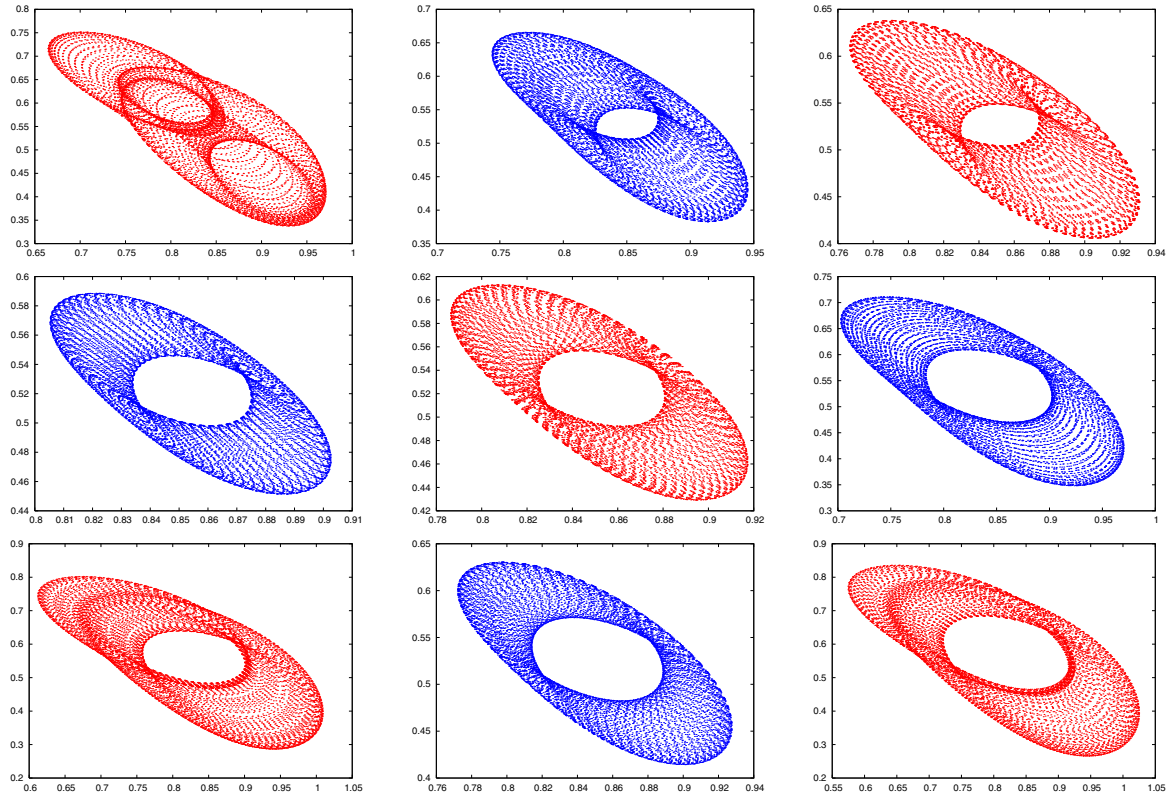


Figure 10. *xy-projections of tori in the plane* $I = 0°$.

Figure 14. Any particular orbit chosen on these tori with nonzero initial inclination can be useful for a spacecraft devoted to performing observations of the asteroid pair from different perspectives. Finally, in Figure 15, we plot the three-dimensional xyz-projections of some tori that appear in Figures 13 and 14. As we mentioned before, for larger inclinations the stability region vanishes, and therefore no quasi-periodic orbits are found near the Lagrangian points for $I \geq 10°$.

5. Conclusions. In this paper, we have developed and studied a model for a satellite orbiting an asteroid pair. First, geometric mechanics was used in the modelization process of the F2BP and in studying the stability of its relative equilibria. Then, an RF3BP was constructed to describe the motion of the satellite, and numerical methods were used to study some of its global stability properties. The main tool for the numerical investigation was the frequency map analysis, which provides a very nice global view of the dynamics of this model. Moreover, this global dynamics can be used in practice to preselect initial conditions for satellite trajectories. Due to the stability properties of these particular orbits, they are very suitable for parking the spacecraft on them in such a way that there is no need to spend extra energy in the station-keeping.

Another interesting result of this paper comes from the analysis of the stability region near the Lagrangian points of the RF3BP and its comparison with the RTBP. We have seen

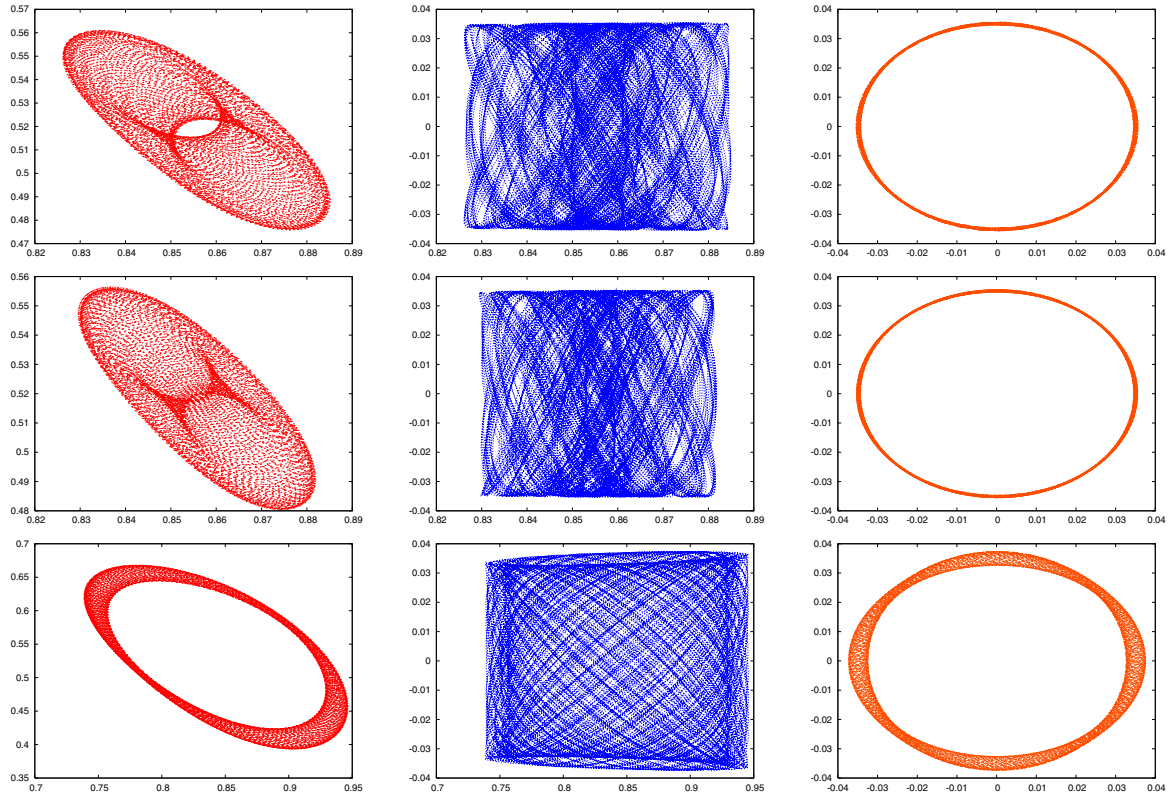


Figure 11. *Invariant tori with initial conditions in the plane $I = 2^\circ$. Left column: xy-projection. Center column: xz-projection. Right column: $z\dot{z}$-projection.*

that the rigid-body effect of one of the primaries is to make this stability region smaller in both the equatorial plane of motion and in inclination. As an example, for $\nu \sim 10^{-3}$ (which is very close to the Sun-Jupiter mass parameter), the stability region in the RF3BP vanishes for inclinations $I \geq 10^\circ$, while in the Sun-Jupiter RTBP (or even in the real solar system) we know there are objects in stable orbits up to inclinations of $I \sim 40^\circ$ [9].

Dissipative mechanisms, such as solar radiation pressure, solar wind drag [42, 22], or the Poynting–Robertson effect, might destroy the stability regions as dissipation-induced instabilities do (see, for instance, [6]). In our case, though, we ignore the effect of solar radiation pressure on the orbiting spacecraft, an approximation that is excellent for a mission to a main-belt binary system. For near-Earth asteroid binary systems, a future study will be done to map out how this additional force modifies the dynamics.

Much work still needs to be done at a theoretical level before sending an actual probe to a binary asteroid. This paper contributes to this problem by applying to it the frequency analysis method and showing a way of finding stable satellite orbits. Future work following this line of investigation could consider several different aspects of the problem: first, one could compute and study similar stability regions for more complex rigid-body gravitational systems (like that in [47], for instance) or for two-body problems that are not in relative equilibrium (e.g., the rigid-body is rotating faster than the relative orbital velocity of the other primary

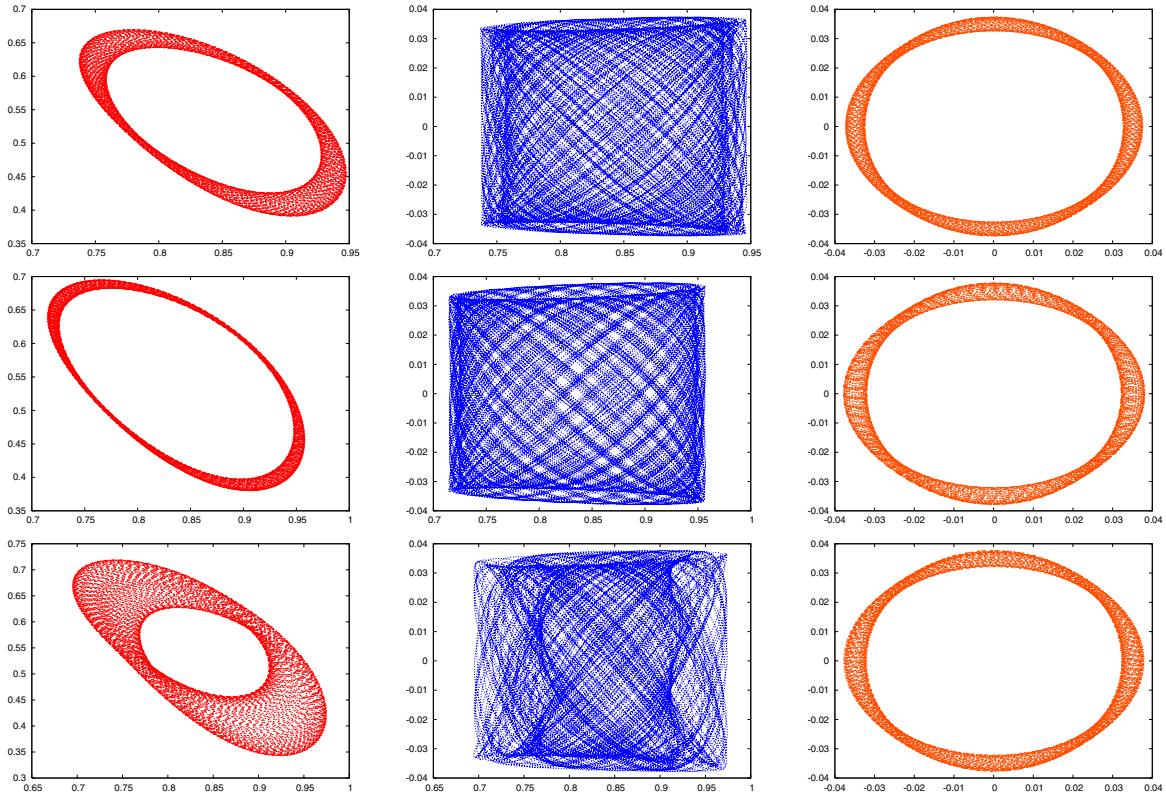


Figure 12. *Invariant tori with initial conditions in the plane $I = 2^\circ$. Left column: xy-projection. Center column: xz-projection. Right column: zż-projection.*

[41]); second, it seems interesting to apply the same method for finding orbits that are not necessarily close to the equilibrium points (orbits that go around one of the primaries, encircle both primaries, retrograde orbits, etc.).

Appendix. Abelian reduction. General setting. In this appendix, we review the reduction process for a system that is invariant under the abelian Lie group $SO(2)$. We perform the reduction on both sides of the problem, Lagrangian and Hamiltonian, and show that they are equivalent via the reduced Legendre transform.

Let us start by assuming that the configuration space Q can be written as a product of the circle S^1 and a manifold B called *shape space*, i.e., $Q = S^1 \times B$ and $q = (q^0, q^\alpha) = (\theta, r^\alpha)$, with $q \in Q$, $\theta \in S^1$ and $r^\alpha \in B \subseteq \mathbb{R}^n$.

Let us assume that the symmetry group $G = SO(2) = S^1$ acts trivially in the following way:

$$\begin{aligned} \Phi : G \times Q &\longrightarrow Q, \\ (\varphi, q) &\longmapsto \Phi_\varphi(q) = \Phi_\varphi(\theta, r^\alpha) = (\theta + \varphi, r^\alpha), \end{aligned}$$

where G is a Lie group with Lie algebra $\mathfrak{g} = \mathbb{R}$ and dual Lie algebra $\mathfrak{g}^* \cong \mathbb{R}$.

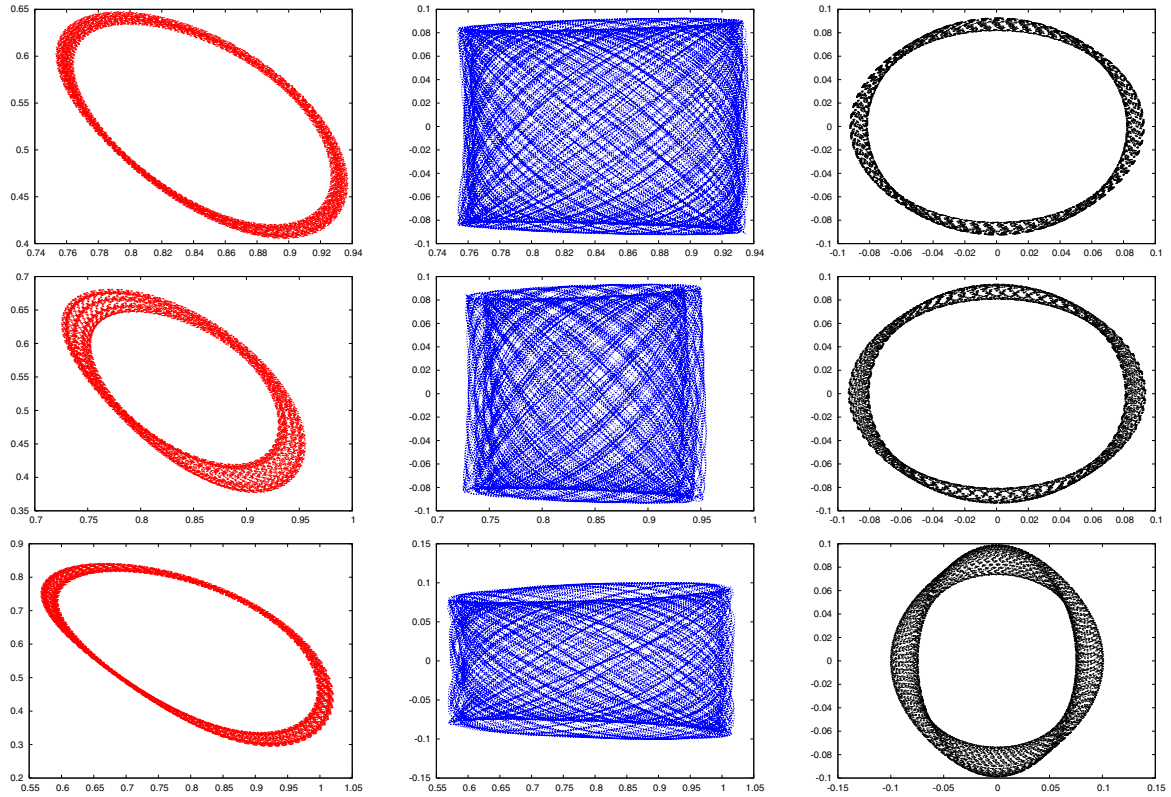


Figure 13. *Invariant tori with initial conditions in the plane $I = 5°$. Left column: xy-projection. Center column: xz-projection. Right column: $z\dot{z}$-projection.*

We also assume that the Lagrangian is of the type *kinetic* minus *potential* energy. Then, it can be written, in a local trivialization of the tangent bundle TQ, as follows:

$$L(q,\dot{q}) = K(q,\dot{q}) - V(q) = \frac{1}{2} g_{ij}\dot{q}^i\dot{q}^j - V(q),$$

where g_{ij} is a Riemannian metric and summations over $i, j = 0, 1, \ldots, n$ are understood. The corresponding Hamiltonian on the cotangent bundle T^*Q is given by

$$H(q,p) = K(\mathbb{F}L(q,\dot{q})) + V(q) = \frac{1}{2} g^{ij} p_i p_j + V(q),$$

where g^{ij} is the inverse of the metric g_{ij}, $(q,p) = \mathbb{F}L(q,\dot{q})$ is the Legendre transform of $(q,\dot{q})$ $((q^i,p_i) = (q^i, g_{ij}\dot{q}^j))$, and the symplectic form is canonical, i.e., $\Omega = dq^i \wedge dp_i$.

A.1. Lagrangian reduction. We start with the Lagrangian rewritten in the following form:

$$L(r^\alpha, \dot{\theta}, \dot{r}^\alpha) = \frac{1}{2} g_{00}\dot{\theta}^2 + g_{0\alpha}\dot{\theta}\dot{r}^\alpha + \frac{1}{2} g_{\alpha\beta}\dot{r}^\alpha\dot{r}^\beta - V(r^\alpha),$$

where $q^0 = \theta \in \mathbb{T}$; $q^j = r^j$, $j = 1 \div n$; and $\alpha, \beta = 1 \div n$. Note that $L \neq L(\theta)$. Thus, θ is a cyclic variable, and the corresponding conjugate momentum $p_\theta = g_{00}\dot{\theta} = \frac{\partial L}{\partial \dot{\theta}}$ is conserved.

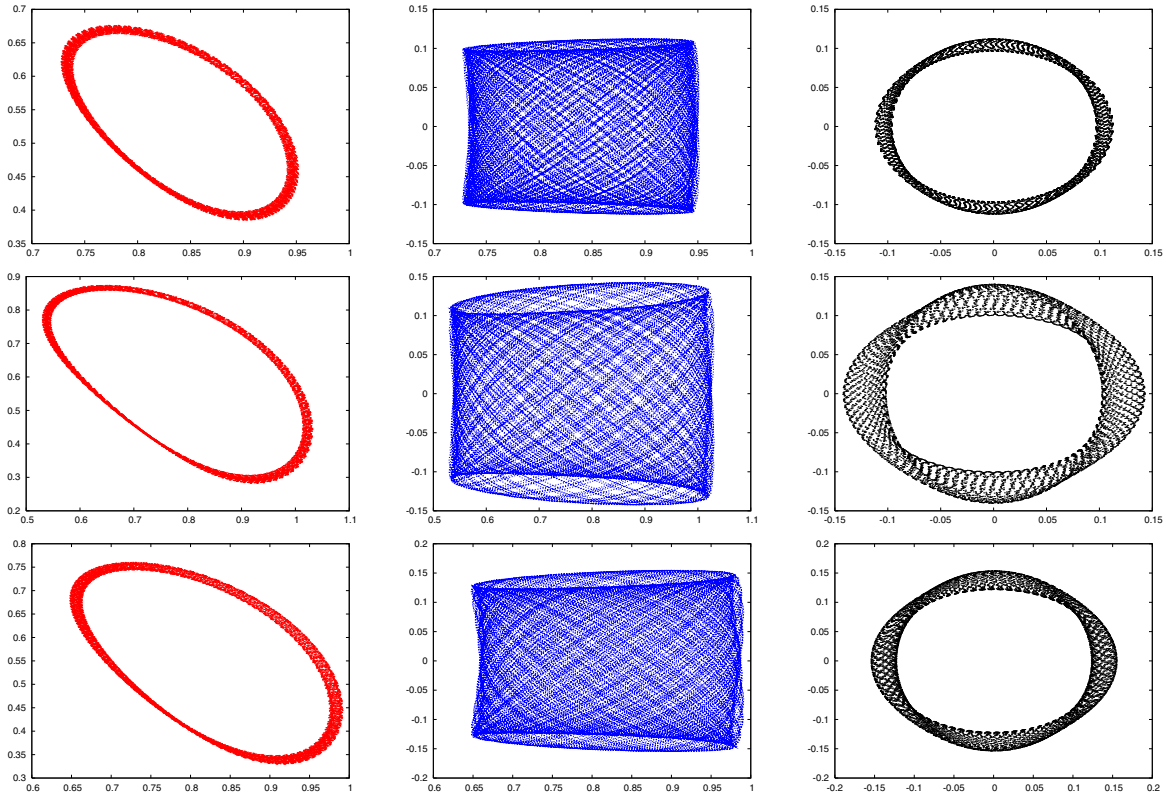


Figure 14. *Invariant tori with initial conditions in the planes $I = 6^\circ$ (first row), $I = 7^\circ$ (second row), and $I = 8^\circ$ (last row). Left column: xy-projection. Center column: xz-projection. Right column: $z\dot{z}$-projection.*

While the classical theory of Routh reduction is valid, we will use modern Routh reduction [30], which applies in a much more general framework.

The ingredients needed in the reduction process (see [30] for details) are the following:

- *Infinitesimal generator:* Given $\xi \in \mathfrak{g}^*$, the infinitesimal generator corresponding to the group action can be computed as follows:

$$\xi_Q(\theta, r^\alpha) = \frac{d}{dt}\left(\exp(t\xi)\cdot(\theta, r^\alpha)\right)|_{t=0} = ((\theta, r^\alpha), (\xi, 0)).$$

- *Lagrangian momentum map:* The associated momentum map $\mathbf{J}_L : TQ \to \mathfrak{g}^*$ is given by

$$\mathbf{J}_L((q,\dot{q}))\cdot\xi_Q = \langle \mathbb{F}L(q,\dot{q}), \xi_Q(q)\rangle = \left\langle (g_{0j}\dot{q}^j, g_{\alpha j}\dot{q}^j), (\xi, 0)\right\rangle = g_{0j}\dot{q}^j\xi.$$

  Thus, $\mathbf{J}_L(q,\dot{q}) = g_{00}\dot{\theta} + g_{0\alpha}\dot{r}^\alpha$.

- *Locked inertia tensor:* The locked inertia tensor is the instantaneous tensor of inertia when the relative motion of the two bodies is locked. If we denote by $\langle\langle\cdot,\cdot\rangle\rangle$ the scalar product induced by the metric g_{ij}, the locked inertia tensor $\mathbb{I}(\theta, r^\alpha) : \mathfrak{g} \longrightarrow \mathfrak{g}^*$ is given (locally) by

$$\langle \mathbb{I}(\theta, r^\alpha)\eta, \xi\rangle = \langle\langle((\theta, r^\alpha), (\eta, 0)), ((\theta, r^\alpha), (\xi, 0))\rangle\rangle = g_{00}\eta\xi.$$

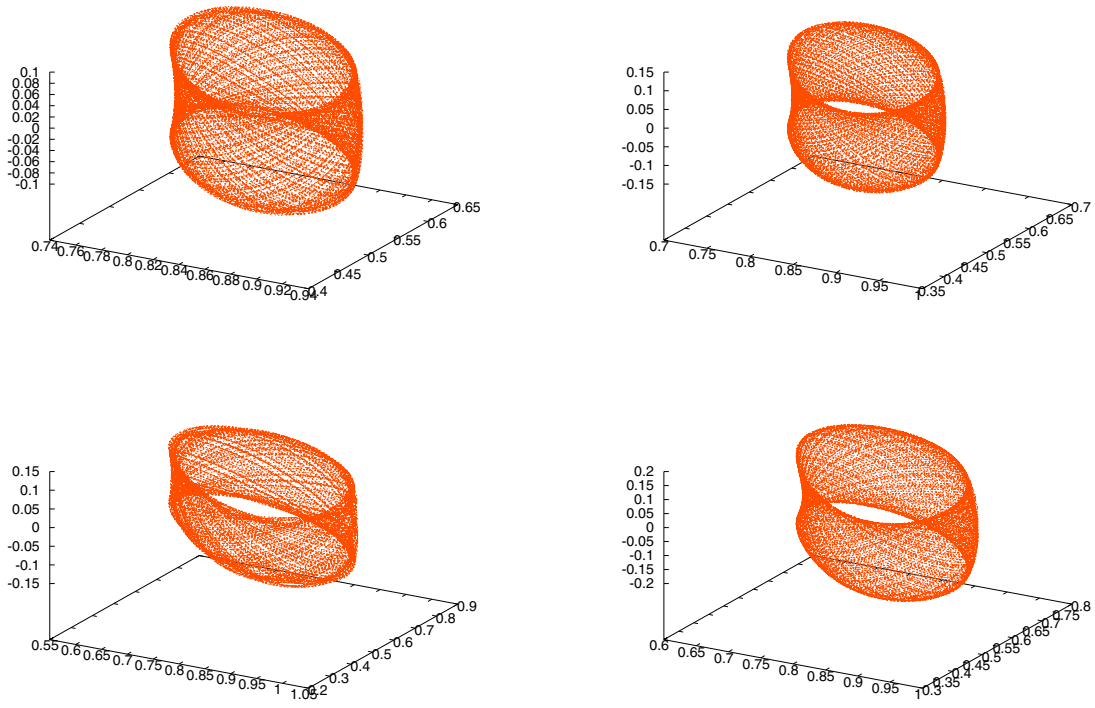


Figure 15. *Left column: xyz-projections of two different invariant tori with initial conditions in the plane $I = 5^\circ$. Right column: Invariant tori with initial conditions in the planes $I = 6^\circ$ (top) and $I = 8^\circ$ (bottom).*

Then, $\mathbb{I}(\theta, r^\alpha) = g_{00}(r^\alpha)$.

- *Mechanical connection:* The connection $\mathcal{A} : TQ \longrightarrow \mathfrak{g}$ can be written (locally) as $\mathcal{A}(\theta, r^\alpha)(\theta, r^\alpha, \dot{\theta}, \dot{r}^\alpha) = \mathbb{I}^{-1}\mathbf{J}(\mathbb{F}L(\theta, r^\alpha, \dot{\theta}, \dot{r}^\alpha)) = g_{00}^{-1} g_{0j}\dot{q}^j$. Thus,

$$\mathcal{A}(\theta, r^\alpha)(\theta, r^\alpha, \dot{\theta}, \dot{r}^\alpha) = \dot{\theta} + g_{00}^{-1} g_{0\alpha}\dot{r}^\alpha.$$

  From $\mathcal{A}$, we can obtain the related one-form: $\mathcal{A}(\theta, r^\alpha) = d\theta + A_\alpha dr^\alpha$, where $A_\alpha = g_{00}^{-1} g_{0\alpha}$, and the curvature $\mathcal{B} = d\mathcal{A} = B_{\alpha\beta} dr^\alpha \wedge dr^\beta$ has components given (locally) by $B_{\alpha\beta} = (\frac{\partial A_\alpha}{\partial r^\beta} - \frac{\partial A_\beta}{\partial r^\alpha})$. For a given $\mu \in \mathfrak{g}^* \cong \mathbb{R}$, the mechanical connection on the fiber $Q \to Q/G$ is

$$\alpha_\mu(\theta, r^\alpha) = \mu d\theta + \mu A_\alpha dr^\alpha.$$

- *Amended potential:* For $\mu \in \mathfrak{g}^*$, the amended potential is defined as

$$V_\mu(q) = V(q) + \frac{1}{2}\left\langle \mu, \mathbb{I}^{-1}(q)\mu \right\rangle = V(q) + \frac{1}{2} g_{00}^{-1}\mu^2.$$

- *Routhian:* The Routhian is a function on TQ defined as

$$R^\mu = \frac{1}{2}\|\mathrm{Hor}(q, \dot{q})\|^2 - V_\mu,$$

where $\mathrm{Hor}(q,\dot{q}) = (-g_{00}^{-1}g_{0\alpha}\dot{r}^{\alpha}, \dot{r}^{\alpha})$ is the horizontal component of $(q,\dot{q})$ and the norm is given by the g_{ij} metric. Then, locally, we can write

$$R^{\mu} = \frac{1}{2}(g_{\alpha\beta} - g_{00}^{-1}g_{0\alpha}g_{0\beta})\dot{r}^{\alpha}\dot{r}^{\beta} - \frac{1}{2}g_{00}^{-1}\mu^2 - V(r^{\alpha}). \tag{A.1}$$

The general reduction theory [28, 30] tells us that if a curve $q(t)$ in Q satisfying $\mathbf{J}_L(q,\dot{q}) = \mu$ is a solution of the Euler–Lagrange equations for the Lagrangian $L(q,\dot{q})$, then the induced curve on Q/G_{μ} satisfies the reduced Lagrangian variational principle; that is, the variational principle of Lagrange and d'Alembert on Q/G_{μ} with magnetic term $\mathcal{B}$ and the Routhian dropped to $T(Q/G_{\mu})$.

Let be $\hat{R}^{\mu}$ the reduced Routhian, that is, the Routhian (A.1) dropped to the reduced space $\mathbf{J}_L^{-1}(\mu)/S^1$. Then (locally),

$$\hat{R}^{\mu} = \frac{1}{2}h_{\alpha\beta}\dot{r}^{\alpha}\dot{r}^{\beta} - V_{\mu}(r^{\alpha}), \tag{A.2}$$

where $h_{\alpha\beta} = g_{\alpha\beta} - g_{00}^{-1}g_{0\alpha}g_{0\beta}$ is a metric in the reduced space and $V_{\mu}(r^{\alpha})$ is the amended potential.

- *Equations of Lagrange–Routh:* The equations of motion in the reduced space $\mathbf{J}_L^{-1}(\mu)/S^1$ are given by

$$\frac{d}{dt}\frac{\partial \hat{R}^{\mu}}{\partial \dot{r}^{\alpha}} - \frac{\partial \hat{R}^{\mu}}{\partial r^{\alpha}} = -\mu B_{\alpha\beta}\dot{r}^{\beta},$$

where $B_{\alpha\beta} = \frac{\partial A_{\alpha}}{\partial r^{\beta}} - \frac{\partial A_{\beta}}{\partial r^{\alpha}}$ and $A_{\alpha} = g_{00}^{-1}g_{0\alpha}$. More concretely,

$$h_{\alpha\beta}\ddot{r}^{\beta} + \left(\frac{\partial h_{\alpha\beta}}{\partial r^{\gamma}} - \frac{1}{2}\frac{\partial h_{\beta\gamma}}{\partial r^{\alpha}}\right)\dot{r}^{\beta}\dot{r}^{\gamma} = -\frac{\partial V_{\mu}(r^{\alpha})}{\partial r^{\alpha}} - \mu B_{\alpha\beta}\dot{r}^{\beta}. \tag{A.3}$$

A.2. Cotangent bundle reduction. Now, we perform the corresponding reduction in the Hamiltonian side [27]. Let us consider that the initial Hamiltonian can be written (locally) as

$$H(r^{\alpha}, p_{\theta}, p_r^{\alpha}) = \frac{1}{2}g^{00}p_{\theta}^2 + g^{0\alpha}p_{\theta}p_{\alpha} + \frac{1}{2}g^{\alpha\beta}p_{\alpha}p_{\beta} + V(r^{\alpha}),$$

where the metric elements g^{ij} correspond to the inverse of the metric g_{ij}. That is, $g_{ij}g^{jk} = \delta_i^k$, where $i,j,k = 0,\ldots,n$, and δ_i^k denotes the Kronecker delta function.

Thus, we assume that the initial Hamiltonian is invariant under the action of the abelian symmetry group $G = SO(2) = S^1$.

Let us perform the computations of all the extra ingredients needed in the reduction in the Hamiltonian side [27], as follows:

- *Momentum map:* The momentum map corresponds to the angular momentum of the system: $\mathbf{J} : T^*Q \longrightarrow \mathfrak{g}^*$

$$\langle \mathbf{J}(\theta, r^{\alpha}, p_{\theta}, p_{r^{\alpha}}), \xi\rangle = \langle (p_{\theta}, p_{\alpha}), (\xi, 0)\rangle = p_{\theta}\xi.$$

Thus, $\mathbf{J}(\theta, r^{\alpha}, p_{\theta}, p_{r^{\alpha}}) = p_{\theta}$.

- *Momentum shift:* In this case, it is convenient to perform a shift of the momenta from $\mathbf{J}^{-1}(\mu)$ to $\mathbf{J}^{-1}(0)$, and also in the corresponding reduced spaces, in the following way:

$$\begin{array}{ccc} \mathbf{J}^{-1}(\mu) = \{(\theta, r^\alpha, \mu, p_\alpha)\} & \xrightarrow{t^\mu} & \mathbf{J}^{-1}(0) = \{(\theta, r^\alpha, 0, \tilde{p}_\alpha)\}, \\ \downarrow & & \downarrow \\ \mathbf{J}^{-1}(\mu)/G_\mu = \mathbf{J}^{-1}(\mu)/S^1 & \xrightarrow{t^\mu_G} & \mathbf{J}^{-1}(0)/S^1 = \mathbf{J}^{-1}(0)/G, \end{array}$$

  where

$$\begin{aligned} t^\mu(\theta, r^\alpha, \mu, p_\alpha) &= (\theta, r^\alpha, \mu, p_\alpha) - (\theta, r^\alpha, \mu, \mu A_\alpha) \\ &= (\theta, r^\alpha, 0, p_\alpha - \mu A_\alpha) = (\theta, r^\alpha, \tilde{p}_\theta, \tilde{p}_\alpha). \end{aligned}$$

  Thus, the shifting map is given by $\tilde{p}_\alpha = p_\alpha - \mu A_\alpha$ and $\tilde{p}_\theta = 0$.
- *Reduced Hamiltonian:* In $\mathbf{J}^{-1}(0)/G$ we have $H_{\alpha_\mu} = \frac{1}{2}\|\tilde{p}\|^2 + V_\mu$, where $\|\cdot\|$ is the norm related to the g^{ij} metric and $V_\mu = V + \frac{1}{2}\mathbb{I}^{-1}\mu^2$ is the amended potential. Thus, recalling that $\tilde{p}_\theta = 0$, the Hamiltonian in the reduced space $\mathbf{J}^{-1}(0)/G$ is

$$H_\mu(r^\alpha, p_\alpha) = \frac{1}{2} g^{\alpha\beta}\tilde{p}_\alpha\tilde{p}_\beta + V(r^\alpha) + \frac{1}{2}\mu^2 g_{00}^{-1} \quad \text{for } \alpha, \beta = 1 \div n.$$

- *Reduced symplectic form:* In general, in the reduced space, the symplectic form is not canonical. The projection is given by the map

$$((T^*Q)_\mu, \Omega_\mu) \xrightarrow{P_\mu} ((T^*(Q/G), \omega - B_\mu),$$

  where the "reduced" symplectic form is

$$\omega_\mu = \omega - B_\mu = dr^\alpha \wedge d\tilde{p}_\alpha - \mu \frac{\partial A_\alpha}{\partial r^\beta} dr^\beta \wedge dr^\alpha.$$

- *Hamiltonian equations:* The Hamiltonian equations are given by [28]

$$\mathbf{i}_{(\dot{r}^\alpha \partial_{r^\alpha} + \dot{\tilde{p}}_\alpha \partial_{\tilde{p}_\alpha})} \omega_\mu = dH_\mu,$$

  where $\mathbf{i}_X \Omega$ denotes the interior product (or contraction) of the vector field X and the 1-form Ω. Computing both sides of this identity,

$$\begin{aligned} \mathbf{i}_{(\dot{r}^\alpha \partial_{r^\alpha} + \dot{\tilde{p}}_\alpha \partial_{\tilde{p}_\alpha})} \omega_\mu &= \dot{r}^\alpha d\tilde{p}_\alpha - \mu \frac{\partial A_\alpha}{\partial r^\beta} \dot{r}^\beta dr^\alpha + \mu \frac{\partial A_\alpha}{\partial r^\beta} \dot{r}^\alpha dr^\beta - \dot{\tilde{p}}_\alpha dr^\alpha, \\ dH_\mu &= \frac{\partial H_\mu}{\partial r^\alpha} dr^\alpha + \frac{\partial H_\mu}{\partial \tilde{p}_\alpha} d\tilde{p}_\alpha, \end{aligned}$$

  we obtain the equations of motion in the reduced $J^{-1}(0)/G$ space:

$$\dot{r}^\alpha = \frac{\partial H_\mu}{\partial \tilde{p}_\alpha}, \qquad \dot{\tilde{p}}_\alpha = -\frac{\partial H_\mu}{\partial r^\alpha} - \mu \left(\frac{\partial A_\alpha}{\partial r^\beta} - \frac{\partial A_\alpha}{\partial r^\beta} \right) \dot{r}^\beta.$$

  Finally, we can write them more explicitly as

$$\dot{r}^\alpha = g^{\alpha\beta}\tilde{p}_\beta, \qquad \dot{\tilde{p}}_\alpha = -\frac{1}{2}\frac{\partial g^{\beta\gamma}}{\partial r^\alpha}\tilde{p}_\beta\tilde{p}_\gamma - \frac{\partial V(r^\alpha)}{\partial r^\alpha} - \frac{1}{2}\mu^2 \frac{\partial g_{00}^{-1}}{\partial r^\alpha} - \mu B_{\alpha\beta} g^{\beta\gamma}\tilde{p}_\gamma. \tag{A.4}$$

A.3. Reduced Legendre transform. The correspondence between the reduced equations of motion on the Hamiltonian and Lagrangian sides is given by the reduced Legendre transform. We start with the reduced Routhian (A.2),

$$\hat{R}^{\mu} = \frac{1}{2}(g_{\alpha\beta} - g_{00}^{-1} g_{0\alpha} g_{0\beta})\dot{r}^{\alpha}\dot{r}^{\beta} - \frac{1}{2} g_{00}^{-1}\mu^2 - V(r^{\alpha}),$$

and the shifted momenta,

$$\tilde{p}_{\alpha} = \frac{\partial \hat{R}^{\mu}}{\partial \dot{r}^{\alpha}} = (g_{\alpha\beta} - g_{00}^{-1} g_{0\alpha} g_{0\beta})\dot{r}^{\beta}.$$

Using the identities

$$\begin{aligned} g_{0\alpha} g^{0\beta} + g_{\alpha\gamma} g^{\gamma\beta} &= \delta_{\alpha}^{\ \beta}, \\ g^{\alpha\beta} g_{\beta 0} + g^{\alpha 0} g_{00} &= 0, \end{aligned} \tag{A.5}$$

we obtain the first equation in (A.4): $g^{\alpha\beta}\tilde{p}_{\beta} = \dot{r}^{\alpha}$.

Now, in order to recover the reduced Lagrange–Routh equations (A.3), we compute the time-derivative of the shifted momenta

$$\dot{\tilde{p}}_{\alpha} = (g_{\alpha\beta} - g_{00}^{-1} g_{0\alpha} g_{0\beta})\ddot{r}^{\beta} + \frac{\partial}{\partial r^{\gamma}}\left(g_{\alpha\beta} - g_{00}^{-1} g_{0\alpha} g_{0\beta}\right)\dot{r}^{\gamma}\dot{r}^{\beta},$$

and the derivative with respect to r^{α} of the identities (A.5),

$$\begin{aligned} &\frac{\partial g_{0\epsilon}}{\partial r^{\alpha}} g^{0\beta} + g_{0\epsilon}\frac{\partial g^{0\beta}}{\partial r^{\alpha}} + \frac{\partial g_{\epsilon\gamma}}{\partial r^{\alpha}} g^{\gamma\beta} + g_{\epsilon\gamma}\frac{\partial g^{\gamma\beta}}{\partial r^{\alpha}} = 0, \\ &\frac{\partial g^{\beta\gamma}}{\partial r^{\alpha}} g_{\gamma 0} + g^{\beta\gamma}\frac{\partial g_{\gamma 0}}{\partial r^{\alpha}} + \frac{\partial g^{\beta 0}}{\partial r^{\alpha}} g_{00} + g^{\beta 0}\frac{\partial g_{00}}{\partial r^{\alpha}} = 0. \end{aligned}$$

If we substitute the last three identities together with (A.5) into the second equation of (A.4), we obtain

$$\begin{aligned} &(g_{\alpha\beta} - g_{00}^{-1} g_{0\alpha} g_{0\beta})\ddot{r}^{\beta} + \frac{\partial}{\partial r^{\gamma}}\left(g_{\alpha\beta} - g_{00}^{-1} g_{0\alpha} g_{0\beta}\right)\dot{r}^{\gamma}\dot{r}^{\beta} \\ &= \frac{1}{2}\frac{\partial}{\partial r^{\alpha}}\left(g_{\beta\gamma} - g_{00}^{-1} g_{0\beta} g_{0\gamma}\right)\dot{r}^{\beta}\dot{r}^{\gamma} - \frac{\partial V(r^{\alpha})}{\partial r^{\alpha}} - \frac{1}{2}\mu^2\frac{\partial g_{00}^{-1}}{\partial r^{\alpha}} - \mu B_{\alpha\beta}\dot{r}^{\beta}, \end{aligned}$$

which exactly corresponds to the Lagrange–Routh equations (A.3).

Acknowledgment. The authors thank J. M. Mondelo for providing them with a program for frequency analysis.

REFERENCES

[1] V. I. Arnold, *Proof of A.N. Kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the Hamiltonian*, Russian Math. Surveys, 18 (1963), pp. 9–36.

[2] J. Bellerose and D. J. Scheeres, *Stability of Equilibrium Points in the Restricted Full Three Body Problem*, preprint, University of Michigan, Ann Arbor, MI, 2006; available online at http://www-personal.engin.umich.edu/∼scheeres/reprints.html.

[3] A. M. Bloch, P. S. Krishnaprasad, J. E. Marsden, and T. S. Ratiu, *Dissipation induced instabilities*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 11 (1994), pp. 37–90.
[4] B. C. Carlson, *Computing elliptic integrals by duplication*, Numer. Math., 33 (1979), pp. 1–16.
[5] H. Cendra and J. E. Marsden, *Geometric mechanics and the dynamics of asteroid pairs*, Dyn. Syst., 20 (2005), pp. 3–21.
[6] M. G. Clerc and J. E. Marsden, *Dissipation-induced instabilities in an optical cavity laser: A mechanical analog near the* 1 : 1 *resonance*, Phys. Rev. E, 64 (2001), paper 067603.
[7] F. Gabern, *On the Dynamics of the Trojan Asteroids*, Ph.D. thesis, Department of Applied Mathematics and Analysis, University of Barcelona, Barcelona, Spain, 2003; available online at http://www.maia.ub.es/dsg/2003/.
[8] F. Gabern and À. Jorba, *A restricted four-body model for the dynamics near the Lagrangian points of the Sun-Jupiter system*, Discrete Contin. Dyn. Syst. Ser. B, 1 (2001), pp. 143–182.
[9] F. Gabern, À. Jorba, and U. Locatelli, *On the construction of the Kolmogorov normal form for the Trojan asteroids*, Nonlinearity, 18 (2005), pp. 1705–1734.
[10] F. Gabern, À. Jorba, and P. Robutel, *On the accuracy of restricted three-body models for the Trojan motion*, Discrete Contin. Dynam. Systems, 11 (2004), pp. 843–854.
[11] F. Gabern, W. S. Koon, and J. E. Marsden, *Spacecraft dynamics near a binary asteroid*, Discrete Contin. Dynam. Systems, supplement (2005), pp. 297–306.
[12] F. Gabern, W. S. Koon, and J. E. Marsden, *Parking a spacecraft near an asteroid pair*, J. Guid. Control Dyn., to appear.
[13] A. Giorgilli and C. Skokos, *On the stability of the Trojan asteroids*, Astron. Astrophys., 317 (1997), pp. 254–261.
[14] G. Gómez, À. Jorba, J. Masdemont, and C. Simó, *Dynamics and Mission Design near Libration Points*, Vol. IV, World Sci. Monogr. Ser. Math. 5, World Scientific Publishing, River Edge, NJ, 2001.
[15] G. Gómez, W. S. Koon, M. Lo, J. E. Marsden, J. Masdemont, and S. D. Ross, *Connecting orbits and invariant manifolds in the spatial restricted three-body problem*, Nonlinearity, 17 (2004), pp. 1571–1606.
[16] G. Gómez, J. M. Mondelo, and C. Simó, *Refined Fourier Analysis: Procedures, Error Estimates and Applications*, preprint, University of Barcelona, Barcelona, Spain, 2001; available online at http://www.maia.ub.es/dsg/2001/.
[17] W. R. Johnston, *Asteroids with Satellites*, available online at Johnston's Archive: http://www.johnstonsarchive.net/astro/asteroidmoons.html (last visit in May 2005).
[18] À. Jorba and J. Masdemont, *Dynamics in the centre manifold of the collinear points of the restricted three body problem*, Phys. D, 132 (1999), pp. 189–213.
[19] A. N. Kolmogorov, *On the persistence of conditionally periodic motions under a small change of the Hamilton function*, Dokl. Akad. Nauk SSSR, 98 (1954), pp. 527–530.
[20] W. S. Koon, J. E. Marsden, S. Ross, M. Lo, and D. J. Scheeres, *Geometric mechanics and the dynamics of asteroid pairs*, Ann. New York Acad. Sci., 1017 (2004), pp. 11–38.
[21] J. Laskar, *Introduction to frequency map analysis*, in Hamiltonian Systems with Three or More Degrees of Freedom, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 533, C. Simó, ed., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 134–150.
[22] J. C. Liou and H. A. Zook, *An asteroidal dust ring of micron-sized particles trapped in* 1 : 1 *mean motion with Jupiter*, Icarus, 113 (1995), pp. 403–414.
[23] R. de la Llave, *A tutorial on KAM theory*, in Smooth Ergodic Theory and Its Applications, Proc. Sympos. Pure Math. 69, AMS, Providence, RI, 2001, pp. 175–292.
[24] A. J. Maciejewski, *Reduction, relative equilibria and potential in the two rigid bodies problem*, Celest. Mech. Dyn. Astron., 63 (1996), pp. 1–28.
[25] W. D. MacMillan, *The Theory of the Potential*, McGraw–Hill, New York, 1930.
[26] J. L. Margot, M. C. Nolan, L. A. M. Benner, S. J. Ostro, R. F. Jurgens, J. D. Giorgini, M. A. Slade, and D. B. Campbell, *Binary asteroids in the near-Earth object population*, Science, 296 (2002), pp. 1445–1448.
[27] J. E. Marsden, *Lectures on Mechanics*, London Math. Soc. Lecture Note Ser. 174, Cambridge University Press, Cambridge, UK, 1992.
[28] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Texts in Appl. Math. 17, Springer-Verlag, New York, 1999.

[29] J. E. Marsden and S. D. Ross, *New methods in celestial mechanics and mission design*, Bull. Amer. Math. Soc. (N.S.), 43 (2006), pp. 43–73.
[30] J. E. Marsden, T. S. Ratiu, and J. Scheurle, *Reduction theory and the Lagrange-Routh equations*, J. Math. Phys., 41 (2000), pp. 3379–3429.
[31] J. E. Marsden and J. Scheurle, *Lagrangian reduction and the double spherical pendulum*, Z. Angew. Math. Phys., 44 (1993), pp. 17–43.
[32] R. McKenzie and V. Szebehely, *Nonlinear stability around the triangular libration points*, Celest. Mech. Dyn. Astron., 23 (1981), pp. 223–229.
[33] W. J. Merline, S. J. Weidenschilling, D. D. Durda, J. L. Margot, P. Pravec, and A. D. Storrs, *Asteroids do have satellites*, in Asteroids III, W. F. Bottke, A. Cellino, P. Paolicchi, and R. P. Binzel, eds., University of Arizona Press, Tucson, AZ, 2002, pp. 289–312.
[34] J. Moser, *On invariant curves of area-preserving mappings of an annulus*, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II, 2 (1962), pp. 1–20.
[35] Y. G. Oh, *A stability criterion for Hamiltonian systems with symmetry*, J. Geom. Phys., 4 (1987), pp. 163–182.
[36] P. Robutel, F. Gabern, and À. Jorba, *The observed Trojans and the global dynamics around the Lagrangian points of the Sun–Jupiter system*, Celest. Mech. Dyn. Astron., 92 (2005), pp. 53–69.
[37] P. Robutel and J. Laskar, *Frequency map and global dynamics in the Solar System* I, Icarus, 152 (2001), pp. 4–28.
[38] D. J. Scheeres, *Dynamics about uniformly rotating triaxial ellipsoids: Applications to asteroids*, Icarus, 110 (1994), pp. 225–238.
[39] D. J. Scheeres, *Stability in the full two-body problem*, Celest. Mech. Dyn. Astron., 83 (2002), pp. 155–169.
[40] D. J. Scheeres, *Stability of relative equilibria in the full two-body problem*, Ann. New York Acad. Sci., 1017 (2004), pp. 81–94.
[41] D. J. Scheeres and J. Bellerose, *The restricted Hill full 4-body problem: Application to spacecraft motion about binary asteroids*, Dyn. Syst., 20 (2005), pp. 23–44.
[42] D. W. Schuerman, *The restricted three-body problem including radiation pressure*, Astrophys. J., 238 (1980), pp. 337–342.
[43] M. B. Sevryuk, *The classical KAM theory at the dawn of the twenty-first century*, Moscow Math. J., 3 (2003), pp. 1113–1144.
[44] J. C. Simo, D. Lewis, and J. E. Marsden, *Stability of relative equilibria.* I. *The reduced energy-momentum method*, Arch. Rational Mech. Anal., 115 (1991), pp. 15–59.
[45] V. Szebehely, *Theory of Orbits*, Academic Press, New York, 1967.
[46] L. V. Vela-Arevalo and J. E. Marsden, *Time-frequency analysis of the restricted three-body problem: Transport and resonance transitions*, Classical Quantum Gravity, 21 (2004), pp. S351–S375.
[47] R. A. Werner and D. J. Scheeres, *Exterior gravitation of a polyhedron derived and compared with harmonic and mascon gravitation representations of asteroid* 4769 *Castalia*, Celest. Mech. Dyn. Astron., 65 (1997), pp. 313–344.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 2, pp. 280–292

Rigorous Computations of Homoclinic Tangencies*

Zin Arai† and Konstantin Mischaikow‡

Abstract. In this paper, we propose a rigorous computational method for detecting homoclinic tangencies and structurally unstable connecting orbits. It is a combination of several tools and algorithms, including the interval arithmetic, the subdivision algorithm, the Conley index theory, and the computational homology theory. As an example, we prove the existence of generic homoclinic tangencies in the Hénon family.

Key words. homoclinic tangency, connecting orbit, Conley index, computational homology

AMS subject classifications. 37B30, 37G25, 37M20

DOI. 10.1137/050626429

1. Introduction.

In this paper, we present a method for proving the existence of homoclinic tangencies and structurally unstable connecting orbits. More precisely, we are interested in proving the existence of *generic* tangencies in a one-parameter family of maps, that is, a quadratic tangency that unfolds generically in the family. The importance of the generic homoclinic tangency comes from the fact that it implies the occurrence of the Newhouse phenomena [17] and strange attractors [14].

To explain how the method works, we apply it to the Hénon family

$$H_{a,b} : \mathbb{R}^2 \to \mathbb{R}^2, \quad (x, y) \mapsto (a - x^2 + by, x). \tag{1.1}$$

Belief in the existence of homoclinic tangencies in the Hénon family is easily obtained by numerical experiments. For example, the plots in Figure 1.1 suggest the existence of tangencies for parameter values close to $a = 1.4$, $b = 0.3$ and $a = 1.3$, $b = -0.3$. Our motivation for this work is to develop a general computationally inexpensive method that provides a mathematically rigorous verification of this numerically induced speculation. In fact, using this technique, we prove the following two results.

Theorem 1.1. *Fix any b_0 sufficiently close to 0.3. Then there exists*

$$a \in [1.392419807915, 1.392419807931]$$

such that the one-parameter family H_{a,b_0} has a generic homoclinic tangency with respect to the saddle fixed point on the first quadrant.

*Received by the editors March 9, 2005; accepted for publication (in revised form) by T. Kaper March 21, 2006; published electronically June 2, 2006. This work was supported in part by NSF grants INT-0089631, DMS-0107396, and DMS-0511115, DOE grant 97891, and DARPA, and by the JSPS Japan-U.S. Cooperative Science Program.
http://www.siam.org/journals/siads/5-2/62642.html
†Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan (arai@math.kyoto-u.ac.jp).
‡Center for Dynamical Systems and Nonlinear Studies, Georgia Institute of Technology, Atlanta, GA 30332 (mischaik@math.gatech.edu).

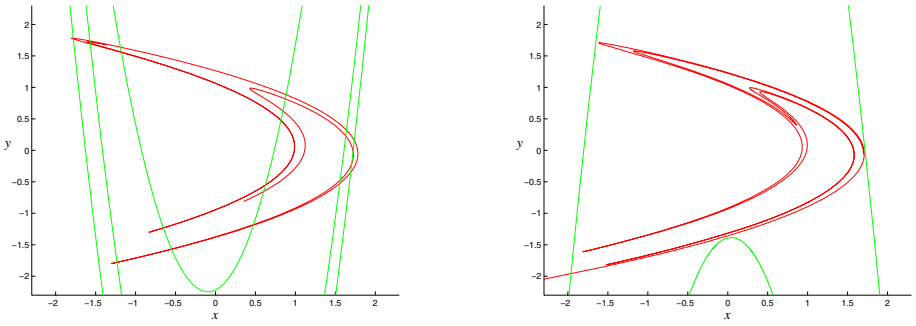


Figure 1.1. *Left*: $a = 1.4, b = 0.3$; *right*: $a = 1.3, b = -0.3$. *Red lines: the unstable manifolds; green lines: the stable manifold.*

Theorem 1.2. *Fix any b_0 sufficiently close to -0.3. Then there exists*

$$a \in [1.314527109319, 1.314527109334]$$

such that the one-parameter family H_{a,b_0} has a generic homoclinic tangency with respect to the saddle fixed point on the third quadrant.

Here we say that a tangency in a one-parameter family is generic if the intersection of unstable and stable manifolds is quadratic, and the intersection is unfolded generically in the family (see section 2 for the precise definition).

Similar results can also be attained by methods of Fornæss and Gavosto [5, 6]. Compared to their technique, which depends on the analyticity of maps, our method is rather geometric and topological, and is designed so that it can be applied to a wider class of maps. Essentially, we require a continuous family of C^2 diffeomorphisms for which we can compute the image of the maps using interval arithmetic. We present a brief overview of our approach; a more detailed description is provided in the following sections.

The essential difficulty of a computer assisted proof in dynamics is that the dynamical system that the computer is capable of representing and evaluating is at best a small perturbation of the system of interest. However, the small perturbations can induce bifurcations which create or destroy the dynamical structure of interest. The Conley index [7, 11, 13] is a powerful tool for this type of problem precisely because it remains constant under perturbations. It is an algebraic topological quantity which can be used to prove the existence of particular dynamical structures including connecting orbits. Using recently developed computational topology tools [9], it is possible to compute the index from the numerically generated data with the guarantee that the index is valid for the original system of interest.

To be more precise, consider $f : X \to X$ a continuous map on a locally compact metric space X. We use the homological Conley index with integer coefficients defined for an isolated invariant set S of f, and denote it by $\mathrm{Con}_*(S, f)$ or simply by $\mathrm{Con}_*(S)$. Recall that $\mathrm{Con}_*(S)$ is the shift equivalence class of the pair of a graded module $CH_*(S)$ and an endomorphism $\chi_*(S)$ on $CH_*(S)$. (See [7, 9] for the concept of shift equivalence and the definition of the Conley index for maps.) By an abuse of notation we write the shift equivalent class $[(CH_*(S), \chi_*(S))]$ simply as $(CH_*(S), \chi_*(S))$.

We say that an orbit $\sigma : \mathbb{Z} \to X$, $f(\sigma(k)) = \sigma(k+1)$ for all k, is a connecting orbit from S_1 to S_2 if its α-limit set is contained in S_1 and its ω-limit set is contained in S_2. The maximal invariant set of $N \subset X$ will be denoted by $\mathrm{Inv}(N)$.

The following theorem (see also [12]), which is proven in section 3, lies at the heart of our algebraic machinery for finding connecting orbits.

Theorem 1.3. *Let N_1, N_2, and N be isolating neighborhoods, and assume that N is the disjoint union of N_1 and N_2. If $f(N_2) \cap N_1 = \emptyset$ and*

$$\mathrm{Con}_*(\mathrm{Inv}(N)) \ncong \mathrm{Con}_*(\mathrm{Inv}(N_1)) \oplus \mathrm{Con}_*(\mathrm{Inv}(N_2))$$

as shift equivalence classes, then there exists a connecting orbit from $\mathrm{Inv}(N_1)$ to $\mathrm{Inv}(N_2)$.

Consider a continuous one-parameter family of C^2 diffeomorphisms $f_\lambda : X \to X$, where the parameter $\lambda \in \mathbb{R}$, and assume that f_0 has a homoclinic tangency. Generically one expects that for $\lambda \neq 0$, f_λ will not posses a homoclinic tangency. Since the Conley index is robust with respect to perturbations, there is no hope that an existence proof can be obtained by a direct application of the index. Thus, we need to recast the problem in such a way that generic homoclinic tangencies become robust isolated objects.

To obtain the isolation observe that at a homoclinic tangency the stable and unstable manifolds share a tangent vector. Let $Pf_\lambda : PX \to PX$ be the induced map on the projective bundle of X. Then a homoclinic tangency of f_0 corresponds to a connecting orbit of Pf_0.

The robustness can be obtained by considering the entire family of maps simultaneously. To do this define

$$\begin{aligned} F : X \times \mathbb{R} &\to X \times \mathbb{R}, \\ (x, \lambda) &\mapsto (f_\lambda(x), \lambda). \end{aligned} \tag{1.2}$$

One now expects that if $\bar{F} : X \times \mathbb{R} \to X \times \mathbb{R}$ of the form $\bar{F}(x, \lambda) = (\bar{f}_\lambda(x), \lambda)$ is induced by a perturbation of f, then there exists $\lambda_0 \approx 0$ such that $\bar{f}_{\lambda_0}$ posses a homoclinic orbit.

With this in mind, one is tempted to apply Theorem 1.3 by restricting F to Λ, a compact interval containing 0, and computing the index of PF. Unfortunately, for technical reasons explained in section 3 this does not work. Instead we compute using PF', where $F' : X \times \mathbb{R} \to X \times \mathbb{R}$ represents a perturbation of F with the property that $F = F'|_{X \times \Lambda}$. Thus a heteroclinic tangency for F' is equivalent to a heteroclinic tangency for F and hence f_λ for some $\lambda \in \Lambda$.

To check that the heteroclinic tangency is indeed quadratic it is sufficient to show that the heteroclinic orbit does not define a connecting orbit for $PPf : PPX \to PPX$, the induced map on the projective bundle of PX.

The details concerning the induced dynamics on the projective bundles are described in section 2. The Conley index tools are described in section 3. Finally, in section 4 we indicate how these techniques are implemented in the context of the Hénon family. All the source files used in the computation can be downloaded from http://www.math.kyoto-u.ac.jp/~arai. To run the computation, one needs software packages GAIO [3, 4] and Computational Homology Programs (CHomP [2]).

2. Tangencies and connecting orbits.

Let f be a diffeomorphism on a manifold X. We denote the tangent bundle of X by TX and the differential of f by Tf.

The dynamical system $f : X \to X$ induces a dynamical system $Pf : PX \to PX$ defined as follows. The space PX is the projective bundle associated with the tangent bundle of X, that is, the fiber bundle on X whose fiber over $x \in X$ is the projective space of T_xX,

$$PX = \coprod_{x\in X} P_xX := \coprod_{x\in X} \{\text{one-dimensional subspace of } T_xX\}.$$

Define Pf to be the map induced from Tf on PX, namely, $Pf([v]) := [Tf(v)]$, where $0 \neq v \in TX$ and $[v]$ is the subspace spanned by v. Identifying X with the image of the zero section of TX, we have the following commutative diagram:

$$\begin{array}{ccc} TX \setminus X & \xrightarrow{Tf} & TX \setminus X \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ PX & \xrightarrow{Pf} & PX \\ {\scriptstyle\pi'}\downarrow & & \downarrow{\scriptstyle\pi'} \\ X & \xrightarrow{f} & X. \end{array}$$

Let $p \in X$ be a hyperbolic fixed point of f and $T_pX = \tilde{E}^s_p \oplus \tilde{E}^u_p$ the corresponding splitting of the tangent space. We denote the stable and unstable manifolds of p by $W^s(p)$ and $W^u(p)$, respectively.

Define $E^s_p := \pi(\tilde{E}^s_p \setminus \{0\})$ and $E^u_p := \pi(\tilde{E}^u_p \setminus \{0\})$. The spaces E^s_p and E^u_p are isolated invariant sets with respect to $Pf : PX \to PX$.

Theorem 2.1 (see [1, Proposition 5.3]). *Let p, q be hyperbolic fixed points of f, and assume that $\dim W^u_f(p) + \dim W^s_f(q) \leq n$. If there exists a connecting orbit from E^u_p to E^s_q under Pf, then $W^u_f(p)$ and $W^s_f(q)$ have a nontransverse intersection.*

Note that if $p = q$, the case of a homoclinic orbit, $\dim W^u_f(p) + \dim W^s_f(p) = n$ always holds. Therefore, the problem of finding homoclinic tangencies is translated to that of finding connecting orbits from $E^u(p)$ to $E^s(p)$ with respect to $Pf : PX \to PX$.

Next, we discuss genericity of tangencies. The definition of genericity is taken from [10]. It is a generalization of the generic (or nondegenerate) tangencies for surface diffeomorphism (see [5, 6, 14, 17]).

Let $\{f_\lambda\}_{\lambda\in\Lambda}$ be a one-parameter family of C^2 diffeomorphism depending smoothly on the parameter $\lambda \in \Lambda \subset \mathbb{R}$. For simplicity, we consider the homoclinic tangency of a family of hyperbolic fixed points $p(\lambda)$ of f_λ. The case for a hyperbolic periodic point is quite similar. Assume that $p(\lambda_0)$ has a homoclinic tangency at $x \in M$ for $\lambda_0 \in \operatorname{int}\Lambda$. For simplicity, we assume that $\lambda_0 = 0$. We say that x is a 1-tangential quadratic homoclinic tangency if the tangent spaces of $W^u_{f_0}(p(0))$ and $W^s_{f_0}(p(0))$ at x have a common subspace of dimension one and the intersection of these manifolds at x is quadratic along this common one-dimensional subspace. Let

$$\mathcal{W}^u_\Lambda = \bigcup_{\lambda\in\Lambda} W^u_{f_\lambda}(p(\lambda)), \qquad \mathcal{W}^s_\Lambda = \bigcup_{\lambda\in\Lambda} W^s_{f_\lambda}(p(\lambda)).$$

These are smooth submanifolds of $X \times \Lambda$ (see [10]), and since we assume that x is a homoclinic tangency for f_0, they have intersect at $(x, 0)$.

Definition 2.2. *The one-parameter family of C^2 diffeomorphisms $\{f_\lambda\}$ has a generic homoclinic tangency at x if x is a 1-tangential quadratic tangency and $\mathcal{W}^u_\Lambda$ and $\mathcal{W}^s_\Lambda$ are transversal in $X \times \Lambda$ at $(x, 0)$.*

Consider the projection $(\pi', id) : PX \times \Lambda \to X \times \Lambda$ and the lifting of a generic homoclinic tangency. Let

$$PF : (x, \lambda) \mapsto (Pf_\lambda(x), \lambda) : PX \times \Lambda \to PX \times \Lambda.$$

It is easy to see that the sets $\mathcal{E}^u_\Lambda := \bigcup_{\lambda\in\Lambda} E^u_{p(\lambda)}$ and $\mathcal{E}^s_\Lambda := \bigcup_{\lambda\in\Lambda} E^s_{p(\lambda)}$ are normally hyperbolic invariant manifolds with respect to PF and

$$W^u_{PF}(\mathcal{E}^u_\Lambda) = \bigcup_{\lambda\in\Lambda} W^u_{Pf_\lambda}(E^u_{p(\lambda)}), \qquad W^s_{PF}(\mathcal{E}^s_\Lambda) = \bigcup_{\lambda\in\Lambda} W^s_{Pf_\lambda}(E^s_{p(\lambda)}).$$

In this setting, the genericity of a tangency is expressed as follows.

Theorem 2.3. *Let f_λ be a one-parameter family of diffeomorphisms with hyperbolic fixed point $p(\lambda)$, and assume that f_0 has a homoclinic tangency with respect to $p(0)$. If the corresponding intersection of $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ is transversal in $PX \times \Lambda$, then the tangency is generic.*

Proof. Denote the unstable and stable dimensions of $p(\lambda)$ by k and ℓ, respectively. Then $E^u_{p(\lambda)}$ and $E^s_{p(\lambda)}$ are normally hyperbolic manifolds of dimension $k-1$ and $\ell-1$ with respect to Pf_λ. Since $E^u_{p(\lambda)}$ is contracting under $Pf_\lambda|_{P_{p(\lambda)}X}$, its unstable dimension is k, and therefore $W^u_{Pf_\lambda}(E^u_{p(\lambda)})$ is a $k-1+k = 2k-1$-dimensional manifold. It follows that $\dim W^u_{PF}(\mathcal{E}^u_\Lambda) = 2k$, and by the same argument that $\dim W^s_{PF}(\mathcal{E}^s_\Lambda) = 2\ell$. Since $k+\ell = \dim X$ and $\dim(PX \times \Lambda) = 2 \cdot \dim X$, the transversal intersection of $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ is 0-dimensional and hence isolated.

Assume that the tangency is not 1-tangential; that is, the dimension of the intersection of the tangent spaces of $W^u_{f_0}(p(0))$ and $W^u_{f_0}(p(0))$ at x is greater than or equal to 2. It follows that the corresponding intersection of $W^u_{Pf_0}(E^u_{p(0)})$ and $W^s_{Pf_0}(E^s_{p(0)})$ must contain a copy of $\mathbb{R}P^k$, where $k \geq 1$. Therefore, the intersection of $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ cannot be an isolated point. This is a contradiction.

Next, assume that the tangency is not quadratic. Then we can take smooth curves c^u on $W^u_{f_0}(p(0))$ and c^s on $W^s_{f_0}(p(0))$ through x such that they have the same first and second derivatives at x. These curves give rise to curves $\tilde{c}^u$ and $\tilde{c}^s$ on $W^u_{Pf_0}(E^u_{p(0)})$ and $W^s_{Pf_0}(E^s_{p(0)})$ that intersect at (x, θ), where θ is the direction of the tangency. Since the second derivatives of c^u and c^s are equal, $\tilde{c}^u$ and $\tilde{c}^s$ are tangent at (x, θ). This is a contradiction.

Finally, since the tangent spaces of $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ span $T_{(x,\theta,0)}(PX \times \Lambda)$ by assumption, it follows that the tangent spaces of $\mathcal{W}^u_\Lambda$ and $\mathcal{W}^s_\Lambda$ span $T_{(x,0)}(X \times \Lambda)$. ∎

3. Method for verifying structurally unstable connecting orbits. In this section, we describe an algebraic-topological method for proving the existence of connecting orbits, especially structurally unstable ones. We begin by proving Theorem 1.3.

Proof. Let $S_1 := \mathrm{Inv}(N_1)$, $S_2 := \mathrm{Inv}(N_2)$, and $S := \mathrm{Inv}(N)$. Suppose there exists no connecting orbit from S_1 to S_2.

Choose an arbitrary $x \in S$. Then there is an orbit $\sigma : \mathbb{Z} \to S$ such that $\sigma(0) = x$. Assume $x \in N_2$. Then its forward orbit is contained in N_2 since $f(N_2) \cap N_1 = \emptyset$. If its backward orbit intersects N_1, then the α-limit set of σ is contained in N_1 because $f(N_2) \cap N_1 = \emptyset$, and thus it follows that σ must be a connecting orbit from S_1 to S_2, contradicting our assumption. Hence $\sigma(\mathbb{Z})$ is contained in N_2, and therefore $x \in \mathrm{Inv}(N_2)$. Similarly, we have $x \in \mathrm{Inv}(N_1)$ if $x \in N_1$.

This means that S is the disjoint union of invariant subsets S_1 and S_2, and it follows from the additivity of the Conley index (see Theorem 3.22 of [13] or Theorem 1.11 of [15], for example) that $\mathrm{Con}_*(S)$ is the direct sum of $\mathrm{Con}_*(S_1)$ and $\mathrm{Con}_*(S_2)$. This is a contradiction. ∎

Here we note that the Conley index is stable under small perturbations, and so are connecting orbits that can be found by Theorem 1.3. Because, if $\mathrm{Con}_*(S, f) \ncong \mathrm{Con}_*(S_1, f) \oplus \mathrm{Con}_*(S_2, f)$, then the same relationship holds for every g sufficiently close to f and the corresponding continuations of S_1, S_2, and S. It follows, therefore, that there also exists a connecting orbit between S_1 and S_2 with respect to g.

This means that we cannot directly apply Theorem 1.3 to find structurally unstable connecting orbits, particularly homoclinic or heteroclinic tangencies. With this in mind, we make the following simple observation: *Having an unstable connection of codimension one is a stable property under small perturbation of one-parameter families.* Thus, our goal is to apply Theorem 1.3 to a set of maps instead of an individual map.

Consider a continuous family of maps $f_\lambda : X \to X$, where λ is a real parameter in a closed interval $\Lambda \subset \mathbb{R}$. Assume that there exist families of isolated invariant sets $S_1(\lambda)$, $S_2(\lambda)$, and $S(\lambda)$ continuing over Λ such that $S_1(\lambda)$ and $S_1(\lambda)$ are invariant subsets of $S(\lambda)$ for each λ.

As in the Introduction, define

$$F : (x, \lambda) \mapsto (f_\lambda(x), \lambda) : X \times \Lambda \to X \times \Lambda.$$

Assume that we have isolating neighborhoods N_1, N_2, and N for $S_1 := \bigcup_{\lambda\in\Lambda} S_1(\lambda)$, $S_2 := \bigcup_{\lambda\in\Lambda} S_2(\lambda)$, and $S := \bigcup_{\lambda\in\Lambda} S(\lambda)$, respectively, such that N is the disjoint union of N_1 and N_2.

It can now be expected that the map F has a connecting orbit from S_1 to S_2 that is stable under small perturbation within the family F, and hence one can hope that Theorem 1.3 can be applied. However, as is shown in the next example, it is often the case that the existence of connecting orbits from S_1 to S_2 is still beyond the scope of Theorem 1.3.

Example 3.1. Consider a one-parameter family of diffeomorphisms f_λ on $\mathbb{R}^3$, as illustrated in Figure 3.1. Let $S_1(\lambda) = x$ and $S_2(\lambda) = y$ be hyperbolic fixed points of f_λ with one and two unstable dimensions, respectively. Assume that $W^u_{f_0}(x)$ intersects $W^s_{f_0}(y)$ at $\lambda = 0$ and therefore that there is a connecting orbit from x to y. Let

$$S(\lambda) = \{x\} \cup \{y\} \cup (W^u_{f_\lambda}(x) \cap W^s_{f_\lambda}(y))$$

and N be a sufficiently small compact neighborhood of $S(0)$. Then N is an isolating neigh-

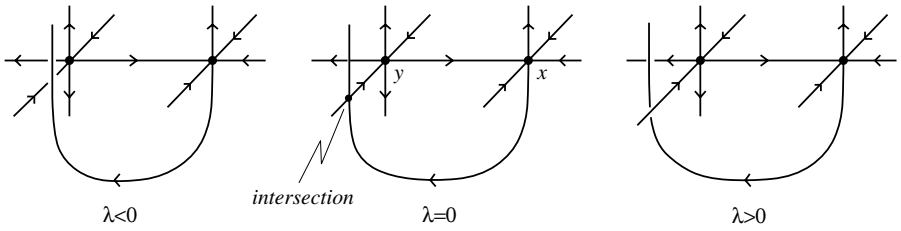


Figure 3.1. *A heterodimensional cycle.*

borhood of $S(\lambda)$ for λ close to 0. It is clear that

$$\mathrm{Con}_*(S_i(\lambda)) \cong \mathrm{Con}_*(S_i) \cong \begin{cases} (\mathbb{Z}, 1) & \text{if } * = i, \\ (0,0) & \text{if } * \neq i. \end{cases}$$

The index for S is obtained by collapsing the exit set of N. Each component of N is collapsed to a space homotopic to S^1, except for the one that contains x, which is collapsed to a space homotopic to the bouquet of two S^1, and the one that contains y, which is collapsed to a space homotopic to S^2. By computing the shift equivalence class, we have

$$\mathrm{Con}_*(S(\lambda)) \cong \mathrm{Con}_*(S) \cong \begin{cases} (\mathbb{Z}, 1) & \text{if } * = 1, 2, \\ (0,0) & \text{other.} \end{cases}$$

Observe that although the connecting orbit from $S_1(0)$ to $S_2(0)$ is structurally unstable, having such a connecting orbit is a stable property with respect to a small perturbation of the family. However, this is an example where the converse of Theorem 1.3 does not hold, and thus we cannot detect the connecting orbit with this theorem. The problem is that the unstable dimensions of S_1 and S_2 are different, and hence they have nontrivial Conley index only at different degrees.

We note that this example illustrates a typical situation that occurs when we consider the projectivization of a homoclinic tangency. Precisely, let p be a hyperbolic saddle fixed point of a surface diffeomorphism f_λ and $x = E^u_p$ and $y = E^s_p$. Then x and y are hyperbolic fixed point of Pf with one- and two-dimensional unstable manifolds. There always exists a connection from $y = E^s_p$ to $x = E^u_p$ that is induced from the action of $P_p f$ on $P_p X$, and there exists a connection from x to y if and only if there exists a homoclinic tangency with respect to p.

To overcome this difficulty, we put an artificial perturbation on F that suspends $\mathrm{Con}_*(S_1)$. Let Λ' be a closed subinterval of Λ such that $\Lambda \setminus \Lambda'$ has two components, and suppose that $F(N_1) \cap N_2$ is included in $X \times \Lambda'$. This implies that there is no connecting orbit for $\lambda \in \Lambda \setminus \Lambda'$.

Define

$$F'(x,\lambda) = \begin{cases} (f_\lambda(x), \lambda + g(\lambda)), & x \in N_1, \\ (f_\lambda(x), \lambda - g(\lambda)), & x \in N_2, \end{cases}$$

where $g : \Lambda \to \mathbb{R}$ is a continuous function that is negative on the left component of $\Lambda \setminus \Lambda'$, vanishing on Λ', and positive on the right component of $\Lambda \setminus \Lambda'$.

After this perturbation, N_1, N_2, and N remain isolating neighborhoods. Define S_1', S_2', and S' to be the maximal invariant sets of N_1, N_2, and N, respectively, with respect to F'. Then by the suspension isomorphism theorem and the homotopy continuation property of the Conley index, we have

$$\mathrm{Con}_*(S_1', F') = \mathrm{Con}_{*-1}(S_1, F), \quad \mathrm{Con}_*(S_2', F') = \mathrm{Con}_*(S_2, F).$$

Note that if we apply this construction to Example 3.1, S_1' has the nontrivial Conley index at degree 2, the same degree at which S_2' has the nontrivial Conley index.

Theorem 3.2. *In the above setting, if*

$$\mathrm{Con}_*(S', F') \ncong \mathrm{Con}_*(S_1', F') \oplus \mathrm{Con}_*(S_2', F'),$$

then there exists $\lambda_0 \in \Lambda'$ *such that there is a connecting orbit from* $S_1(\lambda_0)$ *to* $S_2(\lambda_0)$ *under* f_{λ_0}.

Proof. By Theorem 1.3, there exists a connection from S_1' to S_2' under F'. By our assumption, this connecting orbit must be in $X \times \Lambda'$. However, F' and F are identical on Λ', and hence the theorem follows. ∎

4. Tangencies in the Hénon family. In this section, we verify the existence of generic homoclinic tangencies in the Hénon family (1.1) by applying the ideas developed in sections 2 and 3. We explain the steps of the computation in the case of Theorem 1.1, a tangency close to the classical parameter values $a = 1.4$ and $b = 0.3$. With b fixed at 0.3, $H_{a,0.3}$ is now considered to be a one-parameter family with parameter a. For simplicity and to maintain the notation introduced in the earlier sections we write $f_a := H_{a,0.3}$.

We focus on the fixed point

$$p(a) = \left(\frac{-0.7 + \sqrt{0.49 + 4a}}{2}, \frac{-0.7 + \sqrt{0.49 + 4a}}{2} \right),$$

which lies in the first quadrant. By Theorem 2.1, it is sufficient to show the existence of a connecting orbit from $E^u_{p(a)}$ to $E^s_{p(a)}$ for some a. We conclude that the tangency is generic by checking the transversality of $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ using Theorem 2.3.

First we construct isolating neighborhoods N_1, N_2, and N in $PM \times \Lambda = \mathbb{R}^2 \times S^1 \times \mathbb{R}$, with respect to the dynamical system $PF : (x, a) \mapsto (Pf_a(x), a)$. This is done using cubes, i.e., products of closed intervals, in this case four-dimensional cubes, since TX is homeomorphic to $\mathbb{R}^4$. These isolating neighborhoods are designed so that $S_1 = \mathrm{Inv}(N_1)$ contains $\mathcal{E}^u_\Lambda = \bigcup E^u_{p(a)}$, $S_2 := \mathrm{Inv}(N_2)$ contains $\mathcal{E}^s_\Lambda = \bigcup E^s_{p(a)}$, and $N = N_1 \cup N_2$ contains S_1, S_2, and the connecting orbit of our interest. For simplicity, we write a slice $S \cap (PX \times \{a\})$ of $S \subset PX \times \Lambda$ as $S(a)$, and so forth.

Next we apply the perturbation described in section 2 to the map PF so that the Conley index of S_1 will be suspended. After perturbation, we have three isolated invariant sets S_1', S_2', and S' with respect to PF'.

Here we compute the Conley indexes of S_1', S_2', and S' and apply Theorem 3.2. This proves the existence of a connecting orbit from $S_1(a)$ to $S_2(a)$ for some $a \in \Lambda$. Then we show

that $S_1(a) = E^u_{p(a)}$ and $S_2(a) = E^s_{p(a)}$. It follows that the connecting orbit we found is from $E^u_{p(a)}$ to $E^s_{p(a)}$, which implies the existence of a tangency with respect to f_a.

Finally, we check that $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ are transversal, and conclude that the tangency we found is generic.

The argument above is arranged into the following steps:

Step 1. Construct an initial guess for the location of the connecting orbit.

Step 2. Refine the initial guess up to the desired precision.

Step 3. Modify the refined set to get isolating neighborhoods N_1, N_2, and N.

Step 4. Compute the Conley index and apply Theorem 3.2.

Step 5. Check that $S_1(a) = E^u_{p(a)}$ and $S_2(a) = E^s_{p(a)}$.

Step 6. Check that $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ are transversal.

Before getting into the details of each step, we remark that it is numerically expensive to apply interval arithmetic to trigonometric and inverse trigonometric functions. Therefore, in the following computations, we choose a piecewise linear coordinate $\theta \in (-\pi, \pi]$ for $P_x M = \mathbb{R}P^1 \cong S^1$. This coordinate is not differentiable; nevertheless the Conley index theory is still applicable. To deal with $P(PX \times \Lambda)$, we also take the similar piecewise linear coordinate for $\mathbb{R}P^3$ in the last step.

Step 1. Basically, any method can be used for this step. In our example, we make use of the software package GAIO in this and the next steps. Programs in GAIO are developed for global analysis of invariant objects in dynamical systems by M. Dellnitz, O. Junge and their collaborators; see [3] and the project web page [4]. To construct an initial guess, we simply look at Figure 1.1 and choose cubes that seem to contain the connecting orbit from $E^u_{p(a)}$ to $E^s_{p(a)}$ (Figure 4.1).

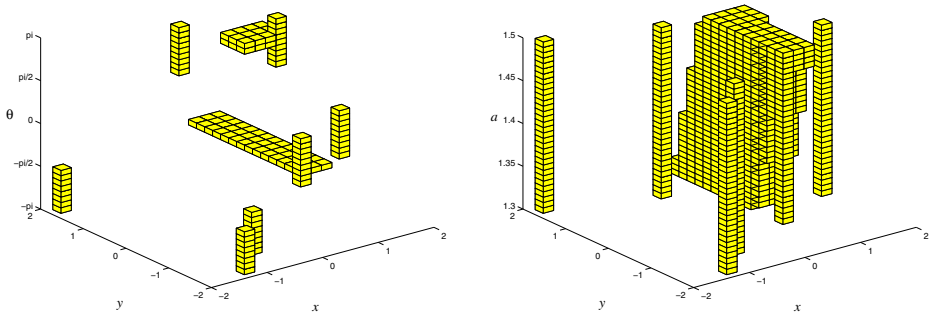


Figure 4.1. *Our initial guess for the connecting orbit. Left: the projection to the x-y-θ space; right: the projection to the x-y-a space.*

Step 2. We refine the initial guess by applying "the subdivision algorithm" [3] of GAIO. In an application of the subdivision algorithm, each cube is divided into two cubes. We make a graph map from the multivalued map induced from PF using interval arithmetic and remove the cubes which do not contain a connecting orbit or a fixed point of the graph map. Since our computation is rigorous, cubes containing a fixed point or a connecting orbit of PF definitely survive this reduction.

After eight applications of the subdivision and reduction procedure, we get the cubes illustrated in Figure 4.2. Cubes after an additional eight applications of the procedure are illustrated in Figure 4.3.

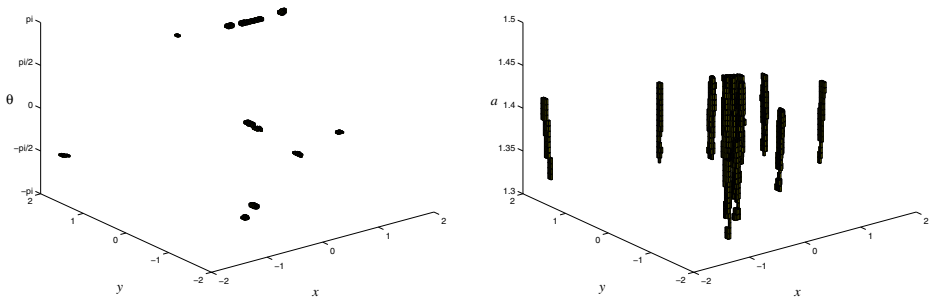


Figure 4.2. *After* 8 *steps of subdivision and reduction procedure. Left: the projection to the* x*-*y*-*θ *space; Right: the projection to the* x*-*y*-*a *space.*

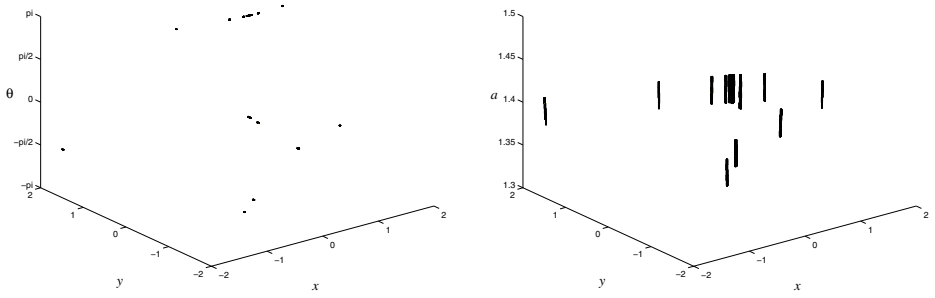


Figure 4.3. *After* 16 *steps of subdivision and reduction procedure. Left: the projection to the* x*-*y*-*θ *space; Right: the projection to the* x*-*y*-*a *space.*

Note that the range of the parameter value a is getting smaller and smaller during this computation. In our example, we apply this procedure 140 times. The resulting set consists of 9029 cubes, and its range of a is smaller than 10^{-10}.

Step 3. Roughly speaking, the algorithm adds cubes to the given set of cubes until it becomes an isolating neighborhood. This is a modification of the algorithm proposed by Junge [8] (see also [9]).

Step 4. To construct index pairs from the isolating neighborhoods found in Step 3, we use the combinatorial index pair algorithm (Algorithm 10.86 of [9]). This gives index pairs for S_1', S_2', and S'.

The Conley index is computed using the Computational Homology Program (CHomP [2]). Application of the program shows that

$$\mathrm{Con}_*(S_1') = \mathrm{Con}_*(S_2') = \begin{cases} (0,0) & \text{if } * \neq 2, \\ (\mathbb{Z},1) & \text{if } * = 2, \end{cases}$$

and

$$\mathrm{Con}_*(S') = \begin{cases} (0,0) & \text{if } * \neq 2, \\ (\mathbb{Z}^{59}, P) & \text{if } * = 2, \end{cases}$$

where P is a 59×59 integer matrix. It can be shown that

$$\mathrm{Con}_2(S') \cong \left(\mathbb{Z}^2, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right) \ncong \left(\mathbb{Z}^2, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \cong \mathrm{Con}_2(S'_1) \oplus \mathrm{Con}_2(S'_2),$$

and therefore, by Theorem 1.3, we conclude that there exists a connecting orbit from $S_1(a)$ to $S_2(a)$ for some $a \in \Lambda'$. In this case, $\Lambda' = [1.392419807915, 1.392419807931]$.

Step 5. We have shown that there exists a parameter value a such that there exists a connecting orbit from $S_1(a)$ to $S_2(a)$. Although $E^u_p(a) \subset S_1(a)$ and $E^s_p(a) \subset S_2(a)$ follow from our construction, it is unknown that whether these sets are equal or not. To show these equalities, we make use of the Hartman–Grobman linearization theorem.

Proposition 4.1. *Let the origin $0 \in \mathbb{R}^n$ be a hyperbolic fixed point of a diffeomorphism f on $\mathbb{R}^n$, and B a ball of radius r and centered at 0. Choose $0 < \mu < 1$ and $\varepsilon > 0$ so that for each eigenvalue λ of $Tf(0)$ we have $|\lambda| < \mu$ or $|\lambda^{-1}| < \mu$, and $\varepsilon + \mu < 1$ and $\varepsilon < m(Tf(0))$ hold. Here m denotes the minimum norm. If the Lipschitz constant of $f - Tf(0)$ restricted to B is less than $\varepsilon/2$, then* $\mathrm{Inv}(B, f) = \{0\}$.

Proof. Let $g := f - Tf(0)$. Define g' by

$$g'(x) = \begin{cases} g(x) & \text{if } x \in B, \\ g(r \cdot x/\|x\|) & \text{if } x \notin B. \end{cases}$$

Then the Lipschitz constant of $g' : \mathbb{R}^n \to \mathbb{R}^n$ is less than ε. Apply the Hartman–Grobman theorem, Theorem 5.7.1 of [16]. (Note that Theorem 5.7.1 of [16] gives the estimate on the size of ε.) ∎

We do not know the exact value of a at which the tangency occurs; therefore we need to show that $S_1(a) = E^u_{p(a)}$ and $S_2(a) = E^s_{p(a)}$ hold for all $a \in \Lambda'$. Note that since we are using interval arithmetic, it suffices to check these equalities for a finite number of intervals that cover Λ'.

Using this proposition, we show that $\pi'(S_1(a))$ and $\pi'(S_2(a))$ coincide with the fixed point $p(a)$. Then it is easy to check $S_1(a) = E^u_{p(a)}$ and $S_2(a) = E^s_{p(a)}$ because the dynamics on the fiber $P_{p(a)}\mathbb{R}^2$ is induced from the linear map $T_{p(a)}f_a$. In practice, we first compute ε of the proposition using interval arithmetic and then check whether the condition of the proposition is satisfied with a ball B containing $\pi'(S_1(a))$ or $\pi'(S_2(a))$. In our example of the Hénon map, we have $(f_a - Tf_a(0))(u, v) = (-u^2, 0)$ by the coordinate change $(x, y) = (u + p(a), v + p(a))$. We then can easily check the condition of the proposition. In general, this check may fail. In that case we apply the subdivision algorithm to $S_1(a)$ and $S_2(a)$ to make these sets smaller, and again check whether the condition of the proposition holds.

Step 6. Recall that $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ are two-dimensional manifolds, and we need to check that these manifolds are transversal in $\mathbb{R}^2 \times S^1 \times \mathbb{R}$.

For this purpose, we apply the procedure of taking projective bundle once again. That is, we construct $PPF : P(PX \times \Lambda) \to P(PX \times \Lambda)$ from $PF : PX \times \Lambda \to PX \times \Lambda$. Recall that $\mathcal{E}^u_\Lambda$ and $\mathcal{E}^u_\Lambda$ are normally hyperbolic invariant manifolds with respect to PF. Let

$$\mathcal{E}^u\mathcal{E}^u_\Lambda := \{v \in P(PX \times \Lambda) \mid v \in P_x(PX \times \Lambda) \text{ where } x \in \mathcal{E}^u_\Lambda, v \in E^u_x\},$$
$$\mathcal{E}^s\mathcal{E}^s_\Lambda := \{v \in P(PX \times \Lambda) \mid v \in P_x(PX \times \Lambda) \text{ where } x \in \mathcal{E}^s_\Lambda, v \in E^s_x\}.$$

These are the set of all unstable and stable vectors based on a point in $\mathcal{E}^u_\Lambda$ and $\mathcal{E}^s_\Lambda$, respectively. Then it follows that if $W^u_{PF}(\mathcal{E}^u_\Lambda)$ and $W^s_{PF}(\mathcal{E}^s_\Lambda)$ are not transversal and hence share a common subspace at the intersection, there must be a connecting orbit from $\mathcal{E}^u\mathcal{E}^u_\Lambda$ to $\mathcal{E}^s\mathcal{E}^s_\Lambda$ with respect to PPF.

Therefore, it suffices to show the nonexistence of such a connecting orbit. This also can be done with interval arithmetics. We subdivide $P(PX \times \Lambda)$ into small cubes and make rigorous coverings (neighborhoods) of $\mathcal{E}^u\mathcal{E}^u_\Lambda$ and $\mathcal{E}^s\mathcal{E}^s_\Lambda$ that consists of cubes. In the case of Hénon map, we can exactly compute $\mathcal{E}^u\mathcal{E}^u_\Lambda$ and $\mathcal{E}^s\mathcal{E}^s_\Lambda$ by hand. In general, we need the help of rigorous interval arithmetics to compute them. Denote these sets of cubes by $\mathcal{U}$ and $\mathcal{S}$, respectively.

By using interval arithmetic, we apply PPF to $\mathcal{U}$ and add the image cubes to $\mathcal{U}$. Then we apply PPF to $\mathcal{U}$ again and repeat this procedure while the number of cubes in $\mathcal{U}$ is increasing. Since the number of the cubes in total space is finite, this procedure stops at some point.

Assume that the number of cubes in $\mathcal{U}$ is the same after one application of PPF. Then it follows that the union of cubes in $\mathcal{U}$ is a rigorous covering of the unstable set of $\mathcal{E}^u\mathcal{E}^u_\Lambda$. Then we check whether $\mathcal{U} \cap \mathcal{S} = \emptyset$. If this holds, then there cannot be a connecting orbit from $\mathcal{E}^u\mathcal{E}^u_\Lambda$ to $\mathcal{E}^s\mathcal{E}^s_\Lambda$, and this is what we wanted to show. If this is not the case, we refine the decomposition of $P(PX \times \Lambda)$ by subdividing all cubes in it and repeat the whole procedure again.

Note that all the discussion in this section is valid for any b sufficiently close to 0.3. This completes the algorithm to prove Theorem 1.1.

The algorithm for Theorem 1.2 is the same, but the computational cost is different as follows:

	$a = 1.4, b = 0.3$	$a = 1.3, b = -0.3$
Step 2	22.2 min	1.9 min
Step 3	153.9 min	22.5 min
Step 4	26.0 min	50.8 min
Step 6	60.8 min	24.1 min

All the computations are done on a PowerMac G5 (2GHz). Since the orbit of tangency is simpler and hence the number of cubes in the isolating neighborhoods is smaller, the computation for the case $a = 1.3, b = -0.3$ is faster. The only exception is Step 4, the computation of homology. The reason for this is the strong expansion rate of the map, which makes the number of the cubes in the image of the isolating neighborhoods significantly large.

Acknowledgments. A part of this work was completed while the first author was visiting Georgia Institute of Technology. He wishes to thank all the people in the Center for Dynamical Systems and Nonlinear Studies for their hospitality. We thank P. Pilarczyk and O. Junge for many important suggestions on computational techniques and programs, and we are also grateful to the referees for suggesting several improvements.

REFERENCES

[1] Z. Arai, *Tangencies and the Conley index*, Ergodic Theory Dynam. Systems, 22 (2002), pp. 973–999.

[2] *Computational Homology Project*, online at http://www.math.gatech.edu/∼chomp/.

[3] M. Dellnitz, F. Gray, and O. Junge, *The algorithms behind GAIO—Set oriented numerical methods for dynamical systems*, in Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems, Springer, Berlin, 2001, pp. 145–174 (and color plates on pp. 805–807).

[4] M. Dellnitz and O. Junge, *The web page of the GAIO project*, online at http://math-www.uni-paderborn.de/∼agdellnitz/gaio/.

[5] J. E. Fornæss and E. A. Gavosto, *Existence of generic homoclinic tangencies for Hénon mappings*, J. Geom. Anal., 2 (1992), pp. 429–444.

[6] J. E. Fornæss and E. A. Gavosto, *Tangencies for real and complex Hénon maps: An analytic method*, Experiment. Math., 8 (1999), pp. 253–260.

[7] J. Franks and D. Richeson, *Shift equivalence and the Conley index*, Trans. Amer. Math. Soc., 352 (2000), pp. 3305–3322.

[8] O. Junge, *Computing specific isolating neighborhood*, in Progress in Analysis, Proceedings of the 3rd International Isaac Congress, Berlin, 2001, Vol. I, II, World Scientific Publishing, River Edge, NJ, 2003, pp. 571–576.

[9] T. Kaczynski, K. Mischaikow, and M. Mrozek, *Computational Homology*, Appl. Math. Sci. 157, Springer-Verlag, New York, 2004.

[10] J.-M. Kleinkauf, *The Numerical Computation and Geometrical Analysis of Heteroclinic Tangencies*, preprint 98-48, SFB 343, University of Bielefeld, Bielefeld, Germany, 1998; available online from http://www.mathematik.uni-bielefeld.de/sfb343/.

[11] K. Mischaikow, *Topological techniques for efficient rigorous computation in dynamics*, in Acta Numer. 11, Cambridge University Press, Cambridge, UK, 2002, pp. 435–478.

[12] K. Mischaikow and M. Mrozek, *Isolating neighborhoods and chaos*, Japan J. Indust. Appl. Math., 12 (1995), pp. 205–236.

[13] K. Mischaikow and M. Mrozek, *The Conley index theory*, in Handbook of Dynamical Systems II, North–Holland, Amsterdam, 2002, pp. 393–460.

[14] L. Mora and M. Viana, *Abundance of strange attractors*, Acta. Math., 171 (1993), pp. 1–71.

[15] M. Mrozek, *Shape index and other indices of Conley type for local maps on locally compact Hausdorff spaces*, Fund. Math., 145 (1994), pp. 15–37.

[16] C. Robinson, *Dynamical Systems; Stability, Symbolic Dynamics, and Chaos*, 2nd ed., CRC Press, Boca Raton, FL, 1999.

[17] C. Robinson, *Bifurcation to infinitely many sinks*, Commun. Math. Phys., 90 (1983), pp. 433–459.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 5, No. 2, pp. 293–312

Random Perturbations of Dynamical Systems with Absorbing States*

Frans Jacobs† and Sebastian J. Schreiber‡

Abstract. Let $F : M \to M$ be a continuous dissipative map of a separable metric space M. Consider a finite collection $\mathcal{A}$ of closed F-forward invariant sets that is closed under intersection and that contains M. For all $\epsilon > 0$, let X^ϵ be a Markov chain for which the elements of $\mathcal{A}$ are absorbing (e.g., extinction boundaries for a population, genotype, or strategy) and such that $d(X^\epsilon_{t+1}, F(X^\epsilon_t)) \leq \epsilon$ for all t. Under an additional nondegeneracy condition (i.e., the noise extends locally in all nonabsorbing directions) and a continuity-like condition on the supports of the random perturbations, we show that for sufficiently small values of ϵ, X^ϵ asymptotically spends all of its time near certain invariant sets of F, so-called absorption preserving chain attractors. Moreover, the weak* limit points of X^ϵ's stationary distributions as $\epsilon \to 0$ are F-invariant probability measures whose supports lie in the absorption preserving chain attractors. Applications to the dynamics of structured and unstructured populations, multispecies interactions, and evolutionary games are given.

Key words. random perturbations of dynamical systems, chain recurrence, absorbing sets, ecological and evolutionary dynamics

AMS subject classifications. 37H10, 92D2, 37B25

DOI. 10.1137/050626417

1. Introduction.

The evolution of many physical and biological processes is governed by a mixture of stochastic forces and nonlinear determinism. For example, ecological and evolutionary systems involve nonlinear interactions that are constantly subject to environmental and demographic fluctuations [1, 17]. When nonlinear determinism dominates, the evolution of these processes is often described by nonlinear dynamical systems in which the current state of the system determines all future states [5]. For these deterministic approximations, it is natural to ask about the correspondence between the behavior of the unperturbed dynamical system and the same system subject to small random perturbations [13]. This correspondence was studied initially in 1933 by Pontryagin, Andronov, and Vitt [15], and more recently by Wentzell and Freidlin [6, 18], Ruelle [16], and Kifer [12]. Each of these studies was primarily motivated by physical processes in which the random perturbations could act locally in all directions of the state space. In many biological systems, however, stochastic forces are limited by biological constraints. These constraints create "absorbing sets" in the state space, which the system cannot leave after entering. For instance, in a closed ecological community

*Received by the editors March 9, 2005; accepted for publication (in revised form) by C. Castillo-Chavez October 30, 2005; published electronically June 30, 2006.
http://www.siam.org/journals/siads/5-2/62641.html

†Section Theoretical Biology, Institute of Biology, Leiden University, Kaiserstraat 63, NL-2311 GP Leiden, The Netherlands (jacobs@rulsfb.leidenuniv.nl). The work of this author was partly supported by National Institutes of Health grant GM56693 and NSF grant DEB 0111613 to S. Gavrilets.

‡Department of Mathematics, The College of William and Mary, Williamsburg, VA 23187-8795 (sjs@math.wm.edu). The work of this author was partially supported by National Science Foundation grants DMS-0077986 and DMS-0517987, and by a summer research grant from the College of William and Mary.

a species that goes extinct remains extinct. It cannot be resurrected by stochastic forces. Alternatively, in structured community models random perturbations may affect the number of individuals in each of the classes or stages, but may not change the structural state or parameters of each of the individuals. Depending on the structure of the model, the collection of absorbing sets may be large or small. For a semelparous population model with age structure, a population missing one or more classes always has one or more age classes missing, a fact that should be respected by random perturbations. For example, the absence of the reproducing age class in the current generation results in the absence of the youngest age class in the next generation. On the other hand, for most iteroparous population models the only absorbing set is extinction [3, 14, 19].

The goal of this article is to investigate random perturbations of dynamical systems with these types of absorbing sets. To achieve this goal, we introduce in section 2 the notion of an *absorbing π-system*, i.e., a collection of closed forward invariant subsets closed under intersection, and random ϵ-perturbations of deterministic maps that preserve a given absorbing π-system. These random ϵ-perturbations act after the deterministic map, preserve the elements of the absorbing π-system (i.e., once the process enters an absorbing set, it remains in that absorbing set), and are no larger than $\epsilon > 0$ in size with respect to a given metric on the community state space. We illustrate these definitions with ecological models accounting for demographic, environmental, and immigrational stochasticity, with replicator equations accounting for demographic, environmental, and mutational stochasticity, and with an age-structured model. To describe the asymptotic behavior of the randomly perturbed system, we prove in section 3 the existence of invariant probability measures μ^ϵ whenever the unperturbed system has at least one attractor. These invariant probability measures describe the long-term statistical behavior of trajectories of the randomly perturbed system. As $\epsilon \downarrow 0$, the work of Khasminskii [11] (see also Kifer [12]) implies that the limit points of these invariant probability measures are invariant probability measures for the unperturbed system. These limit points are *natural invariant measures* for the unperturbed system, as when $\epsilon > 0$ is sufficiently small the long-term statistical behavior of the perturbed system is well approximated by these natural invariant measures.

Since the unperturbed system may have several invariant measures, including ones supported by repellers, sections 4 and 5 determine which invariant sets of the unperturbed system can actually support the natural invariant measures. In section 4 we introduce the notion of an absorption preserving chain attractor. This generalizes the notion of extinction preserving chain attractor as presented in Jacobs and Metz [10], which in turn is a generalization of the concept of chain attractor as derived by Ruelle [16]. Ruelle introduced chain attractors to describe the asymptotic behavior of physical systems, in which the dynamics inherently is influenced by small disturbances. In his construction of chain attractors Ruelle used so-called *pseudo-orbits* to model the effect of limited noise on the orbits of an unperturbed system. In [10] his approach is adapted to unstructured, and in [7] to structured, community-dynamical models as they are studied in ecology, leading to the notion of extinction preserving chain attractors. The present paper extends these ideas to absorbing π-systems that can account for a greater variety of stochastic influences as outlined above. In section 5 we prove that under appropriate assumptions the natural invariant measures are supported by the absorption preserving chain attractors. Moreover, under additional assumptions and $\epsilon > 0$ sufficiently small,

the randomly ϵ-perturbed system eventually remains near *one* of the absorption preserving chain attractors of the unperturbed system.

To illustrate the utility of the theory, we apply our results in section 6 to models of competing species and replicator dynamics. The addition of noise in these models is proven to have significant effects. In section 7 we make some concluding remarks about our results and pose some open questions.

2. Definitions and examples. Let M be a separable metric space with metric d, and $F : M \to M$ a continuous map. Given $A \subset M$ and $x \in M$, let $\text{dist}(x, A) = \inf\{d(x, y) : y \in A\}$, and for $\delta \geq 0$ let $N(A, \delta) = \{y \in M : \text{dist}(y, A) \leq \delta\}$. For notational convenience, when $A = \{x\}$ we write $N(x, \delta)$ instead of $N(\{x\}, \delta)$. We recall a few definitions from dynamical systems theory. Given a subset $S \subset M$, define $F^n(S) = \{F^n(x) : x \in S\}$ and $\omega(S) = \bigcap_{n>0} \overline{\bigcup_{m \geq n} F^m(S)}$, with the notational adaptation to $\omega(x)$ in the case $S = \{x\}$. A set $A \subset M$ is *F-forward invariant* if $F(A) \subset A$, and *F-invariant* if $F(A) = A$. A compact set $A \subset M$ is an *attractor for F* if there exists a compact neighborhood U of A such that $\omega(U) = A$. The *basin of attraction* of a compact F-invariant set $A \subset M$ is the set of points $x \in M$ such that $\omega(x) \subset A$. A point $x \in M$ is *recurrent for F* if $x \in \omega(x)$.

For the map F, specific forward invariant sets may be viewed as absorbing under stochastic perturbations of F. In population models, for example, these forward invariant sets may correspond to the extinction of one or more species, subpopulations, phenotypes, or genotypes. To allow for a mathematical framework flexible enough for structured and unstructured ecological models, replicator equations, and hybrids of these models, we make the following definitions.

Definition 1. *An* absorbing π-system for F *is a finite collection $\mathcal{A}$ of closed F-forward invariant subsets of M which includes the set M and which is closed under intersection (i.e., $A, B \in \mathcal{A}$ implies that $A \cap B \in \mathcal{A}$). An element of $\mathcal{A}$ is called an* absorbing set.

Definition 2. *For $x \in M$ define the* minimal absorbing set for x, *denoted $\mathcal{A}^*(x)$, to be the smallest element in $\mathcal{A}$ containing x.*

Definition 3. *For any set $A \subset M$ and $\delta \geq 0$ define the* ap δ-neighborhood of A *as*

$$N_{ap}(A, \delta) = \bigcup_{x \in A} N(x, \delta) \cap \mathcal{A}^*(x).$$

The index ap refers to absorption preservation. For notational convenience we write $N_{ap}(x, \delta)$ instead of $N_{ap}(\{x\}, \delta)$.

For a given map $F : M \to M$ there are many potential choices of an absorbing π-system, corresponding to different choices about how random perturbations affect the system. Assuming that a π-system $\mathcal{A}$ has been chosen, we make the following definition.

Definition 4. *For $\epsilon \geq 0$, a* random ϵ-perturbation of F respecting the absorbing π-system $\mathcal{A}$ *is a (discrete time) Markov chain X^ϵ, taking values in M and with transition kernel $P^\epsilon_\cdot$,*

$$P^\epsilon_x(\Gamma) = P(X^\epsilon_{t+1} \in \Gamma | X^\epsilon_t = x) \qquad \textit{for all } x \in M \textit{ and for all Borel subsets } \Gamma \subseteq M,$$

which satisfies

H1. $P^\epsilon_x(N_{ap}(F(x), \epsilon)) = 1$.

H1 ensures that the forward invariant sets in $\mathcal{A}$ are absorbing for X^ϵ and that random ϵ-perturbations are ϵ small. When $\mathcal{A} = \{M\}$, we recover random perturbations considered by Kifer [12] and Ruelle [16].

To illustrate choices of random ϵ-perturbations that satisfy H1, we introduce several examples. In these examples, if X_t^ϵ is a vector, then $X_{t,i}^\epsilon$ denotes the ith component of X_t^ϵ.

2.1. Ecological equations. Consider M given by $\mathbf{R}_+^k = \{x = (x_1, \dots, x_k) \in \mathbf{R}^k : x_i \geq 0\}$, where $x = (x_1, \dots, x_k)$ is the vector of population densities. Let $d(x, y) = \max_i |x_i - y_i|$. If $f_i(x)$ denotes the per capita growth rate of the ith population, then $F(x) = (x_1 f_1(x), \dots, x_k f_k(x))$ defines an *ecological difference equation*. For the sake of simplicity, we assume that there exists $\xi \geq 1$ such that $F(\mathbf{R}_+^k) \subset [0, \xi]^k$ (i.e., F is a compact map). For this map we illustrate three choices of noise, corresponding to environmental, demographic, and immigration stochasticity. Combinations of these noises result in different choices of absorbing π-systems.

Environmental stochasticity. Environmental stochasticity occurs when random fluctuations in the environment result in random fluctuations in reproductive or mortality rates. Let $\{Z_t\}_{t\geq 0}$ be a sequence of independent random vectors taking values in $[-1/\xi, 1/\xi]^k$. Let $Z_{t,i}$ denote the ith component of Z_t. Define a random ϵ-perturbation X^ϵ of F by

$$X_{t+1,i}^\epsilon = (1 + \epsilon Z_{t,i}) F_i(X_t^\epsilon).$$

Since $|(1 + \epsilon Z_{t,i})F_i(X_t^\epsilon) - F_i(X_t^\epsilon)| = \epsilon |Z_{t,i} F_i(X_t^\epsilon)| \leq \epsilon$, this choice of X^ϵ satisfies H1 with respect to the absorbing π-system generated by M and all finite intersections of the sets $\{x \in \mathbf{R}_+^k : x_i = 0\}$.

Demographic stochasticity. Demographic stochasticity is the effect that the randomness of birth and death processes has on finite populations. Let $\gamma \gg 1$ denote the habitat size, $N_i = x_i \gamma$ the abundance of population i, $d_i = d_i(\epsilon) \in (0, 1)$ the probability that an individual of population i dies, and $f_i(x)/(1 - d_i)$ the number of progeny produced per individual of population i. If reproduction is deterministic and followed by independent stochastic deaths, the number of surviving individuals of population i is given by a binomial random variable with mean $(1 - d_i)N_i f_i(x)/(1 - d_i) = N_i f_i(x)$ and standard deviation $\sqrt{d_i N_i f_i(x)}$. If we approximate these binomials by appropriately truncated normal random variables $Z_{t,i}(x)$, then $X_{t+1,i}^\epsilon = F_i(X_t^\epsilon) + Z_{t,i}(X_t^\epsilon)$, $i = 1, \dots, k$, satisfies H1 with respect to the absorbing π-system generated by M and all finite intersections of the sets $\{x \in \mathbf{R}_+^k : x_i = 0\}$.

Immigration stochasticity. Suppose that a subset of populations $I \subset \{1, \dots, k\}$ receives a random influx of immigrants. To model this, let $\{Z_t\}_{t\geq 0}$ be a sequence of random vectors with support in $[0, 1]^k$. For populations $i \notin I$ we assume that $Z_{t,i} = 0$, i.e., no immigrants. The random ϵ-perturbation of F given by $X_{t+1}^\epsilon = F(X_t^\epsilon) + \epsilon Z_t$ satisfies H1 with respect to the absorbing π-system generated by M and all finite intersections of the sets $\{x \in \mathbf{R}_+^k : x_i = 0\}$ for $i \notin I$.

Combined random perturbations. In addition to the random perturbations mentioned above, combinations of these random perturbations (e.g., demographic and environmental stochasticity) will also satisfy H1 with respect to the appropriate absorbing π-system.

2.2. Replicator difference equations. Consider $M = \{x \in \mathbf{R}_+^k : \sum_{i=1}^k x_i = 1\}$, where $x = (x_1, \ldots, x_k)$ is the vector of strategy frequencies. Let $d(x, y) = \max_i |x_i - y_i|$. If $f_i(x)$ is the relative fitness of the ith strategy, then

$$F(x) = \left(\frac{x_1 f_1(x)}{\sum_j x_j f_j(x)}, \ldots, \frac{x_k f_k(x)}{\sum_j x_j f_j(x)} \right)$$

defines the distribution of strategies in the next generation, and is called a *replicator equation* (see, e.g., [9]).

Environmental and demographic stochasticity. These forms of random perturbations can be developed for replicator equations in a manner similar to the ecological equations.

Random mutations. Imagine that for every strategy $i \in \{1, \ldots, k\}$ there is a collection of strategies $I_i \subset \{1, \ldots, k\}$ that randomly mutate to strategy i. We assume that I_i always includes i. Let $\{Z_t(i, j)\}_{t \geq 0}$ be a sequence of independent random variables that represent the fraction of strategy i individuals that mutate to strategy j at time t. For each $t \geq 0$ we require $Z_t(i, j) \geq 0$, $\sum_{j=1}^k Z_t(i, j) = 1$, and $Z_t(j, i) > 0$ if and only if $j \in I_i$. Define X^ϵ by

$$X^\epsilon_{t+1,i} = \frac{\sum_{j \in I_i} Z_t(j, i) X^\epsilon_{t,j} f_j(X^\epsilon_t)}{\sum_{j=1}^k X^\epsilon_{t,j} f_j(X^\epsilon_t)}.$$

Under the assumption that $Z_t(i, j) \leq \frac{\epsilon}{k-1}$ for all $i, j \in \{1, \ldots, k\}$ and $i \neq j$, X^ϵ satisfies H1 with respect to the absorbing π-system generated by M and the sets $\{x \in M : x_j = 0 \text{ for all } j \in I_i\}$ with $i = 1, \ldots, k$.

2.3. Age-structured populations. Consider a population with k age classes. Let $x = (x_1, \ldots, x_k)$ be the population vector, where x_i is the density of age class i. A standard model (see, e.g., [2]) for this population is a nonlinear Leslie matrix model

$$F(x) = \begin{pmatrix} f_1(x) & f_2(x) & f_3(x) & \ldots & f_{k-2}(x) & f_{k-1}(x) & f_k(x) \\ s_1(x) & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & s_2(x) & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & s_{k-1}(x) & 0 \end{pmatrix} x,$$

where f_i is the mean number of progeny produced per generation by an individual in age class i, and s_i is the probability that an individual survives from age class i to age class $i+1$. For an age-structured model one can add demographic, environmental, and immigrational stochasticity to each of the age classes. For instance, demographic stochasticity via truncated normal approximations can be used to represent variability in survivorship between age classes and variability in fecundities within each reproductive age class. Depending on the number of reproductive age classes, demographic stochasticity can result in different forms of absorbing π-systems. For example, if the population is semelparous with $f_1 = f_2 = \cdots = f_{k-1} = 0$ and $f_k \neq 0$, then the natural absorbing π-system is given by M, $\cup_i \{x \in \mathbf{R}_+^k : x_i = 0\}$, $\cup_{i_1 > i_2} \{x \in \mathbf{R}_+^k : x_{i_1} = x_{i_2} = 0\}, \ldots, \{0\}$. Alternatively, if the population is significantly iteroparous, e.g., $f_i > 0$ for all i, then the natural absorbing π-system consists of M and $\{0\}$.

3. Empirical and natural invariant measures. Given X^ϵ, define for $t \geq 1$ the *empirical measures* ν_t^ϵ by

$$\nu_t^\epsilon = \frac{1}{t}\sum_{i=1}^{t} \delta_{X_i^\epsilon},$$

where δ_x is a Dirac measure based at the point $x \in M$. One can think of empirical measures as corresponding to plotting a single pixel at each of the points $X_1^\epsilon, \ldots, X_t^\epsilon$, for increasing values of t. As one continues to plot these pixels, certain regions of the phase space M get darkened more than other regions of the phase space. To describe the limiting picture, we consider *weak* limit points* of the sequence $\nu_1^\epsilon, \nu_2^\epsilon, \nu_3^\epsilon, \ldots$. To define a weak* limit point, consider any continuous function $h : M \to \mathbf{R}$. The average of this function with respect to ν_t^ϵ is the average observed value of h up to time t:

$$\int_M h(x)\ d\nu_t^\epsilon(x) = \frac{1}{t}\sum_{i=1}^{t} h(X_i^\epsilon).$$

A weak* limit point of the sequence $\nu_1^\epsilon, \nu_2^\epsilon, \ldots$ is a Borel probability measure μ^ϵ such that there exists an increasing sequence of times $\{t_n\}_{n=1}^\infty$ satisfying

$$\lim_{n\to\infty} \frac{1}{t_n}\sum_{i=1}^{t_n} h(X_i^\epsilon) = \int_M h(x)\ d\mu^\epsilon(x)$$

for all bounded continuous functions $h : M \to \mathbf{R}$.

As we show below, these limit points μ^ϵ are often invariant measures for X^ϵ, i.e., for every Borel set $\Gamma \subset M$,

$$\int_M P_x^\epsilon(\Gamma)\ d\mu^\epsilon(x) = \mu^\epsilon(\Gamma). \tag{1}$$

In the special case of $\epsilon = 0$, (1) simplifies to

$$\mu^0(F^{-1}(\Gamma)) = \mu^0(\Gamma),$$

and μ^0 is an *invariant measure of* F. Another means of defining an invariant measure for X^ϵ is to introduce the operator $\mathcal{P}_\epsilon^*$ on the space of probability measures:

$$\mathcal{P}_\epsilon^*(\mu)(\Gamma) = \int_M P_x^\epsilon(\Gamma)\ d\mu(x),$$

where μ is a Borel probability measure and Γ is a Borel subset of M. An invariant measure μ for X^ϵ is just a fixed point of $\mathcal{P}_\epsilon^*$, i.e., $\mathcal{P}_\epsilon^*(\mu) = \mu$. The importance of this invariance lies in the fact that when $\epsilon > 0$ is sufficiently small, the invariant measures μ^ϵ for X^ϵ obtained as weak* limit points of the sequence $\{\nu_t^\epsilon\}_{t\geq 1}$ will be well approximated by invariant measures of F. In particular, this will mean that if $\epsilon > 0$ is sufficiently small, X^ϵ will spend most of its time near the supports of specific invariant measures of F. Recall that the *support of a probability measure* μ, denoted $\text{supp}(\mu)$, is the intersection of all closed sets K with $\mu(K) = 1$.

Theorem 1. *Let $F : M \to M$ be a continuous map, $\mathcal{A}$ an absorbing π-system for F, and X^ϵ a random ϵ-perturbation of F respecting $\mathcal{A}$. Assume that $A \in \mathcal{A}$ and that B is an attractor for $F|A$. If X_0^ϵ lies in the basin of attraction of B, then for $\epsilon \geq 0$ sufficiently small the sequence $\nu_1^\epsilon, \nu_2^\epsilon, \nu_3^\epsilon, \ldots$ has (with probability one) weak* limit points. These limit points are invariant measures for X^ϵ, with support in B's basin of attraction.*

Remark. Recall that $F : M \to M$ is *dissipative* if it admits a compact global attractor. Theorem 1 ensures the existence of invariant measures for random perturbations X^ϵ of a dissipative F whenever $\epsilon \geq 0$ is sufficiently small.

Proof. By H1 it suffices to prove the theorem when $\mathcal{A} = \{M\}$. Consequently, suppose that $B \subset M$ is an attractor for F and that $x \in M$ is in the basin of attraction of B. Assume that $X_0^\epsilon = x$. Let V be a compact neighborhood of B such that V is contained in B's basin of attraction, and such that $F(V)$ is contained in the interior of V. Since B is an attractor, there exists a natural number T such that $F^T(x)$ is contained in the interior of V. By continuity of F there exists an $\epsilon_1 > 0$ such that if $x_0 = x, x_1, \ldots, x_T \in M$ satisfy $d(x_i, F(x_{i-1})) \leq \epsilon_1$ for $i = 1, \ldots, T$, then x_T is contained in the interior of V. In particular, if $\epsilon \leq \epsilon_1$, then X_T^ϵ is contained in the interior of V with probability one. Choose $\epsilon_2 \in (0, \epsilon_1)$ such that $N(F(V), \epsilon_2) \subset V$. Define

$$U = \bigcup \{x_0 = x, \ldots, x_T : d(x_i, F(x_{i-1})) \leq \epsilon_2 \text{ for } i = 1, \ldots, T\}.$$

Our choice of ϵ_2 implies that, with probability one, $X_t^\epsilon \in U \cup V$ for all $t \geq 0$ and $\epsilon \in [0, \epsilon_2)$. Hence, for $\epsilon \in [0, \epsilon_2)$ the empirical measures ν_t^ϵ, $t \geq 1$, with probability one are supported by the compact set $U \cup V$. By weak* compactness of the Borel probability measures supported in $U \cup V$, there exists a weak* limit point μ^ϵ of the sequence $\{\nu_t^\epsilon\}_{t \geq 1}$ as $t \to \infty$.

To see the invariance of this weak* limit point, let $h : M \to \mathbf{R}$ be any continuous and bounded function. Let $\mathcal{F}_t$ denote the σ-algebra generated by $X_1^\epsilon, \ldots, X_t^\epsilon$. Define sequences $\{Y_t\}_{t \geq 1}$ and $\{Z_t\}_{t \geq 1}$ by

$$Y_t = \frac{1}{t} \left(h(X_t^\epsilon) - E[h(X_t^\epsilon) | X_{t-1}^\epsilon] \right)$$

and

$$Z_t = \sum_{i=1}^{t} Y_i.$$

$\{Z_t\}_{t \geq 1}$ is a martingale with respect to $\{\mathcal{F}_t\}_{t \geq 1}$, as

$$\begin{aligned} E[Z_{t+1} - Z_t | \mathcal{F}_t] &= E[Y_{t+1} | \mathcal{F}_t] \\ &= \frac{1}{t+1} \left(E[h(X_{t+1}^\epsilon) | \mathcal{F}_t] - E[h(X_{t+1}^\epsilon) | \mathcal{F}_t] \right) = 0. \end{aligned}$$

Let $\|h\| = \sup_{x \in M} |h(x)|$. Since $E[Y_{t+1} | \mathcal{F}_t] = 0$, we get that

$$\begin{aligned}
\left|E[Z_{t+1}^2]\right| &= \left|E\left[\left(\sum_{i=1}^{t+1} Y_i\right)^2\right]\right| \\
&= \left|E\left[\sum_i Y_i^2 + 2\sum_{i>j} Y_i Y_j\right]\right| \\
&= \left|\sum_i E[Y_i^2] + 2\sum_{i>j} E[E[Y_i Y_j|\mathcal{F}_{i-1}]\right| \\
&= \left|\sum_i E[Y_i^2] + 2\sum_{i>j} E[Y_j E[Y_i|\mathcal{F}_{i-1}]]\right| \\
&= \left|\sum_i E[Y_i^2]\right| = \left|\sum_i E\left[\frac{1}{i^2}\left(h(X_i^\epsilon) - E[h(X_i^\epsilon)|\mathcal{F}_{i-1}]\right)^2\right]\right| \\
&\leq \sum_{i=1}^{t+1} \frac{1}{i^2} 4\|h\|^2.
\end{aligned}$$

Hence, $\{Z_t\}_{t\geq 1}$ is an L^2 martingale, and Doob's convergence theorem implies that $\lim_{t\to\infty} Z_t$ converges with probability one. By Kronecker's lemma,

$$\lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} \left(h(X_i^\epsilon) - E[h(X_i^\epsilon)|\mathcal{F}_{i-1}]\right) = 0$$

with probability one. Now suppose that $\lim_{i\to\infty} \nu_{t_i}^\epsilon$ converges weakly to μ^ϵ. The previous estimate implies that

$$\begin{aligned}
\int h(x)\, d\mu^\epsilon(x) - \int h(x)\, d\mathcal{P}_\epsilon^*(\mu^\epsilon)(x) &= \lim_{i\to\infty} \int h(x)\, d\nu_{t_i}^\epsilon(x) - \int\int h(y)\, dP_x^\epsilon(y)\, d\nu_{t_i}^\epsilon(x) \\
&= \lim_{i\to\infty} \frac{1}{t_i}\sum_{t=1}^{t_i} \left(h(X_t^\epsilon) - \int h(y)\, dP_{X_t^\epsilon}^\epsilon(y)\right) \\
&= \lim_{i\to\infty} \frac{1}{t_i}\sum_{t=1}^{t_i} \left(h(X_t^\epsilon) - E[h(X_t^\epsilon)|\mathcal{F}_{t-1}]\right) \\
&\qquad + \frac{1}{t_i}\left(E[h(X_1^\epsilon)|\mathcal{F}_0] - E[h(X_{t_i+1}^\epsilon)|\mathcal{F}_{t_i}]\right) \\
&= 0
\end{aligned}$$

with probability one. Since h is an arbitrary continuous bounded function and M is separable, the weak* limit point μ^ϵ is with probability one an invariant Borel probability measure for X^ϵ. ■

In particular this theorem with $\epsilon = 0$ implies that if F has an attractor, then F has an invariant measure. The invariant measures for F may be quite numerous. For instance, any equilibrium or periodic orbit of F, whether stable or unstable, supports an invariant measure.

However, it is natural to assume that some of the invariant measures are more physically or biologically relevant than other invariant measures. For instance, intuition dictates that an invariant measure supported on an unstable fixed point is unlikely to be observed in nature, while an invariant measure supported on a stable fixed point is more likely to be observed. To pick out physically or biologically relevant invariant measures suppose that, for all $\epsilon > 0$ sufficiently small, X^ϵ has an invariant measure μ^ϵ as obtained in Theorem 1. A standard argument (see, e.g., Kifer [12]) implies that the weak* limit points of these μ^ϵ as $\epsilon \downarrow 0$ are F-invariant measures.

Theorem 2. *Let, for each $\epsilon > 0$ sufficiently small, μ^ϵ be an invariant Borel probability measure for X^ϵ. If μ is a weak* limit point of μ^ϵ as $\epsilon \downarrow 0$, then μ is F-invariant.*

We define these weak* limit points as *natural F-invariant measures*. An immediate corollary of Theorems 1 and 2 is the following.

Corollary 1. *If F admits an attractor, then F has a natural F-invariant measure.*

In the next two sections we investigate, under an additional set of assumptions, the supports of these natural invariant measures.

4. Absorption preserving chain attractors. To understand where the dynamics of X^ϵ eventually settles when $\epsilon > 0$ is small, we define in this section absorption preserving chain attractors. These attractors generalize the notion of attractor as introduced by Ruelle in [16], which are obtained for dynamical systems under arbitrarily small perturbations in case the absorbing π-system consists solely of a compact state space. Absorption preserving chain attractors also are a generalization of the extinction preserving chain attractors, defined in [10] for unstructured populations and extended in [7] to the case of structured populations. Extinction preserving chain attractors for a community-dynamical system under arbitrarily small perturbations are equal to absorption preserving chain attractors in case the absorbing π-system is taken to be the community state space together with the collection of all extinction sets for the populations. The derivation below is an adaptation of the derivation of extinction preserving chain attractors as presented in section 3 of [10] to discrete-time dynamical systems and absorbing π-systems. We take the state space M to be compact, e.g., by restricting ourselves to a compact global attractor.

Definition 5. *Let $\epsilon \geq 0$. A sequence $x_0, \ldots, x_n$ of elements in M such that $d(x_{t+1}, F(x_t)) \leq \epsilon$ and $x_{t+1} \in \mathcal{A}^*(F(x_t))$ for all $t \in \{0, \ldots, n-1\}$ is called an* absorption preserving ϵ-pseudo-orbit *(or* ap ϵ-pseudo-orbit*) of F.*

An ap ϵ-pseudo-orbit $x_0, \ldots, x_n$ is said to have length n and to go from x_0 to x_n. Two ap ϵ-pseudo-orbits $x_0, \ldots, x_n$ and $y_0, \ldots, y_m$ of lengths n and m, respectively, and with $d(y_0, F(x_n)) \leq \epsilon$ and $y_0 \in \mathcal{A}^*(F(x_n))$ by concatenation can be combined into the ap ϵ-pseudo-orbit $x_0, \ldots, x_n, y_0, \ldots, y_m$ of length $n + m + 1$ going from x_0 to y_m. The notion of an absorption preserving ϵ-pseudo-orbit reflects the character of irreversibility attached to absorption. In addition, we define *ap-chain recurrency* as follows.

Definition 6. *A point x is* ap-chain recurrent *if for every $\epsilon > 0$ and every $n > 0$ there is an ap ϵ-pseudo-orbit of length $\geq n$ going from x to x. The set of ap-chain recurrent points is called the* ap-chain recurrent set.

Using ap-pseudo-orbits, we introduce a partial ordering on M and a corresponding equivalence relation on M.

Definition 7. *For $x, y \in M$ we define $x \succcurlyeq_{ap} y$ ("x ap-chains to y") if for every $\epsilon > 0$ there exists an ap ϵ-pseudo-orbit going from x to y.*

The relation $\succcurlyeq_{ap}$ (to be called *ap-chaining*) is a preorder on M. Ap-chaining is not necessarily a closed relation: if $(x_i)_{i\geq 0}$ and $(y_i)_{i\geq 0}$ are two sequences in M that converge to x and y, respectively, and are such that $x_i \succcurlyeq_{ap} y_i$ for all i, then not necessarily $x \succcurlyeq_{ap} y$; e.g., take x and y such that $\mathcal{A}^*(x) \bigcap \mathcal{A}^*(y) = \emptyset$.

Definition 8. *For elements $x, y \in M$, define $x \sim_{ap} y$ if $x \succcurlyeq_{ap} y$ and $y \succcurlyeq_{ap} x$.*

Since $\succcurlyeq_{ap}$ is a preorder, $\sim_{ap}$ is an equivalence relation on M, to be called *mutual ap-chaining*. The equivalence class of x under $\sim_{ap}$ is denoted as $[x]_{ap}$, and $\mathcal{M}_{ap}$ denotes the set of equivalence classes in M under $\sim_{ap}$. The expression $x \sim_{ap} y$ ("x and y ap-chain to each other") implies that both x and y belong to $\mathcal{A}^*(x) \cap \mathcal{A}^*(y)$, and consequently $\mathcal{A}^*(x) = \mathcal{A}^*(y)$. In the sense indicated above, the relation $\sim_{ap}$ may not be closed.

Definition 9. *$[x]_{ap}$ is called an* ap-basic class *if x is ap-chain recurrent.*

Three equivalent statements can be made for ap-basic classes, as follows.

Proposition 1. *The following three statements are equivalent:*

1. *$[x]_{ap}$ is an ap-basic class;*
2. *x is a fixed point, or $[x]_{ap}$ contains more than one point;*
3. *for all $t \geq 0$, $F^t([x]_{ap}) = [x]_{ap}$.*

Definition 10. *For elements $[x]_{ap}, [y]_{ap} \in \mathcal{M}_{ap}$ the relation $\geqslant_{ap}$ is defined by $[x]_{ap} \geqslant_{ap} [y]_{ap}$ if $x \succcurlyeq_{ap} y$.*

The relation $\geqslant_{ap}$ (to be called *ap-connecting*) is a partial ordering on the set of equivalence classes of $\sim_{ap}$. By means of $\geqslant_{ap}$ we define ap-chain attractors for dynamical systems with absorbing π-systems.

Definition 11. *$[x]_{ap}$ is an* ap-chain attractor *for $(F, \mathcal{A})$ if it is a minimal element of the ordering $\geqslant_{ap}$.*

An ap-chain attractor is an ap-basic class, and, by Proposition 1, contains the ω-limit sets of all its elements. Existence of ap-chain attractors follows the same line of reasoning that guarantees the existence of attractors as presented by Ruelle in [16], which uses Zorn's lemma and the fact that M is compact. In particular an ap-chain attractor will be closed.

Definition 12. *For an ap-chain attractor $[x]_{ap}$, the collection of points $\{y \in M : y \succcurlyeq_{ap} x\}$ is called the* basin of ap-chain attraction of $[x]_{ap}$.

The basin of ap-chain attraction for an ap-chain attractor $[x]_{ap}$ is not empty, since $x \in [x]_{ap}$. Furthermore, each element in M belongs to the basin of ap-chain attraction of at least one ap-chain attractor.

5. Random perturbations and ap-chain attractors.

In this section we assume that $F : M \to M$ is dissipative. We show that when $\epsilon > 0$ is sufficiently small, X^ϵ spends most of its time near the ap-chain attractors of F. To accomplish this goal, we need to place the following two additional hypotheses on X^ϵ:

H2. For each $\epsilon > 0$ there exists a $\delta = \delta(\epsilon) \in (0, \epsilon)$ such that $P_x^\epsilon(N_{ap}(y, \gamma)) > 0$ for all $x, y \in M$ and $\gamma > 0$ satisfying $N_{ap}(y, \gamma) \subset N_{ap}(F(x), \delta)$.

H3. For each $\epsilon > 0$, if $P_x^\epsilon(K) > 0$ for a closed set $K \subset M$, then there exists a neighborhood U (depending on x and ϵ) of x such that $\inf_{y\in U} P_y^\epsilon(K) > 0$.

H2 ensures that "the noise extends locally in all directions that respect the absorbing π-system." H3 is a crude continuity-like condition on the supports of the random perturbations. Even in the case $\mathcal{A} = \{M\}$, these additional assumptions are weaker than those used by Ruelle [16] and Kifer [12]. The additional assumptions are immediately satisfied, for instance, for ecological or replicator equations with demographic stochasticity as described in sections 2.1 and 2.2. In the case of models X^ϵ with k age classes and demographic stochasticity as described in section 2.3, these additional assumptions are satisfied by replacing the map F with F^k, and the states X_t^ϵ with X_{kt}^ϵ.

Theorem 3. *Let $F : M \to M$ be dissipative, $\mathcal{A}$ be an absorbing π-system for F, and X^ϵ be a Markov chain which satisfies H1–H3. If $x \in M$ does not lie in an ap-chain attractor for $(F, \mathcal{A})$, then there exists a neighborhood U of x such that*

$$P(X^\epsilon \text{ enters } U \text{ infinitely often}) = 0 \tag{2}$$

whenever $\epsilon > 0$ is sufficiently small.

In principle, the proof of the theorem adapts the proof of Kifer [12, Thm. 4.5] to the absorption preserving case. However, due to our formulation of H2 (i.e., the support of P_x^ϵ includes $N_{ap}(F(x), \delta)$ instead of including $F(N_{ap}(x, \delta))$), our proof is shorter and more direct.

Proof. Assume that $x \in M$ does not lie in an ap-chain attractor. We will show that there exist points $y \in M$ and $\epsilon > 0$ such that x ap-chains to y but there exists no ap ϵ-pseudo-orbit from any element in $N_{ap}(y, \epsilon)$ to any element in $N_{ap}(x, \epsilon)$. Since x does not lie in an ap-chain attractor, there exists a point $y \in M$ such that x ap-chains to y but y does not ap-chain to x. Choose $\epsilon > 0$ sufficiently small such that there exists no ap 2ϵ-pseudo-orbit from y to x. We proceed in two cases. First, suppose that $\mathcal{A}^*(F(y)) = \mathcal{A}^*(y)$. Then continuity implies that there exists $\eta > 0$ such that $F(N_{ap}(y, \eta)) \subset N_{ap}(F(y), \epsilon)$. Since there are no ap 2ϵ-pseudo-orbits from $F(y)$ to x, there are no ap ϵ-pseudo-orbits from any element in $N_{ap}(y, \eta)$ to any element in $N_{ap}(x, \epsilon)$. Replacing ϵ with $\min\{\eta, \epsilon\}$ completes this case. For the second case, suppose that $\mathcal{A}^*(F(y)) \subsetneq \mathcal{A}^*(y)$. Since x ap-chains to y, we have $\mathcal{A}^*(y) \subseteq \mathcal{A}^*(x)$ and $x \notin \mathcal{A}^*(F(y))$. By H1, there is no ap ϵ-pseudoorbit from $F(y)$ to x. Since x ap-chains to $F(y)$, replacing y with $F(y)$ completes this case.

Let $\delta = \delta(\epsilon) > 0$ be as given by H2. Let $x_0 = x, x_1, \ldots, x_n = y$ be an ap δ-pseudo-orbit from x to y. H2 and H3 allow us to find a neighborhood U of x, an $\alpha > 0$, and a $\gamma \in (0, \delta)$ such that

$$P_z^\epsilon(N_{ap}(x_1, \gamma)) \geq \alpha$$

for all $z \in U$ and such that

$$P_z^\epsilon(N_{ap}(x_{i+1}, \gamma)) \geq \alpha$$

for all $z \in N_{ap}(x_i, \gamma)$ and $i = 1, \ldots, n-1$. Define $U(0) = U$ and $U(i) = N_{ap}(x_i, \gamma)$ for $i = 1, \ldots, n$. We claim that

$$P(X_n^\epsilon \in U(n) | X_0^\epsilon = z) \geq \alpha^n \quad \text{for all } z \in U(0). \tag{3}$$

To prove (3) notice that

$$P(X_n^\epsilon \in U(n)|X_0^\epsilon = z) = \int \cdots \int P^\epsilon_{z_{n-1}}(U(n))\ dP^\epsilon_{z_{n-2}}(z_{n-1}) \dots dP^\epsilon_z(z_1).$$

Since $P^\epsilon_z(U(n)) \geq \alpha 1_{U(n-1)}(z)$ for all $z \in M$, we get that

$$\begin{aligned} P(X_n^\epsilon \in U(n)|X_0^\epsilon = z) &\geq \alpha \int \cdots \int 1_{U(n-1)}(z_{n-1})\ dP^\epsilon_{z_{n-2}}(z_{n-1}) \dots dP^\epsilon_z(z_1) \\ &= \alpha \int \cdots \int P^\epsilon_{z_{n-2}}(U(n-1))\ dP^\epsilon_{z_{n-3}}(z_{n-2}) \dots dP^\epsilon_z(z_1). \end{aligned}$$

Similarly, applying the estimates $P^\epsilon_z(U(n-i)) \geq \alpha 1_{U(n-i-1)}(z)$ for $i = 1, \dots, n-1$ yields (3). The following standard result in Markov chain theory (see, e.g., [4, Chap. 5, Thm. 2.3]) applied to $X = X^\epsilon$, $B = U(n)$, and $C = U(0)$ yields that

$$P(X^\epsilon \text{ enters } U(0) \text{ infinitely often}) = 0.$$

Theorem 4. *Let X be a Markov chain, and suppose that*

$$P\left(\bigcup_{m=t+1}^{\infty} \{X_m \in B\} \Big| X_t\right) \geq \beta > 0 \text{ on } \{X_t \in C\}.$$

Then

$$P(\{X \text{ enters } C \text{ infinitely often}\} \setminus \{X \text{ enters } B \text{ infinitely often}\}) = 0.$$

In the words of Chung [4, p. 256]: "The intuitive meaning of the preceding theorem has been given by Doeblin as follows: if the chance of a pedestrian's getting run over is greater than $\beta > 0$ each time he crosses a certain street, then he will not be crossing it indefinitely (since he will be killed first)." In our case "the certain street" is the set $U(0)$ and "getting run over" is X^ϵ following an ap ϵ-pseudoorbit from $U(0)$ to $U(n)$. Any time that X^ϵ enters $U(n)$ it will never return to $U(0)$, as there are no ap ϵ-pseudo-orbits back from $U(n)$ to $U(0)$. ■

Recall from Theorem 1 that limit points μ^ϵ of the empirical measures $\nu_t^\epsilon = \frac{1}{t}\sum_{i=1}^t \delta_{X_i^\epsilon}$ as $t \to \infty$ are invariant measures for X^ϵ. Moreover, by Theorem 2, limit points of these μ^ϵ as $\epsilon \downarrow 0$ are natural invariant measures of F. Theorem 3 yields the following corollary, which implies that the natural F-invariant measures are concentrated on F's ap-chain attractors. Consequently, X^ϵ spends most of its time near F's ap-chain attractors when $\epsilon > 0$ is sufficiently small.

Corollary 2. *Let $F : M \to M$ be dissipative, and let $\mathcal{A}$ be an absorbing π-system for F. Let $\{X^\epsilon, \epsilon > 0\}$ be a collection of Markov chains that satisfy H1–H3, and let μ^ϵ denote an invariant Borel probability measure for X^ϵ. All weak* limit points of μ^ϵ as $\epsilon \downarrow 0$ are supported by the ap-chain attractors for $(F, \mathcal{A})$.*

Proof. For all $\epsilon > 0$, let μ^ϵ be an invariant Borel probability measure for X^ϵ. Let $x \in M$ not lie in any ap-chain attractor. Then there exists an open neighborhood U of x such that

$$\mu^\epsilon(U) = 0$$

for $\epsilon > 0$ sufficiently small. Since U is an open set, $\mu(U) = 0$ for any weak* limit point μ of μ^ϵ as $\epsilon \downarrow 0$. Since x was an arbitrary point in the complement of the ap-chain attractors, it follows that μ is supported by the ap-chain attractors. ■

Often, attractors for F break up into a finite number of ap-chain attractors for $(F, \mathcal{A})$. When this occurs, the following result shows that for sufficiently small $\epsilon > 0$, X^ϵ eventually remains in an ap ϵ-neighborhood of one of these ap-chain attractors. Moreover, for $\epsilon > 0$ sufficiently small, X^ϵ reaches an ap-chain attractor with positive probability only if X_0^ϵ ap-chains to an ap-chain attractor.

Theorem 5. *Let $F : M \to M$ be dissipative, $\mathcal{A}$ be an absorbing π-system for F, $\{X^\epsilon, \epsilon > 0\}$ be a collection of Markov chains that satisfy H1–H3, and $X_0^\epsilon = x_0 \in M$ for all $\epsilon > 0$. Assume that there exist k compact subsets $K_1, \ldots, K_k$ of M such that*

- *for each K_i there exists an $A \in \mathcal{A}$ such that $K_i \subset A$ and K_i is an attractor for $F|A$;*
- *$\cup_i K_i$ contains all the ap-chain attractors for $(F, \mathcal{A})$.*

Then for any $\gamma > 0$ and $\epsilon > 0$ sufficiently small

$$P(\text{there exist } s \text{ and } i \text{ such that for all } t \geq s : X_t^\epsilon \in N_{ap}(K_i, \gamma)) = 1. \tag{4}$$

Moreover, if x_0 ap-chains to a point in K_i, then for any $\gamma > 0$ and $\epsilon > 0$ sufficiently small

$$P(\text{there exists an } s \text{ such that for all } t \geq s : X_t^\epsilon \in N_{ap}(K_i, \gamma)) > 0. \tag{5}$$

If x_0 does not ap-chain to any point in K_i, then there is a $\gamma > 0$ such that

$$P(\text{there exists an } s \text{ such that for all } t \geq s : X_t^\epsilon \in N_{ap}(K_i, \gamma)) = 0 \tag{6}$$

whenever $\epsilon > 0$ is sufficiently small.

Proof. The proof of the theorem relies on the following two lemmas.

Lemma 1. *Let $F : M \to M$ be a dissipative map and $\mathcal{A}$ be an absorbing π-system for F. For $A \in \mathcal{A}$, let B be an attractor for $F|A$. Then for every $\gamma > 0$ there exist $\epsilon_0 > 0$ and $\beta \in (0, \gamma)$ such that there is no ap ϵ_0-pseudo-orbit from x to y for all $x \in N_{ap}(B, \beta)$ and $y \in M \setminus N_{ap}(B, \gamma)$.*

Proof. Suppose that the conclusion of the proposition does not hold. Then there is a $\gamma > 0$ for which there do not exist an $\epsilon_0 > 0$ and $\beta \in (0, \gamma)$ such that there is no ap ϵ_0-pseudo-orbit from x to y for all $x \in N_{ap}(B, \beta)$ and $y \in M \setminus N_{ap}(B, \gamma)$. I.e., in that case, for every $\epsilon_0 > 0$ and every $\beta \in (0, \gamma)$ there is an ap ϵ_0-pseudo-orbit from an $x \in N_{ap}(B, \beta)$ to a $y \in M \setminus N_{ap}(B, \gamma)$. Consequently, by letting ϵ_0 and β become infinitesimally small (but positive), it follows that there exists an $x \in B$ that ap-chains to an element $y \in \overline{M \setminus N_{ap}(B, \gamma)}$. Necessarily, any compact neighborhood U of B in A contains part of the set $\{z \in M : x \succcurlyeq_{ap} z\}$ by which x ap-chains to y, and therefore $\omega(U) \supsetneq B$. This contradicts that B is an attractor for $F|A$. ■

Lemma 2. *Let $F : M \to M$ be dissipative with global attractor B, $\mathcal{A}$ be an absorbing π-system for F, $\{X^\epsilon, \epsilon > 0\}$ be a collection of Markov chains that satisfy H1, and $S \subset M$ be a compact set such that $\text{supp}(\{X_0^\epsilon\}) = S$ for all $\epsilon > 0$. Then for every $\gamma > 0$ there exist an $\epsilon_1 > 0$ and $n \geq 0$ such that for all $\epsilon \in (0, \epsilon_1]$*

$$P(\text{for all } t \geq n : X_t^\epsilon \in N(B, \gamma)) = 1.$$

Proof. It suffices to prove the lemma for the case in which $\mathcal{A} = \{M\}$. Since F is dissipative, there exists a compact attractor B whose basin of attraction is M. Let S be as in the lemma. Let $\gamma > 0$ be given, and for this γ choose $\epsilon_0 > 0$, $\beta \in (0, \gamma)$ as given by Lemma 1. Lemma 1 implies that there does not exist an ap ϵ_0-pseudo-orbit from $N_{ap}(B, \beta)$ to $M \setminus N_{ap}(B, \gamma)$. Since S is compact and lies in the basin of attraction of B, there exists an n such that $F^n(S) \subset N(B, \beta/2)$. Compactness of S and continuity of F imply that there exists an $\epsilon_1 \in (0, \min\{\epsilon_0, \frac{\beta}{2}\})$ such that every ap ϵ_1-pseudo-orbit $x_0, \ldots, x_n$ with $x_0 \in S$ satisfies $x_n \in N(B, \beta)$. For any Markov chain X^ϵ that satisfies H1 it follows that if $\epsilon \leq \epsilon_1$, then for any $t \geq 0$ the sequence $X_0^\epsilon, \ldots, X_t^\epsilon$ is an ap ϵ_1-pseudo-orbit. Since $X_0^\epsilon \in S$, we get with probability one that $X_n^\epsilon \in N(B, \beta)$ and $X_t^\epsilon \in N(B, \gamma)$ for $t \geq n$ whenever $\epsilon \in (0, \epsilon_1]$. ∎

Let B be the global attractor of F, and let $\gamma > 0$ be given. Lemma 2 implies that there exist an $\epsilon_0 > 0$ and $n \geq 0$ such that $P(\text{for all } t \geq n : X_t^\epsilon \in K) = 1$ whenever $\epsilon \in (0, \epsilon_0]$. Applying Lemma 1 to each of the K_i implies that there exist $\epsilon_1 \in (0, \epsilon_0)$ and $\beta_1 \in (0, \gamma)$ such that for all $1 \leq i \leq k$ there is no ap ϵ_1-pseudo-orbit from any point in $N_{ap}(K_i, \beta_1)$ to any point in $M \setminus N_{ap}(K_i, \gamma)$. Next, we wish to extend the "absorption preserving" neighborhoods $N_{ap}(K_i, \beta_1)$ of each of the K_i's to full neighborhoods of the K_i's such that X^ϵ cannot enter them too often without getting stuck. Let $\delta(\epsilon)$ be as given by H2. Since K_i is F-invariant, for every $\epsilon > 0$ there exists $\eta = \eta(\epsilon) \in (0, \beta_1)$ such that $F(x) \in N(K_i, \delta(\epsilon)/2)$ whenever $x \in N(K_i, \eta)$. H1–H3 imply that $\inf_{x \in N(K_i, \eta)} P_x^\epsilon(A_i) > 0$. Applying Theorem 4 with $C = N(K_i, \eta) \setminus A_i$ and $B = A_i$ implies that for all $\epsilon > 0$ and each $1 \leq i \leq k$

$$P(X_t^\epsilon \in N(K_i, \eta) \setminus A_i \text{ infinitely often}) = 0. \tag{7}$$

Let $\text{int}(N(K_i, \eta))$ denote the interior of $N(K_i, \eta)$. Since η is strictly positive, every $x \in K \setminus \bigcup_{i=1}^k \text{int}(N(K_i, \eta))$ does not lie in an ap-chain attractor. Theorem 3 implies that for every $x \in K \setminus \bigcup_{i=1}^k \text{int}(N(K_i, \eta))$ there exists a neighborhood U_x of x and $\epsilon_x > 0$ such that for all $\epsilon \in (0, \epsilon_x)$: $P(X_t^\epsilon \in U_x \text{ infinitely often}) = 0$. Compactness of $K \setminus \bigcup_{i=1}^k \text{int}(N(K_i, \eta))$ implies that $K \setminus \bigcup_{i=1}^k \text{int}(N(K_i, \eta))$ is covered by a finite number of these open neighborhoods, say $U_{x_1}, \ldots, U_{x_n}$. Let $\epsilon_2 = \min\{\epsilon_{x_1}, \ldots, \epsilon_{x_n}, \epsilon_1\}$. Since X^ϵ can enter each element of this finite collection of neighborhoods only finitely often,

$$P\left(X_t^\epsilon \in K \setminus \bigcup_{i=1}^k \text{int}(N(K_i, \eta)) \text{ infinitely often}\right) = 0 \tag{8}$$

for all $\epsilon \in (0, \epsilon_2)$.

Equations (7) and (8) imply that

$$P\left(X_t^\epsilon \in K \setminus \bigcup_{i=1}^k N_{ap}(K_i, \eta) \text{ infinitely often}\right) = 0$$

for all $\epsilon \in (0, \epsilon_2)$. Our choice of K implies that

$$P\left(X_t^\epsilon \in M \setminus \bigcup_{i=1}^k N_{ap}(K_i, \eta) \text{ infinitely often}\right) = 0$$

for all $\epsilon \in (0, \epsilon_2)$. It follows that

$$P\left(X_t^\epsilon \in \bigcup_{i=1}^{k} N_{ap}(K_i, \eta) \text{ for some } t \geq 0\right) = 1$$

for all $\epsilon \in (0, \epsilon_2)$. Since there are no ap ϵ_2-pseudo-orbits from $N_{ap}(K_i, \eta)$ to $M \setminus N_{ap}(K_i, \gamma)$ for $i = 1, \ldots, k$, expression (4) follows.

To prove (5), assume that x_0 does ap-chain to $y \in K_i$. Then the arguments leading to (3) imply that $P(X_t^\epsilon$ enters $N_{ap}(K_i, \eta)$ for some $t) > 0$. Since there are no ap ϵ_2-pseudo-orbits from $N_{ap}(K_i, \eta)$ to $M \setminus N_{ap}(K_i, \gamma)$, expression (5) follows.

To prove (6), assume that x_0 does not ap-chain to any point in K_i. For every point $y \in K_i$, there exists an $\epsilon_y > 0$ such that there are no ap $2\epsilon_y$-pseudo-orbits from x_0 to y. Hence, there are no ap ϵ_y-pseudo-orbits from x_0 to any point in $N_{ap}(y, \epsilon_y)$. Since K_i is compact, there are $y_1, \ldots, y_n$ such that $K_i \subset \cup_j N_{ap}(y_j, \epsilon_{y_j})$. Let $\epsilon_3 = \min_j\{\epsilon_{y_j}\}$. Then there are no ap $\frac{\epsilon_3}{2}$-pseudo-orbits from x_0 to any point in $N_{ap}(K_i, \frac{\epsilon_3}{2})$. Hence (6) holds for $0 < \epsilon < \frac{\epsilon_3}{2}$ and $\gamma > 0$ sufficiently small. ■

6. Applications. In this section, we apply the results from the previous sections to models of competing species and replicator dynamics.

6.1. Ecological drift for competing species. When the ecological outcome of competing species is determined by stochastic forces, ecological drift is said to occur. Here we illustrate two scenarios, ecologically equivalent competing species and intermingled basins of competitive exclusion, for which ecological drift occurs. Let x_1 and x_2 be the densities of two competing species. The competing species are ecologically equivalent if the per capita growth of each species is of the form $f(x_1 + x_2)$, in which case the competitive dynamics are given by $F(x_1, x_2) = (x_1 f(x_1 + x_2),\ x_2 f(x_1 + x_2))$. The following proposition proves that, under suitable assumptions on f and the noise, ecological drift occurs in the sense that competitive exclusion of either species occurs with positive probability for all positive initial conditions. Figure 1 illustrates how the probability of exclusion can depend on initial conditions for the map $F(x_1, x_2) = (3.9x_1(1 - x_1 - x_2),\ 3.9x_2(1 - x_1 - x_2))$ with additive uniformly distributed noise on $[-0.01, 0.01]$.

Proposition 2. *Let $m > 0$ (i.e., the maximum density supported by the population) and $M = \{(x_1, x_2) \in \mathbf{R}_+^2 : x_1 + x_2 \leq m\}$. Let $f : [0, m] \to \mathbf{R}_+$ be a continuous decreasing function with $f(0) > 1$, $f(m) < 1$, $f > 0$ on $[0, m)$, and $xf(x) \leq m$ for all $x \in [0, m]$. Let $F(x_1, x_2) = (x_1 f(x_1 + x_2),\ x_2 f(x_1 + x_2))$, $\mathcal{A} = \{M, \{0\} \times [0, m], [0, m] \times \{0\}, (0, 0)\}$, and $\{X^\epsilon, \epsilon > 0\}$ be a collection of Markov chains satisfying H1–H3. Then there exists $a > 0$ such that for all $x_1 x_2 > 0$ and $\epsilon > 0$ sufficiently small*

$$P_x^\epsilon(X_t^\epsilon \in \{0\} \times [a, m] \cup [a, m] \times \{0\} \text{ for } t \text{ sufficiently large}) = 1,$$

$$P_x^\epsilon(X_t^\epsilon \in \{0\} \times [a, m] \text{ for } t \text{ sufficiently large}) > 0,$$

and

$$P_x^\epsilon(X_t^\epsilon \in [a, m] \times \{0\} \text{ for } t \text{ sufficiently large}) > 0.$$

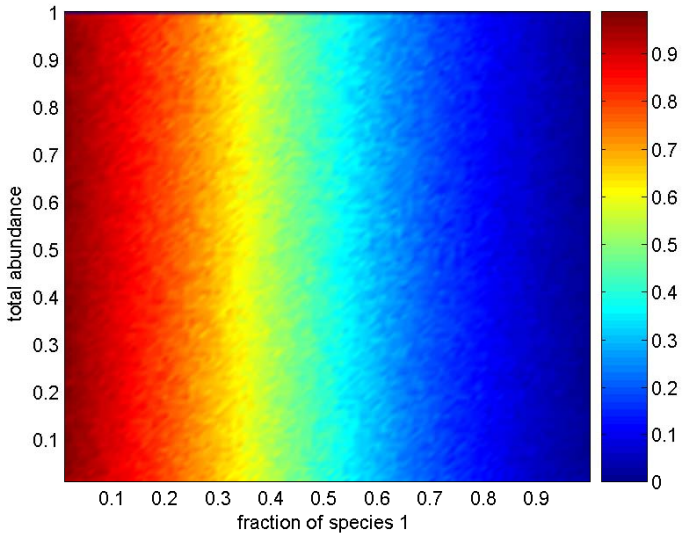


Figure 1. *The probability of extinction of species* 1 *as a function of the initial conditions. These probabilities were computed for the map* $F(x_1, x_2) = (3.9x_1(1-x_1-x_2), 3.9x_2(1-x_1-x_2))$ *with additive uniformly distributed noise on* $[-0.01, 0.01]$.

Proof. Consider the change of variables $u = x_1 + x_2$ and $v = \frac{x_2}{x_1+x_2}$. In this coordinate system, F is given by $G(u, v) = (uf(u), v)$, provided that $u > 0$, and M is given by $[0, m] \times [0, 1]$. Our assumptions of f imply that there is an attractor $A_1 \subset (0, m)$ for $u \mapsto uf(u)$ whose basin of attraction is $(0, m)$. Hence, $A = A_1 \times [0, 1]$ is an attractor for G with a basin of attraction including $(0, m) \times [0, 1]$. The basin of attraction does not include $\{m\} \times [0, 1]$ whenever $f(m) = 0$. Since $G_2(u, v) = v$, all points in $(0, m) \times [0, 1]$ ap-chain to points in $A_1 \times \{0\}$ and to points in $A_1 \times \{1\}$. Applying Theorem 5 completes the proof of the proposition. ■

Another way that ecological drift can occur is when the competing species exhibit *intermingled basins of competitive exclusion*. In the words of [8], this occurs when for "almost all initial conditions one of the two species dies out. But the survivor is unpredictable: The basins of the two chaotic one species attractors are everywhere dense." Hofbauer et al. [8] have proven the existence of an intermingled basin for a class of maps. For systems of this type, Theorem 5 implies the following proposition about ecological drift. Figure 2 illustrates a potential intermingled basin of competitive exclusion for $F(x_1, x_2) = (3.9x_1(1-x_1-x_2)(1+0.1x_2 \sin(2\pi(x_1+x_2))/(x_1+x_2)),\ 3.9x_2(1-x_1-x_2)(1-0.1x_1 \sin(2\pi(x_1+x_2))/(x_1+x_2)))$ and the effect of additive uniformly distributed noise.

Proposition 3. *Let* $m > 0$ *(i.e., the maximum density supported by the population) and* $M = \{(x_1, x_2) \in \mathbf{R}^2_+ : x_1 + x_2 \leq m\}$, $F : M \to M$ *be a continuous map of the form* $F(x_1, x_2) = (x_1 f_1(x_1, x_2),\ x_2 f(x_1, x_2))$, $\mathcal{A} = \{M, \{0\} \times [0, m], [0, m] \times \{0\}, (0, 0)\}$, *and* $\{X^\epsilon, \epsilon > 0\}$ *be a collection of Markov chains satisfying* H1–H3. *Assume that* F *has ap-chain attractors* $A_1 \subset (0, m) \times \{0\}$ *and* $A_2 \subset \{0\} \times (0, m)$ *such that each basin of attraction of* A_i *is dense in* M, A_1*'s basin includes* $(0, m) \times \{0\}$, *and* A_2*'s basin includes* $\{0\} \times (0, m)$. *Then*

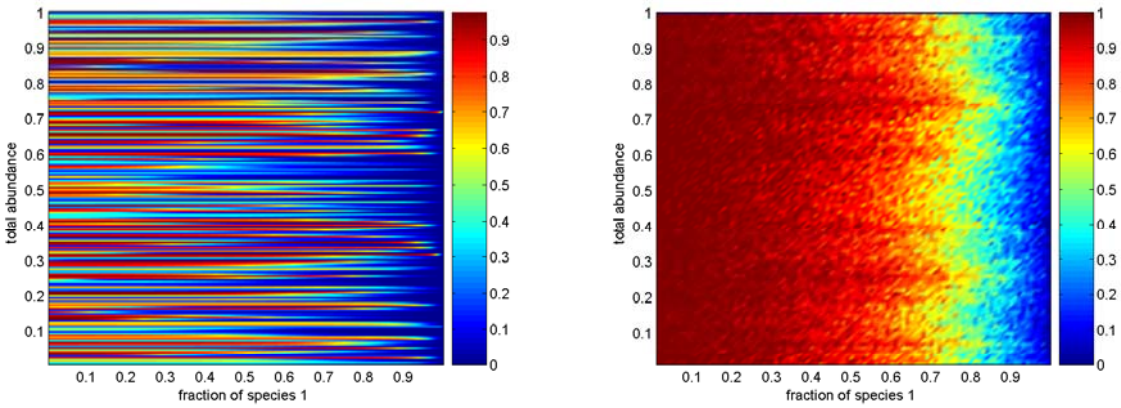


Figure 2. *Intermingled basins of competitive exclusion and the probability of extinction of species* 2 *as a function of the initial conditions. In the left panel, each initial condition for* $F(x_1, x_2) = (3.9x_1(1-x_1-x_2)(1+0.1x_2\sin(2\pi(x_1+x_2))/(x_1+x_2)), 3.9x_2(1-x_1-x_2)(1-0.1x_1\sin(2\pi(x_1+x_2))/(x_1+x_2)))$ *was iterated* 1,000 *time steps, and the final density of species* 2 *is plotted. Warmer colors correspond to higher densities, and cooler colors correspond to lower densities. In the right panel, the map was perturbed by additive uniformly distributed noise on* $[-0.01, 0.01]$*, and the probability of extinction of species* 1 *was computed for a grid of initial conditions.*

there exists $a > 0$ *such that for all* $x_1x_2 > 0$ *and* $\epsilon > 0$ *sufficiently small*

$$P_x^\epsilon(X_t^\epsilon \in \{0\} \times [a, m] \cup [a, m] \times \{0\} \text{ for } t \text{ sufficiently large}) = 1,$$

$$P_x^\epsilon(X_t^\epsilon \in \{0\} \times [a, m] \text{ for } t \text{ sufficiently large}) > 0,$$

and

$$P_x^\epsilon(X_t^\epsilon \in [a, m] \times \{0\} \text{ for } t \text{ sufficiently large}) > 0.$$

6.2. Asymmetric games. In evolution, players in different positions may engage in asymmetric conflicts. For the case of two types of players and two strategies, one can assume without loss of generality that the payoff matrix for one type of player, say females, is of the form

$$\begin{bmatrix} 0 & a_{12} \\ a_{21} & 0 \end{bmatrix},$$

while the payoff matrix for the other type of player, say males, is of the form

$$\begin{bmatrix} 0 & b_{12} \\ b_{21} & 0 \end{bmatrix}.$$

For this payoff structure, the game dynamics is given by

$$\begin{aligned} \frac{dx}{dt} &= x(1-x)(a_{12} - (a_{12}+a_{21})y), \\ \frac{dy}{dt} &= y(1-y)(b_{12} - (b_{12}+b_{21})x), \end{aligned} \tag{9}$$

where x and y are frequencies of strategy 1 for males and females, respectively. Let $M = [0,1] \times [0,1]$, let $(\phi_t)_{t \geq 0}$ denote the flow of the game dynamics, and let $F = \phi_h$ for some $h > 0$. The dynamics of this game is well studied (see Hofbauer and Sigmund [9]) and fall generically into the following four cases:

I. $a_{12}a_{21} < 0$: One of the two strategies of the females dominates the other: x converges monotonically to 0 or 1.

II. $b_{12}b_{21} < 0$: One of the two strategies of the males dominates the other: y monotonically converges to 0 or 1.

III. $a_{12}a_{21} > 0$, $b_{12}b_{21} > 0$, and $a_{12}b_{12} > 0$: There is a unique interior equilibrium

$$(x^*, y^*) = \left(\frac{b_{12}}{b_{12} + b_{21}}, \frac{a_{12}}{a_{12} + a_{21}} \right),$$

which is a saddle, and almost every initial condition converges to opposite corners of M.

IV. $a_{12}a_{21} > 0$, $b_{12}b_{21} > 0$, and $a_{12}b_{12} < 0$: The unique equilibrium (x^*, y^*) is neutrally stable, and all orbits in M are periodic orbits surrounding (x^*, y^*).

Consider random perturbations X^ϵ of F corresponding to demographic stochasticity with or without environmental stochasticity. X^ϵ satisfies H1–H3 with respect to the absorbing π-system generated by M, $\{(0,y) : y \in [0,1]\}$, $\{(1,y) : y \in [0,1]\}$, $\{(x,0) : x \in [0,1]\}$, and $\{(x,1) : x \in [0,1]\}$. For cases I–IV, the only ap-chain attractors are the boundary equilibria $(0,1)$, $(1,0)$, $(0,0)$, and $(1,1)$. Hence, Theorem 5 implies that, with probability 1, X^ϵ converges to one of these boundary equilibria in finite time. Moreover, if $X_0^\epsilon \in (0,1) \times (0,1)$, then for cases I–III, X^ϵ converges with positive probability only to a subset of the boundary equilibria, while for case IV, X^ϵ converges to any boundary equilibrium with positive probability (Figure 3).

7. Discussion. Our analysis studies the effect of localized noise on discrete-time dynamical systems with absorbing sets. Noise is represented by a discrete time Markov chain that in each time step acts on the deterministic image of a state. Certain regions of the state space are assumed to be absorbing, in that the system cannot leave (either deterministically or by a random perturbation) these regions once it has entered such a region, e.g., extinction boundaries in the absence of immigration or mutations. Thus, we assume that noise respects the absorbing sets (H1). We prove that if an unperturbed system has an attractor, then for sufficiently small perturbations the perturbed system has invariant probability measures that describe the asymptotic behavior of the system. Letting the size of the random perturbations go to zero, natural invariant measures for the unperturbed system are obtained as limit points of the invariant measures for the perturbed system. Provided that the random perturbations are sufficiently small, the asymptotic dynamics of the perturbed system is well described by these natural invariant measures.

Adding two more assumptions to our formalism—namely, within each absorbing set noise may locally perturb the dynamics into all admissible directions (H2), and nonzero noise is locally sustained (H3)—allows us to derive that the natural invariant measures of an unperturbed system are supported by the ap-chain attractors. First we show that if a state does not belong to an ap-chain attractor, then under sufficiently small perturbations, this state has a neighborhood which the system cannot enter infinitely often. Next we prove that, given

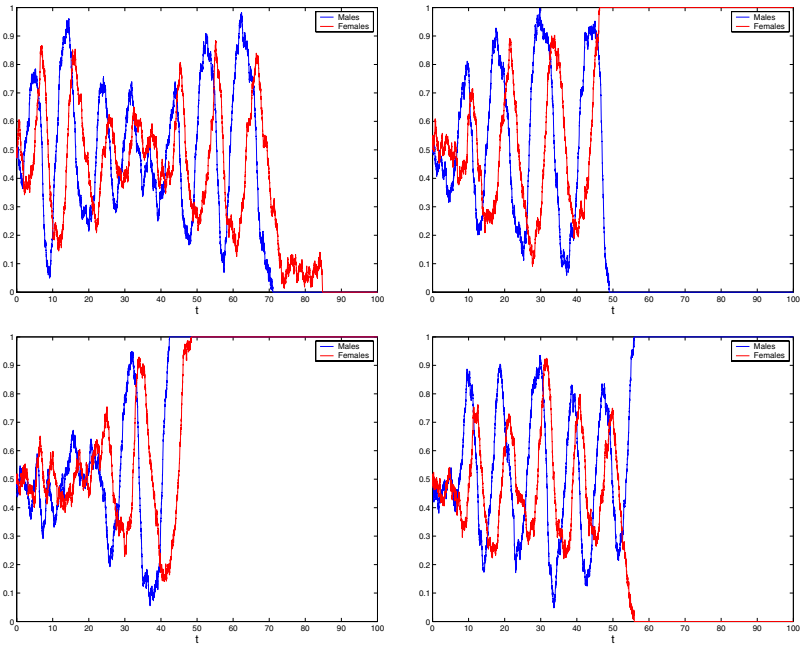


Figure 3. *Four realizations of* (9) *are shown with* $a_{12} = a_{21} = 2$ *and* $b_{12} = b_{21} = -2$ *with demographic stochasticity of size* $\epsilon = 0.01$.

that the ap-chain attractors are contained in the attractors of the unperturbed system, there exists an attractor such that for any neighborhood of that attractor and sufficiently small perturbations, the perturbed system will be restricted to that neighborhood within a finite amount of time. In addition, if small noise is capable of bringing a system into an attractor, then there is a positive probability that the randomly perturbed system will be restricted to an arbitrary neighborhood of that attractor within a finite amount of time.

Although our statements may sound intuitively clear, to our knowledge so far no mathematical proofs have been presented in the literature that support them, given our assumptions H1–H3 on the random perturbations. The papers [10] and [7] introduce the notion of extinction preserving chain attractors, but their relation to the effect of small random perturbations on the dynamics is dealt with only on the intuitive level and does not provide an analysis for the case of random perturbations. Ruelle in his paper [16] derives a result similar to ours if the full state space is the only absorbing set; namely, the randomly perturbed system statistically spends most of its time in a neighborhood of the chain attractors. However, even when the full state space is the only absorbing set, his work differs from ours in that our assumptions on the random perturbations are weaker [16, p. 145]. Moreover, our assumptions H2 and H3 on supports make the proof of our Theorem 3 and its Corollary 2 more straightforward than the proof of the similar statements in Ruelle's setting, as follows from comparison of our proof with, e.g., a proof given by Kifer (the proof of Theorem 4.5 in [12]).

Since small random perturbations are omnipresent in reality, our work explains their effects on community dynamics. A combination of small random perturbations with larger perturbations might lead to a better understanding of fluctuations of population densities due

to the presence of multiple ap-chain attractors. Although sufficiently small perturbations, e.g., those due to environmental stochasticity, are likely to bring the system close to a specific ap-chain attractor, irregular appearances of sufficiently large perturbations might cause the system to change its basin of ap-chain attraction. Large random perturbations are not, however, covered by our framework, and their inclusion in the theory is a possible direction for further research.

REFERENCES

[1] O. N. Bjärnstad and B. T. Grenfell, *Noisy clockwork: Time series analysis of population fluctuations in animals*, Science, 293 (2001), pp. 638–643.

[2] H. Caswell, *Matrix Population Models: Construction, Analysis, and Interpretation*, 2nd ed., Sinauer Associates, Sunderland, MA, 2000.

[3] N. V. Davydova, O. Diekmann, and S. A. van Gils, *Year class coexistence or competitive exclusion for strict biennials?*, J. Math. Biol., 46 (2003), pp. 95–131.

[4] R. Durrett, *Probability: Theory and Examples*, 2nd ed., Duxbury Press, New York, 1996.

[5] S. Elaydi, *An Introduction to Difference Equations*, 3rd ed., Springer, New York, 2005.

[6] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1998.

[7] M. Gyllenberg, F. J. A. Jacobs, and J. A. J. Metz, *On the concept of attractor for community-dynamical processes* II: *The case of structured populations*, J. Math. Biol., 47 (2003), pp. 235–248.

[8] F. Hofbauer, J. Hofbauer, P. Raith, and T. Steinberger, *Intermingled basins in a two species system*, J. Math. Biol., 49 (2004), pp. 293–309.

[9] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, UK, 1998.

[10] F. J. A. Jacobs and J. A. J. Metz, *On the concept of attractor for community-dynamical processes* I: *The case of unstructured populations*, J. Math. Biol., 47 (2003), pp. 222–234.

[11] R. Z. Khasminskii, *The averaging principle for parabolic and elliptical differential equations and Markov processes with small diffusion*, Theory Probab. Appl., 8 (1963), pp. 1–21.

[12] Y. Kifer, *Random Perturbations of Dynamical Systems*, Birkhäuser, New York, 1988.

[13] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise. Stochastic Aspects of Dynamics*, 2nd ed., Appl. Math. Sci. 97, Springer-Verlag, New York, 1994.

[14] M. G. Neubert and H. Caswell, *Density-dependent vital rates and their population dynamic consequences*, J. Math. Biol., 41 (2000), pp. 103–121.

[15] L. S. Pontryagin, A. A. Andronov, and A. A. Vitt, *On statistical considerations of dynamical systems*, J. Exper. Theoret. Phys., 3 (1933), pp. 165–180.

[16] D. Ruelle, *Small random perturbations of dynamical systems and the definition of attractors*, Comm. Math. Phys., 82 (1981), pp. 137–151.

[17] W. M. Schaffer, S. Ellner, and M. Kot, *Effects of noise on some dynamical models in ecology*, J. Math. Biol., 24 (1986), pp. 479–523.

[18] A. D. Wentzell and M. I. Freidlin, *Small random perturbations of dynamical systems*, Uspehi Mat. Nauk, 25 (1970), pp. 3–55.

[19] A. Wikan and M. Einar, *Overcompensatory recruitment and generation delay in discrete age-structured population models*, J. Math. Biol., 35 (1996), pp. 195–239.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 2, pp. 313–363

The Stability of a Stripe for the Gierer–Meinhardt Model and the Effect of Saturation*

Theodore Kolokolnikov†, Wentao Sun‡, Michael Ward§, and Juncheng Wei¶

Abstract. The stability of two different types of stripe solutions that occur for two different forms of the Gierer–Meinhardt (GM) activator-inhibitor model is analyzed in a rectangular domain. For the basic GM model with exponent set (p, q, r, s), representing the powers of certain nonlinear terms in the reaction kinetics, a homoclinic stripe is constructed whereby the activator concentration localizes along the midline of the rectangular domain. In the semistrong regime, characterized by a global variation of the inhibitor concentration across the domain, instability bands with respect to transverse zigzag instabilities and spot-generating breakup instabilities of the homoclinic stripe are determined analytically. In the weak interaction regime, where both the inhibitor and activator concentrations are localized, the spectrum of the linearization of the homoclinic stripe is studied numerically with respect to both breakup and zigzag instabilities. For certain exponent sets near the existence threshold of this homoclinic stripe, where stripe self-replicating behavior is observed, it is shown numerically that a stripe can be stable with respect to a breakup instability but is unstable with respect to a transverse zigzag instability. The zigzag instability is found numerically to be the precursor to a space-filling curve. For a GM model in the semistrong regime that is modified to include a small level of saturation of the activator production, it is shown that a homoclinic stripe solution still exists but, in contrast to the unsaturated GM model, can be stable with respect to breakup instabilities. For larger levels of the saturation, the homoclinic stripe ceases to exist and is replaced by a mesa-stripe, which is composed of two front-back heteroclinic transition layers joined together by an asymptotically flat plateau. In the near-shadow limit of an asymptotically large inhibitor diffusivity, and in a rectangular domain, it is shown analytically that a mesa-stripe is stable to spot-generating breakup instabilities, but can be unstable to either slow zigzag or breather-type instabilities. Finally, the asymptotic and numerical stability results for both homoclinic and mesa-stripes are favorably compared with results obtained from full numerical simulations of the GM model.

Key words. homoclinic stripe, mesa-stripe, zigzag instability, nonlocal eigenvalue problem

AMS subject classifications. 35B25, 35K57, 35P30, 37N25

DOI. 10.1137/050635080

1. Introduction.

In a two-dimensional domain, intricate spatially localized patterns consisting of either spots, stripes, mixed spot-stripe patterns, or space-filling curves have been

*Received by the editors July 4, 2005; accepted for publication (in revised form) by B. Sandstede January 5, 2006; published electronically June 30, 2006.
http://www.siam.org/journals/siads/5-2/63508.html

†Department of Mathematics, Chase Building, Dalhousie University, Halifax, NS, Canada B3H 3J5 (tkolokol@mathstat.dal.edu). This author was supported by a PGS-B graduate scholarship from NSERC Canada.
‡Department of Mathematics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2 and School of Mathematics and Systems Science, Shandong University, China (wtsun@pims.math.ca).
§Department of Mathematics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2 (ward@math.ubc.ca). This author was supported by a grant from NSERC.
¶Department of Mathematics, Chinese University of Hong Kong, New Territories, Hong Kong (wei@math.cuhk.edu.hk). The work of this author was supported by an earmarked grant from the RGC of Hong Kong.

observed in numerical simulations of certain classes of singularly perturbed reaction-diffusion systems. For activator-inhibitor systems, such as the well-known Gierer–Meinhardt (GM) model of biological morphogenesis, spot and stripe patterns are ubiquitous (cf. [6], [13], [21], [22], [46], [11, Chapter 5]). For the Gray–Scott (GS) model of theoretical chemistry, an even greater diversity of spatio-temporal patterns occurs including temporally oscillating spots, spot-replication behavior, spatio-temporal chaos of spot patterns, and labyrinthine patterns of stripes (cf. [32], [26], [27], [44]). In other settings, localized stripe and spot patterns occur for certain hybrid chemotaxis reaction-diffusion models of bacterial pattern formation (cf. [36], [45], [24, Chapter 5]) and of fish skin patterns on growing domains (cf. [18], [30], [31]). Localized patterns also arise in the reaction-diffusion modeling of vegetation patterns in arid environments (cf. [9], [12], [19]).

Most of the previous analyzes of the stability of stripe patterns have been based on a weakly nonlinear theory, where the solution is assumed to be close to some spatially uniform state across the cross section of the stripe. However, for singularly perturbed two-component reaction-diffusion systems, this assumption of near-uniform spatial dependence in the stripe cross section is generally not valid. For such systems in a two-dimensional domain, there are two main types of spatially heterogeneous stripes: homoclinic stripes and mesa-stripes.

A homoclinic stripe results when either one or both of the two solution components becomes localized, or concentrates, on a planar curve in the domain. There are two distinct parameter regimes for homoclinic stripes. The semistrong interaction regime occurs when the ratio of the two diffusivities is asymptotically large, so that only one of the two solution components (the fast component) is localized to form a stripe. In this case, the cross section of the stripe is closely approximated by a homoclinic orbit of a certain scalar ODE problem for the fast subsystem. Although the fast solution component has a negligible interaction with the boundaries of the domain, the slow (global) solution component has a significant interaction with the domain boundary and, possibly, with adjacent stripes. In contrast, in the weak interaction regime, where both diffusivities are asymptotically small and of the same order, both solution components are localized to form a stripe. In this case, the cross section of the stripe is closely approximated by a homoclinic orbit of a coupled ODE system for the two fast components. The term "weak" interaction here refers to the negligible interaction of these two fast components with the boundaries of the domain and with any adjacent stripes. Homoclinic stripes in both the weak and semi strong regimes arise in the basic GM model (cf. [4]) and in certain parameter ranges of the GS model (cf. [15], [16], [23]).

A different type of stripe can occur for bistable singularly perturbed reaction-diffusion systems such as the Fitzhugh–Nagumo model. For such systems, the stripe cross section typically consists of two transition layers, each closely approximated by a heteroclinic solution of the equilibrium problem, which are joined together by an asymptotically flat plateau region. We refer to such a stripe solution as a mesa-stripe.[1] The stability of such stripe solutions was analyzed rigorously in [34] and [35] for certain classes of bistable reaction-diffusion systems. The analyses of [34] and [35] were based on the SLEP (singular limit eigenvalue problem) method developed in [28] and [29] to analyze the stability of mesa-pulses in one spatial dimension. For a generalized Fitzhugh–Nagumo model, the stability and dynamics of mesa-stripes

[1]Mesa means table in Spanish. This term was suggested by P. Fife as referenced in [17].

were analyzed in [7] by a contour dynamics approach. The existence and stability of mesa-stripes for a Brusselator model was studied in [17]. Mesa-stripe solutions can also occur for a modification of the GM model where activator saturation effects are included.

There are two main types of linearized instabilities associated with a homoclinic stripe; varicose (breakup) instabilities and transverse (zigzag) instabilities. For a varicose instability of a homoclinic stripe, the eigenfunction for the perturbation of any fast component is an even function across the stripe cross section and has a normal-mode modulation tangential to the stripe. This has the effect of inducing a ripple on the amplitude of any fast component along the length of the stripe. Since an instability of this type typically leads to the disintegration of the stripe into a sequence of spots, we refer to it here as a breakup instability. Alternatively, for a zigzag instability of a homoclinic stripe, the eigenfunction for the perturbation of any fast component is an odd function across the stripe cross section and has a normal-mode modulation in the direction tangent to the stripe. This has the effect of inducing a ripple on the location of the centerline of the stripe.

For the GM and GS models in the semistrong regime, a homoclinic stripe is typically unstable to a spot-generating breakup instability (cf. [4], [15], [16], [23]). For the GS model in the weak interaction regime, a homoclinic stripe can be destabilized solely by a transverse zigzag instability, which seems to be the precursor to a complicated space-filling curve (cf. [16]). Zigzag and breakup instabilities have also been studied in [8] for a reaction-diffusion system on an unbounded domain with piecewise linear kinetics. Breakup instabilities of localized rings of bacteria, leading to spot formation, have been observed in numerical simulations of certain hybrid chemotaxis reaction-diffusion systems (cf. [36], [45], [24, Chapter 5]).

Based on numerical experiments and a Turing-type linearized stability analysis, the inclusion of saturation effects into the basic GM model is, qualitatively, a well-known way to obtain stable stripe patterns (cf. [13], [21], [22]). Breakup instabilities do not, in general, occur for a mesa-stripe solution, and the instability of this solution typically occurs by two types of transverse instabilities that develop over long time-scales; zigzag instabilities where the edges of the mesa-stripe are in phase, and breather instabilities where these edges are 90° out of phase.

The primary goal of this paper is to give detailed analytical and numerical results for breakup and transverse instability bands of homoclinic and mesa-stripes for the basic GM model, and for the modified GM model where saturation effects are included. Our analysis addresses several key qualitative features observed in numerical simulations of stripe behavior. It determines explicit parameter ranges in these GM models where a homoclinic stripe exists and is stable with respect to spot-generating breakup instabilities. It suggests a common qualitative mechanism responsible for the disappearance of a breakup instability band. It identifies parameter ranges where a transverse instability is the dominant instability mechanism and where labyrinthine patterns are likely to occur.

The basic GM model, where saturation effects are neglected, can be written in the dimensionless form (cf. [10])

$$a_t = \varepsilon_0^2 \Delta a - a + \frac{a^p}{h^q}\,, \quad \tau h_t = D\Delta h - h + \frac{a^r}{\varepsilon_0 h^s}\,, \quad X = (X_1, X_2) \in \Omega\,, \ \ t > 0\,,$$
$$\partial_n a = \partial_n h = 0\,, \quad x \in \partial\Omega\,, \tag{1.1}$$

where the exponent set (p, q, r, s) is assumed to satisfy

$$p > 1\,, \qquad q > 0\,, \qquad r > 1\,, \qquad s \geq 0\,, \qquad \text{with} \qquad \zeta \equiv \frac{qr}{(p-1)} - (s+1) > 0. \tag{1.2}$$

The classical GM model corresponds to the exponent set $(2, 1, 2, 0)$. We consider (1.1) in the rectangular domain

$$\Omega : \ -1 < X_1 < 1\,, \quad 0 < X_2 < d_0\,. \tag{1.3}$$

By rescaling a and h and introducing $X = x/l$, where $l = 1/\sqrt{D}$, (1.1) can be recast into the equivalent form

$$\begin{aligned} &a_t = \varepsilon^2 \Delta a - a + \frac{a^p}{h^q}\,, \quad \tau h_t = \Delta h - h + \frac{a^r}{\varepsilon h^s}\,, \quad x = (x_1, x_2) \in \Omega_l\,, \ \ t > 0\,, \\ &\partial_n a = \partial_n h = 0\,, \quad x \in \partial\Omega_l\,, \end{aligned} \tag{1.4}$$

where

$$\Omega_l : \ -l < x_1 < l\,, \quad 0 < x_2 < d\,, \qquad d \equiv d_0 l\,, \qquad \varepsilon \equiv \varepsilon_0 l\,, \qquad l \equiv 1/\sqrt{D}\,. \tag{1.5}$$

In (1.1), the semistrong regime is characterized by $\varepsilon_0 \ll 1$ and $D = O(1)$, while the weak interaction regime corresponds to $\varepsilon_0 \ll 1$ and $D = O(\varepsilon_0^2) \ll 1$. In terms of (1.4), the weak interaction regime corresponds to the limit $l \to \infty$ with $\varepsilon = \varepsilon_0 l = O(1)$. For the basic GM model (1.4), we will study the stability of a homoclinic stripe of zero curvature that is obtained when a concentrates along the midline $x_1 = 0$ of Ω_l.

Alternatively, the classical GM model (1.1), which is modified to include the effect of saturation of the activator production, can be written in the dimensionless form (cf. [13], [21], [22])

$$a_t = \varepsilon_0^2 \Delta a - a + g(a, h)\,, \qquad g(a, h) \equiv \frac{a^2}{h\,(1 + \kappa a^2)}\,, \qquad \tau h_t = D\Delta h - h + a^2\,. \tag{1.6}$$

Here $\kappa > 0$ is the saturation parameter. For the weak saturation case $\kappa = O(\varepsilon_0^2)$, (1.6) has a homoclinic stripe solution as for the basic GM model (1.1). However, for $\kappa = O(1)$, the numerical simulations of (1.6) in [13], [21], [22], and [46] have suggested the existence of a mesa-stripe solution for (1.6) that is stable to the formation of spots.

The stability of a homoclinic stripe for the basic GM model (1.4) was studied in [4]. In the semistrong interaction regime, the existence of a homoclinic stripe solution for (1.4) on the infinite strip domain $\mathbb{R}^1 \times (0, d)$ was analyzed in [4] using geometric singular perturbation techniques. In addition, by reducing the study of a certain nonlocal eigenvalue problem (NLEP) to computations involving hypergeometric functions, the stability of this stripe with respect to breakup instabilities was analyzed in [4]. In [4], explicit results for the breakup instability bands were obtained for certain exponent sets (p, q, r, s), and it was shown that a homoclinic stripe for (1.4) is stable to breakup instabilities only for asymptotically thin domains of order $d = O(\varepsilon_0)$.

For the semistrong regime of (1.4), we extend this previous work of [4] on breakup instabilities to the finite rectangular domain $[-l,l]\times[0,d]$, and to more general exponent sets (p,q,r,s). Our analysis, which is related to the rigorous analysis of the NLEP in [38] and [39], gives explicit upper and lower bounds for the breakup instability bands for various ranges of the exponents (p,q,r,s) in (1.4). For various exponent sets and domain lengths $l=1/\sqrt{D}$, the most unstable mode from the NLEP is calculated numerically. Our main result, that breakup instabilities always occur unless the domain is $O(\varepsilon_0)$ thin, is given in Proposition 2.3 below. It is an extension of Theorem 4.5 of [4]. Full numerical computations of (1.4) are performed to validate the asymptotic theory.

In the semistrong regime of (1.4), we also analyze the transverse zigzag instabilities of a homoclinic stripe by calculating an explicit formula for the small eigenvalue of order $\lambda=O(\varepsilon_0^2)$ in the spectrum of the linearization. In Principal Result 2.4, we show that there are no unstable zigzag modes in the semistrong regime for exponent sets that satisfy $\gamma\equiv q/(p-1)\le 1$. This range includes the classical GM model, where $(p,q,r,s)=(2,1,2,0)$. For $\gamma>1$, we show that an unstable band of zigzag modes with wave number $m=O(1)$ as $\varepsilon_0\to 0$ exists only when $l=1/\sqrt{D}$ exceeds some critical threshold l_z, which depends on γ. This threshold is calculated numerically. However, since the time-scale for the development of zigzag instabilities in the semistrong regime is $O(\varepsilon_0^{-2})$, they are dominated in this regime by spot-generating breakup instabilities that occur on the more rapid $O(1)$ time-scale.

In the weak interaction regime of (1.1), where $D=\varepsilon_0^2 D_0$ with $D_0=O(1)$, it was proved in [4] that there is a minimum value of D_0, labeled by D_{0c}, for which a homoclinic stripe solution exists. As shown in [14] (see also [4] and Remark 6.2 of [25]), this critical value D_{0c} is a saddle-node value of a bifurcation diagram of the norm $|a|_2$ versus D_0. In the one-dimensional case and for values of D_0 slightly below D_{0c}, a self-replication behavior is observed, whereby a localized initial pulse undergoes a repeated edge-splitting process due to ghost effects of the saddle-node bifurcation point (cf. [4], [14], [25], [26]). A similar stripe self-replication behavior was observed in [4] in the two-dimensional case. For the classical GM model with exponent set $(2,1,2,0)$, and for D_0 slightly above D_{0c}, the full numerical computations in [4] suggested that a stripe can be stable with respect to breakup instabilities for *any* domain width.

We extend this previous work by giving a detailed numerical study of the spectrum of the homoclinic pulse in the weak interaction regime of (1.1). In terms of this eigenvalue problem, we numerically calculate the breakup and zigzag instability bands associated with a homoclinic stripe as a function of D_0 for various exponent sets (p,q,r,s). For certain exponent sets, which include the classical GM set $(2,1,2,0)$, we show numerically that there exists a value D_{0b} of D_0, with $D_{0b}>D_{0c}$, for which the instability thresholds of the breakup instability band coalesce. Therefore, there is no spot-generating breakup instability band on the range $D_{0c}<D_0<D_{0b}$. For the classical GM set this range is $7.17<D_0<8.06$. However, for each of these exponent sets where the breakup band disappears, we show numerically that there is a nontrivial zigzag instability band on this range of D_0. Therefore, for these exponent sets, a homoclinic stripe is destabilized solely by a zigzag instability when D_0 is sufficiently close to D_{0c}. This zigzag instability is found to be the precursor to a space-filling curve. Full numerical simulations of (1.1) are done to confirm the spectral results. Secondary instabilities, such as self-replicating spots, which arise after a spot-generating breakup instability of the stripe, are also illustrated.

For the modified classical GM model (1.6) with saturation parameter $\kappa = O(1)$ and for $\varepsilon \to 0$, we construct a mesa-stripe equilibrium solution for (1.6) in the near-shadow limit $D \gg 1$, where $D = \mathcal{D}/\varepsilon_0$ with $\mathcal{D} = O(1)$. The stability of the mesa-stripe is then studied analytically. In contrast to the spectrum for homoclinic stripe solutions of the basic GM model (1.1), the spectrum of the linearization of an equilibrium mesa-stripe solution for (1.6) contains only the small eigenvalues of order $O(\varepsilon_0^2)$ that correspond to transverse instabilities. For $\mathcal{D}$ above some threshold, we show that the mesa-stripe is stable for all domain widths and for all transverse wave numbers m. However, as $\mathcal{D}$ is decreased below some critical value, it is shown analytically that an unstable zigzag instability band emerges at some critical wave number $m = O(1)$. Upon further decreasing $\mathcal{D}$ below some additional threshold, an additional breather instability is triggered. The critical values of $\mathcal{D}$ and the transverse wave number m at the onset of the zigzag and breather instability are determined explicitly as a function of κ. The asymptotic theory is confirmed with full numerical simulations of (1.6). Our case-study analysis of mesa-stripe stability for (1.6) with $\kappa > 0$ and $\kappa = O(1)$ extends the previous studies of [34] and [35] for generalized Fitzhugh–Nagumo models by providing explicit instability thresholds for zigzag and breather instabilities.

Finally, we consider the modified classical GM model (1.6) with an asymptotically small saturation $\kappa = O(\varepsilon_0^2)$ in the semistrong regime. The study of stripe stability for this problem provides a bridge between the analysis of breakup instabilities of a homoclinic stripe for (1.4) and the analysis of transverse instabilities of a mesa-stripe for (1.6) with $\kappa = O(1)$. We show that for any $k > 0$, where $\kappa = \varepsilon_0^2 k$, (1.6) admits a homoclinic stripe solution for any $D > 0$ with $D = O(1)$. An NLEP governing breakup instabilities of this homoclinic stripe is then derived analytically. From a numerical computation of the spectrum of this NLEP it is shown that the boundaries of the breakup instability band coalesce at some $k = k_d$, so that this instability band disappears for all $k > k_d$. The reason for the disappearance of this breakup instability band for k sufficiently large is related to the ghost effect of a nearby heteroclinic solution, which has the effect of "fattening" the cross section of the homoclinic stripe. This suggests that such "fat" homoclinic stripes can share some of the same qualitative stability properties as mesa-stripe solutions. We remark that a qualitatively similar "fattening" of the homoclinic stripe is also likely responsible for the disappearance of breakup instability bands for certain exponent sets of the basic GM model (1.1) in the weak interaction regime sufficiently close to the existence threshold D_{0c} of the homoclinic stripe. The "fattening" of this other homoclinic stripe is related to the existence of nearby multibump homoclinic solutions that must necessarily exist close to the existence threshold D_{0c} governing self-replication (cf. [14]).

The outline of this paper is as follows. In sections 2.1 and 2.2 we analyze breakup and zigzag instabilities, respectively, for homoclinic stripe solutions of (1.4) in the semistrong interaction regime. In section 2.3 we show that "fat" homoclinic solutions to (1.6) with $\kappa = O(\varepsilon_0^2)$ can be stable to breakup instabilities. In section 3 we numerically study breakup and zigzag instabilities for (1.4) in the weak interaction regime, where both solution components are localized. Finally, in sections 4 and 5 we analyze the existence and linearized stability, respectively, of a mesa-stripe equilibrium solution to the GM model (1.6) when $\kappa > 0$ and $\kappa = O(1)$.

2. The GM model without saturation: Semistrong regime. In the limit $\varepsilon \to 0$ we now construct an equilibrium stripe solution to (1.4), where the stripe is centered on the midline

$x_1 = 0$ of the rectangular domain $\Omega := [-l, l] \times [0, d]$, where $d = ld_0$. Since the cross section of the stripe is a one-dimensional pulse, this solution, with a minor change in notation, was constructed asymptotically in [10] using the method of matched asymptotic expansions. The result is as follows.

Principal Result 2.1 (from [10]). *For $\varepsilon \to 0$, an equilibrium stripe solution to (1.4), labeled by $a_e(x_1)$ and $h_e(x_1)$, is given asymptotically by*

$$a_e(x_1) \sim \mathcal{H}^{\gamma} w\left(\varepsilon^{-1} x_1\right), \qquad h_e(x_1) \sim \mathcal{H} \frac{G_l(x_1)}{G_l(0)}. \tag{2.1}$$

Here $w(y)$ is the unique positive solution to

$$\begin{gathered} w'' - w + w^p = 0, \qquad -\infty < y < \infty, \qquad w \to 0 \quad \text{as} \quad |y| \to \infty, \\ w'(0) = 0, \quad w(0) > 0. \end{gathered} \tag{2.2}$$

The constants $\mathcal{H}$, γ, and $G_l(0)$ in (2.1), for which $h_e(0) = \mathcal{H}$, are defined by

$$\mathcal{H}^{\zeta} \equiv \frac{1}{b_r G_l(0)}, \qquad b_r \equiv \int_{-\infty}^{\infty} [w(y)]^r \, dy, \qquad \gamma \equiv \frac{q}{p-1}, \qquad G_l(0) = \frac{1}{2} \coth l, \tag{2.3}$$

where ζ is defined in (1.2). The Green function $G_l(x_1)$ in (2.1) satisfies

$$G_{lx_1x_1} - G_l = -\delta(x_1), \quad -l < x_1 < l, \qquad G_{lx_1}(\pm l) = 0, \qquad G_l(x_1) = \frac{\cosh(l - |x_1|)}{2 \sinh(l)}. \tag{2.4}$$

To determine the stability of the stripe solution, we introduce the perturbation

$$a = a_e + e^{\lambda t + imx_2} \phi, \qquad h = h_e + e^{\lambda t + imx_2} \eta, \qquad m = \frac{k\pi}{d}, \tag{2.5}$$

where $\phi = \phi(x_1) \ll 1$ and $\eta = \eta(x_1) \ll 1$. The relationship above between m and k results from the Neumann conditions on $x_2 = 0, d$ of $\partial\Omega$. In the analysis below we treat m as a continuous variable. The band of instability with respect to m that is determined below can be mapped to a k-band of instability using (2.5). Substituting (2.5) into (1.4), we obtain the eigenvalue problem

$$\varepsilon^2 \phi_{x_1x_1} - \phi + \frac{p a_e^{p-1}}{h_e^q} \phi - \frac{q a_e^p}{h_e^{q+1}} \eta = \left(\lambda + \varepsilon^2 m^2\right) \phi, \quad -l < x < l, \qquad \phi_x(\pm l) = 0, \tag{2.6a}$$

$$\eta_{x_1x_1} - \left(1 + \tau\lambda + m^2\right) \eta = -\frac{r a_e^{r-1}}{\varepsilon h_e^s} \phi + \frac{s a_e^r}{\varepsilon h_e^{s+1}} \eta, \quad -l < x < l, \qquad \eta_x(\pm l) = 0. \tag{2.6b}$$

There are two classes of eigenvalues and eigenfunctions of (2.6); the large eigenvalues, where $\lambda = O(1)$ as $\varepsilon \to 0$, and the small eigenvalues with $\lambda = O(\varepsilon^2)$ as $\varepsilon \to 0$. For the large

eigenvalues, which determine the stability of the stripe on an $O(1)$ time-scale, the corresponding eigenfunction has the form

$$\phi(x_1) \sim \Phi\left(\varepsilon^{-1}x_1\right), \tag{2.7}$$

where $\int_{-\infty}^{\infty} \Phi(y) w^{r-1}(y)\, dy \neq 0$. This stability problem, treated in section 2.1, involves the analysis of an NLEP. Since unstable eigenfunctions of this type lead to a disintegration of the stripe into spots, we refer to this instability as a breakup instability. Alternatively, the eigenfunction for the small eigenvalues has the form

$$\phi(x_1) \sim w'\left(\varepsilon^{-1}x_1\right) + \varepsilon\phi_1\left(\varepsilon^{-1}x_1\right) + \cdots. \tag{2.8}$$

Since the leading term in (2.8) corresponds to a translation of the spike profile w, unstable modes for this class of eigenvalues lead to zigzag instabilities. This problem is studied below in section 2.2.

2.1. Breakup instabilities: Semistrong regime. We now analyze the spectrum of (2.6) corresponding to breakup instabilities, where $\phi(x_1) \sim \Phi(y)$ with $y = x_1/\varepsilon$ and $\int_{-\infty}^{\infty} \Phi w^{r-1}\, dy \neq 0$. Since the asymptotic derivation of the NLEP for $\Phi(y)$ is similar to that given in section 2 of [38], we give only an outline of this analysis in Appendix A. The result is as follows.

Principal Result 2.2. *Let $\varepsilon \to 0$, and suppose that $\int_{-\infty}^{\infty} \Phi w^{r-1}\, dy \neq 0$. Then $\Phi(y)$ satisfies the NLEP*

(2.9a)
$$L_0\Phi - \chi_m w^p \frac{\int_{-\infty}^{\infty} w^{r-1}\Phi\, dy}{\int_{-\infty}^{\infty} w^r\, dy} = (\lambda + \varepsilon^2 m^2)\Phi \quad -\infty < y < \infty\,; \quad \Phi \to 0 \quad as \quad |y| \to \infty\,,$$

(2.9b)
$$L_0\Phi \equiv \Phi'' - \Phi + pw^{p-1}\Phi\,, \quad C_m(\lambda) \equiv \frac{1}{\chi_m(\lambda)} \equiv \frac{s}{qr} + \frac{\theta_\lambda \tanh(\theta_\lambda l)}{qr\tanh l}\,, \quad \theta_\lambda \equiv \sqrt{1 + m^2 + \tau\lambda}\,.$$

The unique positive eigenvalue ν_0 with eigenfunction Φ_{l0} of the local operator L_0 is (cf. [20], [3, Proposition 5.6])

$$\nu_0 = \frac{1}{4}(p-1)(p+3)\,, \qquad \Phi_{l0} = [w(y)]^{(p+1)/2}\,. \tag{2.10}$$

Equivalently, the eigenvalues of (2.9), with $\int_{-\infty}^{\infty} w^{r-1}\Phi\, dy \neq 0$, are the roots of $g(\lambda) = 0$ defined by

$$g(\lambda) \equiv C_m(\lambda) - f(\lambda + \varepsilon^2 m^2)\,, \qquad f(\mu) \equiv \frac{\int_{-\infty}^{\infty} w^{r-1}\psi\, dy}{\int_{-\infty}^{\infty} w^r\, dy}\,, \qquad \psi \equiv (L_0 - \mu)^{-1} w^p\,. \tag{2.11}$$

To analyze the spectrum of the NLEP (2.9) for the stripe, we modify the rigorous approach developed in [38] for a related NLEP governing the stability of a pulse solution on a finite interval. This analysis, as outlined in Appendix A, leads to the following rigorous characterization of the instability band for (2.9).

Proposition 2.3. *Let (p,q,r,s) satisfy (1.2), let $l = 1/\sqrt{D} > 0$ be fixed, and assume that $\varepsilon \ll 1$. Let m_{b-} be the root of the transcendental equation*

$$G(m) \equiv \sqrt{1+m^2}\left(\frac{\tanh\left[\sqrt{1+m^2}\,l\right]}{\tanh l}\right) = \zeta + 1\,, \qquad \zeta \equiv \frac{qr}{p-1} - (s+1) > 0\,. \tag{2.12}$$

For the near-shadow limit $l \to 0$ we get $m_{b-} \sim \sqrt{\zeta}$, and for $l \to \infty$ we obtain $m_{b-} \sim \sqrt{\zeta^2 + 2\zeta}$. In addition, let m_{b+} be given by

$$m_{b+} = \frac{\sqrt{\nu_0}}{\varepsilon} + O(1)\,, \qquad \nu_0 = \frac{1}{4}(p-1)(p+3)\,. \tag{2.13}$$

Then, when either $r = p = 2$ or $r = p+1$ with $1 < p \le 5$, there is exactly one real positive eigenvalue of (2.9) in $0 < \lambda < \nu_0 - \varepsilon^2 m^2$ for any $\tau \ge 0$ when m is inside the instability band $m_{b-} < m < m_{b+}$. For $r = 2$, this is the only eigenvalue in $Re(\lambda) > 0$. For $m > m_{b+}$, then $Re(\lambda) < 0$ for any $\tau \ge 0$. For $m = m_{b-}$, and if either $r = p = 2$, or $r = p+1$ and $1 < p \le 5$, then there is a unique real positive eigenvalue $\lambda > 0$ when $\tau > \tau_{m-}$, and $\mathrm{Re}(\lambda) \le 0$ when $0 \le \tau \le \tau_{m-}$. Here τ_{m-} is defined by

$$\tau_{m-} = \frac{2qr}{p-1}\left(\frac{1}{p-1} - \frac{1}{2r}\right)\tanh l\left[\frac{\tanh(\theta_- l)}{\theta_-} + l\,\mathrm{sech}^2(\theta_- l)\right]^{-1}\,, \qquad \theta_- \equiv \sqrt{1+m_{b-}^2}\,. \tag{2.14}$$

Finally, suppose that $0 \le m < m_{b-}$. Then, for either $r = p = 2$ or $r = p+1$ and $1 < p \le 5$, there are exactly two real unstable eigenvalues of (2.9) when τ is sufficiently large. In addition, as τ is increased from zero there is a Hopf bifurcation at some point $\tau_H = \tau_H(m)$ (possibly nonunique).

Proof. The proof of this result is given in Appendix A. ∎

To calculate an improved approximation for the upper bound m_{b+} of the instability band we write $m = m_0/\varepsilon$, for $m_0 = O(1)$, and we expand

$$\Phi = \Phi_{l0} + \varepsilon\Phi_1 + \cdots\,, \qquad m_0^2 = \nu_0 + \varepsilon m_1 + \cdots\,. \tag{2.15}$$

Substituting (2.15) into (2.9a) and using $\chi \sim \left(\frac{qr}{m}\right)\tanh l$ for $m \gg 1$, we obtain that Φ_1 satisfies

$$L_0\Phi_1 - \nu_0\Phi_1 = \frac{qr}{\sqrt{\nu_0}}w^p \tanh l\left(\frac{\int_{-\infty}^{\infty} w^{r-1}\Phi_{l0}\,dy}{\int_{-\infty}^{\infty} w^r\,dy}\right) + m_1\Phi_{l0}\,. \tag{2.16}$$

Since L_0 is self-adjoint, m_1 is determined from the solvability condition that the right-hand side of (2.16) be orthogonal to Φ_{l0}. Then, using $m_0 \sim \sqrt{\nu_0} + \frac{\varepsilon m_1}{(2\sqrt{\nu_0})}$, the upper stability bound m_{b+} is given in terms of Φ_{l0} by

$$m_{b+} \sim \frac{\sqrt{\nu_0}}{\varepsilon} - \left(\frac{qr\tanh l}{2\nu_0}\right)\beta\,, \qquad \beta \equiv \frac{\left(\int_{-\infty}^{\infty} w^{r-1}\Phi_{l0}\,dy\right)\left(\int_{-\infty}^{\infty} w^p\Phi_{l0}\,dy\right)}{\left(\int_{-\infty}^{\infty} w^r\,dy\right)\left(\int_{-\infty}^{\infty}\Phi_{l0}^2\,dy\right)}\,. \tag{2.17}$$

Table 1

The lower bound m_{b-} and the upper bound m_{b+} for the breakup instability band for different exponent sets and parameters. Here $m_{b\pm}(n)$ are the full numerical results from the NLEP (2.9), $m_{b-}(a)$ is the asymptotic result from (2.12), and $m_{b+}(a1)$ and $m_{b+}(a2)$ are the one- and two-term asymptotic results from (2.17).

(p,q,r,s)	ε	l	$m_{b-}(n)$	$m_{b-}(a)$	$m_{b+}(n)$	$m_{b+}(a1)$	$m_{b+}(a2)$
(2,1,2,0)	0.100	1.0	1.3241	1.3022	10.351	11.180	10.437
(2,1,2,0)	0.050	1.0	1.3073	1.3022	21.580	22.361	21.618
(2,1,2,0)	0.025	1.0	1.3033	1.3022	43.964	44.721	43.978
(2,1,2,0)	0.010	1.0	1.3022	1.3022	111.07	111.80	111.06
(2,1,2,0)	0.025	$1/\sqrt{10}$	1.0336	1.0332	44.425	44.721	44.423
(2,1,2,0)	0.025	$\sqrt{10}$	1.7267	1.7238	43.721	44.721	43.749
(2,1,3,0)	0.025	1.0	2.1087	2.1029	43.517	44.721	43.560
(2,2,3,3)	0.025	1.0	2.1146	2.1029	42.337	44.721	42.399
(3,2,2,0)	0.025	1.0	1.3024	1.3022	68.589	69.282	68.578
(3,2,3,1)	0.025	1.0	1.3040	1.3022	68.146	69.282	68.139
(4,2,2,0)	0.025	1.0	0.6852	0.6858	91.265	91.652	91.208

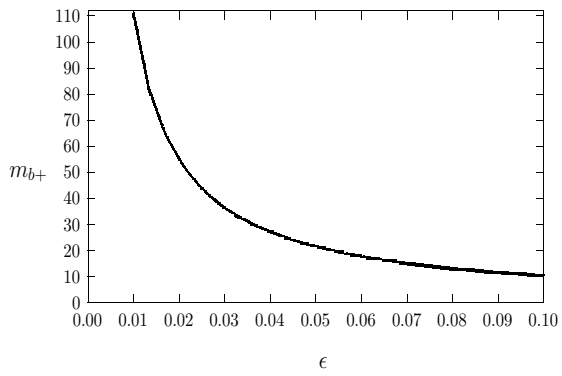
For the classical GM model with exponent set $(p,q,r,s)=(2,1,2,0)$, we use $\Phi_{l0}=\operatorname{sech}^3(y/2)$ from (2.10) to calculate β in (2.17) as $\beta = 3I_5I_7/[2I_4I_6]$, where $I_n \equiv \int_0^\infty \operatorname{sech}^n y\,dy$. By using the recursion relation $I_n=(n-2)I_{n-2}/(n-1)$, together with $I_2=1$ and $I_1=\pi/2$, we readily calculate that $\beta = \frac{\pi^2}{64}(45/16)^2$. Therefore, for the classical GM model where $(p,q,r,s)=(2,1,2,0)$ and $\nu_0=5/4$, we have for $\varepsilon\to 0$ that

$$m_{b+} \sim \frac{\sqrt{5}}{2\varepsilon} - \frac{\pi^2 \tanh l}{80}\left(\frac{45}{16}\right)^2 + \cdots . \tag{2.18}$$

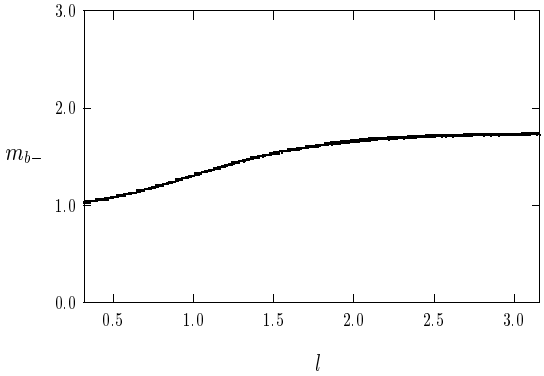
For various exponent sets (p,q,r,s) and values of l and ε, in Table 1 we compare the asymptotic results for the stability thresholds from (2.12) and (2.17) with corresponding results computed numerically from the NLEP (2.9) using finite-difference methods and a discrete eigenvalue solver from LAPACK (cf. [1]). The asymptotic values are found to be very close to the corresponding full numerical results even when $\varepsilon=0.1$. For $l=1$, in Figure 1(a) we compare the full numerical result for the upper threshold m_{b+} with the corresponding asymptotic result (2.17) for a range of values of ε. For $\varepsilon=0.025$, in Figure 1(b) we compare the full numerical result for the lower threshold m_{b-} with the corresponding asymptotic result computed from (2.12) for a range of l values.

For $l=1$, $\tau=0$, and for the classical GM exponent set $(p,q,r,s)=(2,1,2,0)$, in Figure 2(a) we plot the unique unstable eigenvalue λ in the instability band $m_{b-}<m<m_{b+}$ computed numerically from the NLEP (2.9) for three values of ε. From Proposition 2.3, this eigenvalue is necessarily real. In Figure 2(b) we show a similar plot of the unique unstable eigenvalue λ for $\varepsilon=0.025$ for several values of l. Finally, in Figure 3, we plot the numerically computed unique real positive eigenvalue for $\tau=0$ within the instability band $m_{b-}<m<m_{b+}$ for various exponent sets (p,q,r,s). Notice that some of these sets do not satisfy the conditions on the exponents in Proposition 2.3. Therefore, we expect that the conclusions in Proposition 2.3 will hold for a wider range of exponent sets than are listed there.

We now make a few remarks. In [4] breakup instabilities of a stripe for the GM model (1.1) were analyzed using geometric singular perturbation theory for the infinite strip domain

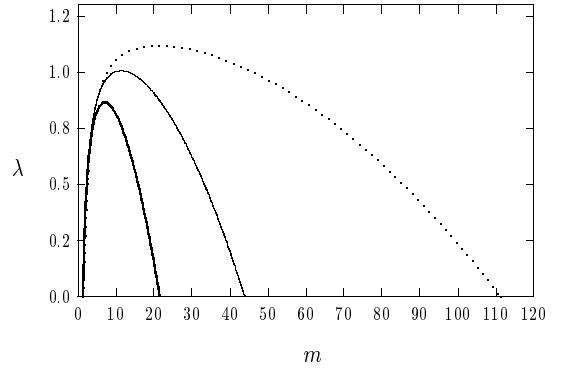


(a) m_{b+} versus ε for $l = 1$

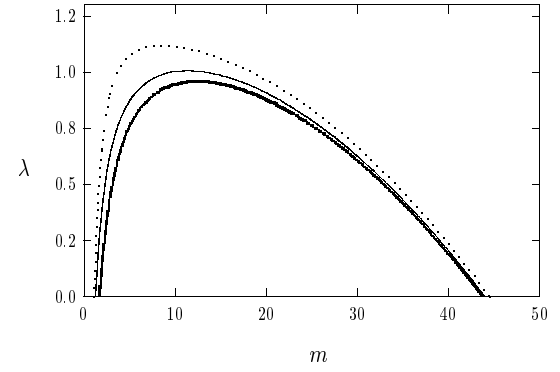


(b) m_{b-} versus l for $\varepsilon = 0.025$

Figure 1. *Instability thresholds for* $(p, q, r, s) = (2, 1, 2, 0)$. (a) *plot of the upper threshold* m_{b+} *versus* ε *for* $l = 1$. *The heavy solid curve is the full numerical result, and the dashed curve is the asymptotic result* (2.17). *They are essentially indistinguishable in this plot.* (b) *plot of the lower threshold* m_{b-} *versus* l *for* $\varepsilon = 0.025$. *The heavy solid curve is the full numerical result, and the dashed curve is the asymptotic result from* (2.12). *They are again indistinguishable.*



(a) λ versus m for $l = 1$



(b) λ versus m for $\varepsilon = 0.025$

Figure 2. *Unstable eigenvalue* λ *in the instability band when* $(p, q, r, s) = (2, 1, 2, 0)$ *and* $\tau = 0$. (a) *plot of* λ *versus* m *when* $l = 1$ *for* $\varepsilon = 0.05$ *(heavy solid curve),* $\varepsilon = 0.025$ *(solid curve), and* $\varepsilon = 0.01$ *(dashed curve).* (b) λ *versus* m *when* $\varepsilon = 0.025$ *for* $l = \sqrt{10}$ *(heavy solid curve),* $l = 1$ *(solid curve), and* $l = 1/\sqrt{10}$ *(dashed curve).*

$\mathbb{R}^1 \times [0, d_0]$. This problem is equivalent to studying (1.4) in the limit $l = 1/\sqrt{D} \to \infty$, but with $\varepsilon \ll 1$. For the classical GM model with exponent set $(p, q, r, s) = (2, 1, 2, 0)$, a detailed stability result for a stripe in such a domain is given in Theorem 4.5 of [4]. Proposition

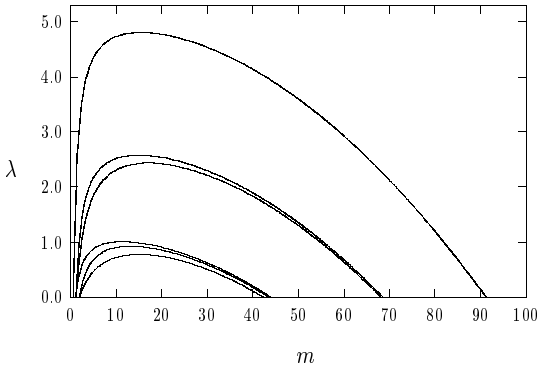


Figure 3. *The unique unstable eigenvalue in the instability band for different exponent sets* (p, q, r, s) *when* $\varepsilon = 0.025$, $l = 1$, *and* $\tau = 0$. *From top to bottom the curves are for* $(4, 2, 2, 0)$, $(3, 2, 2, 0)$, $(3, 2, 3, 1)$, $(3, 1, 2, 0)$, $(2, 1, 3, 0)$, *and* $(2, 2, 3, 3)$.

2.3, given above, generalizes this previous result to allow for a finite rectangular domain and for other exponent sets (p, q, r, s). For $l \to \infty$ and for $(p, q, r, s) = (2, 1, 2, 0)$, the resulting instability band $\sqrt{3} < m < \sqrt{5}/(2\varepsilon)$ is equivalent to that given in Theorem 4.5 of [4]. For other exponent sets (p, q, r, s) some partial stability results for the infinite strip domain were given in [4]. Specifically, the lower threshold $m_{b-} = \sqrt{\zeta^2 + 2\zeta}$ of the breakup band for the limit $l \gg 1$, where ζ is defined in (1.2), is equivalent to that given in (4.17) of Corollary 4.4 of [4]. In addition, the limiting result obtained from (2.14) for τ_{m-} in the limit $l \to \infty$ is readily seen to be equivalent to that given in the unlabeled formula above Remark 4.7 of [4]. For the exponent set $(2, 1, 2, 0)$, we calculate that $\tau_{m-} \to 6$ as $l \to \infty$. An important remark is that a stripe is stable with respect to breakup instabilities only when the inequality $m = k\pi/d < \sqrt{\nu_0}/\varepsilon$ holds for all positive integers k. By using $\varepsilon = \varepsilon_0 l$ and $d = l d_0$, this shows that a stripe for (1.1) is stable only when the domain width d_0 for (1.1) is $O(\varepsilon_0)$ thin and satisfies $d_0 < d_{0b} \equiv \pi\varepsilon_0/\sqrt{\nu_0}$. The same critical domain width was found in Corollary 5.1 of [4] for a stripe solution in the infinite strip $\mathbb{R}^1 \times [0, d]$. Therefore, the effect of lateral boundaries does not influence the critical domain width. However, both the lower threshold m_{b-} for the breakup instability band and the unstable eigenvalue within the band do depend on $l = 1/\sqrt{D}$.

We now test the theoretical predictions for breakup instabilities with full numerical simulations of (1.1) in the square domain $[-1, 1] \times [0, 2]$. The numerical computations are done using a finite-element method with sufficient resolution to accurately compute thin stripes or localized spots.

Experiment 1. We take $(p, q, r, s) = (2, 1, 2, 0)$, $\varepsilon_0 = 0.025$, $D = 1$, and $\tau = 0.1$. For $\tau = 0.1$ there are no Hopf bifurcations for modes with $m < m_{b-}$. From the solid curve in Figure 2(a) with $l = 1/\sqrt{D} = 1$ the most unstable mode is $m \approx 12$, where $\lambda \approx 1.0$. In addition, $\lambda \geq 0.9$ for $5.5 < m < 21.0$. The predicted number N of spots, which corresponds to the number of maxima of the eigenfunction $\cos(my)$ on $0 < y < d_0 = 2$, is $N = m d_0/(2\pi) = m/\pi$. The

initial condition for (1.1) is taken to be a perturbation of (2.1) of the form

$$a = \frac{3\mathcal{H}}{2}\operatorname{sech}^2\left(\frac{x_1}{2\varepsilon_0}\right)(1+\delta v)\,, \quad v = \sum_{k=1}^{20}\cos\left(\frac{k\pi x_2}{2}\right), \qquad h = \frac{\mathcal{H}\cosh[l(1-|x_1|)]}{\cosh l}\,, \tag{2.19}$$

where $\mathcal{H} = \frac{1}{3l}\tanh 1$, $l = 1$, and $\delta = 0.001$. The initial perturbation v covers the entire unstable band in Figure 2(a). In the numerical results shown in the first row of Figure 4 the initial stripe is seen to break up into seven spots on an $O(1)$ time-scale. This corresponds to $m \approx 21.5$, which is near the most unstable mode.

Experiment 2. We choose the same parameter values except that D is now decreased to $D = 0.1$, so that $l = \sqrt{10}$. From the heavy solid curve in Figure 2(b), where $l = \sqrt{10}$, the most unstable mode is $m \approx 12.3$ with $\lambda = 0.96$, and $\lambda \geq 0.9$ for $8.1 < m < 18.7$. The predicted number N of spots is $N = m/\pi \approx 4$. With the initial condition as given in (2.19) with $l = \sqrt{10}$, the resulting numerical solution of (1.1) is shown in the second row of Figure 4. The initial stripe is seen to break up into five spots (a boundary spot is counted as half of a spot).

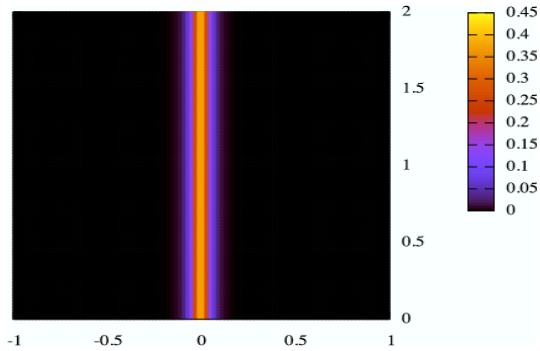
Experiment 3. Next, we take $(p, q, r, s) = (3, 2, 2, 0)$, $\varepsilon_0 = 0.025$, $D = 1$, and $\tau = 0.1$. The initial condition is

$$a = \sqrt{2}\mathcal{H}\operatorname{sech}\left(\frac{x_1}{\varepsilon_0}\right)(1+\delta v)\,, \quad v = \sum_{k=1}^{20}\cos\left(\frac{k\pi x_2}{2}\right), \qquad h = \frac{\mathcal{H}\cosh(1-|x_1|)}{\cosh 1}\,, \tag{2.20}$$

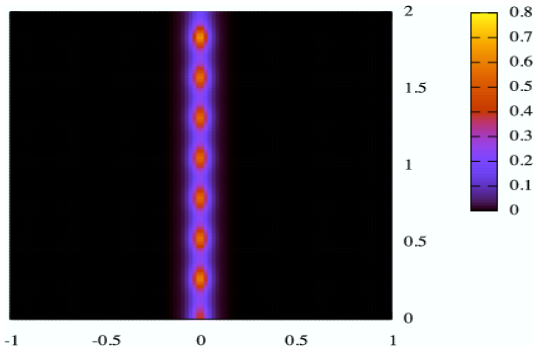
with $\mathcal{H} = \frac{1}{2}\tanh 1$ and $\delta = 0.001$. The most unstable mode from Figure 3 is $m \approx 14.9$ with $\lambda \approx 2.6$. Near this maximum, $\lambda \geq 2.3$ when $6.74 < m < 29.85$. We predict $N = m/\pi$ spots. In the bottom row of Figure 4 we show that the stripe breaks into nine spots at $t = 1.0$, which corresponds to $m \approx 28$. Since the unstable eigenvalue for the exponent set $(3, 2, 2, 0)$ is larger than for $(2, 1, 2, 0)$ (see Figure 4), the time-scale for breakup is quicker than in Experiments 1 and 2.

Although the stability theory is able to predict the initial number of spots that are generated from the break-up of a stripe, it does not account for secondary instabilities relating to a spot competition process that leads to the ultimate annihilation of some of these spots. This secondary instability, which we do not study here, is seen in Figure 4(f) for the exponent set $(3, 2, 2, 0)$ with $D = 1$. A spot competition process (not shown) also occurs in Experiment 1 on the range $t > 5$, where only two spots ultimately remain. However, for the smaller value $D = 0.1$, the five spots that are initially generated from the stripe in Experiment 2 are found to persist for $t \gg 1$. In [10] and [38] a related competition instability was studied analytically for a k-spike solution to the one-dimensional GM model. This analysis showed that for $\tau \ll 1$ there is a threshold value D_k of D for which k-spikes will be stable only when $D < D_k$. Similar thresholds occur for spots, as was shown rigorously in [42].

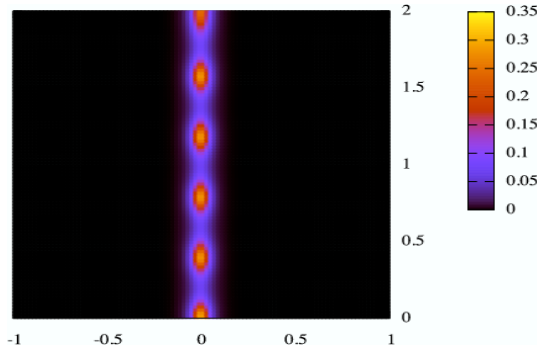
2.2. Zigzag instabilities: Semistrong regime. Next, we analyze zigzag instabilities of a stripe that are associated with unstable eigenfunctions of the form (2.8). Since this analysis is similar to that for a spike given in section 4 of [10], we only outline the key steps of the derivation in Appendix B. In this way, we obtain the following result for the critical eigenvalue $\lambda = O(\varepsilon^2)$.



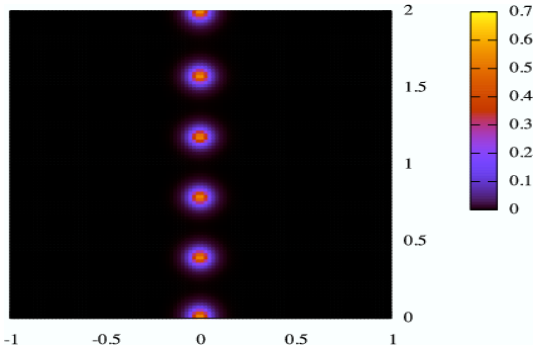
(a) Experiment 1: $t = 0$



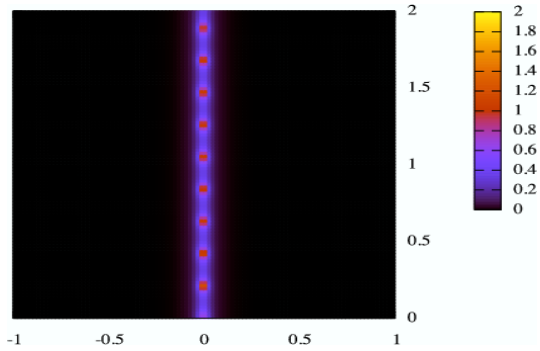
(b) Experiment 1: $t = 9$



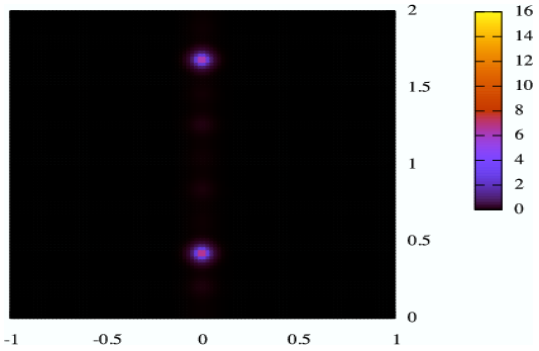
(c) Experiment 2: $t = 12$



(d) Experiment 2: $t = 20$



(e) Experiment 3: $t = 1.0$



(f) Experiment 3: $t = 2.4$

Figure 4. *Breakup instability of a stripe for* (1.1) *with* $\varepsilon_0 = 0.025$, $\tau = 0.1$, *and* $\Omega = [-1, 1] \times [0, 2]$. Top row: Experiment 1. $(p, q, r, s) = (2, 1, 2, 0)$ *and* $D = 1.0$. *The stripe initially breaks into seven spots, which corresponds to a growth rate that is near that of the most unstable mode. However, there is a secondary instability, and eventually only two spots remain (not shown).* Middle row: Experiment 2. $(p, q, r, s) = (2, 1, 2, 0)$ *and* $D = 0.1$. *The stripe breaks up into a five-spot pattern, which corresponds closely to the most unstable mode. There is no secondary instability.* Bottom row: Experiment 3. $(p, q, r, s) = (3, 2, 2, 0)$ *and* $D = 1.0$. *The stripe initially breaks up into a nine-spot pattern. There is a secondary instability, and only two spots remain. The time-scale for spot formation is faster than in Experiment* 1.

Principal Result 2.4. *For $\varepsilon \to 0$ and $\tau \ll O(\varepsilon^{-2})$, the small eigenvalue governing translational instabilities satisfies*

$$\lambda \sim \varepsilon^2 \left[2\gamma\theta_\lambda \tanh l \tanh(\theta_\lambda l) - 2\gamma - m^2\right], \qquad \theta_\lambda \equiv \sqrt{1+m^2+\tau\lambda}, \qquad \gamma \equiv \frac{q}{p-1}. \tag{2.21}$$

Suppose that $\tau \ll O(\varepsilon^{-2})$. Then, there are no unstable zigzag modes when $\gamma \le 1$, which includes the classical GM exponents $(p,q,r,s) = (2,1,2,0)$. Alternatively, when $\gamma > 1$ there is a band $m_{z-} < m < m_{z+}$ of unstable zigzag modes only when the domain half-length l exceeds some critical value l_z or, equivalently, when $D < D_z = l_z^{-2}$. For $l \to \infty$, $m_{z+} \sim 2\sqrt{\gamma(\gamma-1)}$ and $m_{z-} \sim \sqrt{\frac{8\gamma}{\gamma-1}}e^{-l} \ll 1$.

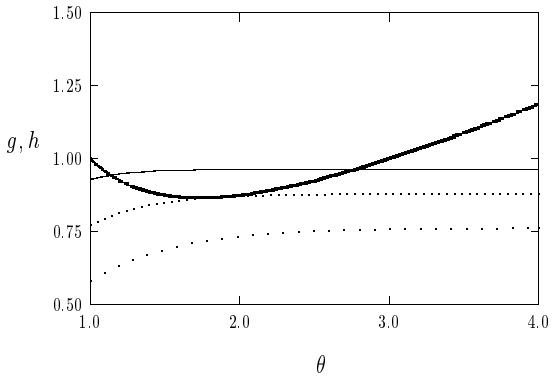
We now derive this result. In Appendix B, (2.21) for λ is obtained by modifying the analysis in section 4 of [10]. To derive the stability result in Principal Result 2.4, we assume $\tau \ll O(\varepsilon^{-2})$, so that $\theta_\lambda \sim \sqrt{1+m^2}$. Then, the stability threshold $\lambda = 0$ in (2.21) corresponds to intersection of the two functions $h(\theta)$ and $g(\theta)$ for $\theta \ge 1$, defined by

$$g(\theta) = h(\theta), \quad g(\theta) = \tanh(\theta l)\tanh l, \quad h(\theta) = \frac{\theta^2 + (2\gamma-1)}{2\gamma\theta}, \quad \theta \equiv \sqrt{1+m^2}. \tag{2.22}$$

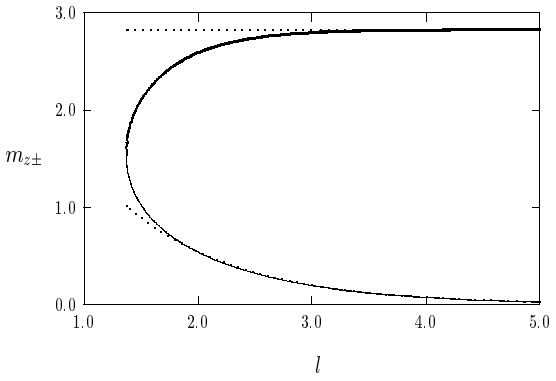
Notice that $\lambda > 0$ whenever $g(\theta) > h(\theta)$. Since $h(1) = 1$ and $g(1) < 1$, the stripe is translationally stable for $m \ge 0$ sufficiently small. Next, we calculate that $h(\theta)$ has a unique minimum at $\theta = \theta_z \equiv \sqrt{2\gamma-1}$ when $\gamma > 1/2$, where $h(\theta_z) = \gamma^{-1}\sqrt{2\gamma-1}$. First, suppose that $0 < \gamma \le 1$. Then, since $\theta_z \le 1$ when $\frac{1}{2} < \gamma \le 1$, and θ_z is undefined when $0 < \gamma < \frac{1}{2}$, we conclude that $h'(\theta) > 0$ for all $\theta \ge 1$. Since $h(1) = 1$ and $g(\theta) < 1$, it follows that there are no roots to (2.22) in $\theta > 1$ when $0 < \gamma \le 1$. Therefore, there are no zigzag instabilities when $0 < \gamma \le 1$. Alternatively, for $\gamma > 1$, we obtain that $\theta_z > 1$ and $h(\theta_z) < 1$. Since for each $\theta > 1$, $g(\theta)$ is an increasing function of l, we conclude that there exists a band of unstable zigzag modes only when $l > l_z$, where l_z is the critical value where $g(\theta)$ and $h(\theta)$ intersect tangentially. Since $l = 1/\sqrt{D}$, this implies that there is an unstable band of zigzag modes only when $D < D_z = l_z^{-2}$. We illustrate this result graphically in Figure 5(a), where we plot $h(\theta)$ for $\gamma = 2$ together with $g(\theta)$ for different values of l. For $\gamma = 2$, in Figure 5(b) we plot the upper and lower zigzag thresholds m_{z+} and m_{z-}, respectively, versus l for $l > l_z$. The dashed lines in this figure are the asymptotic approximations $m_{z+} \sim 2\sqrt{\gamma(\gamma-1)}$ and $m_{z-} \sim \sqrt{\frac{8\gamma}{\gamma-1}}e^{-l}$ for $l \gg 1$, which are readily derived from (2.22). In Figure 6 we plot the critical domain half-length l_z versus γ.

Since $m = k\pi/d$ and $d = ld_0$ with $l = 1/\sqrt{D}$, we conclude from the upper bound $m = m_{z+}$ that a stripe for (1.1) with $\gamma = q/(p-1) > 1$ is stable to zigzag instabilities only when the domain width d_0 for (1.1) satisfies $d_0 < d_{0z} \equiv \pi\sqrt{D}/m_{z+} = O(1)$. For $0 < \gamma < 1$, there are no zigzag instabilities for any domain width.

From the results here and in section 2.1, we conclude that a stripe for (1.1) in the semi-strong regime is stable with respect to both breakup and zigzag instabilities only for thin domains satisfying $d_0 < \pi\varepsilon_0/\sqrt{\nu_0}$, where ν_0, given in (2.10), is the positive eigenvalue of L_0. Since zigzag instabilities develop on a long time $O(\varepsilon^{-2})$ time-scale, they are dominated by any breakup instability that occurs. Finally, we remark that although we have presented only a formal derivation of (2.21), it can be derived rigorously by using a Lyapunov–Schmidt reduction analysis similar to that in [43].



(a) graphical determination of unstable zigzag band



(b) zigzag stability thresholds for $\gamma = 2$

Figure 5. (a) *plot of* $h(\theta)$ *for* $\gamma = 2$ *(heavy solid curve) together with* $g(\theta)$ *for* $l = 1$ *(widely spaced dots),* $l = l_z \approx 1.37$ *(dashed curve), and* $l = 2$ *(solid curve). Here* $g(\theta)$ *and* $h(\theta)$ *are defined in* (2.22)*. The unstable zigzag modes correspond to where* $g(\theta) > h(\theta)$*.* (b) *the upper and lower zigzag stability thresholds* m_{z-} *(solid curve) and* m_{z+} *(heavy solid curve) for* $\gamma = 2$ *as a function of* l *when* $l > l_z \approx 1.37$*. The band disappears when* $m \approx 1.55$*.*

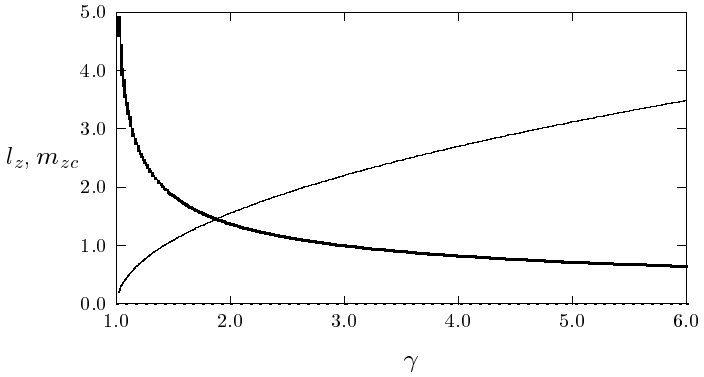


Figure 6. *Plots of the critical domain half-length* l_z *(heavy solid curve) and the mode* m *where the zigzag band disappears (solid curve) versus* $\gamma = \frac{q}{p-1}$ *for* $\gamma > 1$*. For* $l > l_z$*, there is a band of unstable zigzag modes.*

2.3. GM model with small saturation: Fat homoclinics. In the semistrong regime, where $D = O(1)$, we now show that the inclusion of a small amount of saturation in the activator production for the classical GM model can lead to the disappearance of the breakup instability band. The resulting modified classical GM model in a rectangular domain is

$$a_t = \varepsilon_0^2 \Delta a - a + \frac{a^2}{h(1+ka^2)}, \quad \tau h_t = D\Delta h - h + \frac{a^2}{\varepsilon_0}, \tag{2.23}$$
$$X = (X_1, X_2) \in \Omega = \{-1 < X_1 < 1\,, 0 < X_2 < d_0\}\,,$$

with $\partial_n a = \partial_n h = 0$ on $\partial\Omega$. We refer to (2.23) as the small saturation limit of (1.6), since if we replace a and h in (2.23) with $\varepsilon_0 a$ and $\varepsilon_0 h$, we obtain (1.6) with $\kappa = \varepsilon_0^2 k$. Thus, $k = O(1)$ in (2.23) corresponds to $\kappa = O(\varepsilon_0^2)$ in (1.6).

For $\varepsilon \to 0$ we now asymptotically construct an equilibrium homoclinic stripe solution. In the inner region, we let $y = X_1/\varepsilon$ to obtain that $h \sim \mathcal{H}$ and $a \sim \mathcal{H} w(y)$, where $w(y)$ satisfies

$$w'' + f(w) = 0\,, \quad f(w) \equiv -w + \frac{w^2}{1+bw^2}, \qquad -\infty < y < \infty\,, \qquad b = k\mathcal{H}^2. \tag{2.24}$$

For $0 \le b < \frac{1}{4}$ it follows that $w = 0$ and $w = w_\pm$, with $0 < w_- < w_+$, are the rest points of (2.24), where

$$w_\pm = \frac{1}{2}\left[1 \pm \sqrt{1-4b}\right]. \tag{2.25}$$

Both $w = 0$ and $w = w_+$ are saddle points, while w_- is a center. It is readily shown that (2.24) has a homoclinic pulse solution with $w(0) > 0$ and $w(\pm\infty) = 0$, provided that there exists a value $w_m \in (w_-, w_+)$ for which $\mathcal{F}(w_m) = 0$, where $\mathcal{F}(w) \equiv -\int_0^w f(s)\,ds$. Such a value of w_m exists for $0 < b < b_0 < \frac{1}{4}$. However, when $b = b_0$, then (2.24) has a heteroclinic solution with asymptotic end-states $w = 0$ and $w = w_+$. To determine this critical value b_0 of b we set $\mathcal{F}(w_+) = 0$. Upon integrating $f(w)$, we find that w_+ and b_0 are determined uniquely by

$$b_0 = \frac{w_+ - 1}{w_+^2}\,, \qquad 2w_+ \tan^{-1}\left(\sqrt{w_+ - 1}\right) = (w_+ + 1)\sqrt{w_+ - 1}\,. \tag{2.26}$$

By solving (2.26) numerically, we obtain

$$b_0 \approx 0.211376\,, \qquad w_+ \approx 3.295209\,. \tag{2.27}$$

In summary, (2.24) has a unique homoclinic solution with $w(0) > 0$ and $w(\pm\infty) = 0$, provided that $0 < b < b_0$.

In the outer region, we calculate a^2/ε_0 in (2.23) in terms of a Dirac mass. In this way, we obtain

$$h_{X_1X_1} - \theta_0^2 h = -\frac{\beta\mathcal{H}^2}{D}\delta(X_1)\,, \quad -1 < X_1 < 1\,, \quad h_{X_1}(\pm 1) = 0\,, \tag{2.28}$$

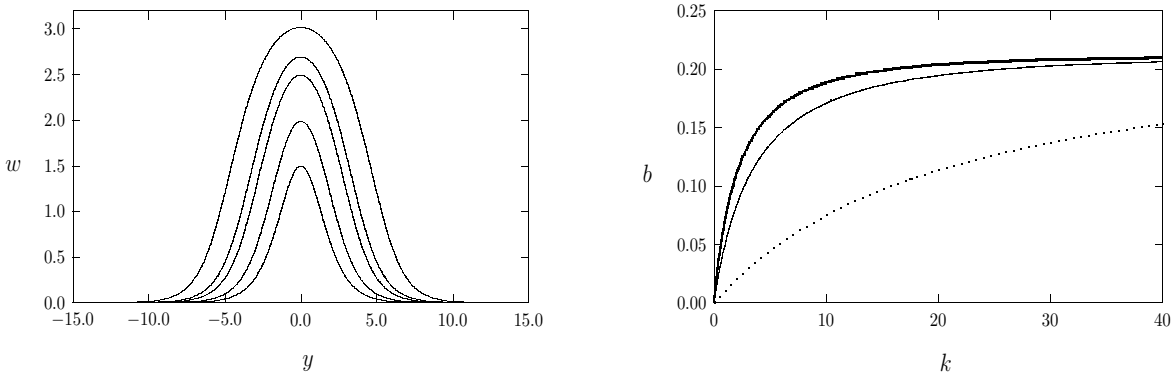
where $\beta \equiv \int_{-\infty}^{\infty} w^2\,dy$ and $\theta_0 \equiv D^{-1/2}$. We solve (2.28) in terms of a Green function, and we impose the matching condition $h(0) = \mathcal{H}$, which determines $\mathcal{H}$. Finally, we recall that $b = k\mathcal{H}^2$. This leads to the following formal result.

Principal Result 2.5. *For $\varepsilon \to 0$ and for any $k > 0$, there is a unique homoclinic stripe solution to (2.23) given by*

$$a_e(X_1) \sim \mathcal{H} w\left(\varepsilon^{-1}X_1\right), \qquad h_e(X_1) \sim \mathcal{H}\frac{G(X_1)}{G(0)}, \qquad G(X_1) \equiv \frac{\cosh\left[\theta_0(1-|X_1|)\right]}{2\theta_0\sinh\theta_0}\,. \tag{2.29}$$

Here, for a fixed value of b in $0 \le b < b_0 \approx 0.2114$, $w(y)$ is the unique positive homoclinic solution to (2.24). The saturation constant k and $h(0) \equiv \mathcal{H}$ are related to b and D by

$$\mathcal{G}(b) \equiv b\beta^2 = 4kD\tanh^2\left(1/\sqrt{D}\right), \quad \mathcal{H} = \frac{2\sqrt{D}}{\beta}\tanh\left(1/\sqrt{D}\right), \quad \beta \equiv \int_{-\infty}^{\infty} w^2\,dy\,. \tag{2.30}$$



(a) w versus y for various b

(b) b versus k

Figure 7. (a) *plots of the homoclinic solution* $w(y)$ *to* (2.24). *From top to bottom the curves are for* $b = 0.211$, $b = 0.2034$, $b = 0.195$, $b = 0.145$, *and* $b = 0.0$. (b) b *versus* k *for* $D = 10$ *(heavy solid curve),* $D = 1$ *(solid curve), and* $D = 0.1$ *(dotted curve).*

In (2.30), the integral β, which is readily computed numerically, depends only on b and satisfies $\beta(b) \to +\infty$ as $b \to b_0^-$. Since $\beta > 0$ and $d\beta/db > 0$ (see Appendix B of [41]), it follows that $\mathcal{G}(b)$ is a monotone increasing in b with $\mathcal{G}(0) = 0$ and $\mathcal{G}(b) \to +\infty$ as $b \to b_0^-$. Therefore, from (2.30), there is a unique value of k for each fixed D and b in $0 \le b < b_0$, and that $k \to \infty$ as $b \to b_0^-$. Therefore, as $k \to \infty$, we have $b \to b_0^-$, and the fattened homoclinic w approaches a mesa pattern that is comprised of a front-back heteroclinic structure connected by an asymptotically flat plateau. In Figure 7(b) we use (2.30) to plot b versus k for several values of D, showing that $k \to \infty$ as $b \to b_0^-$. In Figure 7(a) we plot the solution $w(y)$ to (2.24) for several values of b. Notice that for b slightly below b_0, $w(y)$ becomes rather fat as a result of the ghost effect of the heteroclinic connection that exists when $b = b_0$.

Next, we study breakup instabilities of the homoclinic stripe by deriving an NLEP. Since this derivation is similar to that given in Appendix A, we only highlight its key steps. We introduce a perturbation of the form (2.5). Then, in place of (A.1) and (A.2), we obtain that $\Phi(y)$, with $y = X_1/\varepsilon$, satisfies

(2.31)
$$\Phi'' - \Phi + \frac{2w}{(1+bw^2)^2}\Phi - \frac{\eta(0)w^2}{1+bw^2} = \left(\lambda + \varepsilon^2 m^2\right)\Phi\,, \quad -\infty < y < \infty\,, \quad \Phi \to 0 \quad \text{as} \quad |y| \to \infty\,.$$

Assuming that $\int_{-\infty}^{\infty} \Phi w\, dy \neq 0$, then $\eta(0)$ in (2.31) is to be calculated from

$$\eta_{X_1X_1} - \theta_\lambda^2 \eta = 0\,, \quad -1 < X_1 < 1\,, \qquad \eta_{X_1}(\pm 1) = 0\,, \qquad \theta_\lambda \equiv \sqrt{m^2 + \frac{(1+\tau\lambda)}{D}}\,, \tag{2.32a}$$

$$[\eta] = 0\,, \qquad [\eta_{X_1}] = -\frac{2\mathcal{H}}{D}\int_{-\infty}^{\infty} w\Phi\, dy\,. \tag{2.32b}$$

We solve (2.32) in terms of a Green function, which yields $\eta(0)$ in (2.31). This leads to the following NLEP.

Principal Result 2.6. *Let $\varepsilon \to 0$, and suppose $\int_{-\infty}^{\infty} \Phi w\, dy \neq 0$. Then, for $0 \leq b < b_0$, the stability of the homoclinic stripe for the modified GM model (2.23) on an $O(1)$ time-scale is determined by the spectrum of the NLEP*

(2.33a)

$$L_{0b}\Phi - \frac{\chi_m w^2}{(1+bw^2)} \frac{\int_{-\infty}^{\infty} w\Phi\, dy}{\int_{-\infty}^{\infty} w^2\, dy} = (\lambda + \varepsilon^2 m^2)\Phi\,, \quad -\infty < y < \infty\,, \quad \Phi \to 0 \quad \text{as} \quad |y| \to \infty\,,$$

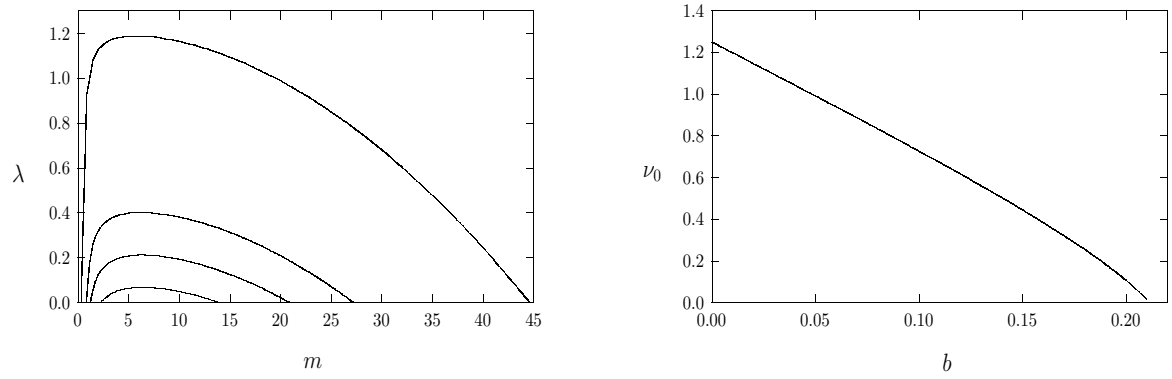
(2.33b)

$$L_{0b}\Phi \equiv \Phi'' - \Phi + \frac{2w}{(1+bw^2)^2}\Phi\,, \quad \chi_m \equiv \frac{2\theta_0 \tanh\theta_0}{\theta_\lambda \tanh\theta_\lambda}\,, \quad \theta_0 \equiv D^{-1/2}\,, \quad \theta_\lambda \equiv \sqrt{m^2 + \frac{(1+\tau\lambda)}{D}}.$$

For $\varepsilon \ll 1$, and in the absence of saturation effects (i.e., $b = 0$), it was shown in section 2.1 that there is always a breakup instability band for any $D > 0$. We now study numerically whether this instability band can disappear for some range of b on the homoclinic existence interval $0 \leq b < b_0$. For simplicity we will consider only the case where $\tau = 0$. Although we are unable to give a rigorous analysis of the spectrum of (2.33) as a function of b, we can readily identify the mechanism for the possible coalescence of the upper and lower breakup instability thresholds. Since $\chi_m \to 0$ as $m \to \infty$ in (2.33), it follows by the same reasoning as in section 2.1 that the upper stability boundary, m_{b+}, satisfies $m_{b+} \sim \sqrt{\nu_0}/\varepsilon_0$ for $\varepsilon_0 \ll 1$, where $\nu_0 = \nu_0(b) > 0$ is the unique positive eigenvalue of the local operator $L_{0b}\Psi = \nu\Psi$. However, ν_0 must tend to zero as $b \to b_0^- \approx 0.2114$, or equivalently as $k \to \infty$, since we necessarily must have $\nu_0 = 0$ for the heteroclinic orbit where $b = b_0$. This behavior of ν_0 is confirmed in Figure 8(b), where we plot the numerically computed curve $\nu_0 = \nu_0(b)$. Therefore, for any fixed ε small, it follows that $m_{b+} \to 0$ as $b \to b_0^-$. In addition, since a homoclinic pulse with $m = 0$ for (2.33) is always stable when $\tau = 0$ (cf. [41]), it follows that if the lower instability boundary exists, it will satisfy $m_{b-} > 0$. Hence, the instability band $m_{b-} < m < \sqrt{\nu_0}/\varepsilon_0$ must become narrower as b increases towards b_0, or equivalently as $k \to \infty$.

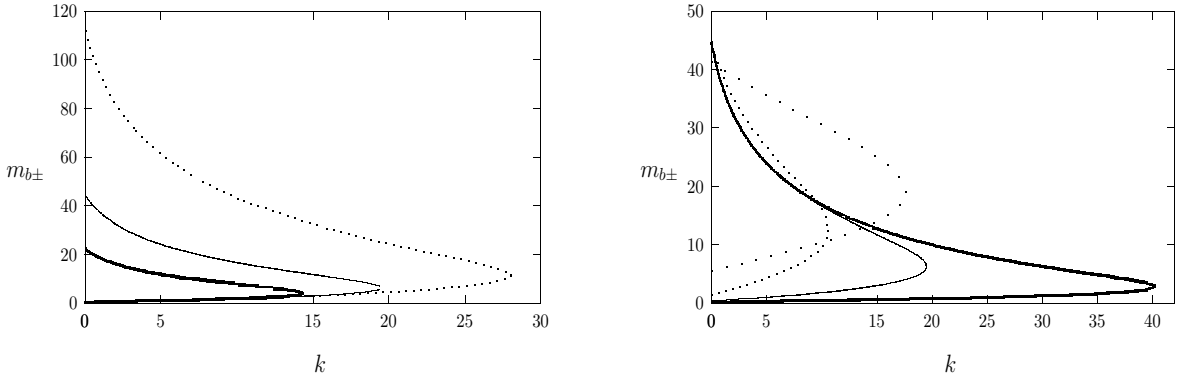
To numerically calculate breakup instability bands, we first solve (2.24) numerically and then discretize (2.33) using centered differences and the trapezoidal rule. The eigenvalues of the resulting matrix eigenvalue problem are found using LAPACK [1], and a quasi-Newton method is used to locate the edges of any instability band. Continuation in b is then used starting from $b = 0$, where the instability band is known from Proposition 2.3. Finally, the relation (2.30) between b and k determines the instability band with respect to the saturation parameter k in (2.23).

For $D = 10$ and $\varepsilon_0 = 0.025$, in Figure 8(a) we plot the unique positive eigenvalue within the instability band for four values of k. As k increases, the band becomes narrower, until it finally pinches off at the critical value $k \approx 19.4$. This coalescence of the edges of the breakup instability band for k sufficiently large occurs for other values of D and ε_0. In particular, when $D = 10$, in Figure 9(a) we show the merging of the upper and lower thresholds $m_{b\pm}$ at some value of k for three values of ε_0. In Figure 9(b) we show a similar merging behavior when $\varepsilon_0 = 0.025$ for four values of D.



(a) λ versus m (b) ν_0 versus b

Figure 8. (a) *the unstable eigenvalue* λ *versus* m *within an instability band when* $D = 10$ *and* $\varepsilon_0 = 0.025$. *From top to bottom the curves are for* $k = 0$, $k = 3.663$, $k = 6.802$, *and* $k = 12.519$. (b) *the principal eigenvalue* ν_0 *of the local operator* L_{0b} *versus* b, *with* $\nu_0 \to 0$ *as* $b \to 0.2114$.



(a) $m_{b\pm}$ versus k for $D = 10$ (b) $m_{b\pm}$ versus k for $\varepsilon_0 = 0.025$

Figure 9. (a) *plots of* $m_{b\pm}$ *versus* k *when* $D = 10$ *for* $\varepsilon_0 = 0.05$ *(heavy solid curve),* $\varepsilon_0 = 0.025$ *(solid curve), and* $\varepsilon_0 = 0.01$ *(dotted curve).* (b) *plots of* $m_{b\pm}$ *versus* k *when* $\varepsilon_0 = 0.025$ *for* $D = 100$ *(heavy solid curve),* $D = 10$ *(solid curve),* $D = 1$ *(dotted curve), and* $D = 0.1$ *(widely spaced dots).*

Qualitatively, these numerical results show that the breakup instability disappears on some range $k > k_{0d}$ (or, equivalently, $b_{0d} < b < b_0 \approx 0.2114$) when the homoclinic solution w to (2.24) is sufficiently broad (see Figure 7(a)). This widening of the homoclinic is a result of the ghost effect of the heteroclinic connection that exists when $b = b_0$. Therefore, we suggest that the stability properties of "fat" homoclinic stripes can be similar to those of mesa-stripe

solutions, such as those that occur for the Fitzhugh–Nagumo model (cf. [34], [35]) and the modified GM model of sections 4–5 with large saturation, where spot-generating breakup instabilities do not occur.

3. The GM model without saturation: Weak interaction regime. In the weak interaction regime we now show numerically that, depending on the exponent set (p, q, r, s), a stripe for (1.1) can be stable with respect to breakup instabilities for any domain width d_0 but is unstable to zigzag instabilities unless $d_0 = O(\varepsilon_0)$ thin. As in section 2.3, this disappearance of the breakup instability band is again related to the "fattening" of a homoclinic solution as a parameter is varied. In (1.4) the weak interaction regime corresponds to the limit $l = 1/\sqrt{D} \gg 1$ with $\varepsilon = \varepsilon_0 l = O(1)$. Equivalently, in (1.1) we write $D = D_0\varepsilon_0^2$ for some $D_0 = O(1)$, and we let $\varepsilon_0 \to 0$. In terms of $y = x_1/\varepsilon_0$, and upon rescaling a and h in (1.1), the resulting equilibrium problem for (1.1) is to look for even homoclinic solutions to the coupled system

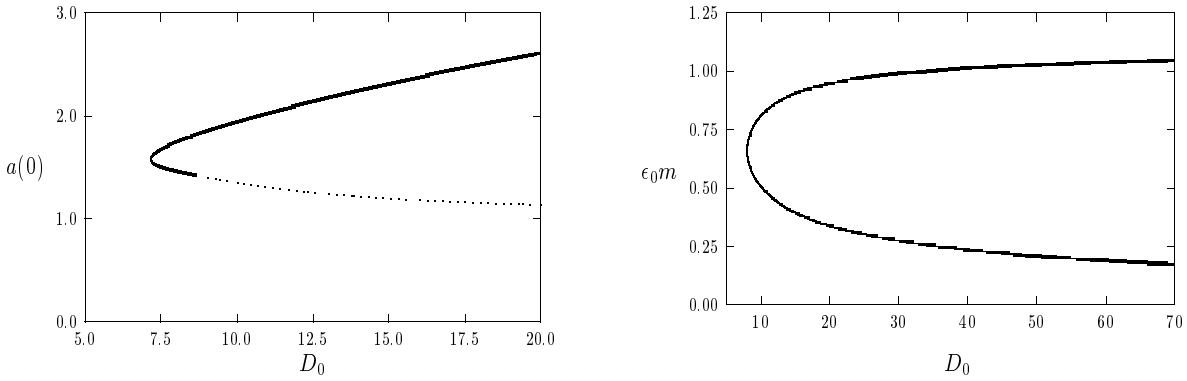
(3.1)
$$a_{yy} - a + \frac{a^p}{h^q} = 0\,, \quad D_0 h_{yy} - h + \frac{a^r}{h^s} = 0\,, \quad \infty < y < \infty\,, \qquad a \to 0\,, \quad h_y \to 0\,, \quad |y| \to \infty\,.$$

In [14] numerical solutions to (3.1) are computed by using the boundary-value solver COLSYS [2] together with path-following in D_0. As D_0 is decreased from some initially large value it was shown for various exponent sets in [14] that the bifurcation diagram of $a(0)$ versus D_0 for (3.1) has a saddle-node bifurcation at some critical value D_{0c} and that there are no homoclinic solutions to (3.1) when $0 < D_0 < D_{0c}$. The existence of such a fold-point value D_{0c} was proved in [4] using geometric singular perturbation theory. For the classical GM model, where $(p, q, r, s) = (2, 1, 2, 0)$, a plot of $a(0)$ versus D_0 is shown in Figure 10(a). When viewed as a pulse solution in one dimension, the upper branch of this bifurcation diagram is stable when τ is below some $O(1)$ Hopf bifurcation threshold $\tau_H = \tau_H(D_0)$, and the lower branch is unstable for any $\tau \geq 0$ (cf. [14]). The dashed portion along the lower solution branch of Figure 10(a) is where $a(y)$ has two distinct local maxima, with one on either side of the symmetry point $y = 0$. This two-bump structure begins at the point where $a_y(0) = 0$, which we label by $D_0 = D_{0m}$.

A stripe solution a_e, h_e for the two-dimensional GM model (1.1) is obtained by taking the homoclinic solution of (3.1) as the cross-sectional profile of the stripe. The stripe, with a width $O(\varepsilon_0)$, is then localized along the midline $x_1 = 0$ of the rectangular domain. For various exponent sets we then determine the stability of this stripe solution along each point on the upper branch of the $a(0)$ versus D_0 bifurcation diagram by writing

(3.2)
$$a = a_e\left(\varepsilon_0^{-1}x_1\right) + \Phi\left(\varepsilon_0^{-1}x_1\right)e^{\lambda t}\cos(mx_2)\,, \qquad h = h_e\left(\varepsilon_0^{-1}x_1\right) + N\left(\varepsilon_0^{-1}x_1\right)e^{\lambda t}\cos(mx_2)\,.$$

By substituting (3.2) into (1.1), we obtain the following eigenvalue problem for $\Phi(y)$ and $N(y)$



(a) $a(0)$ versus D_0 for $(2,1,2,0)$

(b) Breakup instability band for $(2,1,2,0)$

Figure 10. (a) *bifurcation diagram of* $a(0)$ *versus* D_0 *for* (3.1) *for* $(p,q,r,s)=(2,1,2,0)$. *On the dashed portion of the lower branch* $a(y)$ *has a multibump structure.* (b) *the breakup instability band versus* D_0 *for* $(p,q,r,s)=(2,1,2,0)$. *For modes* $\varepsilon_0 m$ *within the band the stripe is unstable. The band terminates before the saddle-node value* $D_{0c}=7.17$.

on $0 \le y < \infty$:

$$\begin{aligned} &\Phi_{yy} - (1+\mu)\Phi + \frac{p a_e^{p-1}}{h_e^q}\Phi - \frac{q a_e^p}{h_e^{q+1}} N = \lambda \Phi\,, \\ &D_0 N_{yy} - (1+D_0\mu)\, N + \frac{r a_e^{r-1}}{h_e^s}\Phi - \frac{s a_e^r}{h_e^{s+1}} N = \tau \lambda N\,. \end{aligned} \tag{3.3}$$

Here $\mu \equiv \varepsilon_0^2 m^2$, where m is either the breakup mode or the zigzag mode transverse to the stripe.

To study breakup instabilities we compute the spectrum of (3.3) for even eigenfunctions Φ and N so that $\Phi_y(0) = N_y(0) = 0$. We first compute the homoclinic solution of (3.1) on a long interval $0 < y < L$, where $L \gg 1$. We then discretize (3.3) on the same interval $[0, L]$ by using centered differences, and we label $\boldsymbol{\Phi_0} = (\Phi_0(y_1), \dots, \Phi_0(y_n))^t$ and $\boldsymbol{N_0} = (N_0(y_1), \dots, N_0(y_n))^t$, where $y_j = jh$ for $j = 0, \dots, n$ with $h = L/n$. We also impose that $\Phi_y(L) = N_y(L) = 0$. In this way, we obtain the block matrix eigenvalue problem

$$\begin{pmatrix} \mathcal{M} - (1+\mu)I + \Lambda_1 & -\Lambda_2 \\ \Lambda_3 & D_0\mathcal{M} - (1+D_0\mu)I - \Lambda_4 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Phi_0} \\ \boldsymbol{N_0} \end{pmatrix} = \lambda \begin{pmatrix} I & 0 \\ 0 & \tau I \end{pmatrix} \begin{pmatrix} \boldsymbol{\Phi_0} \\ \boldsymbol{N_0} \end{pmatrix}. \tag{3.4}$$

Here Λ_j for $j = 1, \dots, 4$ are $n \times n$ diagonal matrices, and $\mathcal{M}$ is a tridiagonal matrix defined

Table 2

The second and third columns are the saddle-node bifurcation values D_{0c} and $\alpha_c \equiv a(0)$ for the existence of a stripe. The fourth column gives the values D_{0b} of D_0 for the lower bound of the breakup instability band. A stripe is stable to breakup instabilities when $D_{0c} < D_0 < D_{0b}$. For $(3,2,2,0)$ and $(2,2,3,3)$ the band does not terminate before D_{0c}, and so D_{0b} is undefined. The fifth column gives the values D_{0z} of D_0 for the upper bound of the zigzag instability band in the weak-interaction regime. For $(2,2,3,3)$, where $\gamma = q/(p-1) = 2$, the band continues into the semistrong regime. The sixth column has the smallest values D_{0m} of D_0, where $a(y)$ has a multibump structure on the lower branch of the $a(0)$ versus D_0 bifurcation diagram. The seventh column gives the saddle-node values D_{0s}, computed in [33], representing the smallest value of D_0 where a radially symmetric spot solution exists in $\mathbb{R}^2$.

(p,q,m,s)	D_{0c}	α_c	D_{0b}	D_{0z}	D_{0m}	D_{0s}
$(2,1,2,0)$	7.17	1.58	8.06	24.0	8.92	9.82
$(2,1,3,0)$	10.35	1.42	19.14	30.0	12.36	16.31
$(3,2,2,0)$	3.91	1.62	$--$	32.3	5.08	5.23
$(3,2,3,1)$	4.41	1.53	5.13	28.0	5.36	5.97
$(2,2,3,3)$	33.7	2.28	$--$	$--$	41.80	85.52
$(4,2,2,0)$	0.89	1.36	1.00	27.9	1.06	0.89

by

(3.5)

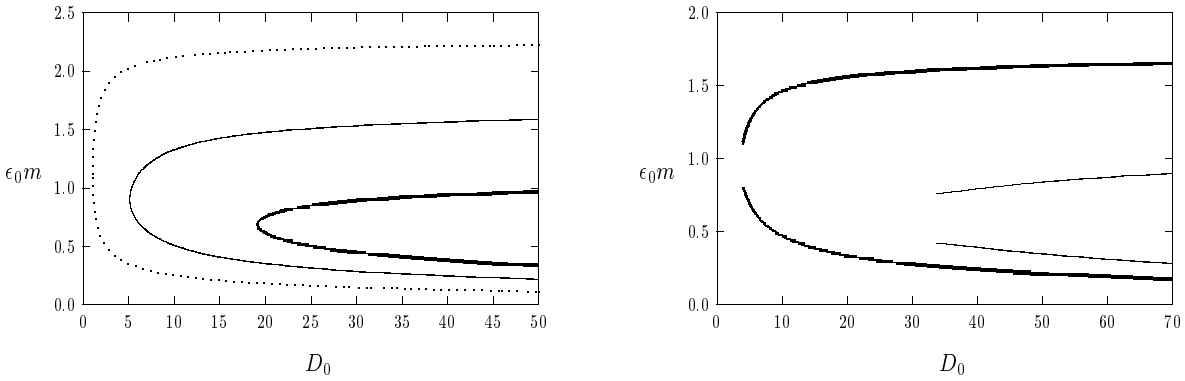
$$\Lambda_{1jj} = \frac{pa_e^p(y_j)}{h_e^q(y_j)}, \quad \Lambda_{2jj} = \frac{qa_e^p(y_j)}{h_e^{q+1}(y_j)},$$

$$\Lambda_{3jj} = \frac{ra_e^{r-1}(y_j)}{h_e^s(y_j)}, \quad \Lambda_{4jj} = \frac{sa_e^r(y_j)}{h_e^{s+1}(y_j)},$$

$$\mathcal{M} \equiv \frac{1}{h^2}\begin{pmatrix} -2 & 2 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & \ddots & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & 1 & -2 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 2 & -2 \end{pmatrix}.$$

For $\tau = 0.01$, we numerically determine the range of values of μ for which (3.4) has unstable eigenvalues. For D_0 sufficiently large, and for all of the exponent sets in Table 2, our computational results from LAPACK [1] with $n = 250$ meshpoints and $L = 15$ show that there are threshold values μ_1 and μ_2 for which there is a unique real positive eigenvalue λ_0 in the breakup instability band $\sqrt{\mu_1} < \varepsilon_0 m < \sqrt{\mu_2}$, and that $\mathrm{Re}(\lambda) < 0$ for $0 \le \varepsilon_0 m < \sqrt{\mu_1}$ and $\varepsilon_0 m > \sqrt{\mu_2}$. However, as D_0 is decreased towards the existence threshold D_{0c}, our results show that the instability band disappears for some of these exponent sets at some critical value $D_{0b} > D_{0c}$ on the upper branch. Numerical values for D_{0b} are given in Table 2. Increasing n and L did not change the results in Table 2 significantly.

In Figure 10(b) we plot the upper and lower thresholds for the breakup instability band for the exponent set $(p,q,r,s) = (2,1,2,0)$, showing the coalescence of the thresholds when $D_0 = D_{0b} = 8.06$. This critical value compares reasonably well with the corresponding critical value $D_0 \approx 1/(.12) = 8.33$ estimated in [4, see p. 99] based on full numerical computations and on an extrapolation of stability results from the semistrong regime into the weak interaction regime (see Figure 5.2 of [4]). Therefore, for the classical GM model, a stripe solution for



(a) breakup bands: $(4,2,2,0)$, $(3,2,3,1)$, $(2,1,3,0)$

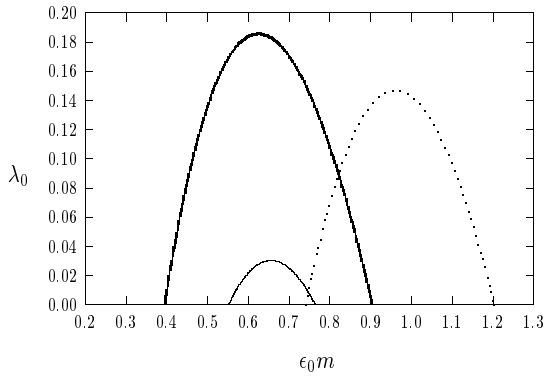
(b) breakup bands: $(3,2,2,0)$, $(2,2,3,3)$

Figure 11. (a) *breakup instability bands that terminate before the saddle-node value* D_{0c}. *The heavy solid curve is for* $(p,q,r,s) = (2,1,3,0)$, *the solid curve is for* $(3,2,3,1)$, *and the dashed curve is for* $(4,2,2,0)$. (b) *breakup bands that do not terminate before* D_{0c}. *The heavy solid and solid curves are for* $(3,2,2,0)$ *and* $(2,2,3,3)$, *respectively.*

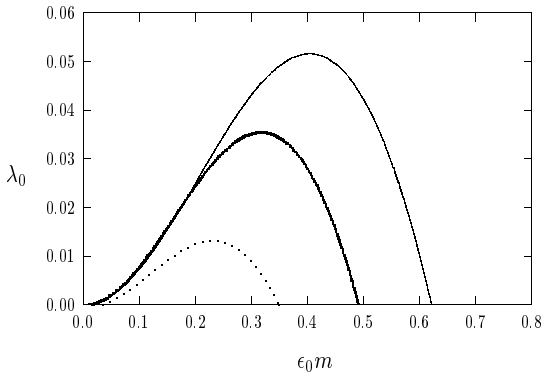
(1.1) exists and is stable with respect to breakup instabilities when $7.17 < D_0 < 8.06$. In Figure 11(a) we plot similar upper and lower stability thresholds of the breakup instability band for the exponent sets $(4,2,2,0)$, $(3,2,3,1)$, and $(2,1,3,0)$. For these exponent sets the breakup instability band disappears below some threshold value D_{0b} larger than D_{0c}, so that a stripe solution for (1.1) is stable with respect to breakup on the range $D_{0c} < D_0 < D_{0b}$. For three parameter sets, in Figure 12(a) we plot the unique unstable real eigenvalue within an instability band. As seen from this figure, the most unstable mode occurs roughly in the middle of this band.

However, as shown in Figure 11(b), the breakup instability band does not disappear at some D_0 value greater than the existence threshold D_{0c} for the exponent sets $(3,2,2,0)$ and $(2,2,3,3)$. Consequently, for these exponent sets, a stripe solution to (1.1) will always be unstable to breakup instabilities when the domain width d_0 is $O(1)$.

Next, we study zigzag instabilities by calculating the spectrum of (3.3) for odd eigenfunctions Φ and N so that $\Phi(0) = N(0) = 0$. The discrete eigenvalue problem has the same form as in (3.4) and (3.5) except that now $y_j = jh$ for $j = 1, \dots, n$ with $h = L/n$, and where $\mathcal{M}_{12} = 1$ replaces the corresponding entry in the matrix $\mathcal{M}$ in (3.5). The spectrum of the resulting discrete eigenvalue problem is then computed numerically for $\tau = 0.01$, $L = 15$, and with $n = 250$. In Figure 13(a) we show the unstable zigzag band for three exponent sets for which $\gamma = \frac{q}{p-1} = 1$. Recall that for $\gamma = 1$ the theory of section 2.2 showed that there is no unstable zigzag band in the semistrong interaction regime. Our computational results show that as D_0 is decreased towards the existence threshold D_{0c}, an unstable zigzag band first emerges at some critical value D_{0z} with $D_{0z} > D_{0c}$. Numerical values for D_{0z} are given in Table 2. Within the zigzag band there is a unique unstable real eigenvalue when $\tau = 0.01$.

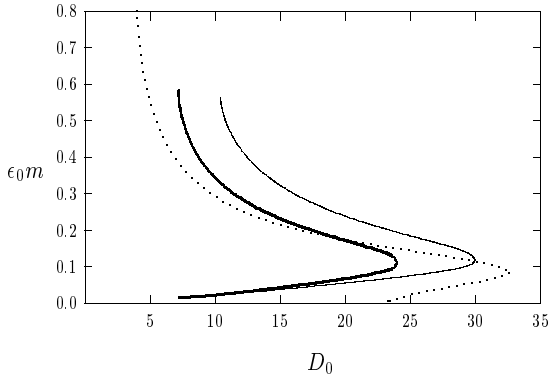


(a) unstable breakup eigenvalue

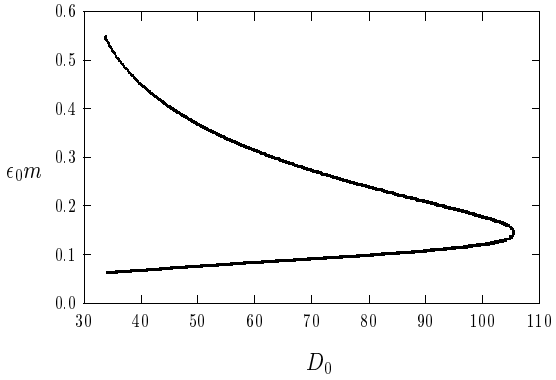


(b) unstable zigzag eigenvalue

Figure 12. (a) *the unstable breakup eigenvalue* λ_0 *within an instability band. The heavy solid curve is for* $(2,1,2,0)$ *with* $D_0 = 15.0$*, the solid curve is for* $(2,1,2,0)$ *with* $D_0 = 9.01$*, and the dashed curve is for* $(3,2,2,0)$ *with* $D_0 = 4.5$. (b) *the unstable zigzag eigenvalue* λ_0 *within an instability band. The heavy solid curve is for* $(2,1,2,0)$ *with* $D_0 = 7.60$*, the solid curve is for* $(3,2,2,0)$ *with* $D_0 = 4.5$*, and the dashed curve is for* $(2,1,3,0)$ *with* $D_0 = 14.0$.



(a) zigzag bands: $(2,1,2,0)$, $(2,1,3,0)$, $(3,2,2,0)$



(b) zigzag band: $(2,2,3,3)$

Figure 13. (a) *the zigzag instability band for* $(2,1,2,0)$ *(heavy solid curve), for* $(2,1,3,0)$ *(solid curve), and for* $(3,2,2,0)$ *(dashed curve), plotted for* $D_0 > D_{0c}$*. For these exponent sets there is a value* D_{0z} *of* D_0 *for which there is an unstable band for* $D_{0c} < D_0 < D_{0z}$*. Within the band there is a unique and real unstable eigenvalue.* (b) *the unstable zigzag band for* $(2,2,3,3)$*. Since* $\gamma = q/(p-1) = 2 > 1$*, this band continues into the semistrong interaction regime and terminates there (see section 2.2).*

For three parameter sets, this eigenvalue is plotted in Figure 12(b) within an instability band. From section 2.2 we recall that there is an unstable zigzag band in the semistrong regime

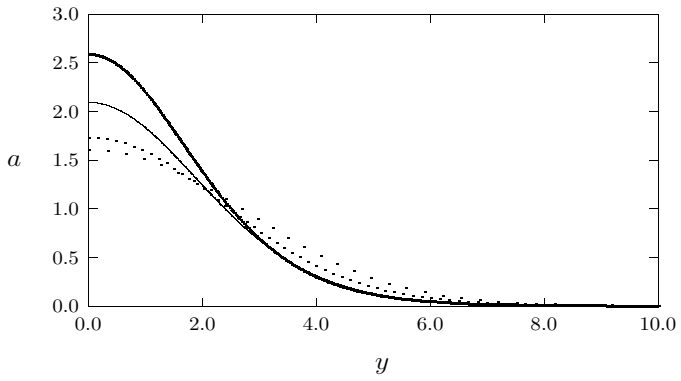


Figure 14. *Plots of $a(y)$ for various D_0 computed from (3.1) for the exponent set $(2,1,2,0)$. The heavy solid, solid, dotted, and widely spaced dotted curves are for $D_0 = 19.6$, $D_0 = 11.9$, $D_0 = 7.9$, and $D_0 = 7.2$, respectively. Notice that the homoclinic becomes broader as D_0 is decreased.*

$D = O(1)$ when $\gamma = \frac{q}{p-1} > 1$. For the exponent set $(2,2,3,3)$, where $\gamma = 2$, in Figure 13(b) we show the continuation of this band into the weak interaction regime. For a domain length $L = 15$, or equivalently $\varepsilon_0 = 1/15$, we obtain from Figure 13(b) that the zigzag band exists for $D_0 < D_{0z} \approx 107$. This value corresponds to $D = D_0\varepsilon_0^2 \approx 0.4755$, or equivalently $l = 1/\sqrt{D} = 1.45$. This critical value of l for $\gamma = 2$ agrees well with the critical value $l_z \approx 1.4$ obtained from Figure 6 for when the zigzag band first forms in the semistrong interaction regime.

We now qualitatively summarize our conclusions regarding the stability of the stripe. An important conclusion is that for the exponent sets (p,q,m,s), where the breakup instability disappears at some value above the existence threshold as D_0 is decreased, the stripe will always be unstable with respect to zigzag instabilities for domain widths d_0 that are $O(1)$ as $\varepsilon_0 \to 0$. Since the upper zigzag threshold m_{z+} satisfies $m_{z+} = O(\varepsilon_0^{-1})$, the zigzag instability can only be suppressed near the existence threshold D_{0c} by taking the domain width d_0 to be $O(\varepsilon_0)$ thin. We also observe from Figure 12 that the time-scale for breakup instabilities is generally faster than for zigzag instabilities. However, both time-scales are independent of ε_0. By comparing Figures 10(b) and 11 with Figure 13, we observe that whenever a breakup instability band exists the breakup and zigzag bands overlap in such a way that there are no domain widths d_0 where a zigzag instability is not accompanied by a breakup instability.

Remark 3.1. It appears to be difficult to provide a rigorous study of the eigenvalue problem (3.3) to theoretically confirm the possible coalescence of the breakup instability band and the emergence of the zigzag band near the existence threshold D_{0c}. However, the shape of the homoclinic $a(y)$ near the existence threshold gives some indication of the reason for the change in the dominant instability mechanism. For D near D_{0c}, the region near the maximum of $a(y)$ is generally wider than it is for larger values of D_0. This is shown numerically in Figure 14 for the exponent set $(2,1,2,0)$. In fact, on the unstable branch, but near D_{0c}, the cross section of the "fattened" homoclinic stripe develops a multibump structure at some value D_{0m} (see Table 2). In addition to the general pulse-splitting criteria of [5], this multibump structure

also appears to be an essential factor for self-replication behavior (cf. [14]). Therefore, near the existence threshold, the cross section of the homoclinic stripe becomes fatter in a similar way as was studied in section 2.3 for the small-saturation GM model (2.23) in the semistrong regime. This suggests that the stability problem near the existence threshold D_{0c} can be similar to that of a bistable system where zigzag instabilities are the dominant instability mechanism (cf. [34]) and where breakup instabilities do not occur.

We now perform a few numerical experiments on (1.1) to illustrate and validate the spectral results for breakup and zigzag instabilities. For each of the experiments below, we solve (1.1) for $\tau = 0.01$ in the square domain $\Omega = [-1, 1] \times [0, 2]$, with an initial condition of the form

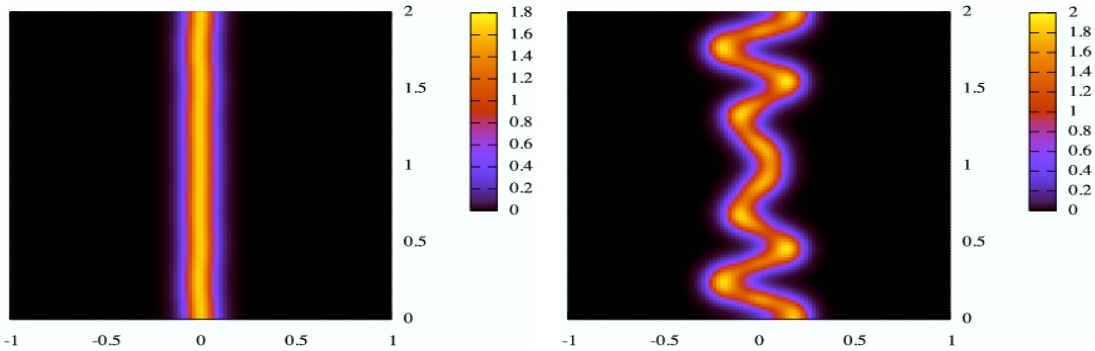
$$a(x_1, x_2, 0) = A\text{sech}^2\left(\frac{x_1}{\varepsilon_0}\right), \qquad h(x_1, x_2, 0) = H\text{sech}^2\left(\frac{x_1}{2\varepsilon_0}\right). \tag{3.6}$$

Whenever the equilibrium stripe exists, in the experiments below we have taken $A = a(0)$ and $H = h(0)$, where a and h are the numerical solution of (3.1) for the specified value of D_0.

Experiment 4. In Experiment 4 we take $(p, q, r, s) = (2, 1, 2, 0)$, $\varepsilon_0 = 0.025$, and $D_0 = 7.6$. The initial condition for (1.1) is (3.6) with $A = 1.69$ and $H = 1.45$. Since $D_{0c} < D_0 < D_{0b}$ there is no breakup instability band, and we predict that the initially straight stripe will not break up into spots. From the heavy solid curve in Figure 12(b) the most unstable zigzag mode is $\varepsilon_0 m \approx 0.315$ with a corresponding growth rate $\lambda_0 \approx 0.035$. Therefore, the number of zigzag crests is theoretically predicted to be $N = \frac{m}{\pi} = \frac{0.315}{\pi\varepsilon_0} \approx 4$. The full numerical results from (1.1) are shown in Figure 15. We observe that, initially, there is indeed no breakup instability and that when $t = 300$ the stripe develops a noticeable zigzag instability with four crests. However, the wriggled stripe then undergoes a breakup instability near the points of its maximum curvature, leading to spot formation. Since $D_0 = 7.6$ is below the existence threshold $D_{0s} = 9.82$ for a locally radially symmetric spot solution (see Table 2), these spots then undergo a repeated self-replication process which fills the entire domain. The solution at time $t = 1000$, shown in Figure 15(d), is near an equilibrium state and more closely resembles a Turing-type pattern than a pattern with isolated spots.

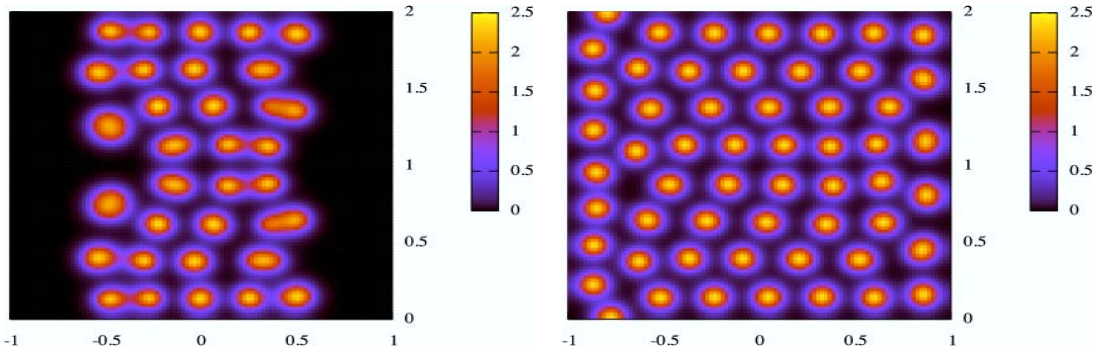
Experiment 5. We take $(p, q, r, s) = (2, 1, 2, 0)$, $\varepsilon_0 = 0.025$, and we increase D_0 to $D_0 = 15.0$. The initial condition for (1.1) is (3.6) with $A = 2.3$ and $H = 1.8$. Since $D_0 > D_{0b} = 8.06$ and $D_0 < D_{0z} = 24.0$ the initially straight stripe is unstable to a breakup and a zigzag instability. From the heavy solid curve in Figure 12(a) the most unstable breakup mode is $\varepsilon_0 m \approx 0.623$ with a growth rate of $\lambda_0 \approx 0.186$. Therefore, the theoretically predicted number of spots is $N = \frac{m}{\pi} = \frac{0.623}{\pi\varepsilon_0} \approx 8$. From the full numerical results shown in Figure 16(a) it is observed that the stripe initially breaks up into eight spots. The eight spots in Figure 16(a) are not in perfect vertical alignment, owing to the zigzag instability of the initial stripe. Then, as shown in Figure 16(b), the repulsive spot interactions accentuate the broken vertical symmetry. Since $D_0 = 15$ is well above the spot-existence threshold of $D_{0s} = 9.82$ given in Table 2, there is no spot self-replication behavior. Instead there is an exponentially slow, or metastable, drift of the spots towards a stable hexagonal equilibrium configuration (cf. Figure 16(c) and Figure 16(d)).

Experiment 6. Next, we take $(p, q, r, s) = (2, 1, 2, 0)$, $\varepsilon_0 = 0.025$, and $D_0 = 6.8$. The initial condition for (1.1) is (3.6) with $A = 1.6$ and $H = 1.4$. Since $D_0 < D_{0c}$, there is no equilibrium stripe solution. For the related problem of a pulse on a one-dimensional interval with $D_0 <$



(a) Experiment 4: $t = 200$

(b) Experiment 4: $t = 300$



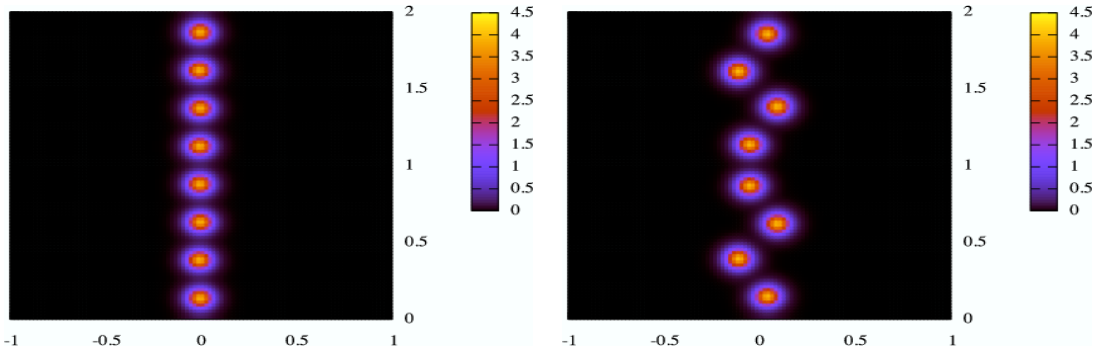
(c) Experiment 4: $t = 400$

(d) Experiment 4: $t = 1000$

Figure 15. Experiment 4. *The numerical solution to* (1.1) *for* $(p, q, r, s) = (2, 1, 2, 0)$ *with* $\varepsilon_0 = 0.025$, $D_0 = 7.6$, $\tau = 0.01$, *and in a square domain* $\Omega = [-1, 1] \times [0, 2]$. *The initially straight stripe develops a zigzag instability. Spots are formed near the region of maximum curvature of the wriggled stripe. These spots then undergo a repeated self-replication process leading to a Turing-type pattern. See also the accompanying animation* (63508_01.gif [650KB]).

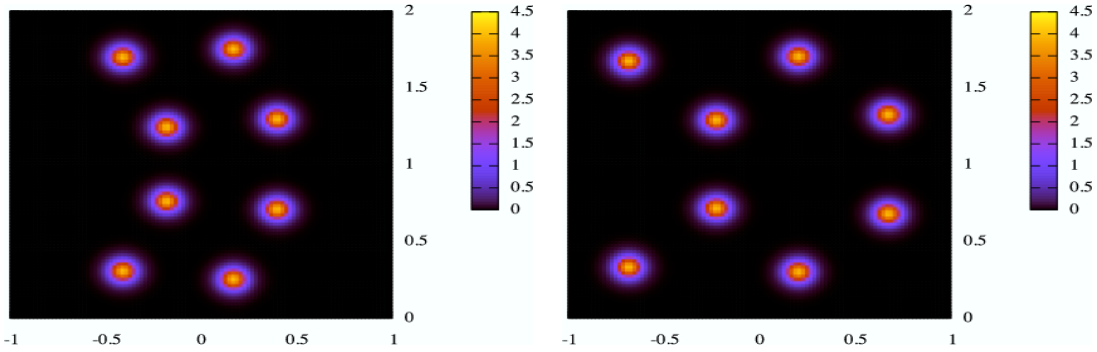
D_{0c}, an initial one-pulse profile undergoes an edge-splitting pulse-replication process leading to a Turing-type pattern (cf. [4], [14], [25, Remark 6.2]). The numerical results for the two-dimensional GM model are shown in Figure 17. We observe that the stripe first splits into two and then develops a zigzag instability. The wriggled stripes undergo a breakup instability near their points of maximal curvature. Since $D_0 < D_{0s}$ the emerging spots then undergo a spot-splitting process. For short times, stripe-replication behavior was computed in Figure 5.3 of [4] for $D_0 = 1/(.14) \approx 7.14$ before any zigzag instabilities occur.

Experiment 7. We now take $(p, q, r, s) = (3, 2, 2, 0)$, $\varepsilon_0 = 0.025$, and $D_0 = 4.5$. The initial condition for (1.1) is (3.6) with $A = 1.82$ and $H = 0.74$. For this exponent set, in addition to the zigzag instability, a breakup instability is guaranteed since the breakup band does not terminate in the weak interaction regime. From the dashed curve in Figure 12(a) the most unstable breakup mode is $\varepsilon_0 m \approx 0.96$ with a growth rate of $\lambda_0 \approx 0.147$. With



(a) Experiment 5: $t = 500$

(b) Experiment 5: $t = 800$



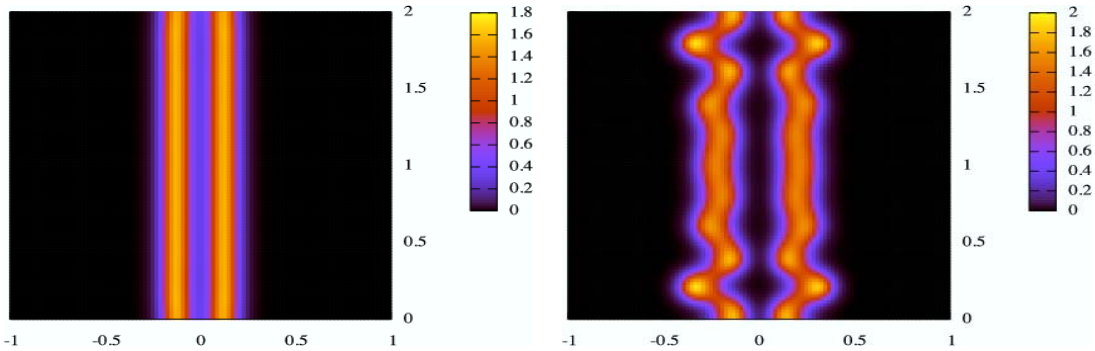
(c) Experiment 5: $t = 3000$

(d) Experiment 5: $t = 30000$

Figure 16. Experiment 5. *The numerical solution to* (1.1) *for* $(p, q, r, s) = (2, 1, 2, 0)$ *with* $\varepsilon_0 = 0.025$, $D_0 = 15.0$, $\tau = 0.01$, *and in a square domain* $\Omega = [-1, 1] \times [0, 2]$. *The stripe initially breaks up into eight spots that are not in perfect vertical alignment, owing to the zigzag instability of the initial stripe. There is no self-replication behavior and there is an exponentially slow, or metastable, evolution of the eight spots towards a stable equilibrium configuration consisting of a hexagonal structure.*

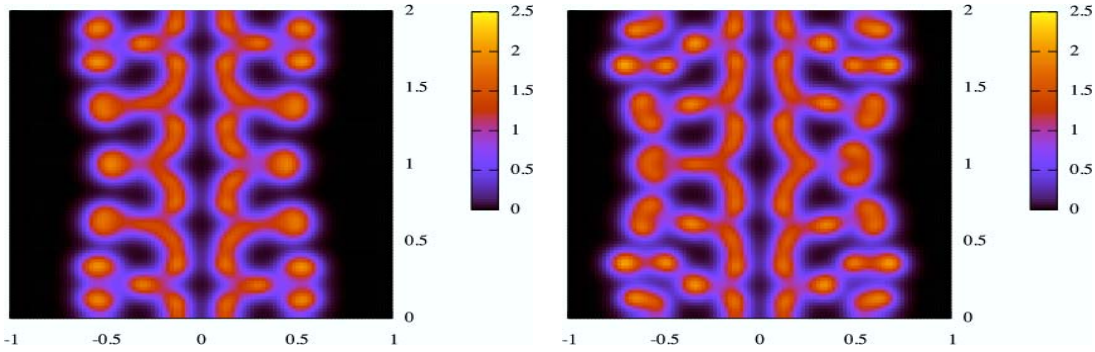
this most unstable mode, we predict that the stripe will break up into $N = \frac{m}{\pi} = \frac{0.96}{\pi\varepsilon_0} \approx 12$ spots. Alternatively, from the solid curve in Figure 12(b), the most unstable zigzag mode is $\varepsilon_0 m \approx 0.43$ with a corresponding growth rate $\lambda_0 \approx 0.051$. This corresponds to a most unstable zigzag mode with $N = \frac{m}{\pi} = \frac{0.43}{\pi\varepsilon_0} \approx 5$ crests. In the numerical results shown in Figure 18 we observe that the initially straight stripe breaks up into fourteen spots (see Figure 18(a)) and then develops a zigzag instability with six crests, which breaks the vertical symmetry of the array of spots (see Figure 18(b)). Since $D_0 < D_{0s} = 5.23$ the resulting spots then undergo a spot-splitting process (cf. Figure 18(c)) leading to a final state that closely resembles a Turing-type pattern (cf. Figure 18(d)).

Experiment 8. Finally, we consider $(p, q, r, s) = (2, 1, 3, 0)$, $\varepsilon_0 = 0.01$, and $D_0 = 14.0$. The initial condition for (1.1) is (3.6) with $A = 1.6$ and $H = 1.3$. For this exponent set there are no breakup instabilities on the rather wide range $10.35 < D_0 < 19.14$ (see Table 2). Therefore,



(a) Experiment 6: $t = 75$

(b) Experiment 6: $t = 175$



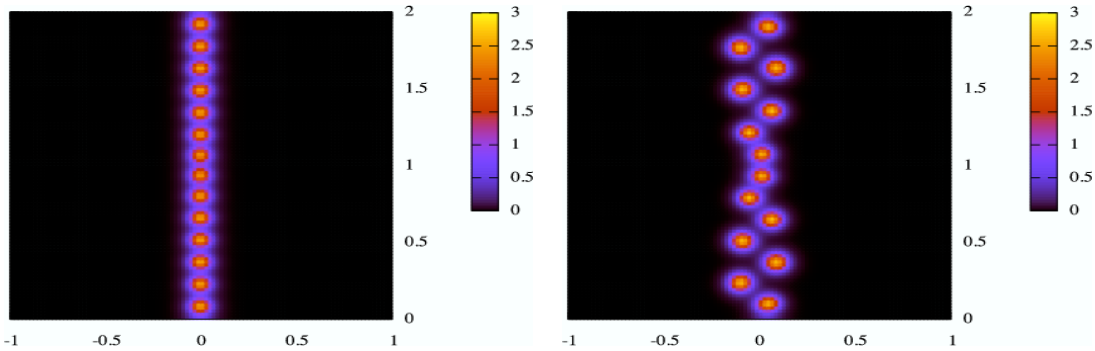
(c) Experiment 6: $t = 225$

(d) Experiment 6: $t = 250$

Figure 17. Experiment 6. *The numerical solution to* (1.1) *for* $(p, q, r, s) = (2, 1, 2, 0)$ *with* $\varepsilon_0 = 0.025$, $D_0 = 6.8$, $\tau = 0.01$, *and in a square domain* $\Omega = [-1, 1] \times [0, 2]$. *Since* $D_0 < D_{0c}$, *the initially straight stripe splits into two. The two stripes become wriggled as a result of a zigzag instability. Spots are formed near the region of maximum curvature of the wriggled stripes. These spots then undergo a self-replication process. See also the accompanying animation* (63508_02.gif [512KB]).

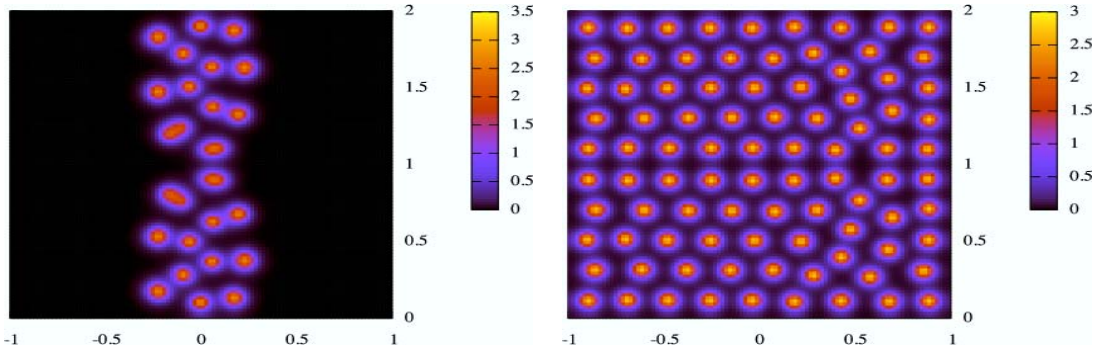
we expect no breakup instability. From the dashed curve in Figure 12(b) the most unstable zigzag mode is $\varepsilon_0 m \approx 0.23$, with a corresponding growth rate $\lambda_0 \approx 0.0131$, and the expected number of zigzag crests is $N = \frac{m}{\pi} = \frac{0.23}{\pi\varepsilon_0} \approx 7$. In the numerical results shown in Figure 19(b) a zigzag instability with exactly seven crests is observed. In contrast to Experiment 4, where D_0 was only slightly below the breakup threshold D_{0b}, for this example D_0 is significantly below the breakup threshold, and the wriggled stripe in Figure 19 is less vulnerable than in Experiment 4 (see Figure 15) to a breakup instability occurring near a local maxima of its curvature. The final pattern in Figure 19(d) is composed almost exclusively of stripes.

4. The GM model with saturation: A mesa-stripe solution in the near-shadow limit. For the modified GM model (1.6) with saturation parameter κ, we now construct a different type of equilibrium stripe solution centered along the midline of the rectangular domain $\Omega :=$



(a) Experiment 7: $t = 100$

(b) Experiment 7: $t = 200$



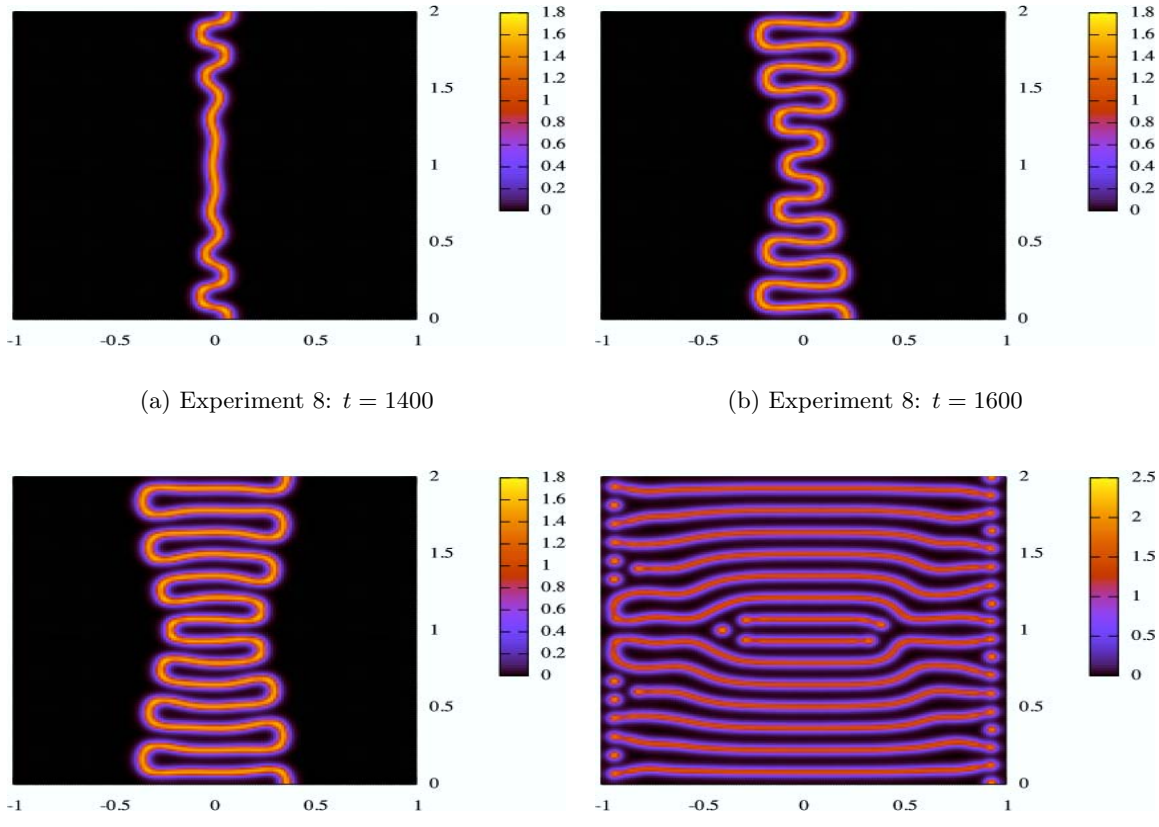
(c) Experiment 7: $t = 250$

(d) Experiment 7: $t = 1000$

Figure 18. Experiment 7. *The numerical solution to* (1.1) *for* $(p, q, r, s) = (3, 2, 2, 0)$ *with* $\varepsilon_0 = 0.025$, $D_0 = 4.5$, $\tau = 0.01$, *and in a square domain* $\Omega = [-1, 1] \times [0, 2]$. *Since the breakup instability band exists up to* D_{0c}, *the initially straight stripe breaks up into spots. It then develops a twist as a result of a zigzag instability. Since* $D_0 < D_{0s} = 5.23$, *the spots then undergo a repeated self-replication process leading to a Turing-type pattern.*

$[0, 1] \times [0, d_0]$, with $\partial_n a = \partial_n h = 0$ on $\partial\Omega$. In section 5 we analyze the stability of this solution. For our analysis of (1.6) we assume that $\varepsilon_0 \ll 1$ and $\kappa > 0$, where κ is independent of ε_0. Recall that the case $\kappa = O(\varepsilon_0^2)$ was considered in section 2.3.

Our analysis is limited to a near-shadow limit $D \gg 1$, where $D = \mathcal{D}/\varepsilon_0$ with $\mathcal{D} = O(1)$. A similar restriction was made in [35] in their study of interface stability for a generalized Fitzhugh–Nagumo system. The analysis of mesa-stripes in the regime $D = O(1)$ is not a straightforward extension of the analysis for the near-shadow limit $D = O(\varepsilon^{-1})$. For $D = O(1)$ a new phenomenon related to the self-replication of mesa-stripes can occur. This behavior is shown in Experiment 11 below. Since a detailed study of mesa-splitting for the regime $D = O(1)$ is expected to be rather involved, and is not related to the goal of this paper of characterizing breakup and zigzag instabilities of a stripe, we do not attempt to study the regime $D = O(1)$ here.



(a) Experiment 8: $t = 1400$

(b) Experiment 8: $t = 1600$

(c) Experiment 8: $t = 1800$

(d) Experiment 8: $t = 4300$

Figure 19. Experiment 8. *The numerical solution to* (1.1) *for* $(p, q, r, s) = (2, 1, 3, 0)$ *with* $\varepsilon_0 = 0.01$, $D_0 = 14.0$, $\tau = 0.01$, *and in a square domain* $\Omega = [-1, 1] \times [0, 2]$. *There is no breakup instability of the straight stripe, and the wriggled stripe is less vulnerable than in Experiment* 4 *to a breakup instability occurring near local maxima of the curvature. The zigzag instability of the straight stripe is seen to be the precursor to a large-scale deformation of the stripe. See also the accompanying animation* (63508_03.gif [362KB]).

In the near-shadow limit, (1.6) admits an equilibrium stripe solution in the form of a front-back transition layer structure, where the layers are connected by an asymptotically flat plateau. We refer to such a solution as a mesa-stripe. This type of solution is distinctly different from the homoclinic stripe solutions of sections 2 and 3. Since $D \gg O(1)$, then $h \sim \mathcal{H}$ uniformly on $0 \le x_1 \le 1$, where $\mathcal{H}$ is a constant. Therefore, from the inhibitor equation of (1.6), we get

$$\mathcal{H} \sim \int_0^1 a^2 \, dx_1 \, . \tag{4.1}$$

The transition layer solution centered at some $x_1 = \xi_l$ is given by $a \sim \mathcal{H} w(y)$, with $y = \varepsilon_0^{-1}(x_1 - \xi_l)$. From (1.6) we obtain that w satisfies (2.24) with $b = \kappa \mathcal{H}^2$. It was shown in section 2.3 that (2.24) has a heteroclinic orbit connecting $w = 0$ and $w = w_+$ when $b = b_0$.

Here w_+ and b_0 are given in (2.27). For $b = b_0$, the heteroclinic solution to (2.24) satisfies

(4.2a)

$$w'' - w + g_0(w) = 0\,, \quad -\infty < y < \infty\,, \quad g_0(w) \equiv \frac{w^2}{1 + b_0 w^2}\,, \quad b_0 = \kappa \mathcal{H}^2\,,$$

(4.2b)

$$w(y) \sim c_- \exp(y)\,, \quad y \to -\infty\,, \quad w(y) \sim w_+ - c_+ \exp(-\nu_+ y)\,, \quad y \to +\infty\,, \quad \nu_+ \equiv \sqrt{1 - \frac{2}{w_+}}$$

for some positive constants $c_\pm$. To break the translation invariance we impose the condition $w(0) = w_+/2$. For the stability analysis in section 5, we must evaluate $\beta \equiv \int_{-\infty}^{\infty} (w')^2\, dy$. A simple calculation gives

(4.3)

$$\beta \equiv \int_{-\infty}^{\infty} (w')^2\, dy = \int_0^{w_+} \sqrt{2\mathcal{F}(w)}\, dw \approx 1.49882\,, \qquad \mathcal{F}(w) \equiv \frac{w^2}{2} - \frac{w}{b_0} + \frac{1}{b_0^{3/2}} \tan^{-1}\left(w\sqrt{b_0}\right).$$

For $b = b_0$ and $\varepsilon_0 \ll 1$, a composite expansion for the mesa-stripe solution has the form

(4.4)

$$a \sim \mathcal{H}\left[w_l(y_l) + w_r(y_r) - w_+\right], \qquad w_l(y_l) \equiv w\left[\varepsilon_0^{-1}(x_1 - \xi_l)\right], \qquad w_r(y_r) \equiv w\left[\varepsilon_0^{-1}(\xi_r - x_1)\right].$$

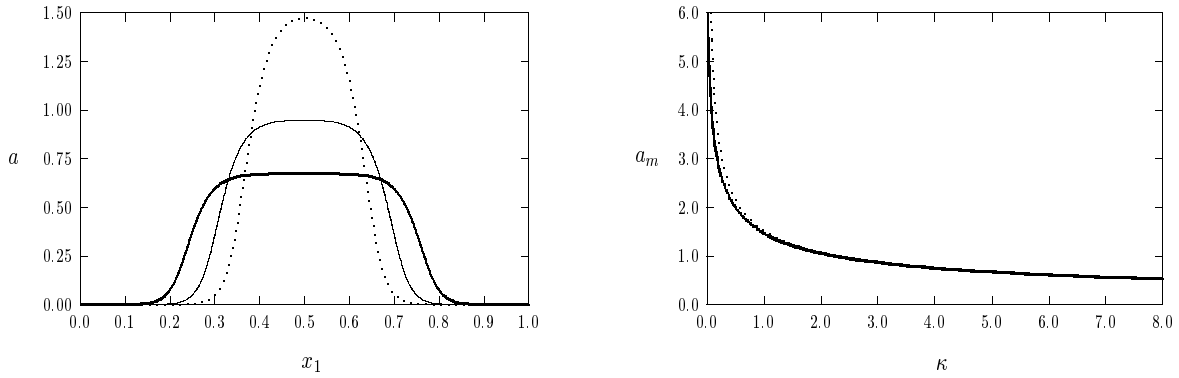
By using (4.1) and (4.4) , we obtain that $\mathcal{H} \sim \mathcal{H}^2 w_+^2 L + O(\varepsilon_0)$, where $L = \xi_r - \xi_l$. Then, from the relation $b = \mathcal{H}^2 \kappa$ of (4.2a), we obtain that

$$\mathcal{H} \sim \frac{1}{w_+^2 L} + O(\varepsilon_0)\,, \qquad L \sim \frac{\sqrt{\kappa}}{\sqrt{b_0} w_+^2} < 1\,. \tag{4.5}$$

In Figure 20(a) we plot the numerically computed equilibrium solution a_e versus x for several values of κ when $\varepsilon_0 = 0.02$ and $D = 10$. This solution was computed using COLSYS [2]. Notice that as κ increases, the maximum value of a_e decreases, while the spatial extent of the plateau also increases. For a range of κ, in Figure 20(b) we show a very favorable comparison between the numerically computed value for the plateau height $a_e(1/2)$ and the corresponding asymptotic result $a_e(1/2) \sim \mathcal{H} w_+ = (w_+ L)^{-1}$. As a remark, the analysis of homoclinic solutions in section 2 showed that $a_e\,(1/2) = O(\varepsilon_0^{-1})$ when $\kappa = 0$. In Figure 21 we plot the relation (4.5) between the saturation parameter κ and the plateau length L. The condition $L < 1$ ensures that the plateau fits within the unit length of the rectangle. Although ξ_l and ξ_r are undetermined at this stage, we anticipate by symmetry that the plateau is centered in the middle of the interval $[0, 1]$ so that $\xi_l = (1 - L)/2 = (1 - \xi_r)$. This result is derived analytically below.

Although (4.4) and (4.5) give the leading-order solution, a higher-order construction of the equilibrium solution is required for the stability analysis of section 5. This is done by first considering the outer regions, comprised of the plateau region $\xi_l < x_1 < \xi_r$ together with near-boundary regions $0 < x_1 < \xi_l$ and $\xi_r < x_1 < 1$. In these outer regions we expand

$$h \sim \mathcal{H} + \varepsilon_0 h_1 + \varepsilon_0^2 h_2 + \cdots\,. \tag{4.6}$$



(a) a_e versus x (b) $a_m \equiv a_e\,(1/2)$ versus κ

Figure 20. (a) *the numerically computed equilibrium solution* a_e *versus* x *for* $\varepsilon_0 = 0.02$ *and* $D = 10$ *when* $\kappa = 1$ *(dashed curve),* $\kappa = 2.5$ *(solid curve), and* $\kappa = 5.0$ *(heavy solid curve). The length* L *of the plateau increases with increasing* κ. (b) *comparison of the asymptotic plateau value* $a_m \equiv a_e\,(1/2) \sim \mathcal{H} w_+$ *(dashed curve) versus* κ *with the corresponding full numerical result for* $a_e\,(1/2)$ *(heavy solid curve).*

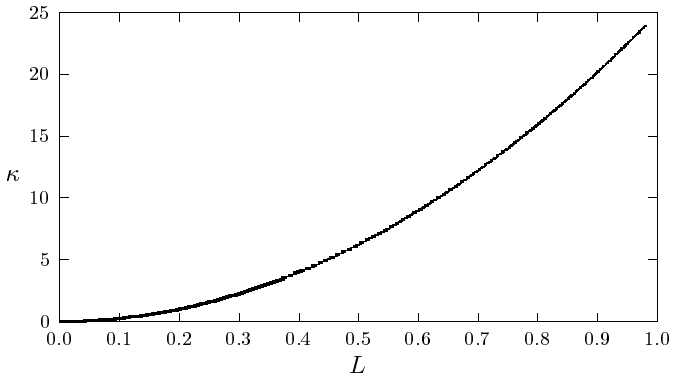


Figure 21. *The length* L *of the plateau of the mesa versus the saturation parameter* κ *of the activator kinetics.*

In the near-boundary regions a is exponentially small as $\varepsilon_0 \to 0$, while in the plateau region $a \sim \mathcal{H} w_+ + O(\varepsilon_0)$. Therefore, by substituting (4.6) into the inhibitor equation of (1.6) with $D = \mathcal{D}/\varepsilon_0$, we obtain that

$$\mathcal{D} h_{1x_1x_1} = \begin{cases} \mathcal{H}\,, & 0 < x_1 < \xi_l\,, \\ \mathcal{H} - \mathcal{H}^2 w_+^2\,, & \xi_l < x_1 < \xi_r\,, \\ \mathcal{H}\,, & \xi_r < x_1 < 1\,, \end{cases} \tag{4.7}$$

with $h_{1x_1}(0) = h_{1x_1}(1) = 0$. The conditions for h_1 at the transition layers ξ_l and ξ_r are found

below upon matching h_1 to appropriate inner solutions. In the plateau region we expand a as

$$a \sim \mathcal{H} w_+ + \varepsilon_0 \mathcal{A}_1 + \cdots , \qquad \xi_l < x_1 < \xi_r \, . \tag{4.8}$$

By substituting (4.8) into the activator equation of (1.6), we obtain that $\mathcal{A}_1$ satisfies

$$\mathcal{A}_1 = g_a \left(\mathcal{H} w_+ , \mathcal{H} \right) \mathcal{A}_1 + g_h \left(\mathcal{H} w_+ , \mathcal{H} \right) h_1 \, , \qquad \xi_l < x_1 < \xi_r \, . \tag{4.9}$$

By using $g_a(\mathcal{H} w_+, \mathcal{H}) = 2/w_+$ and $g_h(\mathcal{H} w_+, \mathcal{H}) = -w_+$, we calculate $\mathcal{A}_1$ from (4.9) as

$$\mathcal{A}_1 = \frac{w_+^2 h_1}{2 - w_+} \, . \tag{4.10}$$

In the inner region near $x_1 = \xi_l$, we let $y = \varepsilon_0^{-1}(x_1 - \xi_l)$, and we expand $a(y)$ and $h(y)$ as

$$a = a_0 + \varepsilon_0 a_1 + \varepsilon_0^2 a_2 + \cdots , \qquad h = \mathcal{H} + \varepsilon_0 \mathcal{H}_1 + \varepsilon_0^2 \mathcal{H}_2 + \cdots \, . \tag{4.11}$$

Upon substituting (4.11) into the equilibrium problem for (1.6), and letting $D = \mathcal{D}/\varepsilon_0$, we obtain that $a_0 = \mathcal{H} w$ and that a_1 satisfies

$$L a_1 \equiv a_1'' - a_1 + g_0'(w) a_1 = g_0(w) \mathcal{H}_1 \, , \qquad \mathcal{H}_1'' = 0 \, , \qquad -\infty < y < \infty \, . \tag{4.12}$$

Here $g_0(w)$ is defined in (4.2a). This yields that $\mathcal{H}_1 = \mathcal{H}_{10} + y \mathcal{H}_{11}$, for some unknown constants $\mathcal{H}_{10}$ and $\mathcal{H}_{11}$. However, since $h = \mathcal{H} + O(\varepsilon_0)$ in the plateau and near-boundary regions, we require that $\mathcal{H}_{11} = 0$ in order to match the inner and outer solutions for h. Therefore, $\mathcal{H}_1 = \mathcal{H}_{10}$. To determine $\mathcal{H}_{10}$ we use a solvability condition. Since $L w' = 0$, the solvability condition for (4.12) gives $\mathcal{H}_{10} \int_{-\infty}^{\infty} g_0(w) w' \, dy = 0$, which yields $\mathcal{H}_{10} = 0$. Therefore, since $L a_1 = 0$, we get $a_1 = c_1 w'$ for some constant c_1. However, since $a(0) = \mathcal{H} w_+/2$ defines the transition layer location, we require that $a_1(0) = 0$. Since $w'(0) \neq 0$, this implies that $c_1 = 0$. Therefore, $a_1 = \mathcal{H}_1 \equiv 0$, and we must proceed to the next order in ε_0 to find the first nonvanishing correction term.

Upon substituting (4.11) into (1.6), we obtain that a_2 and $\mathcal{H}_2$ satisfy

$$L a_2 \equiv a_2'' - a_2 + g_0'(w) a_2 = g_0(w) \mathcal{H}_2 \, , \qquad \mathcal{H}_2'' = 0 \, , \qquad -\infty < y < \infty \, . \tag{4.13}$$

Therefore, $\mathcal{H}_2 = \mathcal{H}_{20} + \mathcal{H}_{21} y$ for some constants $\mathcal{H}_{20}$ and $\mathcal{H}_{21}$. The solvability condition for (4.13) gives one relation between these two constants in the form

$$\int_{-\infty}^{\infty} \left(\mathcal{H}_{20} + y \mathcal{H}_{21} \right) g_0(w) w' \, dy = 0 \, . \tag{4.14}$$

The other relationship between these constants is determined from the matching condition that $h \sim \mathcal{H} + \varepsilon_0^2 \mathcal{H}_2 + \cdots$ agrees asymptotically as $y \to \pm\infty$ with the behavior of the outer solution $h \sim \mathcal{H} + \varepsilon_0 h_1(x_1) + \varepsilon_0^2 h_2(x_1) + \cdots$ as $x \to \xi_l^{\pm}$. This matching condition readily yields that

$$h_1(\xi_l^{\pm}) = 0 \, , \qquad \mathcal{H}_{21} = h_{1x_1}(\xi_l^-) = h_{1x_1}(\xi_l^+) \, , \tag{4.15}$$

and $h_2(\xi_l^{\pm}) = \mathcal{H}_{20}$. The condition $h_{1x_1}(\xi_l^-) = h_{1x_1}(\xi_l^+)$ is found below to determine the equilibrium transition layer location ξ_l uniquely as $\xi_l = (1-L)/2$. The conditions (4.15) and (4.14) determine $\mathcal{H}_{20}$ and $\mathcal{H}_{21}$ uniquely.

Finally, we derive a key identity needed in section 5. We differentiate (4.13) with respect to y to obtain

$$La_2' = -g_0''(w)w' a_2 + g_0(w)\mathcal{H}_2' + g_0'(w)\mathcal{H}_2 w' . \tag{4.16}$$

Since $Lw' = 0$, the solvability condition for (4.16) yields the identity

$$-\mathcal{H}_2' \int_{-\infty}^{\infty} g_0(w)w'\,dy = \int_{-\infty}^{\infty} \left[g_0'(w)\mathcal{H}_2 - g_0''(w)a_2 \right] (w')^2\,dy\,, \tag{4.17}$$

where $\mathcal{H}_2' = h_{1x_1}(\xi_l^+)$ by the matching condition (4.15).

A similar analysis can be done near the other transition layer at $x_1 = \xi_r$ to obtain that $h_1(\xi_r^{\pm}) = 0$. By solving (4.7) with $h_{1x_1}(0) = h_{1x_1}(1) = 0$ and $h_1(\xi_l^{\pm}) = h_1(\xi_r^{\pm}) = 0$, we obtain that

$$h_1 = \begin{cases} \frac{1}{2\mathcal{D}w_+^2 L}\left(x_1^2 - \xi_l^2\right), & 0 < x_1 < \xi_l, \\ \frac{\mathcal{H}(L-1)}{2\mathcal{D}L}\left[(x_1-\xi_l)^2 - (\xi_r - \xi_l)(x_1 - \xi_l) \right], & \xi_l < x_1 < \xi_r, \\ \frac{1}{2\mathcal{D}w_+^2 L}\left[(1-x_1)^2 - (1-\xi_r)^2 \right], & \xi_r < x_1 < 1. \end{cases} \tag{4.18}$$

From this solution, we calculate the following one-sided derivatives, which are also needed in section 5 below:

$$h_{1x_1}(\xi_l^+) = \frac{\mathcal{H}(1-L)}{2\mathcal{D}} = -h_{1x_1}(\xi_r^-)\,, \qquad h_{1x_1}(\xi_l^-) = \frac{\mathcal{H}\xi_l}{\mathcal{D}}\,, \qquad h_{1x_1}(\xi_r^+) = -\frac{\mathcal{H}(1-\xi_r)}{\mathcal{D}}\,. \tag{4.19}$$

The condition that h_{1x_1} is continuous across $x_1 = \xi_l$ and $x_1 = \xi_r$ determines the equilibrium layer locations uniquely in terms of the length L of the plateau as $\xi_l = (1-L)/2$ and $1 - \xi_r = (1-L)/2$.

5. The GM model with saturation: Stability analysis in the near-shadow limit. We now study the stability of the mesa-stripe equilibrium solution $a_e(x_1)$, $h_e(x_1)$ constructed in section 4. Since there are Neumann boundary conditions on the sides $x_2 = 0, d_0$ of the rectangular domain, the perturbation takes the form

$$a = a_e + e^{\lambda t + imx_2}\phi\,, \qquad h = h_e + e^{\lambda t + imx_2}\psi\,, \qquad m = \frac{k\pi}{d_0}\,, \quad k = 1, 2, \dots\,, \tag{5.1}$$

where $\phi = \phi(x_1) \ll 1$ and $\psi = \psi(x_1) \ll 1$. The bands of instability with respect to the continuous variable m derived below can be mapped to integer k-bands of instability using (5.1).

Substituting (5.1) into (1.6) and using $D = \mathcal{D}/\varepsilon_0$, we obtain the eigenvalue problem

$$L_\varepsilon \phi + g_h(a_e, h_e)\psi = \bar{\lambda}\phi\,, \quad 0 < x_1 < 1\,, \qquad \phi_{x_1}(0) = \phi_{x_1}(1) = 0\,, \tag{5.2a}$$

$$\psi_{x_1x_1} - m^2\psi = \frac{\varepsilon_0}{\mathcal{D}}(1+\tau\lambda)\psi - \frac{2\varepsilon_0}{\mathcal{D}} a_e \phi\,, \quad 0 < x_1 < 1\,, \qquad \psi_{x_1}(0) = \psi_{x_1}(1) = 0\,. \tag{5.2b}$$

Here $\bar{\lambda}$ and the operator L_ε are defined by

$$L_\varepsilon \phi \equiv \varepsilon_0^2 \phi_{x_1 x_1} - \phi + g_a(a_e, h_e)\phi\,, \qquad \bar{\lambda} \equiv \lambda + \varepsilon_0^2 m^2\,. \tag{5.2c}$$

In the analysis below we will show that $\lambda = O(\varepsilon_0^2)$ when $m > 0$ and $m = O(1)$. Therefore, for $\varepsilon_0 \ll 1$, the term $\tau\lambda$ in (5.2b) is asymptotically negligible when $\tau = O(1)$ and is consequently neglected.

We first determine the asymptotic form of the eigenfunction ϕ corresponding to $\bar{\lambda} \ll 1$. In the plateau region, where $a_e \sim \mathcal{H} w_+$, $g_a(a_e, h_e) \sim 2w_+^{-1}$, and $g_h(a_e, h_e) \sim -w_+$, we obtain from (5.2a) that

$$\phi \sim \mu\psi\,, \qquad \mu \equiv \frac{w_+^2}{2 - w_+}\,, \qquad \xi_l < x_1 < \xi_r\,. \tag{5.3}$$

In the near-boundary regions ϕ is exponentially small as $\varepsilon_0 \to 0$. Near the transition layers at ξ_l and ξ_r, ϕ is proportional to the derivative w' of the heteroclinic orbit. Therefore, this motivates the asymptotic form

$$\phi \sim \begin{cases} c_l\big(w'(y_l) + O(\varepsilon_0) + \cdots\big)\,, & y_l \equiv \varepsilon_0^{-1}(x_1 - \xi_l) = O(1)\,, \\ \phi_i \equiv \mu\psi\,, & \xi_l < x_1 < \xi_r\,, \\ c_r\big(w'(y_r) + O(\varepsilon_0) + \cdots\big)\,, & y_r \equiv \varepsilon_0^{-1}(\xi_r - x_1) = O(1)\,, \end{cases} \tag{5.4}$$

for some unknown constants c_l and c_r to be found. Here μ is defined in (5.3).

Since ϕ is localized near the transition layers, we use (5.4) to calculate in the sense of distributions that

$$\frac{2\varepsilon_0 a_e \phi}{\mathcal{D}} \sim \frac{\varepsilon_0^2 \mathcal{H} w_+^2}{\mathcal{D}}\left[c_l \delta(x_1 - \xi_l) + c_r \delta(x_1 - \xi_r)\right] + \left(\frac{2\varepsilon_0}{\mathcal{D}}\right)\mathcal{H} w_+ \mu\psi \chi_{[\xi_l, \xi_r]}\,, \tag{5.5}$$

where $\mathcal{H} \sim 1/(w_+^2 L)$. Here $\chi_{[\xi_l,\xi_r]}$ is the indicator function defined to be unity for $\xi_l \le x_1 \le \xi_r$ and zero outside this plateau region. Substituting (5.5) into (5.2b), we obtain that ψ satisfies

$$\begin{aligned} &\psi_{x_1 x_1} - \theta^2 \psi = -\frac{\varepsilon_0^2 \mathcal{H} w_+^2}{\mathcal{D}}\left[c_r \delta(x_1 - \xi_r) + c_l \delta(x_1 - \xi_l)\right], \quad 0 < x_1 < 1, \\ &\psi_{x_1}(0) = \psi_{x_1}(1) = 0\,. \end{aligned} \tag{5.6a}$$

Here θ is the piecewise constant function

$$\theta = \begin{cases} \theta_- \equiv \left[m^2 + \frac{\varepsilon_0}{\mathcal{D}}\right]^{1/2}\,, & 0 < x_1 < \xi_l\,, \quad \xi_r < x_1 < 1\,, \\ \theta_+ \equiv \left[m^2 + \frac{\varepsilon_0}{\mathcal{D}}\left(1 + \frac{2w_+}{L(w_+ - 2)}\right)\right]^{1/2}\,, & \xi_l < x_1 < \xi_r\,. \end{cases} \tag{5.6b}$$

The jump conditions for (5.6) are that

$$\psi_{x_1}(x_l^+) - \psi_{x_1}(x_l^-) = -\frac{\varepsilon_0^2 \mathcal{H} w_+^2 c_l}{\mathcal{D}}\,, \qquad \psi_{x_1}(x_r^+) - \psi_{x_1}(x_r^-) = -\frac{\varepsilon_0^2 \mathcal{H} w_+^2 c_r}{\mathcal{D}}\,. \tag{5.6c}$$

Next, we derive a matrix eigenvalue problem for $\bar{\lambda}$. We substitute (5.4) into (5.2a) and multiply the resulting expression by $w_l^{'}$, where $w_l^{'} \equiv w^{'}(y_l)$. Since $w_l^{'}$ is localized, we then obtain for $\varepsilon_0 \ll 1$ that

$$c_l \left(w_l^{'}, L_\varepsilon w_l^{'} \right) + \left(w_l^{'}, g_h(a_e, h_e)\psi \right) \sim c_l \bar{\lambda} \left(w_l^{'}, w_l^{'} \right) . \tag{5.7}$$

Here we have defined $(f, g) \equiv \int_0^1 fg\, dx_1$. We use (4.2) to estimate the second and third terms in (5.7) as

$$\left(w_l^{'}, g_h(a_e, h_e)\psi \right) \sim -\varepsilon_0 \psi(\xi_l) \int_{-\infty}^{\infty} w^{'} g_0(w)\, dy = -\varepsilon_0 \psi(\xi_l) \int_{-\infty}^{\infty} (w - w^{''}) w^{'}\, dy = -\varepsilon_0 \psi(\xi_l) \frac{w_+^2}{2} , \tag{5.8a}$$

$$\left(w_l^{'}, w_l^{'} \right) \sim \varepsilon_0 \int_{-\infty}^{\infty} (w^{'})^2\, dy = \varepsilon_0 \beta , \qquad \beta \equiv \int_{-\infty}^{\infty} (w^{'})^2\, dy . \tag{5.8b}$$

To calculate the first term in (5.7) we first use the inner solution $a_e \sim \mathcal{H} w + \varepsilon_0^2 a_2$ and $h_e \sim \mathcal{H} + \varepsilon_0^2 \mathcal{H}_2$ to obtain

$$g_a(a_e, h_e) \sim g_a\left(\mathcal{H} w, \mathcal{H}\right) + \varepsilon_0^2 \left[g_{aa}\left(\mathcal{H} w, \mathcal{H}\right) a_2 + g_{ah}\left(\mathcal{H} w, \mathcal{H}\right) \mathcal{H}_2 \right] + \cdots , \tag{5.9a}$$

$$g_a(a_e, h_e) \sim g_0^{'}(w) + \frac{\varepsilon_0^2}{\mathcal{H}} \left[g_0^{''}(w) a_2 - g_0^{'}(w) \mathcal{H}_2 \right] + \cdots . \tag{5.9b}$$

By using (5.9b), and upon differentiating (4.2) with respect to y, we readily obtain that

$$L_\varepsilon w^{'} \sim \frac{\varepsilon_0^2}{\mathcal{H}} \left[g_0^{''}(w) a_2 - g_0^{'}(w) \mathcal{H}_2 \right] w^{'} . \tag{5.10}$$

Then, using (5.10) and the identity (4.17), we derive that

$$\left(w_l^{'}, L_\varepsilon w_l^{'} \right) \sim \frac{\varepsilon_0^3}{\mathcal{H}} \int_{-\infty}^{\infty} \left[g_0^{''}(w) a_2 - g_0^{'}(w) \mathcal{H}_2 \right] (w^{'})^2\, dy = \frac{\varepsilon_0^3 \mathcal{H}_2^{'}}{\mathcal{H}} \int_{-\infty}^{\infty} g_0(w) w^{'}\, dy \tag{5.11a}$$

$$\sim \frac{\varepsilon_0^3 \mathcal{H}_2^{'}}{\mathcal{H}} \int_{-\infty}^{\infty} (w - w^{''}) w^{'}\, dy = \frac{\varepsilon_0^3 \mathcal{H}_2^{'} w_+^2}{2\mathcal{H}} = \frac{\varepsilon_0^3 h_{1x_1}(\xi_l^+) w_+^2}{2\mathcal{H}} . \tag{5.11b}$$

Finally, upon substituting (5.8) and (5.11b) into (5.7), we obtain that

$$\bar{\lambda} \varepsilon_0 c_l \beta \sim \frac{\varepsilon_0^3 c_l}{2\mathcal{H}} h_{1x_1}(\xi_l^+) w_+^2 - \frac{\varepsilon_0}{2} \psi(\xi_l) w_+^2 . \tag{5.12a}$$

In a similar way, we obtain from the transition layer solution at $x_1 = \xi_r$ that

$$\bar{\lambda} \varepsilon_0 c_r \beta \sim -\frac{\varepsilon_0^3 c_r}{2\mathcal{H}} h_{1x_1}(\xi_r^-) w_+^2 - \frac{\varepsilon_0}{2} \psi(\xi_r) w_+^2 . \tag{5.12b}$$

The next step in the analysis is to reduce (5.12) to an explicit matrix eigenvalue problem. To do so, we first solve (5.6) for ψ in order to calculate $\psi(\xi_l)$ and $\psi(\xi_r)$. A simple calculation shows that

$$\begin{pmatrix} \psi(\xi_l) \\ \psi(\xi_r) \end{pmatrix} = \frac{\varepsilon_0^2 \mathcal{H} w_+^2}{\mathcal{D}} \mathcal{G} \boldsymbol{c} , \qquad \boldsymbol{c} \equiv \begin{pmatrix} c_l \\ c_r \end{pmatrix} , \tag{5.13}$$

where $\mathcal{G}$ is the Green's function matrix defined by

(5.14)
$$\mathcal{G} \equiv \frac{1}{d^2-e^2}\begin{pmatrix} d & e \\ e & d \end{pmatrix}, \quad d \equiv \theta_+ \coth(\theta_+ L) + \theta_- \tanh\left(\frac{\theta_-(1-L)}{2}\right), \quad e \equiv \theta_+ \operatorname{csch}(\theta_+ L).$$

Substituting (5.13) and (4.19) into (5.12), and recalling that $\bar{\lambda} = \lambda + \varepsilon_0^2 m^2$ and $\mathcal{H} w_+^2 \sim 1/L$ from (5.2c) and (4.5), we obtain that λ is an eigenvalue of the matrix eigenvalue problem

$$\alpha(\lambda + \varepsilon_0^2 m^2)\boldsymbol{c} \sim \varepsilon_0^2 \left[\frac{1}{2}L(1-L)I - \mathcal{G}\right]\boldsymbol{c}. \tag{5.15}$$

Here I is the identity matrix, and $\alpha \equiv 2\beta L\mathcal{D}/w_+^2$.

The spectrum $\mathcal{G}\boldsymbol{v} = \sigma\boldsymbol{v}$ is readily calculated as

$$\boldsymbol{v}_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \sigma_+ = \frac{1}{d-e} = \left[\theta_+ \tanh\left(\frac{\theta_+ L}{2}\right) + \theta_- \tanh\left(\frac{\theta_-(1-L)}{2}\right)\right]^{-1} \tag{5.16a}$$

and

$$\boldsymbol{v}_- = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \sigma_- = \frac{1}{d+e} = \left[\theta_+ \coth\left(\frac{\theta_+ L}{2}\right) + \theta_- \tanh\left(\frac{\theta_-(1-L)}{2}\right)\right]^{-1}. \tag{5.16b}$$

Here $\theta_\pm$ are defined in (5.6b). Combining (5.16) and (5.15), we obtain the explicit eigenvalues

(5.17)
$$\lambda_\pm \sim \frac{\varepsilon_0^2}{\alpha}\left[-\alpha m^2 + \frac{L}{2}(1-L) - \sigma_\pm\right], \quad \boldsymbol{c}_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{c}_- = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \alpha \equiv \frac{2\beta L\mathcal{D}}{w_+^2}.$$

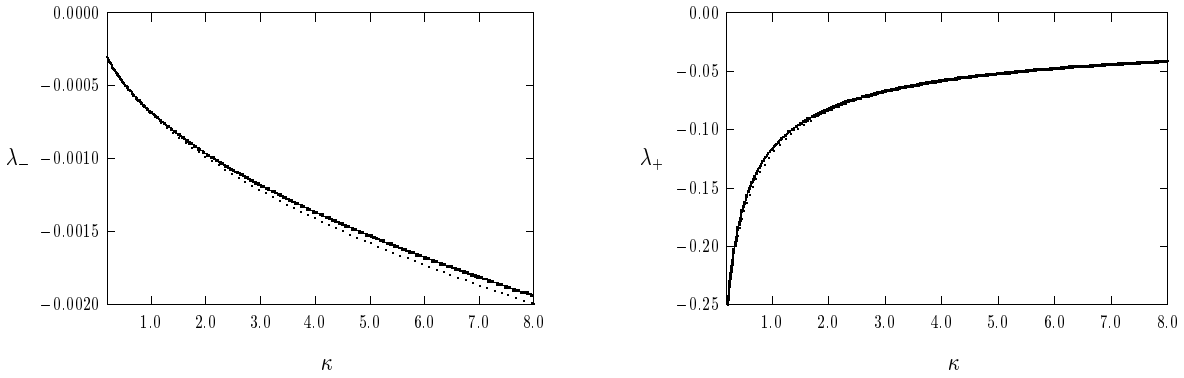
The two eigenvectors $\boldsymbol{c}_\pm$ in (5.17), which determine the eigenfunction ϕ in (5.4), lead to two different types of transverse instability of the mesa-stripe. The zigzag mode, where ξ_l and ξ_r are perturbed in the same direction, corresponds to $\boldsymbol{c}_- = (1,-1)^t$ because the signs of x_1 in y_l and y_r in (5.4) are different. Alternatively, the mode $\boldsymbol{c}_+ = (1,1)^t$ corresponds to a breather-type instability. For α sufficiently large, it is easy to see that $\lambda_\pm < 0$ for any $m \geq 0$. Therefore, a mesa-stripe is stable when $\mathcal{D}$ is above some threshold. However, as $\mathcal{D}$ is decreased (or equivalently as α is decreased), first the zigzag mode and then the breather mode admits a nontrivial band of unstable wave numbers m. The fact that the zigzag mode becomes unstable before the breather mode as $\mathcal{D}$ is decreased arises from the inequality $\sigma_- < \sigma_+$. Since $\lambda_\pm = O(\varepsilon_0^2)$ the time-scale for the development of these instabilities is $O(\varepsilon_0^{-2})$.

We first consider zigzag and breather instabilities for a one-dimensional pulse where $m = 0$. For $\varepsilon_0 \to 0$, a simple calculation using (5.16) for $m = 0$ and $\varepsilon_0 \ll 1$ gives

$$\sigma_- \sim \frac{L}{2} + O(\varepsilon_0), \qquad \sigma_+ \sim \frac{2\mathcal{D}}{\varepsilon_0}\left[1 - \frac{2w_+}{2-w_+}\right]^{-1}. \tag{5.18}$$

Then, from (5.17), we obtain for $m = 0$ and $\varepsilon_0 \ll 1$ that

(5.19)
$$\lambda_- \sim -\frac{\varepsilon_0^2 L^2}{2\alpha} = -1.8112\frac{\varepsilon_0^2 L}{\mathcal{D}} < 0, \qquad \lambda_+ \sim -\frac{\varepsilon_0 w_+^2}{\beta L}\left[1 - \frac{2w_+}{2-w_+}\right]^{-1} = -1.1899\frac{\varepsilon_0}{L} < 0.$$



(a) λ_- versus κ for $m = 0$

(b) λ_+ versus κ for $m = 0$

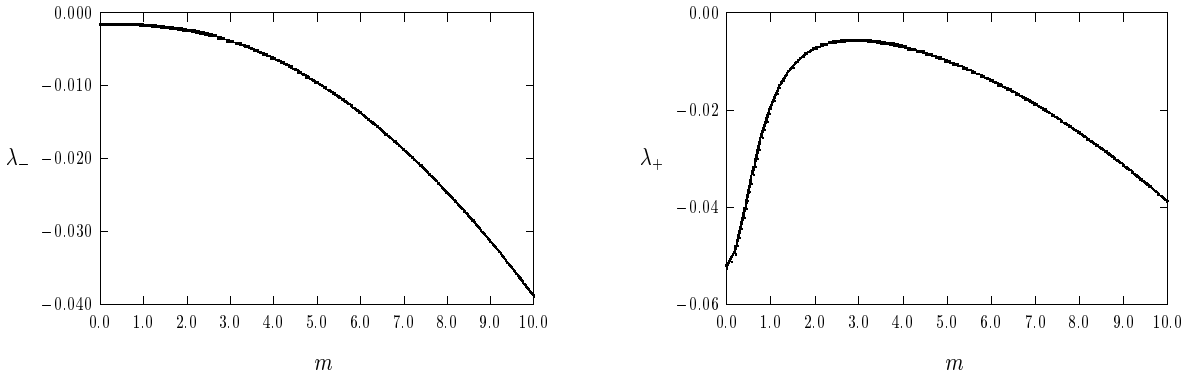
Figure 22. *Comparison of asymptotic (heavy solid curves) and full numerical results (dashed curves) for* $\lambda_\pm$ *versus* κ *for* $m = 0$ *when* $\varepsilon_0 = 0.02$ *and* $D = 10.0$. (a) *the zigzag eigenvalue* λ_-. (b) *the breather eigenvalue* λ_+.

Therefore, when $L > 0$, we have $\lambda_\pm < 0$ for $m = 0$ and $\varepsilon_0 \ll 1$, with $\lambda_- = O(\varepsilon_0^2)$ and $\lambda_+ = O(\varepsilon_0)$. This shows that a one-dimensional pulse solution is always stable under the effect of activator saturation. For the parameter set $m = 0$, $\varepsilon_0 = 0.02$, and $D = 10$, for which $\mathcal{D} = 0.2$, in Figure 22(a) we compare the asymptotic result for the zigzag eigenvalue λ_-, given in (5.17), with the corresponding full numerical result computed from (5.2). The numerical result is obtained by first discretizing (5.2) using centered differences and then using LAPACK [1] to compute the relevant eigenvalue of a matrix eigenvalue problem. For the same parameter set $m = 0$, $\varepsilon_0 = 0.02$, and $D = 10$, in Figure 22(b) we show a similar favorable comparison between the asymptotic result for the breather eigenvalue λ_+ of (5.2) and the corresponding full numerical result for λ_+ computed from (5.2).

A similar favorable agreement between the asymptotic and numerical results for $\lambda_\pm$ occurs for $m > 0$. As a function of m, in Figures 23(a) and 23(b) we show a favorable comparison between the asymptotic results for λ_- and λ_+, respectively, and the corresponding full numerical results computed from (5.2) when $\varepsilon_0 = 0.02$ and $\kappa = 5.0$.

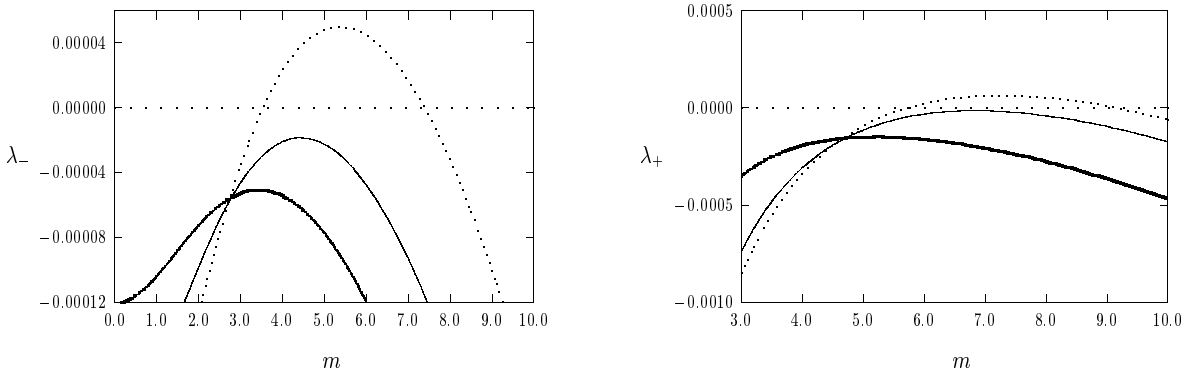
For both λ_- and λ_+ we have $\lambda_\pm < 0$ when $m = 0$ and $\lambda_\pm < 0$ for $m \gg 1$. This latter inequality is readily seen from the estimate $\sigma_\pm = O(m^{-1})$ for $m \gg 1$ obtained from (5.16). For α sufficiently small, or equivalently for $\mathcal{D}$ sufficiently small, it follows from (5.17) that there will be a band of unstable zigzag and breather modes where $m = O(1)$. For three values of $\mathcal{D}$, in Figure 24(a) we plot λ_- versus m, computed from the asymptotic result (5.17), for $\kappa = 2.0$ and $\varepsilon_0 = 0.0025$. A similar plot is shown in Figure 24(b) for $\kappa = 4.0$ and $\varepsilon_0 = 0.0025$. For a fixed value of the saturation parameter, these figures show the emergence of $O(1)$ instability bands as $\mathcal{D}$ is decreased.

For the zigzag mode a nontrivial band of unstable modes emerges at the value $\alpha = \alpha_z$ and $m = m_z$, where the tangency conditions $\lambda_- = 0$ and $d\lambda_-/dm = 0$ are satisfied. From these



(a) λ_- versus m (b) λ_+ versus m

Figure 23. (a) *Comparison of asymptotic (heavy solid curve) and full numerical results (dashed curve) for* λ_- *versus* m *when* $\kappa = 5.0$ *and* $\varepsilon_0 = 0.02$. (b) *Comparison of asymptotic (heavy solid curve) and full numerical results (dashed curve) for* λ_+ *versus* m *when* $\kappa = 5.0$ *and* $\varepsilon_0 = 0.02$.



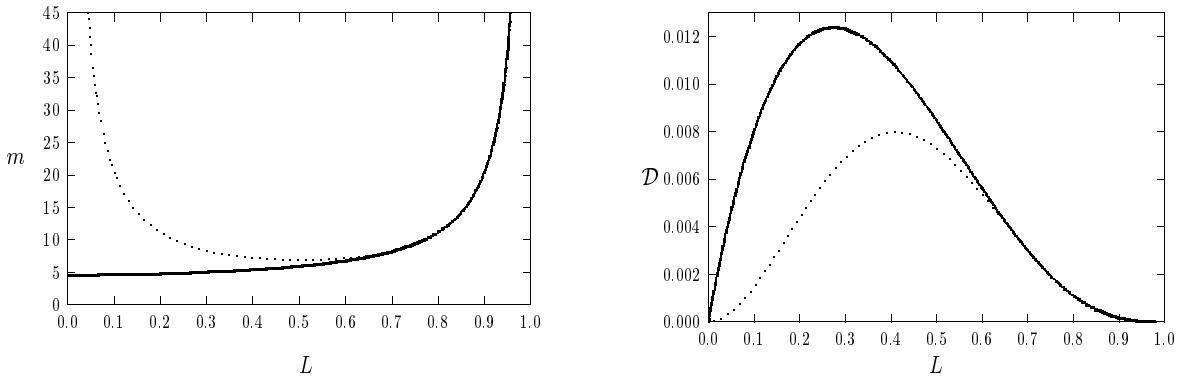
(a) λ_- versus m (b) λ_+ versus m

Figure 24. (a) *the asymptotic result* λ_- *of* (5.17) *versus* m *for* $\kappa = 2.0$ *and* $\varepsilon_0 = 0.0025$ *when* $D = 10.0$ *(heavy solid curve),* $D = 6.0$ *(solid curve), and* $D = 4.0$ *(dashed curve). An instability band exists when* $D = 4.0$ *since* $\mathcal{D} = D\varepsilon_0 = 0.010 < \mathcal{D}_z$. (b) *the asymptotic result* λ_+ *versus* m *of* (5.17) *for* $\kappa = 4.0$ *and* $\varepsilon = 0.0025$ *when* $D = 10.0$ *(heavy solid curve),* $D = 3.5$ *(solid curve), and* $D = 2.8$ *(dashed curve). An instability band exists when* $D = 2.8$ *since* $\mathcal{D} = D\varepsilon_0 = 0.007 < \mathcal{D}_b$.

conditions, we obtain that m_z is the root of

$$\frac{m}{2}\frac{d\sigma_-}{dm} = \sigma_- - \frac{L}{2}(1-L)\,. \tag{5.20}$$

Here m_z depends on the plateau length L, which depends on the saturation parameter κ



(a) m_z and m_b versus L

(b) $\mathcal{D}_z$ and $\mathcal{D}_b$ versus L

Figure 25. (a) *the critical modes m_z and m_b versus L for the emergence of a zigzag (heavy solid curve) and a breather (dashed curve) instability band, respectively.* (b) *the critical diffusivities $\mathcal{D}_z$ (heavy solid curve) and $\mathcal{D}_b$ (dashed curve) at the modes m_z and m_b, where an unstable zigzag and breather instability respectively, emerge. The mesa-stripe solution is stable for values of $\mathcal{D}$ that lie above both curves in the right figure.*

from (4.5). In terms of $m = m_z$, the critical value α_z is $\alpha_z = m_z^{-2}\left[\frac{L}{2}(1-L) - \sigma_-\right]$. Since α is related to $\mathcal{D}$ by (5.17), this latter formula defines a critical value $\mathcal{D}_z$ for the inhibitor diffusivity as a function of either L or κ as

$$\mathcal{D}_z = \frac{w_+^2}{2\beta L m_z^2}\left[\frac{L}{2}(1-L) - \sigma_-\right]. \tag{5.21}$$

To compute the curve m_z at each fixed L we use Newton's method coupled to a continuation procedure in L starting from $L \ll 1$. For $L \ll 1$, a simple calculation using (5.16b) for σ_- shows that

$$\sigma_- \sim \frac{L}{2} - \frac{mL^2}{4}\tanh\left(\frac{m}{2}\right) + O(L^3). \tag{5.22}$$

Therefore, from (5.20) we obtain for $L \ll 1$ that $m_z \approx 4.5298$ is the unique root of

$$-\frac{m}{4}\tanh\left(\frac{m}{2}\right) + \frac{m^2}{8}\mathrm{sech}^2\left(\frac{m}{2}\right) = -1. \tag{5.23}$$

In Figure 25(a) we plot m_z versus L, and in Figure 25(b) we plot the critical diffusivity $\mathcal{D}_z$ versus L. The relation between L and κ in Figure 21 then determines these critical values in terms of the saturation parameter. In Figure 25(b), the maximum value of the curve $\mathcal{D}_z$ versus L occurs at $\mathcal{D}_z \approx 0.0124$ and $m_z = 4.92$ when $L \approx 0.277$, or equivalently $\kappa = 1.92$ from Figure 21.

A similar calculation can be done for the breather instability corresponding to λ_+. To determine the value $m = m_b$ and $\mathcal{D} = \mathcal{D}_b$ where a nontrivial breather instability band first

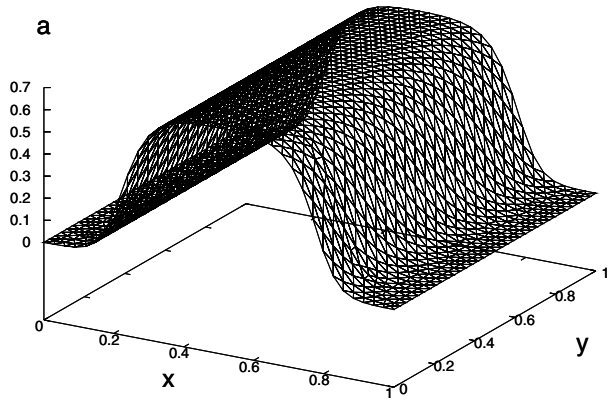


Figure 26. Experiment 9. *The numerical solution to* (1.6) *in a square domain* $[0,1]\times[0,1]$ *at time* $t = 1000$ *for the parameter set* $\varepsilon_0 = 0.03$, $\kappa = 5.0$, *and* $D = 10$ *Since* $\mathcal{D} = D\varepsilon_0 = 0.3 > \mathcal{D}_z = .0124$, *the mesa-stripe solution is stable to both zigzag and breather instabilities.*

emerges, we set $\lambda_+ = d\lambda_+/dm = 0$ to obtain

$$\frac{m}{2}\frac{d\sigma_+}{dm} = \sigma_+ - \frac{L}{2}(1-L)\,, \qquad \mathcal{D}_b = \frac{w_+^2}{2\beta L m_z^2}\left[\frac{L}{2}(1-L) - \sigma_+\right]\,. \tag{5.24}$$

In Figure 25(a) we plot m_b versus L, and in Figure 25(b) we plot the critical diffusivity $\mathcal{D}_b$ versus L. Notice that $\mathcal{D}_z > \mathcal{D}_b$, which implies that a zigzag instability occurs before a breather instability as $\mathcal{D}$ is decreased. From Figure 25(a) we also observe that $m_b \gg 1$ for $L \ll 1$. This is readily seen by using (5.16a) for σ_+ in the transcendental relation (5.24) for m_b. In Figure 25(b), the maximum value of the curve $\mathcal{D}_b$ versus L occurs at $\mathcal{D}_b \approx 0.00798$ and $m_b \approx 7.09$ when $L \approx 0.406$, or equivalently $\kappa = 4.11$ from Figure 21.

Finally, we perform a few full numerical simulations on (1.6) to confirm the asymptotic stability theory.

Experiment 9. Consider (1.6) in the square $[0,1]\times[0,1]$ for the parameter set $\varepsilon_0 = 0.03$, $\kappa = 5.0$, and $D = 10$. For these values, (4.5) yields that $L \approx 0.45$ and $\mathcal{H} \approx 0.2056$. The initial condition for (1.6) is taken to be

$$a = \frac{\mathcal{H}w_+}{2}\left(\tanh\left[\frac{\varepsilon_0^{-1}(x_1-\xi_l)}{2}\right] + \tanh\left[\frac{\varepsilon_0^{-1}(\xi_r - x_1)}{2}\right]\right)\,, \qquad h = \mathcal{H}\,, \tag{5.25}$$

where $\xi_l = 0.275$ and $x_r = 0.725$. Since $\mathcal{D} = D\varepsilon_0 = 0.3 > \mathcal{D}_z = .0124$, the asymptotic theory predicts that the mesa-stripe solution is stable to both zigzag and breather instabilities. This is confirmed in Figure 26, where we plot the numerical solution to (1.6) at time $t = 1000$, showing its convergence to a stable mesa-stripe solution.

Experiment 10. Next, we consider (1.6) in the square $[0,1]\times[0,1]$ with $\varepsilon_0 = 0.01$ and $\kappa = 1.92$, and for various values of D. The initial condition for (1.6) is (5.25) with $L \approx 0.28$,

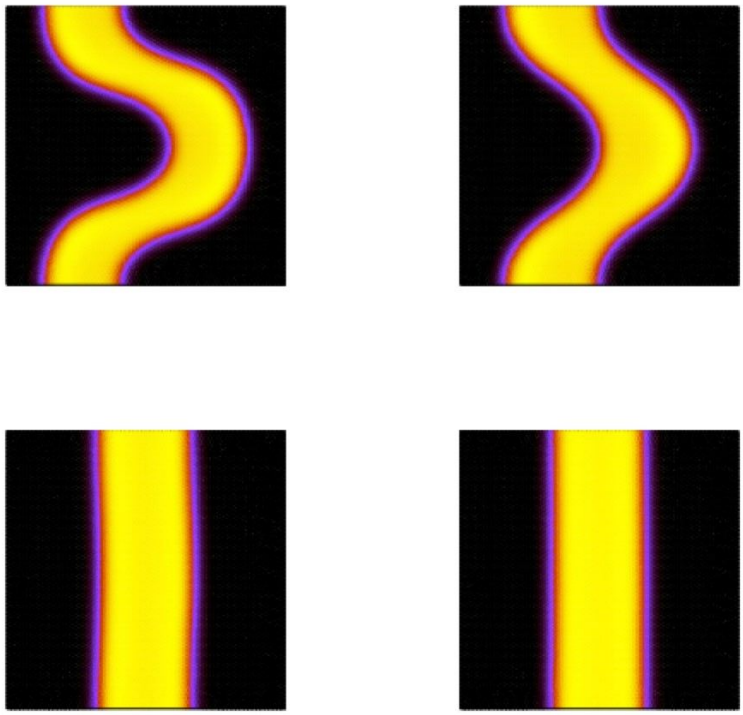


Figure 27. Experiment 10. *The numerical solution to* (1.6) *in a square domain* $[0,1]\times[0,1]$ *for* $\varepsilon_0 = 0.01$ *and* $\kappa = 1.92$. *Top left:* $D = 0.6$ *at time* $t = 10,000$. *Top right:* $D = 0.8$ *at time* $t = 10,000$. *Bottom left:* $D = 1.0$ *at time* $t = 20,000$. *Bottom right:* $D = 1.4$ *at time* $t = 20,000$. *A zigzag instability occurs in each case, except for the value* $D = 1.4$, *which is above the zigzag instability threshold.*

$\mathcal{H} = 0.33$, $\xi_l = 0.36$, and $\xi_r = 0.64$. From Figure 25(b), $\kappa = 1.92$ corresponds to the maximum point $\mathcal{D}_z \approx 0.0124$ of the curve $\mathcal{D}_z$ versus L. Since $\varepsilon_0 = 0.01$ and $D = \mathcal{D}/\varepsilon_0$, the asymptotic theory predicts that a zigzag instability occurs when $D < D_z = 1.24$. Although this asymptotic stability result was derived in the limit $D \gg 1$, we now show that it is in reasonable quantitative agreement with full numerical simulations even when $D \approx 1$. In Figure 27 we plot the numerical solution to (1.6) for $D = 0.6$, $D = 0.8$, $D = 1.0$, and $D = 1.4$, at the times shown in the figure caption. The mesa-stripe has a pronounced zigzag instability for $D = 0.6$ and $D = 0.8$, but only a very slight instability for $D = 1.0$. For these parameter values and for the domain width $d_0 = 1$, the asymptotic theory predicts that the unstable zigzag mode has exactly one crest. For $D = 1.4$, which is above the zigzag threshold $D = 1.24$, the mesa-stripe is found to be stable. We believe that the saturation of the zigzag instability leading to the apparent steady-state solution in Figure 27 is a result of the interaction of the global inhibitor field h with the lateral boundaries of the rectangle. We remark that in order to give a more precise test of the instability threshold, one would have to compute numerical solutions of (1.6) with a value of ε_0 that is a decade smaller than $\varepsilon_0 = 0.01$. With such a small value of ε_0, it is challenging to obtain sufficient numerical resolution to resolve the transition layers at the edges of the mesa-stripe.

Experiment 11. Finally, we consider (1.6) in the square $[0,1]\times[0,1]$ with $k = 2.0$. We

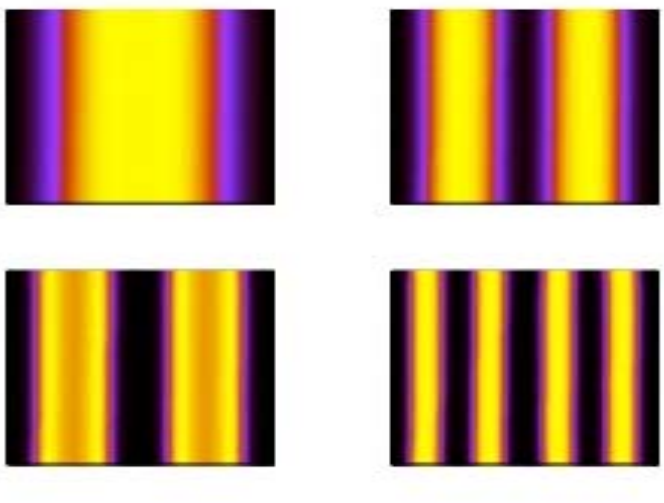


Figure 28. Experiment 11. *The numerical solution to* (1.6) *in a square domain* $[0,1]\times[0,1]$ *for* $k=2.0$ *when* ε_0 *and* D *are slowly decreasing functions of time given in* (5.26). *The initial condition is given in* (5.25) *with* $\mathcal{H}=0.21$, $\xi_l=0.36$, *and* $\xi_r=0.64$. *The mesa-stripe is found to undergo a self-replication process leading to a multistripe pattern. Top left:* $t=100$. *Top right:* $t=140$. *Bottom left:* $t=160$. *Bottom right:* $t=180$. *See also the accompanying animation* (63508_04.gif [731KB]).

take ε_0 and D to be slowly decreasing functions of time given by

$$\varepsilon_0(t)=0.2D(t)\,,\qquad D(t)=0.2e^{-0.002t}\,. \tag{5.26}$$

The initial condition for (1.6) is (5.25) with $\mathcal{H}=0.21$, $\xi_l=0.36$, and $\xi_r=0.64$. Since D is not asymptotically large, the theory developed in sections 4 and 5 does not apply for this example. The numerical results in Figure 28 show a new phenomenon whereby the initial mesa-stripe splits into two, with the two daughter stripes undergoing a further splitting at later times. Since the time-scale for splitting is much less than that for the development of transverse instabilities, there are no zigzag instabilities observed in Figure 28. This stripe-replication phenomenon is significantly more robust than that observed in Experiment 6 of section 3 for the unsaturated GM model in the weak interaction regime. An analysis of this mesa self-replication phenomena of Figure 28 is an open problem.

6. Conclusion. We have analyzed the stability of a stripe for two different forms of the GM activator-inhibitor model in a rectangular domain. For the basic GM model, where saturation effects are neglected, the stability of a homoclinic stripe was analyzed with respect to spot-generating breakup instabilities and transverse zigzag instabilities. The wave number instability bands for each of these mechanisms were found to depend sensitively on the asymptotic range of the inhibitor diffusivity D. In the semistrong regime, where $D=O(1)$, the homoclinic stripe typically disintegrates into an array of spots unless the domain width is asymptotically small. In contrast, in the weak interaction regime, where the activator and inhibitor diffusivities have the same asymptotic order, there are certain exponent sets associated with the nonlinear kinetics where the homoclinic stripe can be destabilized solely by

a transverse zigzag instability. In the semistrong regime it was also shown that a homoclinic stripe can be stabilized with respect to breakup instabilities upon allowing for an asymptotically small level of activator saturation. For larger levels of the saturation, the homoclinic stripe ceases to exist and is replaced by a mesa-stripe, whose cross section consists of front-back transition layers joined by an asymptotically flat plateau. For an asymptotically large inhibitor diffusivity, it was shown that such a mesa-stripe can be stable with respect to both zigzag and breakup instabilities.

There are some open problems suggested by this study. For homoclinic stripe solutions of (1.4), a key open problem is to provide an analytical theory that characterizes the intricate nature of the zigzag and breakup instability bands for a homoclinic stripe in the weak interaction regime. For a mesa-stripe solution of (1.6), an interesting open problem is to rigorously study the transition behavior in the stability properties of a stripe as the saturation parameter κ decreases. In particular, for $\kappa = O(\varepsilon_0^2)$ in (2.23), where "fattened" homoclinic stripes occur, it would be interesting to give a rigorous analytical confirmation of the disappearance of the spot-generating breakup instability band shown numerically in section 2.3. Additionally, it would be interesting to construct multiple mesa-stripe equilibria to (1.6) when $D = O(1)$ in order to study the global bifurcation properties of these solutions. Such a bifurcation diagram is likely to be crucial for an analysis of the mesa-stripe self-replication behavior observed in Experiment 11 of section 5. Other important open problems include providing a weakly nonlinear theory for zigzag and breakup instabilities of homoclinic stripes, and studying the stability of multistripe patterns.

Finally, it would be interesting to extend the stability analyzes given here to investigate breakup and zigzag instabilities of stripes in the hybrid chemotaxis reaction-diffusion systems of [36], [45], [30], and [31], and in the models of [9] and [19] for the spatial patterning of vegetation in arid environments.

Appendix A. The nonlocal eigenvalue problem: Semistrong regime. We first outline the derivation of the NLEP (2.9). In terms of the inner variable $y = x_1/\varepsilon$, we use (2.1) to calculate $pa_e^{p-1}/h_e^q \sim pw^{p-1}$ and $qa_e^p/h_e^{q+1} \sim q\mathcal{H}^{\gamma p-(q+1)}w^p$. Therefore, from (2.6a), $\Phi(y)$ satisfies

$$\Phi'' - \Phi + pw^{p-1}\Phi - q\mathcal{H}^{\gamma p-(q+1)}w^p\eta(0) = \left(\lambda + \varepsilon^2 m^2\right)\Phi\,, \quad -\infty < y < \infty\,, \qquad \Phi \to 0 \quad \text{as} \quad |y| \to \infty\,. \tag{A.1}$$

In (2.6b), η is not singularly perturbed, and so we can determine $\eta(0)$ in (A.1) from the outer solution for $\eta(x_1)$. To do this, we use (2.1) and (2.3) to calculate the coefficients in (2.6b) in terms of Dirac masses. This leads to

$$\eta_{x_1x_1} - \theta_\lambda^2\eta = 0\,, \quad -l < x_1 < l\,, \qquad \eta_{x_1}(\pm l) = 0\,, \qquad \theta_\lambda \equiv \sqrt{1 + m^2 + \tau\lambda}\,, \tag{A.2a}$$

$$[\eta] = 0\,, \qquad [\eta_{x_1}] = \frac{s\eta(0)}{G_l(0)} - \frac{r\mathcal{H}^{-\gamma+1}}{G_l(0)}\left(\frac{\int_{-\infty}^{\infty} w^{r-1}\Phi\, dy}{\int_{-\infty}^{\infty} w^r\, dy}\right)\,. \tag{A.2b}$$

Here $[f] \equiv f(0^+) - f(0^-)$. To solve (A.2), we introduce the Green function $G_\lambda(x_1)$ satisfying

(A.3)
$$G_{\lambda x_1 x_1} - \theta_\lambda^2 G_\lambda = -\delta(x_1)\,, \quad |x_1| < l\,, \qquad G_{\lambda x_1}(\pm l) = 0\,, \qquad G_\lambda(x_1) = \frac{\cosh\left[\theta_\lambda\left(l - |x_1|\right)\right]}{2\theta_\lambda \sinh\left(\theta_\lambda l\right)}\,.$$

The solution $\eta(x_1)$ can be written in terms of $G_\lambda(x_1)$, and in this way we determine $\eta(0)$ as

$$\eta(0) = r\mathcal{H}^{1-\gamma}\left[s + \frac{\theta_\lambda \tanh(\theta_\lambda l)}{\tanh l}\right]^{-1}\left(\frac{\int_{-\infty}^{\infty} w^{r-1}\Phi\, dy}{\int_{-\infty}^{\infty} w^r\, dy}\right)\,. \tag{A.4}$$

Upon substituting (A.4) into (A.1), we obtain the eigenvalue problem (2.9) in Principal Result 2.2.

In the remainder of this appendix we prove Proposition 2.3 for the spectrum of (2.9). We begin with a key lemma, which is an extension of a result of [41].

Lemma A.1 (from [38]). *Let $f(\mu)$ and $C_m(\lambda)$ be as defined in (2.11) and (2.9b), and let $\nu_0 > 0$ be the unique positive eigenvalue of the local operator L_0 in (2.9). Then, when μ and λ are real, the following properties hold:*

(A.5a)
$$\text{(i)}\quad f(0) = \frac{1}{p-1}\,, \qquad f'(0) = \frac{1}{p-1}\left[\frac{1}{p-1} - \frac{1}{2r}\right]\,,$$

(A.5b)
$$\text{(ii)}\quad f'(\mu) > 0 \ \text{for}\ 0 < \mu < \nu_0\,, \quad \text{when either}\ r = 2\,,\ 1 < p \le 5\,, \quad \text{or} \quad r = p+1\,,\ p > 1\,,$$

(A.5c)
$$\text{(iii)}\quad f''(\mu) > 0 \ \text{for}\ 0 < \mu < \nu_0\,, \quad \text{when either}\ r = 2\,,\ p = 2\,, \quad \text{or} \quad r = p+1\,,\ 1 < p \le 5\,,$$

(A.5d)
$$\text{(iv)}\quad f(\mu) \to +\infty \ \text{as}\ \mu \to \nu_0^-\,, \qquad f(\mu) < 0 \ \text{for}\ \mu > \nu_0\,.$$

In addition, $C_m'(\lambda) > 0$, $C_m''(\lambda) < 0$, $\frac{dC_m}{dm}(0) > 0$, and $C_m(\lambda) = O(m)$ for $m \gg 1$, where $C_m(\lambda)$ is given in (2.9b).

Proof. The proof of (A.5) is given in Proposition 3.5 of [38]. The positivity of $C_m'(\lambda)$, the concavity of $C_m(\lambda)$, and the positivity of $\frac{dC_m(0)}{dm}$, all follow from a simple direct calculation using the expression for $C_m(\lambda)$ in (2.9b). ∎

The proof of Proposition 2.3 is given in two parts. In part 1, Lemma A.1 is used to analyze the spectrum of the NLEP (2.9) on the positive real axis. In part 2 a winding number criterion locates any complex unstable spectrum.

We begin with part 1. By calculating $C_m(0)$ in (2.9b) and by using $\frac{dC_m}{dm}(0) > 0$, we obtain that $C_m(0) > 1/(p-1)$ when $m > m_{b-}$, where m_{b-} is the unique root of the transcendental equation (2.12) in Proposition 2.3. Then, with $\varepsilon m \ll 1$, it follows from (A.5) and the conditions $C_m'(\lambda) > 0$ and $C_m''(\lambda) < 0$, that the curves $C_m(\lambda)$ and $f(\lambda + \varepsilon^2 m^2)$ intersect exactly once in $0 < \lambda < \nu_0$ for any $\tau \ge 0$ when the condition (A.5c) on the exponents r and p are satisfied. Therefore, under this condition, we conclude for $m > m_{b-}$ and $\varepsilon m \ll 1$ that there

is a unique real root in $0 < \lambda < \nu_0$ to $g(\lambda) = 0$ defined in (2.11). Under these conditions, we obtain a unique unstable real eigenvalue of (2.9).

Next, we consider real spectrum on the range $0 < m < m_{b-} = O(1)$, for which $C_m(0) < f(\varepsilon^2 m^2) \sim \frac{1}{p-1}$. Since $C_m(\lambda) = O(\tau^{1/2})$ for $\tau \gg 1$, it follows from (A.5) that for $\tau > \tau_m$ there are exactly two real roots to $g(\lambda) = 0$ in $0 < \lambda < \nu_0$ when the condition (A.5c) on the exponents holds. This yields two unstable real eigenvalues. Alternatively, for $\tau < \tau_m$, there are no real roots to $g(\lambda) = 0$, and hence no unstable real eigenvalues. For $m = m_{b-}$ a simple calculation shows that $\lambda = 0$ is a double zero eigenvalue when $C_m'(0) = f'(\varepsilon^2 m^2)$. By using (2.9b) for C_m, (A.5a) for $f'(0)$, and assuming that $p-1 < 2r$, this condition yields the critical value $\tau_{m-} > 0$ of τ, defined in (2.14) of Proposition 2.3. For $m = m_{b-}$ and $\tau > \tau_{m-}$ there is a unique unstable real eigenvalue of (2.9).

Next, we consider real spectrum when $m > \sqrt{\nu_0}/\varepsilon$. For this range of m, (A.5d) shows that $f(\lambda + \varepsilon^2 m^2) < 0$ for any $\lambda > 0$. Thus, since $C_m(0) > 0$ and $C_m'(\lambda) > 0$, there are no real roots to $g(\lambda) = 0$ in $\lambda > 0$ for any $\tau > 0$, and consequently no unstable real eigenvalues of (2.9). Finally, suppose that $m = \sqrt{\beta}/\varepsilon$ with $0 < \beta < \nu_0$, so that $g(\lambda) = C_m(\lambda) - f(\lambda + \beta)$. Then, since $C_m(0) = O(m)$ for $m \gg 1$, we get $C_m(0) = O(\varepsilon^{-1}) > f(\beta)$ for $0 < \beta < \nu_0$. Therefore, under the condition (A.5c) on the exponents r and p, there is a unique root to $g(\lambda) = 0$ in $0 < \lambda < \nu_0 - \beta$ for any $\tau \geq 0$, and consequently a unique unstable real eigenvalue to (2.9).

In part 2 of the proof we must count the number N of complex eigenvalues in the right half-plane $\mathrm{Re}(\lambda) > 0$. To do so, we proceed as in section 3 of [38] by using a winding number criterion that determines N in terms of the change in the argument of $g(i\lambda_I)$ along the positive imaginary axis Γ_I, denoted by $[\arg g]_{\Gamma_I}$, traversed in the downwards direction. For any $\tau > 0$, a slight modification of Proposition 3.3 of [38] shows that

$$N = \frac{5}{4} + \frac{1}{\pi}[\arg g]_{\Gamma_I}\,, \quad 0 < m < \frac{\sqrt{\nu_0}}{\varepsilon}\,, \qquad N = \frac{1}{4} + \frac{1}{\pi}[\arg g]_{\Gamma_I}\,, \quad m > \frac{\sqrt{\nu_0}}{\varepsilon}\,. \tag{A.6}$$

Assume that the exponents satisfy $r = 2$ and $p > 1$. Then, by adapting the proof of Proposition 3.4 of [38], we conclude that $[\arg g]_{\Gamma_I} = -\pi/4$ when $m > m_{b-}$ and $\tau > 0$. This yields $N = 1$ when $m_{b-} < m < \sqrt{\nu_0}/\varepsilon$ and $N = 0$ when $m > \sqrt{\nu_0}/\varepsilon$. Therefore, the unstable eigenvalue for $m_{b-} < m < \sqrt{\nu_0}/\varepsilon$ is the real positive eigenvalue obtained in part 1 of the proof. For the range $0 < m < m_{b-}$, Proposition 3.4 of [38] can be applied directly, and for $r = 2$ and $p > 1$ we conclude that $[\arg g]_{\Gamma_I} = 3\pi/4$ when τ is sufficiently large and $[\arg g]_{\Gamma_I} = -5\pi/4$ when τ is sufficiently small. For $0 < m < m_{b-}$ with $r = 2$, this shows that $N = 2$ when τ is sufficiently large and $N = 0$ when τ is sufficiently small. Therefore, for this range of the parameters, there is a Hopf bifurcation as τ is increased past some critical value, which generates unstable complex conjugate eigenvalues. Our results above for the positive real axis $\lambda > 0$ show that these unstable complex eigenvalues must merge onto this axis when τ is sufficiently large. This completes the proof of Proposition 3.3. ■

Appendix B. Zigzag eigenvalue: Semistrong regime. In this appendix we outline the derivation of (2.21). We first write (2.6a) in terms of an operator L_ε as

$$L_\varepsilon \phi - \frac{q a_e^p}{h_e^{q+1}}\eta = (\lambda + \varepsilon^2 m^2)\phi\,, \quad -l < x < l\,, \quad \phi_x(\pm l) = 0\,, \qquad L_\varepsilon \phi \equiv \varepsilon^2 \phi_{xx} - \phi + \frac{p a_e^{p-1}}{h_e^q}\phi\,. \tag{B.1}$$

We differentiate the equilibrium problem for a_e in (1.4) with respect to x_1 to get $L_\varepsilon a_{ex_1} = q a_e^p h_{ex_1}/h_e^{q+1}$, where L_ε is defined in (B.1). Therefore, since $a_e \sim \mathcal{H}^\gamma w$, we obtain

$$L_\varepsilon w' \sim \frac{\varepsilon q \mathcal{H}^q w^p}{h_e^{q+1}} h_{ex_1} \,. \tag{B.2}$$

This suggests that we expand ϕ and η as

$$\phi = w' + \varepsilon\phi_1 + \cdots \,, \qquad \eta(x_1) = \varepsilon\eta_0(x_1) + \cdots \,. \tag{B.3}$$

We substitute (B.3) into (B.1) and use (B.2) with $\lambda = O(\varepsilon^2)$ and $m = O(1)$. This yields that $\phi_1(y)$ satisfies

$$L_\varepsilon \phi_1 \sim \frac{q w^p \mathcal{H}^q}{h_e^{q+1}} f(\varepsilon y) \,, \qquad f(x_1) \equiv \mathcal{H}^\gamma \eta_0(x_1) - h_{ex_1}(x_1) \,. \tag{B.4}$$

By substituting (B.3) and $\eta = \varepsilon\eta_0$ into (2.6b) and labeling $\theta_\lambda \equiv \sqrt{1 + m^2 + \tau\lambda}$, we get that η_0 satisfies

$$\eta_{0x_1x_1} - \theta_\lambda^2 \eta_0 = -\frac{r a_e^{r-1}}{\varepsilon^2 h_e^s}\left(w' + \varepsilon\phi_1\right) + \frac{s a_e^r}{\varepsilon h_e^{s+1}}\eta_0 \,, \quad -l < x < l \,, \qquad \eta_{0x_1}(\pm l) = 0 \,. \tag{B.5}$$

The term proportional to w' on the right-hand side of (B.5) behaves like a dipole as $\varepsilon \to 0$. Therefore, for $\varepsilon \to 0$, it can be represented as a multiple of $\delta'(x_1)$, where $\delta(x_1)$ is the delta function. Thus, η_0 is discontinuous across $x_1 = 0$. However, $f(x_1)$ defined in (B.4) is continuous across $x_1 = 0$. To see this, we differentiate (1.4) for h_e with respect to x_1 and then subtract appropriate multiples of the resulting equation and (B.5) to find that the dipole term cancels exactly. Thus, $f(x_1)$ is continuous across $x_1 = 0$, and so $\langle f \rangle = f(0)$. Since $\langle h_{ex} \rangle = 0$ from (2.1), we get $f(0) = \mathcal{H}^\gamma \langle \eta_0 \rangle$. Here and below we have defined $\langle \xi \rangle \equiv (\xi(0^+) + \xi(0^-))/2$ and $[\xi] \equiv \xi(0^+) - \xi(0^-)$, where $\xi(0^\pm)$ are the one-sided limits of $\xi(x_1)$ as $x_1 \to 0^\pm$. Therefore, for $\varepsilon \ll 1$, ϕ_1 in (B.4) satisfies

$$L_\varepsilon \phi_1 \sim q w^p \mathcal{H}^{\gamma-1} \langle \eta_0 \rangle \,. \tag{B.6}$$

Since $L_\varepsilon w = (p-1)w^p + O(\varepsilon)$, the solution to (B.6) is simply

$$\phi_1(y) = \frac{q}{p-1} w(y) \mathcal{H}^{\gamma-1} \langle \eta_0 \rangle + O(\varepsilon) \,. \tag{B.7}$$

Next, we use (2.1), (2.3), (B.3), and (B.7) to calculate the coefficients in (B.5) in the sense of distributions. With $\tilde{\eta}_0$ defined by $\eta_0 = H^{1-\gamma}\tilde{\eta}_0$, and by using $G_l(0) = \frac{1}{2}\coth l$ as given in (2.3), we obtain that $\tilde{\eta}_0$ satisfies

$$\tilde{\eta}_{0x_1x_1} - \theta_\lambda^2 \tilde{\eta}_0 = 0 \,, \quad -l < x_1 < l \,, \qquad \tilde{\eta}_{0x_1}(\pm l) = 0 \,, \qquad \theta_\lambda \equiv \sqrt{1 + m^2 + \tau\lambda} \,, \tag{B.8a}$$

$$[\tilde{\eta}_0] = -\frac{1}{G_l(0)} \,, \qquad [\tilde{\eta}_{0x_1}] = \left(s - \frac{qr}{p-1}\right)\frac{\langle \tilde{\eta}_0 \rangle}{G_l(0)} \,. \tag{B.8b}$$

The remaining part of the derivation proceeds as in (4.16)–(4.23) of [10]. This leads to

$$(\lambda+\varepsilon^2 m^2) \sim \frac{\varepsilon^2 q J}{p+1}\left(\langle \tilde{\eta}_{0x_1}\rangle - \frac{h_{ex_1x_1}(0)}{\mathcal{H}}\right), \qquad J \equiv \frac{\int_{-\infty}^{\infty} w^{p+1}\,dy}{\int_{-\infty}^{\infty} w'^2\,dy}. \tag{B.9}$$

From Appendix A of [37] we calculate $J = 2(p+1)/(p-1)$. Then, from (2.1), we obtain $h_{ex_1x_1}(0) = \mathcal{H}$. Finally, by solving (B.8) explicitly, we calculate

$$\langle \tilde{\eta}_{0x_1}\rangle = \theta_\lambda \tanh l \tanh(\theta_\lambda l)\,. \tag{B.10}$$

Upon substituting these formulae into (B.9), we obtain that the small eigenvalue $\lambda = O(\varepsilon^2)$ satisfies (2.21).

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK User's Guide: Third Edition*, SIAM, Philadelphia, 1999.

[2] U. Ascher, R. Christiansen, and R. Russell, *Collocation Software for Boundary Value ODE's*, Math. Comp., 33 (1979), pp. 659–679.

[3] A. Doelman, R. A. Gardner, and T. Kaper, *Large stable pulse solutions in reaction-diffusion equations*, Indiana Univ. Math. J., 50 (2001), pp. 443–507.

[4] A. Doelman and H. van der Ploeg, *Homoclinic stripe patterns*, SIAM J. Appl. Dyn. Sys., 1 (2002), pp. 65–104.

[5] S. Ei, Y. Nishiura, and K. Ueda, 2^n *Splitting or edge splitting?: A manner of splitting in dissipative systems*, Japan J. Indust. Appl. Math., 18 (2001), pp. 181–205.

[6] A. Gierer and H. Meinhardt, *A theory of biological pattern formation*, Kybernetik, 12 (1972), pp. 30–39.

[7] R. E. Goldstein, D. J. Muraki, and D. M. Petrich, *Interface proliferation and the growth of labyrinths in a reaction-diffusion system*, Phys. Rev. E., 53 (1996), pp. 3933–3957.

[8] P. Hirschberg and E. Knobloch, *Zigzag and varicose instabilities of a localized stripe pattern*, Chaos, 3 (1993), pp. 713–721.

[9] J. von Hardenberg, E. Meron, M. Shachak, and Y. Zarmi, *Diversity of vegetation patterns and desertification*, Phys. Rev. Lett., 87 (2001), paper 198101.

[10] D. Iron, M. J. Ward, and J. Wei, *The stability of spike solutions to the one-dimensional Gierer–Meinhardt model*, Phys. D, 150 (2001), pp. 25–62.

[11] B. S. Kerner and V. V. Osipov, *Autosolitions: A New Approach to Problem of Self-Organization and Turbulence*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[12] C. A. Klasmeier, *Regular and irregular patterns in semiarid vegetation*, Science, 284 (1999), pp. 1826–1828.

[13] A. J. Koch and H. Meinhardt, *Biological pattern formation from basic mechanisms to complex structures*, Rev. Modern Phys., 66 (1994), pp. 1481–1507.

[14] T. Kolokolnikov, M. J. Ward, and J. Wei, *Pulse-splitting for some reaction-diffusion systems in one-space dimension*, Stud. Appl. Math., 114 (2005), pp. 115–165.

[15] T. Kolokolnikov and J. Wei, *On ring-like solutions for the Gray-Scott model: Existence, instability, and self-replicating regimes*, European J. Appl. Math., 16 (2005), pp. 201–237.

[16] T. Kolokolnikov, M. J. Ward, and J. Wei, *Zigzag and breakup instabilities of stripes and rings for the two-dimensional Gray–Scott model*, Stud. Appl. Math., 116 (2006), pp. 35–95.

[17] T. Kolokolnikov, T. Erneux, and J. Wei, *Mesa-type patterns in the one-dimensional Brusselator and their stability*, Phys. D, 214 (2006), pp. 63–77.

[18] S. Kondo and R. Asai, *A viable reaction-diffusion wave on the skin of pomacanthus, a marine angelfish*, Nature, 376 (1995), pp. 765–768.

[19] E. Meron, E. Gilad, J. von Hardenberg, M. Shachak, and Y. Zarmi, *Vegetation patterns along a rainfall gradient*, Chaos Solitons and Fractals, 19 (2004), pp. 367–376.
[20] C. S. Lin, W. M. Ni, and I. Takagi, *Large amplitude stationary solutions to a chemotaxis system*, J. Differential Equations, 72 (1988), pp. 1–27.
[21] H. Meinhardt, *Models of Biological Pattern Formation*, Academic Press, London, 1982.
[22] H. Meinhardt, *The Algorithmic Beauty of Sea Shells*, Springer-Verlag, Berlin, 1995.
[23] D. Morgan and T. Kaper, *Axisymmetric ring solutions of the Gray–Scott model and their destabilization into spots*, Phys. D, 192 (2004), pp. 33–62.
[24] J. D. Murray, *Mathematical Biology* II: *Spatial Models and Biomedical Applications*, Interdiscip. Appl. Math. 18, Springer, New York, 2003.
[25] Y. Nishiura, *Far-From-Equilibrium Dynamics*, Trans. Math. Monogr. 209, AMS, Providence, RI, 2002.
[26] Y. Nishiura and D. Ueyama, *A skeleton structure of self-replicating dynamics*, Phys. D, 130 (1999), pp. 73–104.
[27] Y. Nishiura and D. Ueyama, *Spatio-temporal chaos for the Gray–Scott model*, Phys. D, 150 (2001), pp. 137–162.
[28] Y. Nishiura and H. Fujii, *Stability of singularly perturbed solutions to systems of reaction-diffusion equations*, SIAM J. Math. Anal., 18 (1987), pp. 1726–1770.
[29] Y. Nishiura, *Singular limit approach to stability and bifurcation for bistable reaction-diffusion systems*, Rocky Mountain J. Math., 21 (1991), pp. 727–767.
[30] K. J. Painter, P. K. Maini, and H. G. Othmer, *Complex patterns in a hybrid chemotaxis reaction-diffusion model*, J. Math. Biol., 41 (2000), pp. 285–314.
[31] K. J. Painter, P. K. Maini, and H. G. Othmer, *Stripe formation in juvenile pomacanthus explained by a generalized turing mechanism with chemotaxis*, Proc. Natl. Acad. Sci. USA, 96 (1999), pp. 5549–5554.
[32] J. E. Pearson, *Complex patterns in a simple system*, Science, 216 (1993), pp. 189–192.
[33] W. Sun, M. J. Ward, and R. Russell, *Localized Solutions, Dynamic Bifurcations, and Self-Replication of Spots for the Gierer–Meinhardt Model in Two Spatial Dimensions*, in preparation.
[34] M. Taniguchi and Y. Nishiura, *Instability of planar interfaces in reaction-diffusion systems*, SIAM J. Math. Anal., 25 (1994), pp. 99–134.
[35] M. Taniguchi and Y. Nishiura, *Stability and characteristic wavelength of planar interfaces in the large diffusion limit of the inhibitor*, Proc. Roy. Soc. Edinburgh Sect. A, 126 (1996), pp. 117–145.
[36] R. C. Tyson, *Pattern Formation by E. Coli; Mathematical and Numerical Investigation of a Biological Phenomenon*, Ph. D. thesis, Department of Applied Mathematics, University of Washington, Seattle, WA, 1996.
[37] M. J. Ward, D. McInerney, P. Houston, D. Gavaghan, and P. Maini, *The dynamics and pinning of a spike for a reaction-diffusion system*, SIAM J. Appl. Math., 62 (2002), pp. 1297–1328.
[38] M. J. Ward and J. Wei, *Hopf bifurcations and oscillatory instabilities of spike solutions for the one-dimensional Gierer–Meinhardt model*, J. Nonlinear Sci., 13 (2003), pp. 209–264.
[39] M. J. Ward and J. Wei, *Hopf bifurcation of spike solutions for the shadow Gierer–Meinhardt model*, European J. Appl. Math., 14 (2003), pp. 677–711.
[40] J. Wei, *On single interior spike solutions for the Gierer–Meinhardt system: Uniqueness and stability estimates*, European J. Appl. Math., 10 (1999), pp. 353–378.
[41] J. Wei and M. Winter, *On the Gierer–Meinhardt system with saturation*, Commun. Contemp. Math., 6 (2004), pp. 259–277.
[42] J. Wei and M. Winter, *Spikes for the two-dimensional Gierer–Meinhardt system: The weak coupling case*, J. Nonlinear Sci., 11 (2001), pp. 415–458.
[43] J. Wei and M. Winter, *Existence and stability analysis of asymmetric patterns for the Gierer–Meinhardt system*, J. Math. Pures Appl. (9), 83 (2004), pp. 433–476.
[44] R. Williams, XMORPHIA, website http://www.cacr.caltech.edu/ismap/image.html, Concurrent Supercomputing Facilities, Caltech, Pasadena, CA, 1998.
[45] D. E. Woodward, R. C. Tyson, J. D. Murray, E. O. Budrene, and H. Berg, *Spatio-temporal patterns generated by salmonella typhimurium*, Biophys. J., 68 (2005), pp. 2181–2189.
[46] F. M. Ytreberg and S. Mckay, *Second mechanism for transitions in a reaction-diffusion system*, Phys. Rev. E., 59 (1999), pp. 3376–3381.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 3, pp. 365–402

Accurately Model the Kuramoto–Sivashinsky Dynamics with Holistic Discretization*

T. MacKenzie† and A. J. Roberts†

Abstract. We analyze the nonlinear Kuramoto–Sivashinsky equation to develop accurate discretizations modeling its dynamics on coarse grids. The analysis is based upon center manifold theory, so we are assured that the discretization accurately models the dynamics and may be constructed systematically. The theory is applied after dividing the physical domain into small elements by introducing isolating internal boundaries which are later removed. Comprehensive numerical solutions and simulations show that the holistic discretizations excellently reproduce the steady states and the dynamics of the Kuramoto–Sivashinsky equation. The Kuramoto–Sivashinsky equation is used as an example to show how holistic discretization may be successfully applied to fourth-order, nonlinear, spatio-temporal dynamical systems. This novel center manifold approach is holistic in the sense that it treats the dynamical equations as a whole, not just as the sum of separate terms.

Key words. Kuramoto–Sivashinsky equation, low-dimensional modeling, computational discretizations, multi-scale modeling

AMS subject classifications. 37M99, 37L65, 65M20

DOI. 10.1137/050627733

1. Introduction.

The Kuramoto–Sivashinsky equation, here

$$\frac{\partial u}{\partial t}+4\frac{\partial^4 u}{\partial x^4}+\alpha\left(u\frac{\partial u}{\partial x}+\frac{\partial^2 u}{\partial x^2}\right)=0, \tag{1}$$

was introduced by Sivashinsky [32] as a model of instabilities on interfaces and flame fronts, and by Kuramoto [16] as a model of phase turbulence in chemical oscillations. It receives considerable attention as a model of complex spatio-temporal dynamics [13, 21, 5, 12, e.g.]. In the form (1), with 2π periodic boundary conditions, α is a bifurcation parameter that depends upon the size of the typical pattern [31]. The Kuramoto–Sivashinsky equation includes the mechanisms of linear negative diffusion αu_{xx}, high-order dissipation $4u_{xxxx}$, and nonlinear advection/steepening $\alpha u u_x$. The PDE (1) has strong dissipative dynamics arising from the fourth-order dissipation. Many modes of this PDE decay rapidly because of this strong dissipation. Thus the dynamics are dominated by a relatively few large scale modes. We create and explore the macroscopic modeling of the Kuramoto–Sivashinsky dynamics using holistic discretization as initiated by MacKenzie and Roberts [18].

We study the Kuramoto–Sivashinsky equation here for several reasons. First, the PDE is fourth-order and therefore, following the example of Burgers' equation [25], provides a further

*Received by the editors March 3, 2005; accepted for publication (in revised form) by D. Barkley February 10, 2006; published electronically July 17, 2006.
http://www.siam.org/journals/siads/5-3/62773.html
†Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, Queensland 4352, Australia (tony.mackenzie@gmail.com; aroberts@usq.edu.au, http://www.sci.usq.edu.au/staff/aroberts).

test case for the application of the holistic approach to higher-order dissipative PDEs. Second, the Kuramoto–Sivashinsky equation has analogies with the Navier–Stokes equations of fluid dynamics. Holmes, Lumley, and Berkooz [12] argued that these analogies exist on two levels: in the energy source and dissipation terms of both dynamical systems, and in the reflection and translational symmetries of the Kuramoto–Sivashinsky equation and the spanwise symmetries of the Navier–Stokes equations in the boundary layer. This analogy between symmetries suggests that the Fourier series and corresponding modal interactions are comparable for these two problems. Third, Cross and Hohenberg [5] describe how the Kuramoto–Sivashinsky equation exhibits the complexities of weak turbulence or spatio-temporal chaos. The complex dynamics of the Kuramoto–Sivashinsky equation (1) is a searching test of the performance of the holistic approach to coarse grained modeling of dynamical systems.

Approximate inertial manifolds and variants [11, 9, 10, 1, 14, e.g.] capture the long-term low-dimensional behavior of the Kuramoto–Sivashinsky equation. Most constructions of approximate inertial manifolds are based upon nonlinear Galerkin methods [22, 20, 14, 10, e.g.]. Approximate inertial manifolds are generally constructed by finding *global* eigenfunctions of the linear dynamics. Our approach is similar to these methods, in that we project onto natural solutions of the PDE, and performs nearly as well; see section 4.3. However, in contrast, the holistic approach undertaken here bases analysis upon the *local* dynamics within and between finite elements, and thus we contend it will be more useful in applications; for example, the approach is readily adapted to the modeling of a wide variety of physical boundary conditions [27].

Our approach is to divide the spatial domain into disjoint elements of finite size h (section 2.1). Initially these finite elements are decoupled, and so dissipation would cause solutions to exponentially quickly become constant in each element. We then couple the elements together so that information is exchanged between elements—parameterized by a coupling parameter γ so that $\gamma = 1$ recovers the original Kuramoto–Sivashinsky dynamics. The coupling drives the evolution of the field in each element. Solving the Kuramoto–Sivashinsky PDE within each element, and with the interelement coupling, hence resolves subgrid scale dynamics and their interactions with nearby elements. Crucially, such solutions are determined in constructing the holistic model and so are done only once; the holistic model may then be used many times without any further analysis of subgrid structure, as the subgrid structure has already been incorporated into the closure of the holistic model. Center manifold theory [3, 23, e.g.] provides rigorous support for holistic models as introduced by Roberts [25] for Burgers' equation and discussed in section 2.2.

For example, a low-order analysis, reported in section 3.1, of the Kuramoto–Sivashinsky equation (1) favors the discretization

$$\begin{aligned}&\frac{du_j}{dt}+\frac{4u_{j+2}-16u_{j+1}+24u_j-16u_{j-1}+4u_{j-2}}{h^4}\\&+\alpha\left(\frac{-u_{j+2}+16u_{j+1}-30u_j+16u_{j-1}-u_{j-2}}{12h^2}\right)\\&+\alpha\left(u_j\frac{u_{j+1}-u_{j-1}}{4h}+\frac{u_{j+1}^2-u_{j-1}^2}{4h}-\frac{u_{j+2}u_{j+1}-u_{j-2}u_{j-1}}{12h}\right)\approx 0,\end{aligned}\tag{2}$$

where the u_j's are grid values spaced h apart (one for each element). The first two lines of the holistic discretization (2) show that the holistic method generates conventional centered finite difference approximations for the linear terms $4u_{xxxx}$ and αu_{xx}. The third line details a specific nonstandard approximation for the nonlinear term $\alpha u u_x$: it is a mix of three valid approximations to uu_x; the holistic analysis determines the specific mix through the subgrid scale modeling of physical processes; see section 3.2. The holistic discretization is not constructed by discretizing the Kuramoto–Sivashinsky equation (1) term-by-term; rather, the subgrid scale dynamics of (1) together with interelement coupling generate the specific holistic discretization (2).

The discretization (2) is a low-order approximation. Center manifold theory provides systematic refinements. Analysis of the center manifold to higher orders in nonlinearity or interelement interaction, discussed in section 3, gives further refinement to the discretization. The higher-order terms come from resolving more subgrid scale processes and interactions. One effect of resolving systematically the interactions between elements is that the holistic model and the PDE agree to high order in the element size h: for example, the low-order model (2) is equivalent to the PDE

$$\frac{\partial u}{\partial t} = -\alpha\left(u\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2}\right) - 4\frac{\partial^4 u}{\partial x^4} - \frac{2h^2}{3}\frac{\partial^6 u}{\partial x^6} + \mathcal{O}\big(h^4\big) \tag{3}$$

as $h \to 0$. Thus the nonlinear processes are modeled with an error that decreases quadratically to zero with element size. Further higher-order analyses lead to higher-order consistency, as element size $h \to 0$, between the equivalent PDEs, such as (3), of the holistic discretizations and the Kuramoto–Sivashinsky PDE (see section 3.3). Such consistency is further justification for our approach, *in addition* to the support provided by center manifold theory for finite element size h.

The bulk of this paper is then a comprehensive comparative study of the various models of the Kuramoto–Sivashinsky dynamics; even further details are reported by MacKenzie [19]. A detailed numerical study of the holistic predictions for the steady states of the Kuramoto–Sivashinsky equation is the focal point of section 4, followed by an exploration of the holistic predictions for the time dependent phenomena of the Kuramoto–Sivashinsky equation in section 5. We compare the predicted steady states, their stability and bifurcation diagrams; the dynamics near the steady states; Hopf bifurcations leading to period doubling sequences; and the spatio-temporal patterns at relatively large nonlinearity parameter α. We find that the holistic models have excellent performance on coarse grids, thus enabling simulations to use large time steps. The excellent performance detailed herein is further evidence that the holistic approach is a robust and useful method for discretizing PDEs.

2. Use a homotopy in interelement coupling. The construction of a discretization is based upon breaking the spatial domain into disjoint finite elements and then joining them together again. We control this process by a coupling parameter γ that smoothly parametrizes the transition between decoupled elements and fully coupled elements for which we recover a model for the original PDE. Furthermore, we construct the model using solutions of the PDE within each element and hence resolve subgrid scale dynamics. Center manifold theory [3, 23, e.g.] provides rigorous support for holistic models, as introduced by Roberts [25] for Burgers' equation.

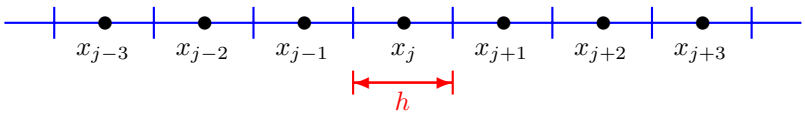


Figure 1. *An example of the one-dimensional grid with regular elements of width h. The jth element is centered about the grid point x_j. The vertical blue lines form the element boundaries, which for the jth element are located at $x_{j\pm1/2} = (j \pm 1/2)h$.*

2.1. Introduce internal boundaries between elements. Establish the spatial discretization by dividing the domain into m elements of equal and finite width h and introducing an equispaced grid of collocation points, $x_j = jh$, at the center of each element; see Figure 1.[1] Express the subgrid field in the jth element by $u = v_j(x,t)$—we solve the Kuramoto–Sivashinsky PDE (1) with interelement coupling introduced via artificial internal boundary conditions (IBCs). We introduce a homotopy in an interelement coupling parameter γ: when $\gamma = 0$ the elements are effectively isolated from each other, providing the basis for the application of center manifold theory; whereas when evaluated at $\gamma = 1$ the elements are fully coupled together, and hence the discretized model applies to the original PDE. Since the Kuramoto–Sivashinsky PDE is fourth-order we require four IBCs for each element to ensure satisfactory coupling between neighboring elements. Here we use the nonlocal IBCs

$$\delta_x v_j(x,t) = \gamma\, \delta v_{j\pm1/2}(x,t) \quad \text{at } x = x_{j\pm1/2}, \tag{4}$$

$$\delta_x^3 v_j(x,t) = \gamma^2 \delta^3 v_{j\pm1/2}(x,t) \quad \text{at } x = x_{j\pm1/2}, \tag{5}$$

which are an extension of the nonlocal IBCs explored by Roberts [26] for Burgers' equation; alternative local IBCs were explored by MacKenzie [19] but are generally inferior. The nonlocal IBCs (4)–(5) involve the centered difference operators δ and δ_x: the operator δ_x denotes a centered difference in x only, with step h, whereas the operator δ denotes a centered difference applied to the grid index j with step 1; so for example, the first of the IBCs, (4), is

$$v_j(x_{j\pm1},t) - v_j(x_j,t) = \gamma[v_{j+1}(x_{j\pm1},t) - v_j(x_j,t)]. \tag{6}$$

Note that the field $v_j(x,t)$ extends analytically to at least $x_{j\pm2}$ to allow the application of the nonlocal IBCs (5). The physical interpretation of these IBCs is not obvious. First, when $\gamma = 0$, (4)–(5) ensure that the first and third differences in x of the field v_j centered about the element boundaries $x_{j\pm1/2}$ are zero. These isolate each element from its neighbors, as there is then no coupling between them. In each element $v_j(x,t) =$ constant is an equilibrium. It is dynamically attractive, provided that the instability controlled by α/h^2 is not too large compared with the dissipation of order $1/h^4$. This simple class of piecewise constant solutions provide the basis for analyzing the dynamics when the elements are coupled together with nonzero γ. Second, the nonlocal IBCs evaluated at $\gamma = 1$ require that the field $v_j(x,t)$, when extrapolated to $x_{j\pm1}$ and $x_{j\pm2}$, is equal to the grid point value of the subgrid field of that element, $u_{j\pm1}$ and $u_{j\pm2}$, respectively. See the schematic representation in Figure 2 of

[1] In principle, elements may be of unequal size. However, to simplify the analysis, herein all elements are of equal width h.

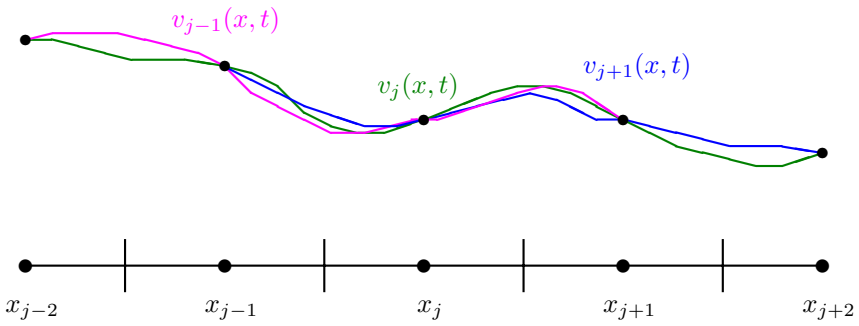


Figure 2. *Schematic diagram of the fields $v_j(x,t)$, $v_{j+1}(x,t)$, and $v_{j-1}(x,t)$ for the nonlocal* IBCs (4)–(5) *with $\gamma = 1$. See the fields pass through neighboring grid values u_j and $u_{j\pm1}$, and also $u_{j\pm2}$ when appropriate.*

these nonlocal boundary conditions evaluated at $\gamma = 1$. This requirement restores sufficient continuity to ensure that the holistic model applies to the original PDE.

The interelement coupling parameter γ controls the flow of information between neighboring elements. We construct solutions as power series expansions in the coupling parameter γ.[2] When $\mathcal{O}(\gamma^2)$ terms are neglected in the holistic model, the field in the jth element involves information about the fields in the $j \pm 1$ elements. Similarly, when $\mathcal{O}(\gamma^3)$ terms are neglected in the approximation, the field in the jth element involves information about the fields in the $j \pm 1$ and $j \pm 2$ elements. Consequently, the order of γ retained in the holistic model controls the stencil width of the discretization.

Roberts [26] argued that this particular form of the nonlocal IBCs ensures that these holistic models are consistent with any given PDE to high order in the grid size h as $h \to 0$. Interestingly, Roberts and Kevrekidis [29] show that closely related coupling boundary conditions provide high-order consistency for multiscale simulations using the gap-tooth scheme.

2.2. Center manifold theory supports the discretization. The existence, relevance, and approximation theorems [3, 4, e.g.] of center manifold theory apply to the Kuramoto–Sivashinsky PDE (1) with IBCs (4)–(5). Similar to the application to Burgers' equation by Roberts [25], the result here is support for a low-dimensional discrete model for the Kuramoto–Sivashinsky dynamics at finite grid size.

Theoretical support is based upon the piecewise constant solutions obtained when all the elements are insulated from each other. Adjoin to the Kuramoto–Sivashinsky PDE (1) the dynamically trivial equations for the coupling parameter γ and the nonlinearity parameter α,

$$\frac{\partial \gamma}{\partial t} = \frac{\partial \alpha}{\partial t} = 0, \tag{7}$$

and consider the dynamics in the extended state space $(u(x), \gamma, \alpha)$. Adjoining such trivial equations for parameters is commonly used to unfold bifurcations [3, section 1.5]. In this extended space there is a subspace of fixed points with u = constant in each element and

[2]Such homotopies are used successfully in other numerical methods. For example, Liao [17] proposed a homotopy in his general boundary element method from auxiliary linear operators whose fundamental solutions are well known. Here the homotopy is only in the IBCs.

$\gamma = \alpha = 0$.[3] Linearizing the PDE and IBCs about each fixed point, $u = \text{constant} + u'(x,t)$ for small u', gives

$$\frac{\partial u'}{\partial t} = -\frac{\partial^4 u'}{\partial x^4} \quad \text{such that} \quad \delta_x u'(x,t)\big|_{x=x_{j\pm1/2}} = \delta_x^3 u'(x,t)\big|_{x=x_{j\pm1/2}} = 0.$$

Solving these linear equations, the nth linear eigenmode associated with each element is

$$\alpha = \gamma = 0, \qquad u' \propto e^{\lambda_n t} \cos\left[\frac{n\pi}{h}(x - x_{j-1/2})\right] \tag{8}$$

for the nonlocal IBCs (4)–(5), where $n = 0, 1, 2, \ldots$ and the eigenvalue $\lambda_n = -n^4\pi^4/h^4$. There are also the trivial modes $\gamma = \text{const}$ and $\alpha = \text{const}$. Therefore, in a spatial domain of m elements there are $m+2$ zero eigenvalues: one associated with each of the m elements and two from the trivial (7). All other eigenvalues are negative, $\le -\pi^4/h^4$. Thus, the existence theorem (see [4, p. 281] or [33, p. 96]) guarantees that an $m+2$-dimensional center manifold $\mathcal{M}$ exists for the Kuramoto–Sivashinsky PDE (1) with the trivial (7) and IBCs (4)–(5).

We parametrize the $(m+2)$-dimensional center manifold $\mathcal{M}$ by the $m+2$ parameters γ, α, and the grid values u_j.[4] Define $\boldsymbol{u}$ as the vector of the m grid values. Thus for some function v to be determined, the center manifold $\mathcal{M}$ is

$$u(x,t) = v(x; \boldsymbol{u}, \gamma, \alpha). \tag{9}$$

However, we find it convenient to view the center manifold (9) as the union of the set of subgrid fields $v_j(x; \boldsymbol{u}, \gamma, \alpha)$ over the physical domain. The corresponding amplitude condition, that the field in each element has to pass through its grid value, is

$$u_j = v_j(x_j; \boldsymbol{u}, \gamma, \alpha). \tag{10}$$

The existence theorem [4] also asserts that on the center manifold the grid values u_j evolve deterministically in time according to the system of ODEs

$$\dot{u}_j = du_j/dt = g_j(\boldsymbol{u}, \gamma, \alpha), \tag{11}$$

where g_j is the restriction of the Kuramoto–Sivashinsky PDE (1) with the trivial (7) and IBCs (4)–(5) to the center manifold $\mathcal{M}$. It is this evolution (11) of the grid values that gives the holistic discretization.

Note that the center manifold $\mathcal{M}$ is global in u but local in γ and α. When the parameters $\gamma = \alpha = 0$ the Kuramoto–Sivashinsky PDE has an m-dimensional center subspace $\mathcal{E}$ of fixed points with the field u being independently constant in each element; these are fixed points for all $\boldsymbol{u}$. When the parameters γ and α are nonzero this subspace is "bent" to the curved center

[3] In principle we need not base the analysis about a zero nonlinearity parameter. In practice, allowing nonzero α as a base makes the algebra too complicated, as the subgrid structure would then be determined by recursively solving equations of the form $-4u'_{xxxx} - \alpha(u'_{xx} + u_j u'_x) = \text{RHS}$ for corrections u'. Thus the practical route for this modeling of *nonlinear* PDEs is to adjoin $\partial\alpha/\partial t = 0$ so we treat the nonlinear parameter α as small.

[4] These grid values are one choice to represent the magnitude of the field u in each element. Other choices to represent the local field are possible, but the grid values appear most convenient.

manifold $\mathcal{M}$. Thus the models we construct are valid for small enough γ and α, although we use them at finite γ and α, but are formally valid for all $|\boldsymbol{u}|$. Numerical solutions of the center manifold models, such as those in section 3.2, indicate that parameter values as large as $\gamma = 1$ and $\alpha = 20$–50 are indeed within the range of validity of our approach, even on relatively coarse grids.

We now support the claim that the evolution of the discrete grid values (11) actually models the Kuramoto–Sivashinsky PDE (1). The relevance theorem of center manifolds (see [4, p. 282] or [33, p. 128]) guarantees that all solutions of the Kuramoto–Sivashinsky PDE (1) with (7) and the IBCs (4)–(5) that remain in some neighborhood of the center subspace $\mathcal{E}$ in $(u(x), \gamma, \alpha)$ space are exponentially quickly attracted to the center manifold $\mathcal{M}$ and thence to a solution of the m discrete ODEs (11). For our application of center manifold theory to the holistic model we seek regimes where this neighborhood includes $\gamma = 1$ and α of interest. We estimate the rate of attraction by the leading negative eigenvalue, here $\lambda_1 = -\pi^4/h^4$. The actual rate of attraction may be less due to the difference between center manifold $\mathcal{M}$ and the center subspace $\mathcal{E}$, but λ_1 is the correct order of magnitude. This exponentially quick attraction ensures the so-called asymptotic completeness [30]: after the exponentially quick transients of the approach to $\mathcal{M}$ by any trajectory, the evolution of the discretization (11) on $\mathcal{M}$ accurately models the dynamics of the Kuramoto–Sivashinsky PDE (1).

2.3. Approximate the shape of the center manifold. Having established that we may find a low-dimensional description (9)–(11) of the interacting elements that is relevant to the Kuramoto–Sivashinsky PDE (1), we need to construct the shape of the center manifold and the corresponding evolution on the manifold. The approximation theorem of Carr and Muncaster [4, p. 283] assures us that upon substituting the ansatz (9)–(11) into the complete system and solving to some order of error in α and γ, then the center manifold $\mathcal{M}$ and the evolution thereon will be approximated to the same order. However, we need to evaluate the approximations at the coupling parameter $\gamma = 1$ because it is only then that the artificial internal boundaries are removed. Thus the actual error of the model due to the evaluation at $\gamma = 1$ is not estimated. However, the holistic method for discretizing the Kuramoto–Sivashinsky equation is supported in three ways: first, the smooth homotopy from $\gamma = 0$ with large spectral gap to the gravest decaying mode with decay rate $\approx -\pi^4/h^4$; second, the holistic models are consistent with the Kuramoto–Sivashinsky PDE to high order in grid size h (see section 3.3); third, we see in sections 4–5 that the holistic discretization resolves accurately both steady-state solutions and time dependent phenomena of the Kuramoto–Sivashinsky PDE.

To construct the center manifold, we solve for the field v_j in each element. For definiteness, here we consider domains periodic in space, or equivalently elements far from the influence of any physical boundary. By translational symmetry of the Kuramoto–Sivashinsky PDE (1) the subgrid field in each element is identical, except for the appropriate shift in the grid index j. Thus we construct the subgrid field and evolution for a general jth element; section 3 gives some examples.

The algebraic details of the derivation of the center manifold model (9)–(11) are handled by computer algebra. In an algorithm introduced by Roberts [24], iteration drives to zero the residuals of the governing PDE (1) and its IBCs (4)–(5) and amplitude condition (10). Since the algebraic details of the construction are tedious, they are not given; instead, see the computer

algebra procedure of [28].

This computer algebra is based upon driving the residuals of the governing equations to zero in the following manner. Recall from section 2.2 that the grid values $\boldsymbol{u}$ parametrize the center manifold (9) and that (11) gives the evolution of the grid values. Thus substitute these into the Kuramoto–Sivashinsky PDE (1) and seek to solve

$$\frac{\partial v_j}{\partial t} = \sum_k \frac{\partial v_j}{\partial u_k} g_k = -4\frac{\partial^4 v_j}{\partial x^4} - \alpha\left(\frac{\partial^2 v_j}{\partial x^2} + v_j \frac{\partial v_j}{\partial x}\right), \tag{12}$$

together with the nonlocal IBCs(4)–(5) and the amplitude equation (10), to some order in parameters γ and α. The iteration is that, given any approximation, denoted by $\widetilde{\cdot}$, we seek corrections, denoted by primes, such that $v_j = \widetilde{v}_j + v'_j$ and $g_j = \widetilde{g}_j + g'_j$ better satisfy the Kuramoto–Sivashinsky PDE. Thus in each iteration we solve a problem of the form

$$-4\frac{\partial^4 v'_j}{\partial x^4} = g'_j + \text{Residual}, \tag{13}$$

where from (12) the

$$\text{Residual} = \sum_k \frac{\partial \widetilde{v}_j}{\partial u_k} g_k + 4\frac{\partial^4 \widetilde{v}_j}{\partial x^4} + \alpha\left(\frac{\partial^2 \widetilde{v}_j}{\partial x^2} + \widetilde{v}_j \frac{\partial \widetilde{v}_j}{\partial x}\right), \tag{14}$$

together with the IBCs, for the corrections, primed quantities, to the subgrid field and the evolution of the grid values. Note that the residual in (14) is the residual of the Kuramoto–Sivashinsky PDE for the current approximation. The iteration scheme starts with the linear solution in each element, namely $v_j(x, \boldsymbol{u}, \gamma, \alpha) = u_j$ and $g_j(\boldsymbol{u}, \gamma, \alpha) = 0$. The iteration terminates when the residuals of the Kuramoto–Sivashinsky PDE (12), and the IBCs, are zero to some order in (γ, α). Then theory assures us that the subgrid field in each element and the evolution of the grid values are correct to the same order in (γ, α); that is, when the residuals are of order $\mathcal{O}(\gamma^m, \alpha^n)$, theory assures us the errors are also of order $\mathcal{O}(\gamma^m, \alpha^n)$. This assurance holds for both steady states and time dependent dynamics.

3. Various holistic models. Here we record holistic models of the Kuramoto–Sivashinsky PDE (1), to various orders in coupling parameter γ, governing the width of the numerical stencil, and in the nonlinearity parameter α. In order to be used, the models need to be evaluated at $\gamma = 1$, as then the nonlocal IBCs (4)–(5) ensure sufficient continuity in the solution field. We write the models in terms of the centered difference and mean operators,

$$\delta u_j = u_{j+1/2} - u_{j-1/2} \quad \text{and} \quad \mu u_j = (u_{j+1/2} + u_{j-1/2})/2,$$

respectively. A REDUCE program [28] constructs all models. We present in detail here only holistic models to errors $\mathcal{O}(\alpha^2)$, as the level of complexity increases enormously with the order of nonlinearity α.

3.1. Some holistic discretizations. In order to represent the spatial fourth derivative in the Kuramoto–Sivashinsky equation, we need at least a five-point stencil approximation. Thus we determine the interactions between at least next-nearest neighboring elements by obtaining up to at least quadratic terms in the coupling parameter γ. Computer algebra readily determines higher-order expressions in the coupling parameter γ:

$$
\begin{aligned}
\dot{u_j} = & -\frac{\gamma\alpha}{h^2}\delta^2 u_j - \frac{\gamma\alpha}{h} u_j \delta\mu u_j - \frac{4\gamma^2}{h^4}\delta^4 u_j + \frac{\gamma^2\alpha}{12h^2}\delta^4 u_j \\
& + \frac{\gamma^2\alpha}{12h}\left(2u_j\delta^3\mu u_j + \delta^2 u_j \delta^3 \mu u_j + \delta^4 u_j \delta\mu u_j\right) \qquad (15) \\
& + \frac{2\gamma^3}{3h^4}\delta^6 u_j - \frac{\gamma^3\alpha}{90h^2}\delta^6 u_j \\
& - \frac{\gamma^3\alpha}{480h}\left(16\, u_j\delta^5\mu u_j + 30\,\delta^4 u_j \delta^3\mu u_j + 40\,\delta^2 u_j\delta^3\mu u_j\right. \\
& \qquad + 40\,\delta^4 u_j\delta\mu u_j + 28\,\delta^2 u_j\delta^5\mu u_j + 14\,\delta^6 u_j\delta\mu u_j \\
& \qquad \left. + 7\delta^4 u_j\delta^5\mu u_j + 7\delta^6 u_j\delta^3\mu u_j\right) \qquad (16) \\
& - \frac{7\gamma^4}{60h^4}\delta^8 u_j + \frac{\gamma^4\alpha}{560h^2}\delta^8 u_j \\
& + \frac{\gamma^4\alpha}{60480h}\left(432\, u_j\delta^7\mu u_j + 3528\,\delta^2 u_j\delta^5\mu u_j + 1507\,\delta^2 u_j\delta^7\mu u_j\right. \\
& \qquad + 3780\,\delta^4 u_j\delta^3\mu u_j + 3951\,\delta^4 u_j\delta^5\mu u_j + 984\,\delta^4 u_j\delta^7\mu u_j \\
& \qquad + 1764\,\delta^6 u_j\delta\mu u_j + 3419\,\delta^6 u_j\delta^3\mu u_j + 1414\,\delta^6 u_j\delta^5\mu u_j \\
& \qquad + 164\,\delta^6 u_j\delta^7\mu u_j + 523\,\delta^8 u_j\delta\mu u_j + 656\,\delta^8 u_j\delta^3\mu u_j \\
& \qquad \left. + 164\,\delta^8 u_j\delta^5\mu u_j\right) + \mathcal{O}\left(\gamma^5, \alpha^2\right). \qquad (17)
\end{aligned}
$$

We have ordered (and colored) the terms in this discretization in increasing powers of coupling parameter γ in order to discuss the three different truncations labeled by the three different equation numbers.

The $\mathcal{O}(\gamma^3, \alpha^2)$ *holistic discretization* is formed by truncating the above model to the (green) terms before and at (15). Evaluated at coupling parameter $\gamma = 1$, this discretization forms the basic five-point stencil approximation (2) discussed in the Introduction. The first line of (15), when evaluated at $\gamma = 1$, gives a second-order centered difference approximation for the hyperdiffusion term $4u_{xxxx}$, a fourth-order centered difference approximation to the linear growth term αu_{xx}, and a second-order centered difference approximation to the nonlinear advection term $\alpha u u_x$. The second line modifies the nonlinear discretization to account for interaction with nonlinear effects caused by the next-nearest neighbor elements.

The holistic discretization (15) contains the approximation

$$
uu_x|_{x_j} \approx \left(u_j \frac{u_{j+1} - u_{j-1}}{4h} + \frac{u_{j+1}^2 - u_{j-1}^2}{4h} - \frac{u_{j+2}u_{j+1} - u_{j-2}u_{j-1}}{12h} \right) \qquad (18)
$$

when evaluated at $\gamma = 1$. This is a $1/2 : 1 : -1/2$ mix of the approximations

$$uu_x|_{x_j} \approx u_j \frac{u_{j+1} - u_{j-1}}{2h} \approx \frac{u_{j+1}^2 - u_{j-1}^2}{4h} \approx \frac{u_{j+2}u_{j+1} - u_{j-2}u_{j-1}}{6h}, \tag{19}$$

respectively. This particular nonstandard approximation (18) to the nonlinear term $\alpha u u_x$ arises due to the modeling of subgrid scale interactions between the Kuramoto–Sivashinsky equation and the interelement coupling. Such nonstandard approximations generated through this approach can have robust numerical characteristics [26, section 2].

The $\mathcal{O}(\gamma^4, \alpha^2)$ *holistic discretization* is formed by truncating the above model to the (green and olive green) terms before and at (16). This discretization forms a seven-point stencil approximation, involving u_j, $u_{j\pm1}$, $u_{j\pm2}$, and $u_{j\pm3}$. The first and third lines of (16), when evaluated at $\gamma = 1$, give a fourth-order centered difference approximation to the hyperdiffusion term, a sixth-order centered difference approximation to the linear growth term, and a second-order centered difference approximation to the nonlinear advection term. The second, fourth, fifth, and sixth lines account for higher-order subgrid scale dynamics of the nonlinearity and its interelement coupling to generate a fourth-order centered difference approximation to the nonlinearity uu_x.

The $\mathcal{O}(\gamma^5, \alpha^2)$ *holistic discretization* (17) (all colored terms) forms a nine-point stencil approximation, involving only u_j, $u_{j\pm1}$, $u_{j\pm2}$, $u_{j\pm3}$, and $u_{j\pm4}$. The first, third, and seventh lines of (17) when evaluated at $\gamma = 1$ give a sixth-order centered difference approximation for the hyperdiffusion term, an eighth-order centered difference approximation for the linear growth term, and a second-order centered difference approximation for the nonlinear advection term. The remaining lines provide modifications to model the nonlinear uu_x to sixth order through resolving subgrid scale dynamics.

We do not code these discretizations manually. Instead, the computer algebra program of [28] is used with the UNIX editor `sed` to automatically write the discretization in a form suitable for MATLAB simulation.

Compare with conventional centered difference models. Traditional direct finite differences generate the following discretization of the Kuramoto–Sivashinsky PDE (1):

$$\dot{u}_j = -\frac{\alpha}{h} u_j \delta\mu u_j - \frac{\alpha}{h^2}\delta^2 u_j - \frac{4}{h^4}\delta^4 u_j \tag{20}$$

$$+ \frac{\alpha}{h}\frac{1}{6} u_j \delta^3 \mu u_j + \frac{\alpha}{h^2}\frac{1}{12}\delta^4 u_j + \frac{4}{h^4}\frac{1}{6}\delta^6 u_j \tag{21}$$

$$- \frac{\alpha}{h}\frac{1}{30} u_j \delta^5 \mu u_j - \frac{\alpha}{h^2}\frac{1}{90}\delta^6 u_j - \frac{4}{h^4}\frac{7}{240}\delta^6 u_j + \mathcal{O}\left(h^6\right). \tag{22}$$

Truncate the above discretization to the terms before and at

- (20) to obtain a five-point conventional discretization;
- (21) to obtain a seven-point conventional discretization;
- (22) to obtain a nine-point conventional discretization.

Consider the different view of the errors for the discretizations: the centered difference approximations (20)–(22) are justified by consistency as grid size $h \to 0$, whereas the holistic discretizations (15)–(17) are supported by center manifold theory at finite grid size h. The

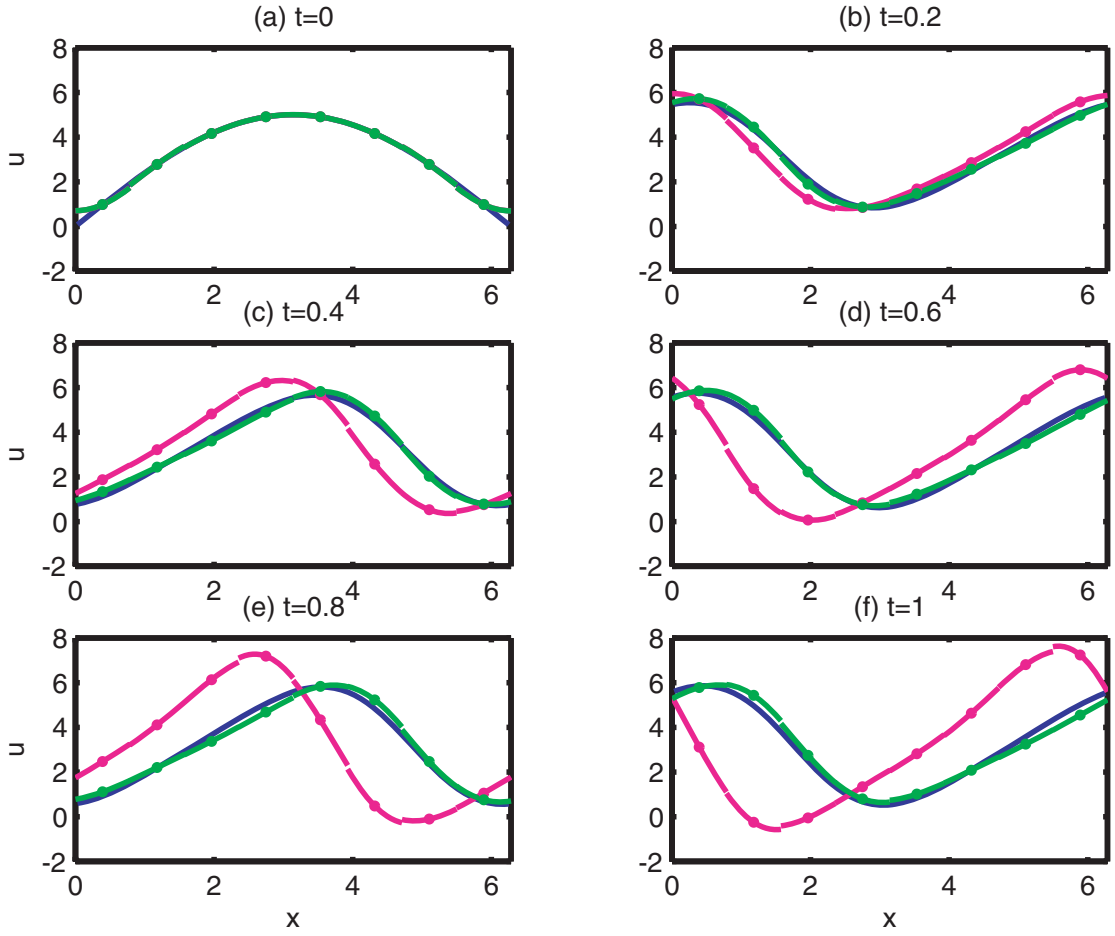


Figure 3. $\alpha = 5$. *Wave-like solutions at* $t = 0, 0.2, 0.4, 0.6, 0.8, 1$ *for the* $\mathcal{O}(\gamma^3, \alpha^2)$ *holistic model* (15)*, shown in green, and the second-order centered difference approximation* (20)*, in magenta, on coarse grids of eight elements on* $[0, 2\pi]$*. The accurate solution is shown in blue.*

errors in the center manifold approach are due to the truncation of dependence in the interelement coupling parameter γ and the nonlinearity parameter α. However, as argued by Roberts [26] for linear systems and as demonstrated in section 3.3, the particular choice of the IBCs (4)–(5) ensures that the holistic discretizations *are also consistent as* $h \to 0$ with the Kuramoto–Sivashinsky PDE (1).

3.2. Modeling the subgrid field powers our methodology. Recall that the set of subgrid fields over the physical domain form a state on the center manifold. Here we plot some example subgrid fields for various holistic models. For example, Figure 3 shows the evolving subgrid fields of wave-like solutions of the holistic discretization (15) compared with an accurate solution and the Lagrangian interpolation of the second-order centered difference model (20). Such plots reinforce the link between the abstract center manifold description of the dynamics and the physical subgrid fields for the low-order holistic models. Recall that

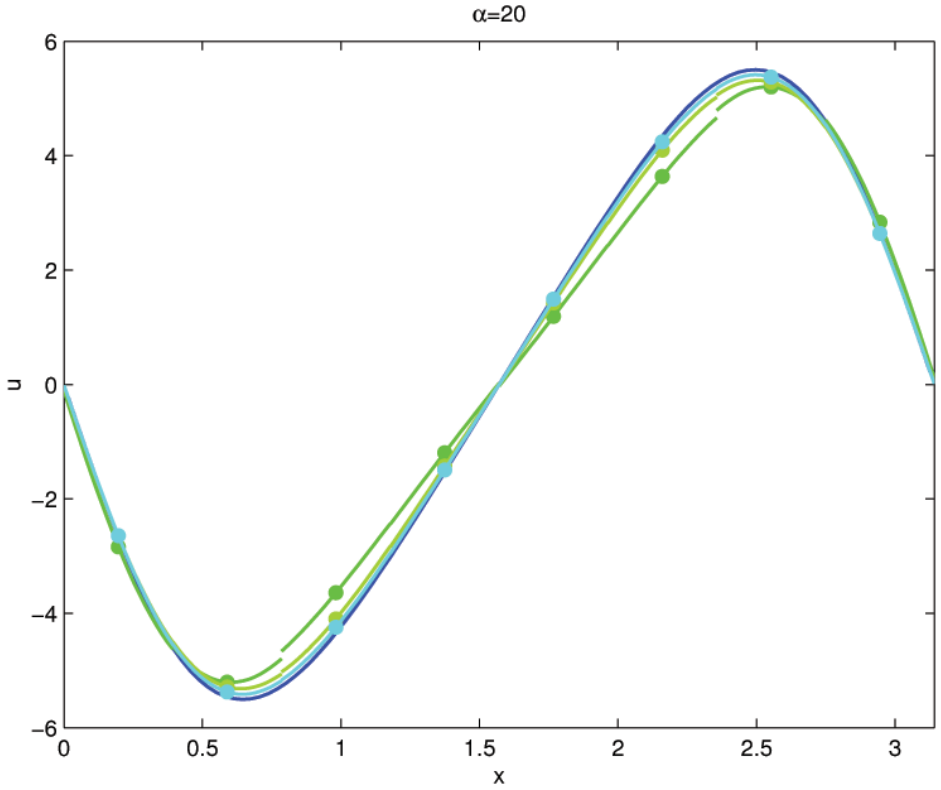


Figure 4. *Subgrid fields of the holistic models with errors $\mathcal{O}(\gamma^3, \alpha^2)$ (15) (green), $\mathcal{O}(\gamma^4, \alpha^2)$ (16) (olive green), and $\mathcal{O}(\gamma^5, \alpha^2)$ (17) (cyan), for a steady state of the Kuramoto–Sivashinsky equation at $\alpha = 20$, with eight elements on $[0, \pi]$. An accurate solution is also plotted in blue.*

the key methodology difference with conventional finite differences, and finite elements, is that the subgrid fields of the holistic models are constructed by actual local solutions of the Kuramoto–Sivashinsky PDE; see section 2.3.

Until section 5 we restrict our attention to odd symmetric solutions that are 2π periodic. This restriction is to compare results with the numerical investigations of Jolly, Kevrekidis, and Titi [14], which we consider in more detail in sections 4 and 5. Typically we use a grid of eight equi-spaced elements on the interval $[0, \pi]$. The subgrid fields are plotted for approximations to the steady states of the Kuramoto–Sivashinsky equation (1) with these periodic boundary conditions, computed using holistic discretizations at the high values of the nonlinearity parameter $\alpha = 20$ and $\alpha = 50$.

Higher-order holistic models improve the accuracy and continuity of the subgrid field. Figure 4 displays the subgrid fields of three holistic models for a steady state of the Kuramoto–Sivashinsky PDE for $\alpha = 20$. Observe that the collection of subgrid fields forms the field u, which is a state on the center manifold. The $\mathcal{O}(\gamma^3, \alpha^2)$ holistic model (15) (green) is the least accurate and has the largest jump at element boundaries. But note that Lagrangian interpolation of the steady state of the second-order conventional approximation (20) is significantly less accurate [19]. The $\mathcal{O}(\gamma^4, \alpha^2)$ (16) model (olive green) displays improvement over

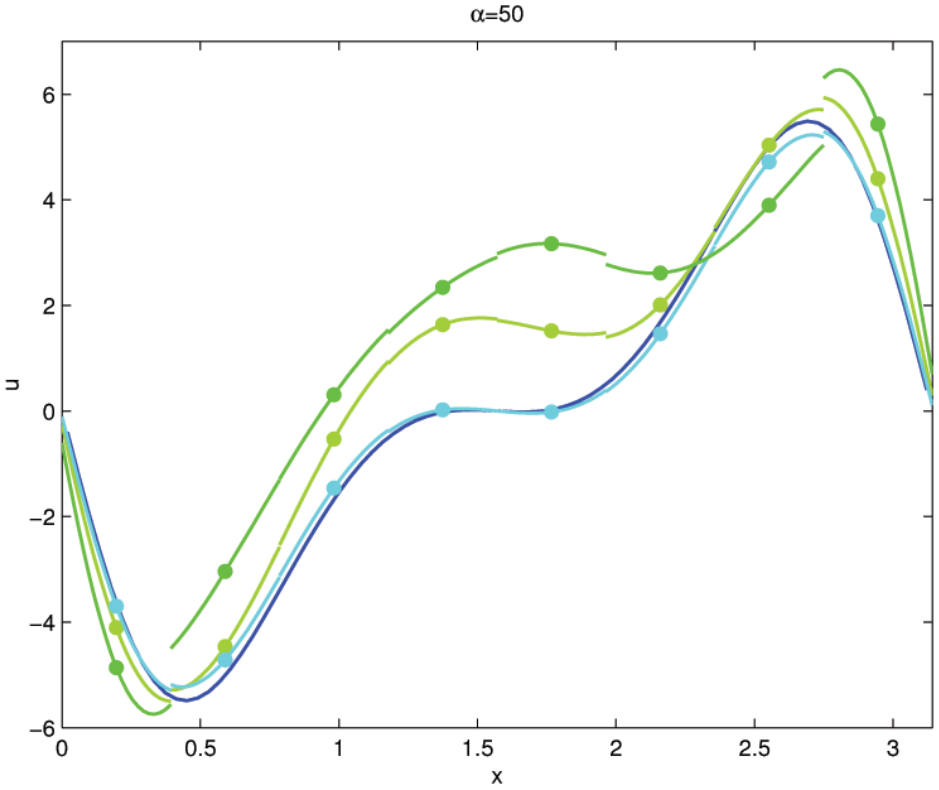


Figure 5. *Subgrid fields of the holistic models with errors $\mathcal{O}(\gamma^3, \alpha^2)$ (15) (green), $\mathcal{O}(\gamma^4, \alpha^2)$ (16) (olive green), and $\mathcal{O}(\gamma^5, \alpha^2)$ (17) (cyan), for a steady state of the Kuramoto–Sivashinsky equation at $\alpha = 50$, with eight elements on $[0, \pi]$. An accurate solution is also plotted in blue.*

the holistic $\mathcal{O}(\gamma^3, \alpha^2)$ approximation. The $\mathcal{O}(\gamma^5, \alpha^2)$ (17) model (cyan) is the most accurate, being almost indistinguishable from the correct solution (dark blue).

Figure 5 shows a steady state of the Kuramoto–Sivashinsky PDE at $\alpha = 50$. The accurate field is symmetric (blue curve). For this value of the nonlinearity there is no steady-state solution for centered difference approximations of either second (20), fourth (21), or sixth order (22) on this coarse grid of eight elements on $[0, \pi]$. However, the five-point stencil holistic approximation with errors $\mathcal{O}(\gamma^3, \alpha^2)$ (15) (green) models this steady state of the Kuramoto–Sivashinsky equation even for such a large value of the nonlinearity on this coarse grid. This $\mathcal{O}(\gamma^3, \alpha^2)$ holistic solution has significant jumps across the subgrid field at element boundaries; moreover, the subgrid field is not symmetric and is most inaccurate near the center of the spatial domain considered here. The seven-point stencil holistic approximation with errors $\mathcal{O}(\gamma^4, \alpha^2)$ (16) (olive green) is more accurate with smaller jumps between neighboring subgrid fields, but is also not symmetric. The nine-point stencil holistic approximation with errors $\mathcal{O}(\gamma^5, \alpha^2)$ (17) (cyan) is the most accurate of the holistic models illustrated here; it is symmetric, and the jumps between neighboring subgrid fields are almost indiscernible.

These illustrations of the subgrid fields of steady states of the Kuramoto–Sivashinsky equation at $\alpha = 20$ and $\alpha = 50$ indicate that the holistic models perform well even at such

Table 1

Errors E of the equivalent PDE *of holistic and conventional discretizations; see* (24)*. MacKenzie* [19] *gave more details.*

Stencil	Holistic (15)–(17)	Conventional (20)–(22)
5 point	$-\frac{2h^2}{3}\frac{\partial^6 u}{\partial x^6}+\mathcal{O}(h^4)$	$-h^2\left(\frac{\alpha}{6}\frac{\partial^3 u}{\partial x^3}\frac{\partial u}{\partial x}+\frac{\alpha}{12}\frac{\partial^4 u}{\partial x^4}+\frac{2}{3}\frac{\partial^6 u}{\partial x^6}\right)+\mathcal{O}(h^4)$
7 point	$-\frac{7h^4}{60}\frac{\partial^8 u}{\partial x^8}+\mathcal{O}(h^6)$	$-h^4\left(\frac{\alpha}{30}\frac{\partial^5 u}{\partial x^5}\frac{\partial u}{\partial x}+\frac{\alpha}{90}\frac{\partial^6 u}{\partial x^6}+\frac{7}{60}\frac{\partial^8 u}{\partial x^8}\right)+\mathcal{O}(h^6)$
9 point	$-\frac{41h^6}{1890}\frac{\partial^{10} u}{\partial x^{10}}+\mathcal{O}(h^8)$	$-h^6\left(\frac{\alpha}{140}\frac{\partial^7 u}{\partial x^7}\frac{\partial u}{\partial x}+\frac{\alpha}{560}\frac{\partial^8 u}{\partial x^8}+\frac{41}{1890}\frac{\partial^{10} u}{\partial x^{10}}\right)+\mathcal{O}(h^8)$

large values of a supposedly small parameter. The performance of the holistic models are explored further in section 4 for steady states and section 5 for time dependent phenomena.

3.3. The holistic discretizations are consistent. Holistic models constructed by implementing the IBCs (4)–(5) have dual justification [26]: they are supported by center manifold theory for small enough α and γ, as well as being justified by their consistency as the grid size $h\to 0$. We explore consistency as a well-established feature of numerical analysis.[5]

Here we examine the equivalent PDEs for the holistic discretizations (15)–(17) evaluated at $\gamma=1$, and the centered difference approximations (20)–(22). These equivalent PDEs establish the $\mathcal{O}(h^{2p-2})$ consistency with the Kuramoto–Sivashinsky PDE for holistic models constructed with residuals $\mathcal{O}(\gamma^{p+1})$.

Roberts [26] proved that using IBCs of the form introduced in section 2 and retaining terms up to γ^p in the holistic approximations results in approximations which are consistent with the *linear* terms of the Kuramoto–Sivashinsky equation (1) to $\mathcal{O}(h^{2p-2})$, provided $p\geq 2$. However, accumulating evidence indicates the IBCs (4)–(5) also ensure $\mathcal{O}(h^{2p-2})$ consistency for the nonlinear dynamics. As yet no formal proof exists of this nonlinear consistency, but all holistic models of the Kuramoto–Sivashinsky equation, containing terms up to γ^7 and α^4 and constructed using (4)–(5), are nonlinearly consistent (although not all are recorded here).

Find the equivalent PDEs for the various discretizations by expanding the discretizations in grid size h about a grid point x_j. That is, write

$$u_{j\pm m}=u_j\pm mh\frac{\partial u_j}{\partial x}+m^2\frac{h^2}{2}\frac{\partial^2 u_j}{\partial x^2}+\sum_{k=3}^{\infty}(\pm m)^k\frac{h^k}{k!}\frac{\partial^k u_j}{\partial x^k}, \tag{23}$$

to whatever order in h is required. Computer algebra performs the tedious details. As for the example equivalent PDE (3), all equivalent PDEs have the form

$$\frac{\partial u}{\partial t}=-\alpha\left(u\frac{\partial u}{\partial x}+\frac{\partial^2 u}{\partial x^2}\right)-4\frac{\partial^4 u}{\partial x^4}+E, \tag{24}$$

where E denotes the error between the equivalent PDE of the discretization and the Kuramoto–Sivashinsky PDE (1). Table 1 lists the errors for the holistic discretizations (15)–(17) and the conventional finite difference discretizations (20)–(22). Observe that for each stencil width,

[5]However, note that high-order consistency is not a primary goal of this holistic approach, since we aim to develop and support models for finite element size h.

both the holistic and the conventional discretizations are consistent to the same order in grid spacing h. However, the holistic discretizations have fewer error terms than their corresponding conventional finite difference approximations. The holistic discretizations better model the nonlinear terms through the resolution of subgrid scale structures.

4. Holistic models accurately give steady states. The relevance of our holistic models is rigorously supported by center manifold theory for sufficiently small parameters γ and α. However, the holistic models must be evaluated at coupling parameter $\gamma = 1$ to model the Kuramoto–Sivashinsky dynamics. The important question is: Does evaluating the holistic models at $\gamma = 1$ provide useful and accurate numerical models? Numerical experiments detailed in this and the next section provide strong support that it does.

In this section we explore the accuracy of the holistic models by constructing bifurcation diagrams of the various holistic discretizations and comparing them to conventional explicit centered difference approximations and to the bifurcation diagrams presented by Jolly, Kevrekidis, and Titi [14] for various traditional Galerkin and nonlinear Galerkin approximations.

We restrict our exploration to solutions that are both 2π periodic and odd: thus

$$u(x,t) = u(x+2\pi,t) \quad \text{and} \quad u(x,t) = -u(2\pi - x,t). \tag{25}$$

We also restrict the nonlinearity parameter to the range $0 \le \alpha \le 70$. These restrictions are to compare our results to those of [14] for approximate inertial manifold methods. For this range of nonlinearity α, the trivial solution $u = 0$ undergoes pitchfork bifurcations at $\alpha = 4, 16, 36, 64$, leading to the unimodal, bimodal, trimodal, and quadrimodal branches, respectively; see the bifurcation diagram Figure 6.

Such bifurcation diagrams usefully summarize qualitative and quantitative information for a large range of the nonlinearity parameter α. We use the software package XPPAUT [8], which incorporates the continuation software AUTO [7], to calculate the bifurcation information. The information is then filtered through a function written in MATLAB to draw the bifurcation diagram. The input to XPPAUT is a text `.ode` file describing the set of ODEs. Because the holistic models contain a large number of terms, the `.ode` files are generated automatically using REDUCE and MATLAB; see [19] for more details.

4.1. The reference bifurcation diagram. Here we introduce accurate solutions for the steady states of the Kuramoto–Sivashinsky PDE (1) over the range $0 \le \alpha \le 70$, as summarized in the bifurcation diagram of Figure 6. The sixth-order centered difference approximation (22) with 48 grid points on $[0, \pi]$ provides these accurate solutions. These provide the reference for the approximations on coarse grids, and serve to also introduce the conventions we adopt in bifurcation diagrams.

All the bifurcation diagrams plot a signed solution norm versus the nonlinearity parameter α. Incorporating the sign is a difference from the plots of Jolly, Kevrekidis, and Titi [14] but empowers us to investigate more detail by separating positive and negative branches—stability differs along these branches. For example, Figure 6 shows that the negative bimodal branch is stable for $16.140 < \alpha < 22.556$, whereas the positive bimodal branch is unstable. The solution norm is signed corresponding to the sign of the first grid value, $u_1 = u(x_1)$. The blue curves are branches of stable fixed points, and the red curves are branches of unstable

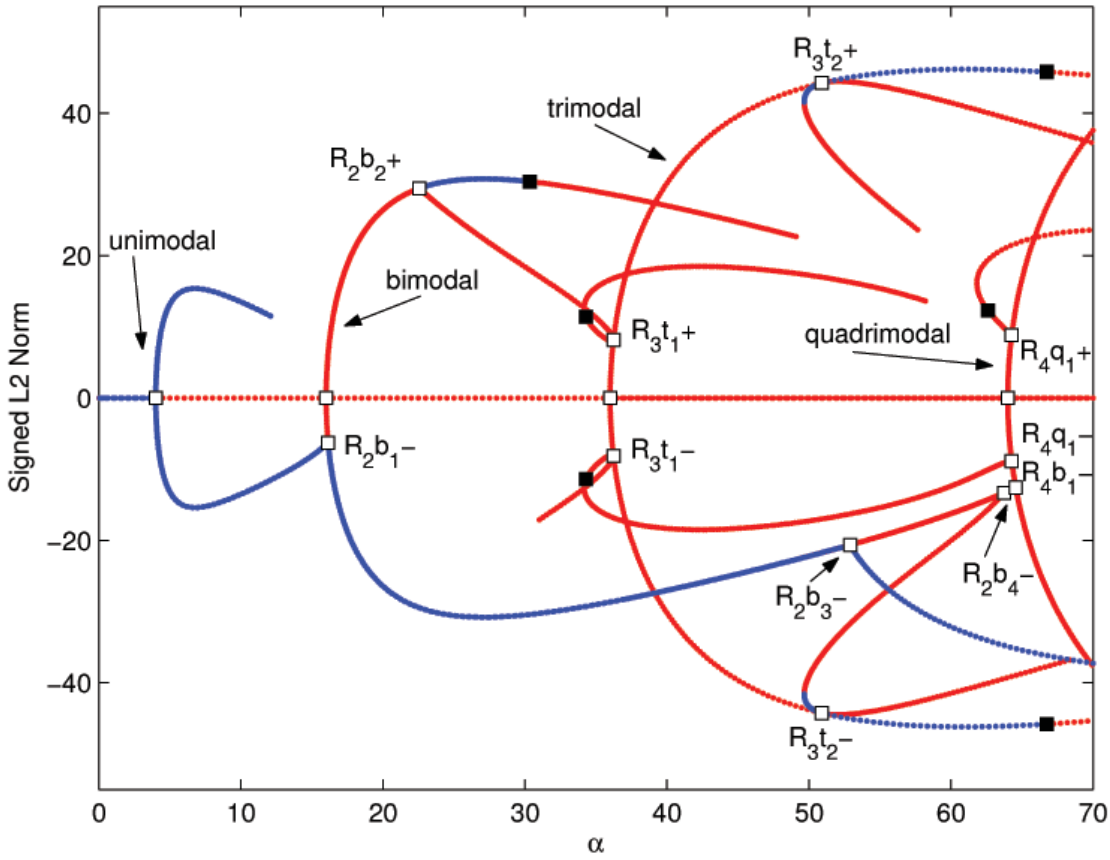


Figure 6. *Accurate bifurcation diagram $0 \leq \alpha \leq 70$ for the Kuramoto–Sivashinsky equation, using a sixth-order centered difference approximation with 48 points on the interval $[0, \pi]$. A signed L^2 norm is plotted against α.*

fixed points. The open squares denote pitchfork bifurcations, and the black squares denote Hopf bifurcations.

The labeling scheme used in Figure 6 follows that of [14] and [31], with the addition of a plus or minus sign, depending upon the sign of u_1. For example, the secondary bifurcation on the negative bimodal branch is labeled R_2b_1- from the labeling scheme of Scovel, Kevrekidis, and Nicolaenko, with the addition of the $-$ sign because it occurs on the negative branch. Figure 6 appears to show several discontinuities. For example, the positive unimodal branch ends at approximately $\alpha = 12$. This apparent discontinuity arises due to the convention adopted here of taking the sign of u_1 to sign the norm: actually there is a continuous transformation as the positive unimodal branch and the negative unimodal branch transform into the negative bimodal branch. It is straightforward to sign the branch near the trivial solution, but away from the trivial solution the distinction between positive and negative may be ambiguous and occasionally leads to jumps in the bifurcation diagram.

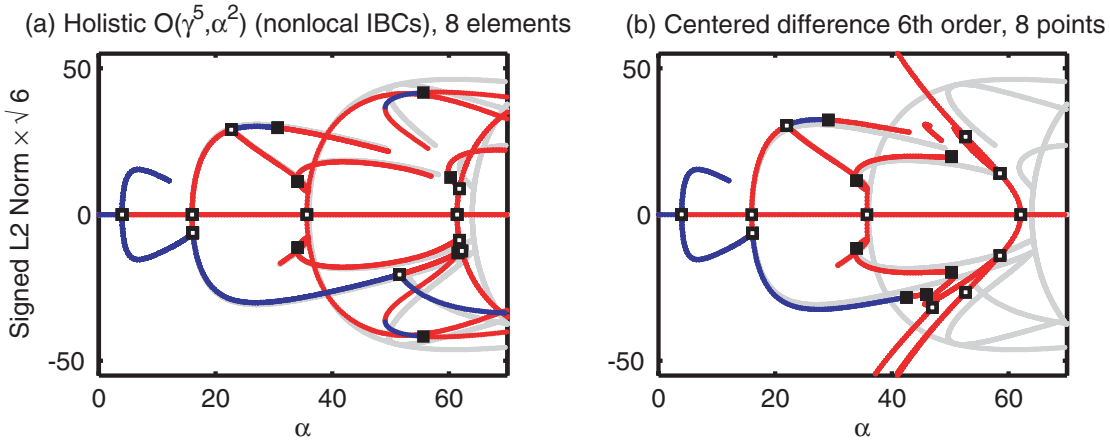


Figure 7. *Bifurcation diagrams for coarse grid approximations with eight elements on* $[0, \pi]$ *for* (a) *holistic model* (17)*,* $\mathcal{O}(\gamma^5, \alpha^2)$*,* (b) *centered difference sixth order* (22)*.*

4.2. Holistic models are accurate on coarse grids. We investigate the reproduction of the bifurcation diagram, Figure 6, of the Kuramoto–Sivashinsky dynamics using coarse grids on the interval $[0, \pi]$ by both the holistic and centered difference discretizations. Bifurcation diagrams are analogous to skeletons for the dynamics: dynamically evolving solutions fit around the steady states of a bifurcation diagram. Reproducing the "skeleton" of the bifurcation diagram is crucial to accurate modeling.

Note that holistic discretizations such as (15)–(17) provide reasonable solutions for significantly higher nonlinearity α than do the centered difference discretizations (20)–(22). We discuss bifurcation diagrams for $0 < \alpha < 70$.

4.2.1. Bifurcation diagrams show success. Now turn to the bifurcation diagram to obtain a more comprehensive view. We see that the holistic model has good bifurcation diagrams on a coarse grid of eight elements.

Figure 7 shows a side-by-side comparison of the holistic model with errors $\mathcal{O}(\gamma^5, \alpha^2)$ with eight elements on $[0, \pi]$ and the sixth-order centered difference approximation with eight grid points on $[0, \pi]$. These approximations are both nine-point stencil approximations. Plotted in grey, but without any stability information, is the accurate bifurcation diagram. The signed L_2 norms for the bifurcation diagrams on the coarse grid of eight elements are adjusted by a factor of $\sqrt{6}$ to allow comparison to the accurate bifurcation diagram constructed with 48 grid points on $[0, \pi]$. When comparing bifurcation diagrams of different grid resolutions, the signed L_2 norms are adjusted this way to provide a consistent reference. Figure 7(a) shows that the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model gives good agreement with the accurate bifurcation diagram for $\alpha < 40$ and qualitatively reproduces most of the bifurcation picture for $40 < \alpha < 70$. The $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model does not detect the bifurcation points $R_3 t_2 \pm$ on this coarse grid, and the bifurcation points $R_3 t_1 \pm$ are incorrectly identified as fold points. However, the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model finds all of the other bifurcation points in this range of nonlinearity α. Figure 7(b) shows that the sixth-order centered difference approximation gives good

Table 2

*α values at which bifurcation points occur for the various coarse grid approximations; * denotes a bifurcation point identified as a fold point.*

Approximation	R_2b_1	R_2b_2	R_2b_3	R_2b_4	R_3t_1	R_3t_2	R_4b_1	R_4q_1
Accurate 48 pts								
6th-order	16.14	22.56	52.89	63.74	36.23	50.91	64.56	64.28
Holistic 8 elements								
$\mathcal{O}(\gamma^3,\alpha^2)$ (15)	14.64	20.36	39.34	44.96	29.28*	—	45.28	44.87
$\mathcal{O}(\gamma^3,\alpha^3)$	14.65	20.52	39.66	45.16	29.33*	—	45.47	44.96
$\mathcal{O}(\gamma^3,\alpha^4)$	14.65	20.53	39.72	45.21	29.33*	—	45.51	44.97
$\mathcal{O}(\gamma^4,\alpha^2)$ (16)	16.00	22.56	48.62	57.38	34.73*	—	57.89	57.49
$\mathcal{O}(\gamma^4,\alpha^3)$	16.00	22.56	48.25	56.84	34.73*	—	57.45	57.28
$\mathcal{O}(\gamma^4,\alpha^4)$	16.00	22.57	48.10	56.63	34.73*	—	57.30	57.21
$\mathcal{O}(\gamma^5,\alpha^2)$ (17)	16.13	22.72	51.54	61.54	35.89*	—	62.20	61.78
$\mathcal{O}(\gamma^5,\alpha^3)$	16.13	22.73	51.53	61.37	35.91*	—	62.04	61.70
$\mathcal{O}(\gamma^5,\alpha^4)$	16.13	22.73	51.60	61.38	35.91*	—	62.02	61.69
Centered 8 pts								
2nd-order (20)	15.30	19.81	—	—	—	—	—	—
4th-order (21)	16.02	21.55	—	—	35.94*	—	—	—
6th-order (22)	16.12	21.99	—	—	35.83*	—	—	—
Holistic 12 elements								
$\mathcal{O}(\gamma^3,\alpha^2)$ (15)	15.45	21.67	45.96	53.94	32.86	45.98	54.49	54.17
$\mathcal{O}(\gamma^3,\alpha^3)$	15.45	21.69	46.05	54.00	32.87	46.33	54.55	54.20
$\mathcal{O}(\gamma^4,\alpha^2)$ (16)	16.11	22.62	51.93	62.10	35.90	50.92	62.83	62.52
$\mathcal{O}(\gamma^4,\alpha^3)$	16.11	22.62	51.94	62.10	35.90	50.94	62.83	62.52
Centered 12 pts								
2nd-order (20)	15.77	21.68	48.33	57.63	34.36	44.70	58.34	58.36
4th-order (21)	16.12	22.37	51.74	62.33	35.98	48.62	63.11	62.98

agreement with the accurate bifurcation diagram only for $\alpha < 20$ and qualitatively reproduces the bifurcation diagram for $20 < \alpha < 40$. The sixth-order centered difference approximation performs poorly for $\alpha > 40$. Table 2 lists the values of α at which the bifurcation points occur and confirms that the $\mathcal{O}(\gamma^5,\alpha^2)$ holistic model performs more accurately than the sixth-order centered difference approximation on this coarse grid of eight elements.

MacKenzie [19] also showed how the holistic model reproduces more of the bifurcation diagram than the sixth-order centered difference approximation even on a coarser grid of just six elements.

We also investigate various holistic models for the Kuramoto–Sivashinsky PDE by comparing bifurcation diagrams of holistic models of higher orders. We examine bifurcation diagrams for holistic models with errors $\mathcal{O}(\gamma^p,\alpha^q)$, for $3 \le p \le 5$ and $2 \le q \le 4$, and find that retaining terms of higher order in coupling parameter γ, corresponding to wider stencil approximations, gives much greater improvement in accuracy than retaining terms of higher order in the nonlinearity parameter α.

Figure 8 shows the bifurcation diagrams for the holistic models up to and including the $\mathcal{O}(\gamma^5,\alpha^4)$ holistic model. Surveying across the columns of Figure 8, note the bifurcation diagrams for holistic models of increasing order of coupling parameter γ, corresponding to approximations of increasing stencil width. For example, the top row of Figure 8 shows the bifurcation diagrams for the holistic models (15), (16), and (17), respectively. Lower rows of

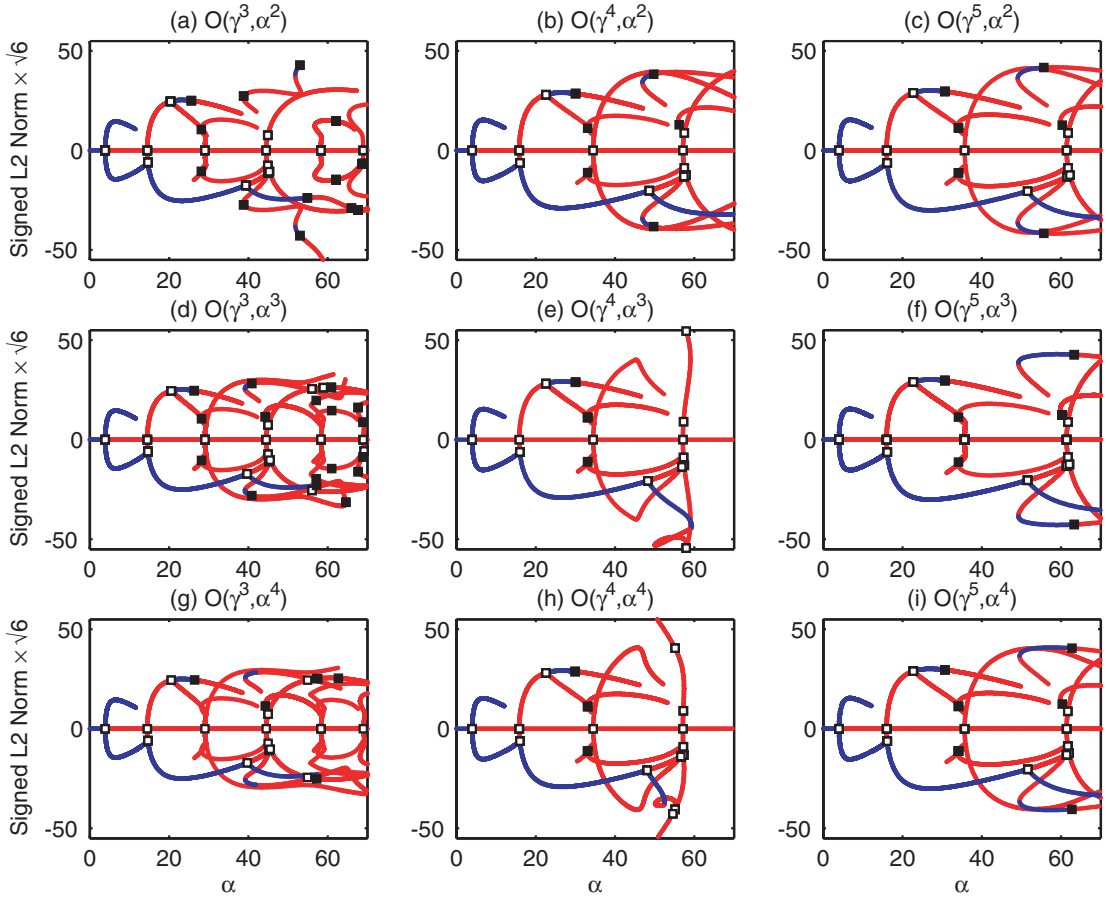


Figure 8. *Bifurcation diagrams for the holistic models with eight elements on the interval* $[0, \pi]$ *up to and including the* $\mathcal{O}(\gamma^5, \alpha^4)$ *holistic model.*

Figure 8 display the bifurcation diagrams for increasing orders of the nonlinearity parameter α. Figure 8 illustrates the improvement in accuracy of the higher-order holistic models. Note first the dramatic improvement in accuracy gained by moving from left to right across Figure 8, corresponding to approximations of higher orders in the coupling parameter γ.

Second, less improvement is gained by moving from top to bottom in Figure 8, corresponding to approximations of higher order in the nonlinearity parameter α. There are some peculiarities about this series of bifurcation pictures for holistic models of increasing order in α. For the five-point stencil approximations displayed in the left column of Figure 8, higher orders in α appear to gain some improvement. In particular, Figures 8(d) and (g) show that the $\mathcal{O}(\gamma^3, \alpha^3)$ and $\mathcal{O}(\gamma^3, \alpha^4)$ holistic models reproduce the unstable trimodal branches that were missing from the $\mathcal{O}(\gamma^3, \alpha^2)$ bifurcation diagram shown in Figure 8(a). However, for the seven-point stencil approximations displayed in the second column of Figure 8, holistic models of higher orders in α lose some features of the Kuramoto–Sivashinsky PDE. The correct behavior of the unstable trimodal and quadrimodal branches is reproduced for the $\mathcal{O}(\gamma^4, \alpha^2)$

model shown in Figure 8(b), but not reproduced for the higher-order $\mathcal{O}(\gamma^4, \alpha^3)$ and $\mathcal{O}(\gamma^4, \alpha^4)$ models shown in Figures 8(e) and (h) respectively. For the nine-point stencil approximations, displayed in the third column of Figure 8, the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model shown in Figure 8(c) reproduces the unstable trimodal branch, whereas the higher-order $\mathcal{O}(\gamma^5, \alpha^3)$ model shown in Figure 8(f) does not reproduce the unstable trimodal branch. These peculiarities suggest that while we have observed excellent performance of the holistic models constructed with the nonlocal IBCs on coarse grids, it may be possible that modifications could be made to the nonlocal IBCs such that higher-order approximations in the nonlinear parameter are improved. Further research will explore such possible modifications.

4.2.2. Holistic models outperform centered differences. Section 4.2.1 shows that the performance of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (17) constructed with nonlocal IBCs is far superior to the explicit sixth-order centered difference approximation (22). To complete the comparison of holistic models to explicit centered difference schemes, we compare the $\mathcal{O}(\gamma^3, \alpha^2)$ (15) and $\mathcal{O}(\gamma^4, \alpha^2)$ (16) holistic models to the second-order (20) and fourth-order (21) centered difference approximations; these are five-point and seven-point discretizations respectively.

The first row of Figure 9 is a side-by-side comparison of the $\mathcal{O}(\gamma^3, \alpha^2)$ holistic model and the second-order centered difference approximation with eight elements on $[0, \pi]$. The second row of Figure 9 is a side-by-side comparison of the $\mathcal{O}(\gamma^4, \alpha^2)$ holistic model and the fourth-order centered difference approximation on the same coarse grid. The accurate bifurcation diagram is plotted in grey without any stability information.

Although comparing Figures 9(b) and (d) shows that some improvement is gained by taking higher-order centered difference approximations, this improvement is not as pronounced as for the holistic models on this coarse grid, as shown in Figures 9(a) and (c). Both the second-order and fourth-order centered difference approximations fail to reproduce the correct behavior of the unstable trimodal and quadrimodal branches. In contrast, even the five-point stencil $\mathcal{O}(\gamma^3, \alpha^2)$ holistic approximation qualitatively reproduces the trimodal and quadrimodal branches on the same coarse grid. The values at which the bifurcation points occur are listed in Table 2 and confirm that these holistic models outperform the centered difference approximations on this coarse grid of eight elements on $[0, \pi]$.

4.2.3. Grid refinement improves accuracy. Since the equivalent PDEs (24) for our holistic models are of $\mathcal{O}(h^2)$, $\mathcal{O}(h^4)$, and $\mathcal{O}(h^6)$ consistency, respectively (see Table 1), grid refinement should result in improved accuracy.

Figure 10 shows the bifurcation diagrams of the holistic models up to and including the $\mathcal{O}(\gamma^4, \alpha^3)$ model on a finer grid of 12 elements on $[0, \pi]$. Compare Figure 10 with Figure 8 to confirm the improved accuracy for the holistic models on this refined grid. Table 2 also shows that the bifurcation points are more accurately reproduced for the holistic models on this refined grid.

4.3. Comparison to Galerkin approximations. Here we investigate the traditional Galerkin and nonlinear Galerkin approximations [14] for the Kuramoto–Sivashinsky equation (1) with the periodic and odd conditions (25). We find that the holistic models compare well with the Galerkin methods. While the Galerkin methods are of superior accuracy for solving the Kuramoto–Sivashinsky PDE (1) with periodic boundary conditions, because of their global

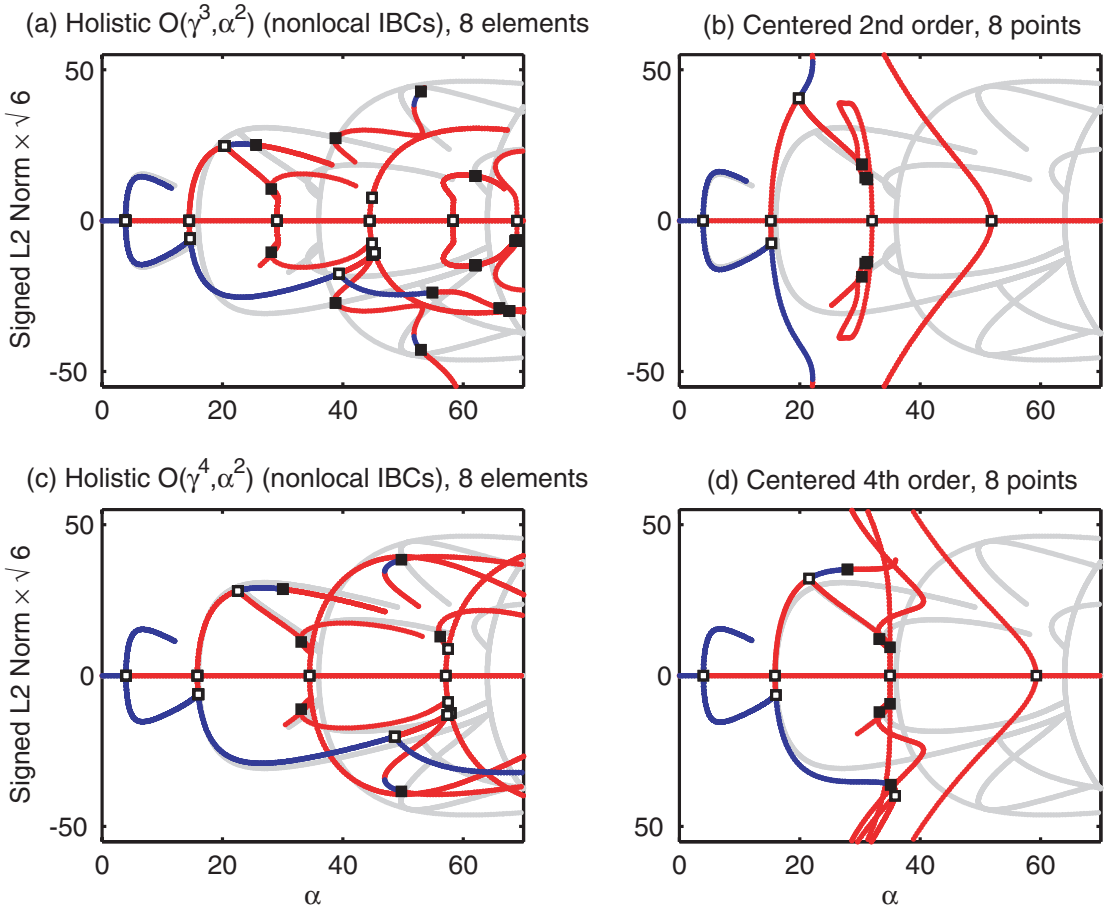


Figure 9. *Bifurcation diagrams for* (a) $\mathcal{O}(\gamma^3, \alpha^2)$ *holistic model,* (b) *second-order centered difference,* (c) $\mathcal{O}(\gamma^4, \alpha^2)$ *holistic model, and* (d) *fourth-order centered difference all with eight elements on the interval* $[0, \pi]$.

nature they lack the flexibility of the local nature of the holistic models. Although not explored here, this local nature of the holistic models empowers its use with physical boundary conditions [27] other than periodic.

Galerkin methods seek solutions in the form which is dominantly the superposition of m periodic, global modes:

$$u(x,t) = \sum_{k=1}^{m} b_k(t) \sin(kx). \tag{26}$$

The m-mode traditional Galerkin approximation [14] is

$$\frac{db_k}{dt} \approx \left(-4k^4 + \alpha k^2\right) b_k - \alpha \beta_k^m, \quad 1 \le k \le m, \tag{27}$$

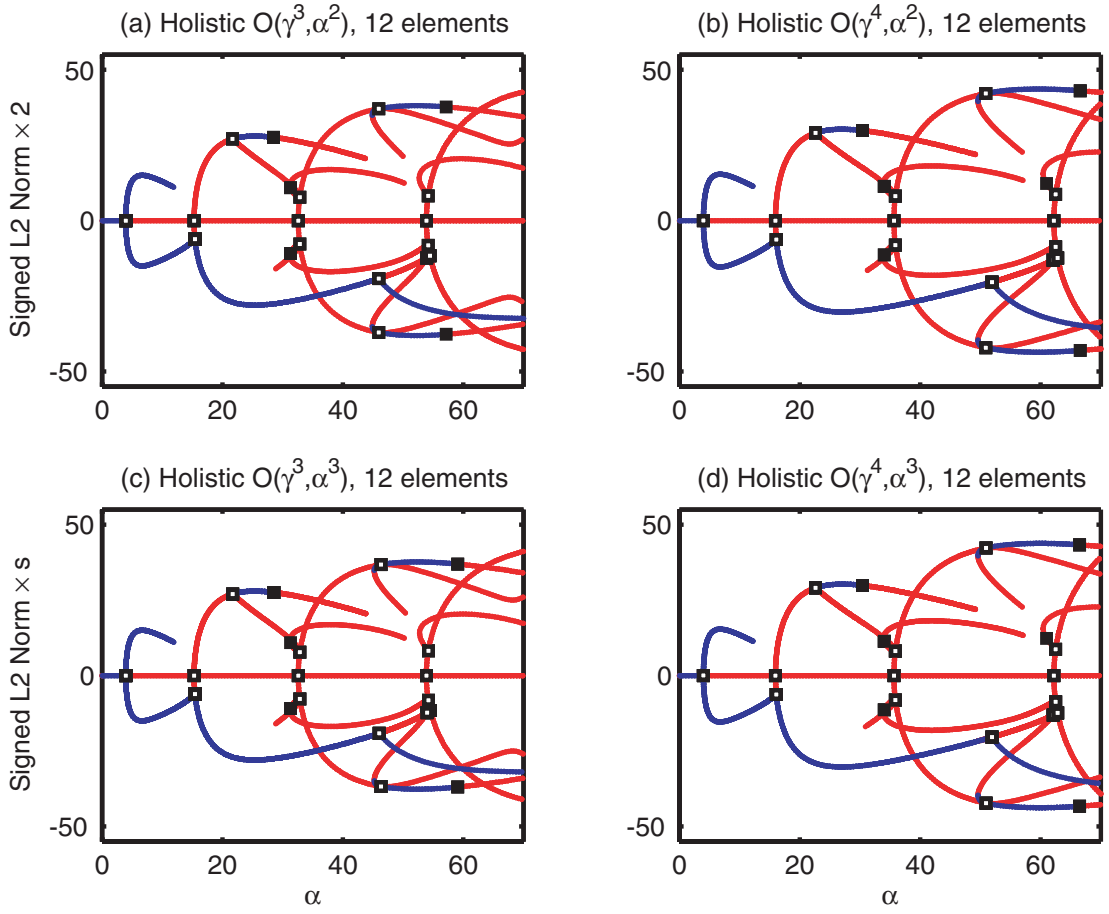


Figure 10. *Bifurcation diagrams for the holistic models with* 12 *elements on the interval* $[0, \pi]$. *Compare with Figure* 8 *with* 8 *elements.*

where

$$\beta_k^m(b_1, \ldots, b_m) = \frac{1}{2} \sum_{j=1}^{m} j b_j \left[b_{k+j} + \operatorname{sign}(k-j) b_{|k-j|} \right]. \tag{28}$$

The m-mode first iterate nonlinear Galerkin approximation [14] is based upon the adiabatic approximation (30) for higher wavenumber modes $k = m+1 : 2m$, namely,

$$\frac{db_k}{dt} \approx \left(-4k^4 + \alpha k^2\right) b_k - \alpha \beta_k^{2m}(b_1, \ldots, b_m, \phi_{m+1}, \ldots, \phi_{2m}), \tag{29}$$

for $1 \le k \le m$, where

$$\phi_j = -\frac{\alpha}{4j^4} \beta_j^{2m}\left(b_1, \ldots, b_m, 0, \ldots, 0\right) \tag{30}$$

for $m+1 \le j \le 2m$ and β_j^{2m} is given by (28).

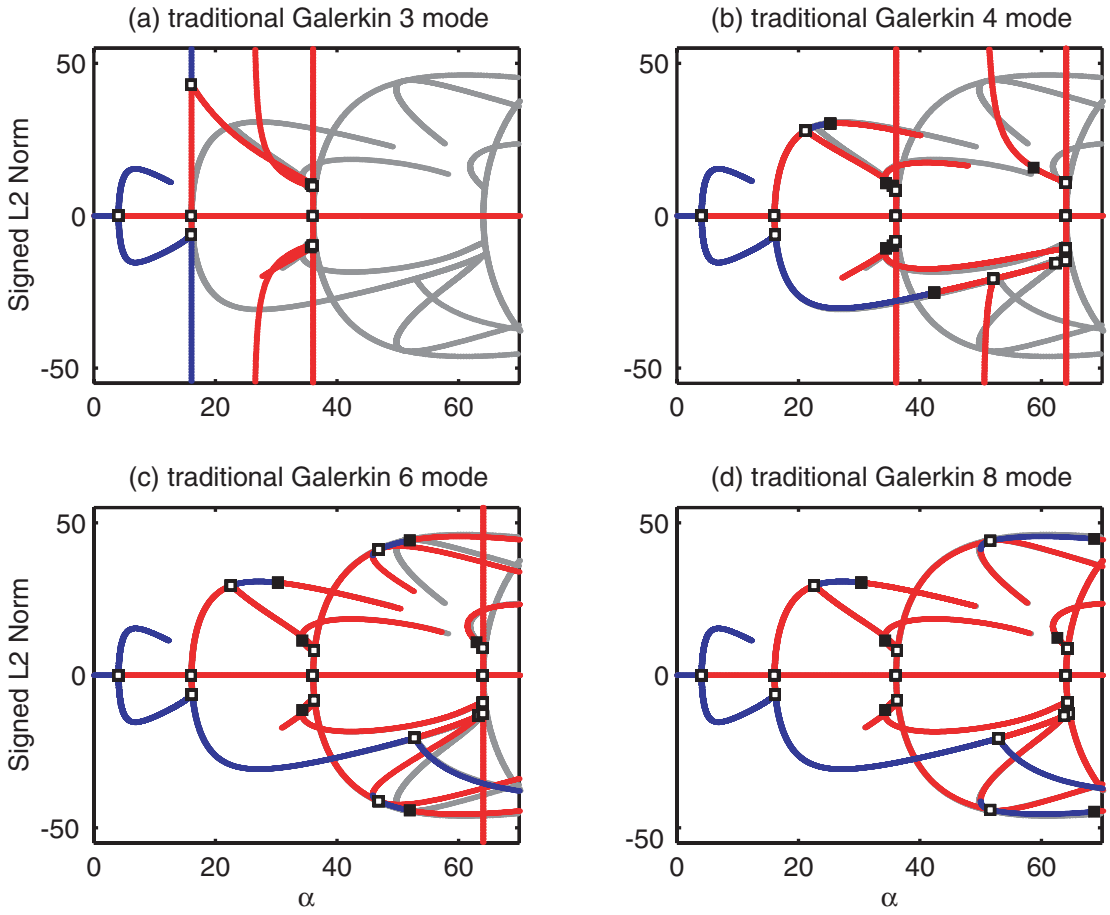


Figure 11. *Bifurcation diagrams for* (a) 3 *mode,* (b) 4 *mode,* (c) 6 *mode, and* (d) 8 *mode traditional Galerkin approximations on* $[0, \pi]$.

Obtain higher-order nonlinear Galerkin approximations [22] through recognizing time derivatives of these and even higher wavenumber modes. We do not explore such higher-order Galerkin approximations.

Now examine the bifurcation diagrams of the two Galerkin approximations (26)–(30) for $0 \leq \alpha \leq 70$, and compare them with the bifurcation diagrams of the holistic models on coarse grids, presented in section 4.2. Figure 11 shows the bifurcation diagrams for the 3 mode, 4 mode, 6 mode, and 8 mode traditional Galerkin approximations on $[0, \pi]$. See that at least four modes are needed to qualitatively reproduce the behavior of the stable bimodal branch. The $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model with eight elements from Figure 7(a) and the 8 mode traditional Galerkin approximation qualitatively model most steady-state dynamics. However, the 8 mode traditional Galerkin approximation is more accurate.

Figure 12 shows the bifurcation diagrams for the 3 mode, 4 mode, 6 mode, and 8 mode first iterate nonlinear Galerkin approximations (29) on $[0, \pi]$. Note the impressive accuracy for the low mode first iterate nonlinear Galerkin approximations. The 6 mode nonlinear

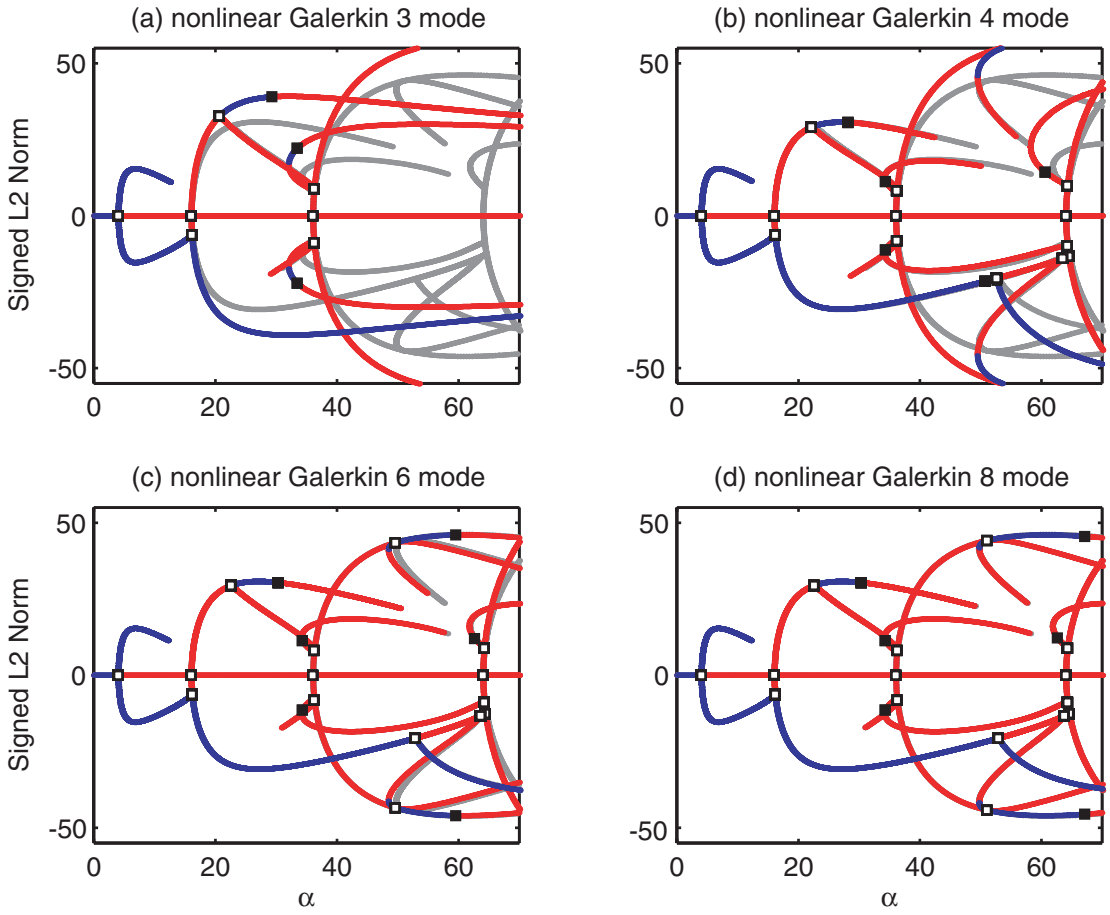


Figure 12. *Bifurcation diagrams for* (a) 3 *mode,* (b) 4 *mode,* (c) 6 *mode, and* (d) 8 *mode first iterate nonlinear Galerkin approximations on* $[0, \pi]$.

Galerkin approximation reproduces all of the steady-state dynamics for the range $0 \leq \alpha \leq 70$. There is no discernible difference between the bifurcation diagram of the 8 mode nonlinear Galerkin approximation and the accurate bifurcation diagram for this range of α. Table 3 lists the values of nonlinearity parameter α at which bifurcation points occur for the coarse grid holistic models and the Galerkin approximations [14]. The low mode first iterate nonlinear Galerkin approximations are impressively accurate.

This evidence suggests that the holistic models are competitive with traditional Galerkin approximations, but that nonlinear Galerkin models are significantly better. However, recall that the holistic models are based upon analysis of local dynamics, and thus we expect them to be more flexibly useful in applications than the global methods of these Galerkin approximations, which rely on not only global eigenfunctions being known, here the sine functions, but also knowing closed-form expressions for nonlinear combinations of the eigenfunctions. Furthermore, even when the nonlinear combinations of the eigenfunctions are known, the operation count of the nonlinear Galerkin method is high. For example, evaluating the time

Table 3

α values at which bifurcation points occur for the various coarse grid holistic models and low mode Galerkin approximations. ("k-m" denotes the k mode Galerkin.)

Approximation	R_2b_1	R_2b_2	R_2b_3	R_2b_4	R_3t_1	R_3t_2	R_4b_1	R_4q_1
Accurate 48 pts								
6th-order	16.14	22.56	52.89	63.74	36.23	50.91	64.56	64.28
Holistic 8 elements								
$\mathcal{O}(\gamma^3, \alpha^2)$ (15)	14.64	20.36	39.34	44.96	29.28*	—	45.28	44.87
$\mathcal{O}(\gamma^3, \alpha^3)$	14.65	20.52	39.66	45.16	29.33*	—	45.47	44.96
$\mathcal{O}(\gamma^3, \alpha^4)$	14.65	20.53	39.72	45.21	29.33*	—	45.51	44.97
$\mathcal{O}(\gamma^4, \alpha^2)$ (16)	16.00	22.56	48.62	57.38	34.73*	—	57.89	57.49
$\mathcal{O}(\gamma^4, \alpha^3)$	16.00	22.56	48.25	56.84	34.73*	—	57.45	57.28
$\mathcal{O}(\gamma^4, \alpha^4)$	16.00	22.57	48.10	56.63	34.73*	—	57.30	57.21
$\mathcal{O}(\gamma^5, \alpha^2)$ (17)	16.13	22.72	51.54	61.54	35.89*	—	62.20	61.78
$\mathcal{O}(\gamma^5, \alpha^3)$	16.13	22.73	51.53	61.37	35.91*	—	62.04	61.70
$\mathcal{O}(\gamma^5, \alpha^4)$	16.13	22.73	51.60	61.38	35.91*	—	62.02	61.69
Holistic 12 elements								
$\mathcal{O}(\gamma^3, \alpha^2)$ (15)	15.45	21.67	45.96	53.94	32.86	45.98	54.49	54.17
$\mathcal{O}(\gamma^3, \alpha^3)$	15.45	21.69	46.05	54.00	32.87	46.33	54.55	54.20
$\mathcal{O}(\gamma^4, \alpha^2)$ (16)	16.11	22.62	51.93	62.10	35.90	50.92	62.83	62.52
$\mathcal{O}(\gamma^4, \alpha^3)$	16.11	22.62	51.94	62.10	35.90	50.94	62.83	62.52
Galerkin [14]								
3-m Euler–Galerkin	16.10	20.59	246.14	—	36.21	—	—	—
3-m Pseudo-stdy II	16.13	21.93	102.90	—	36.21	—	—	—
3-m Pseudo-stdy	16.13	22.01	93.91	—	36.24	63.91	—	—
12-m traditional	16.14	22.56	52.89	63.74	36.23	50.91	64.56	64.28
6-m traditional	16.14	22.55	52.72	63.28	36.23	46.85	64.00	64.00
3-m traditional	16.14	16.00	16.0	16.0	36.00	36.0	—	—

derivatives in (29) costs roughly $3m^2$ operations. Being quadratic in the number m of modes, this evaluation is not scalable in the size of the domain. However, a holistic model with m elements to some order $\mathcal{O}(\gamma^p, \alpha^q)$ will evaluate all necessary time derivatives in a time proportional to m and is thus scalable. True, the constant of proportionality may be rather large for any given order of truncation, due to the modeling of the subgrid and interelement interactions. Nonetheless, holistic discretization is scalable, whereas Galerkin modeling is not.

4.4. Coarse grids allow large time steps. A major benefit of accurate models on coarse grids is that larger time steps are possible while maintaining numerical stability. Section 4.2 shows the remarkable accuracy of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (17) on a coarse grid of eight elements. Here we investigate the maximum stable time step for *explicit* Runge–Kutta time integration on various holistic models—implicit integration schemes are not considered here.

In particular, we compare approximations of similar accuracy but different grid resolutions to demonstrate the superior performance of the holistic models. MacKenzie [19] showed that the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (15) on eight elements is of similar accuracy to the 16-point second-order centered difference approximation (20). Thus we compare the computability of these two schemes.

Numerical experiments used the fourth-order Runge–Kutta scheme to estimate the maximum stable time step for different holistic models and centered difference approximations

Table 4
Approximate maximum time steps for stability of the fourth-order Runge–Kutta scheme.

Approximation	$\alpha = 10$	$\alpha = 20$	$\alpha = 30$
Holistic 8 elements			
$\mathcal{O}(\gamma^3, \alpha^2)$ (15)	.0011	.0014	.0017
$\mathcal{O}(\gamma^3, \alpha^3)$	.0011	.0014	.0017
$\mathcal{O}(\gamma^3, \alpha^4)$	.0011	.0014	.0017
$\mathcal{O}(\gamma^4, \alpha^2)$ (16)	.0006	.0007	.0008
$\mathcal{O}(\gamma^4, \alpha^3)$	.0006	.0007	.0008
$\mathcal{O}(\gamma^4, \alpha^4)$	.0006	.0007	.0008
$\mathcal{O}(\gamma^5, \alpha^2)$ (17)	.0005	.0005	.0006
$\mathcal{O}(\gamma^5, \alpha^3)$	.0005	.0005	.0006
$\mathcal{O}(\gamma^5, \alpha^4)$	.0005	.0005	.0006
Centered 8 points			
2nd-order (20)	.0011	.0012	—
4th-order (21)	.0006	.0007	.0008
6th-order (22)	.0005	.0005	.0006
Centered 16 points			
2nd-order (20)	.00006	.00006	.00006

at various values of nonlinearity parameter α. Table 4 lists the approximate maximum time steps that maintain numerical stability along both the negative unimodal branch at $\alpha = 10$ and the negative bimodal branch at $\alpha = 20$ and $\alpha = 30$. For the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model with eight elements, the maximum time step maintaining numerical stability is approximately 10 times larger than the corresponding time step for the second-order centered difference approximation with 16 grid points. The $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model requires approximately 3 times the number of floating point operations per grid value at each time step compared to the second-order centered difference approximation. However, on a coarse grid of 16 points the second-order centered difference approximation must be applied at twice as many grid points. Thus the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model can be integrated an order of magnitude faster than the second-order centered difference approximation while maintaining similar accuracy.

Note that Table 4 shows that the higher-order terms in the nonlinearity α, generated by the holistic method, do not reduce numerical stability. Wider stencil holistic approximations reduce the maximum stable time step somewhat, but so do the wider stencil conventional centered difference approximations. Thus, bear in mind that we need to balance the accuracy gained by using higher-order approximation in γ, that is, wider stencil approximations, with the reduction in numerical stability and the increase in computation per grid value.

5. Holistic models are accurate for time dependent phenomena. The Kuramoto–Sivashinsky equation (1) has rich dynamics [16, 14, 15, 31, 5, 12, 6, 2]. Having established the excellent performance of the holistic models in reproducing the steady states of the Kuramoto–Sivashinsky PDE in section 4, we now investigate the holistic model's performance at reproducing time dependent phenomena. The Kuramoto–Sivashinsky PDE exhibits complex time dependent behavior such as limit cycles, period doubling, and spatio-temporal chaos. This provides us with an example to explore the holistic approach to modeling time dependent phenomena with relatively coarse discretizations.

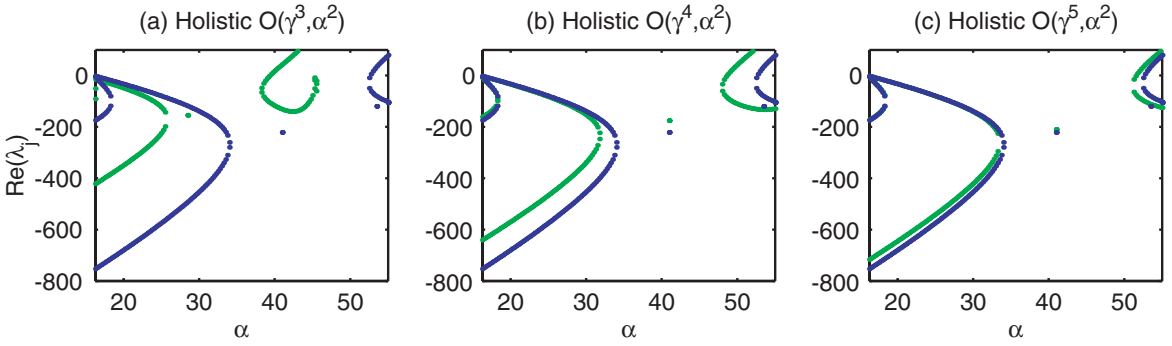


Figure 13. *The four largest (least negative) eigenvalues along the stable bimodal branch for the* (a) $\mathcal{O}(\gamma^3, \alpha^2)$ (15), (b) $\mathcal{O}(\gamma^4, \alpha^2)$ (16), (c) $\mathcal{O}(\gamma^5, \alpha^2)$ (17) *holistic models shown in green for eight elements on* $[0, \pi]$. *Accurate eigenvalues are shown in blue.*

We restrict our attention to 2π periodic solutions,

$$u(x,t) = u(x + 2\pi, t). \tag{31}$$

Initially we restrict ourselves further to solutions with odd symmetry, as in the previous section, which exhibit (see Figure 6) Hopf bifurcations to limit cycle solutions, and subsequent period doubling bifurcations apparently leading to low-dimensional chaos [14, 15, 31]. In section 5.1 we examine the dynamics of the holistic models on coarse grids through the eigenvalues of the models near the steady states. For example, we see that the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model reproduces much of the eigenvalue information for $0 \leq \alpha \leq 70$ on a coarse grid of eight elements. In section 5.2 we explore the bifurcation diagrams near the first Hopf bifurcation and capture the stable limit cycles and period doubling sequence. The holistic models more accurately model the dynamics than centered difference approximations of equal stencil width. Subsequently we just require spatial periodicity, whence stable traveling waves appear followed by, at higher values of the nonlinearity parameter α, more complex spatio-temporal chaos, as investigated by Holmes, Lumley, and Berkooz [12] and Dankowicz et al. [6]. In section 5.3 we find that the holistic discretizations more accurately model the amplitude and wave speed of traveling wave solutions, and predict better space time plots and time averaged power spectra, than the corresponding centered difference approximations.

5.1. Dynamics near steady states are reproduced. Consider the eigenvalues of the Kuramoto–Sivashinsky PDE (1) linearized about the steady states and restricted to odd symmetry. Accurate modeling of the eigenvalues near the steady states is a necessary condition for the accurate modeling of the dynamics.

Compare eigenvalues along the bimodal branch. Consider the real part of the four largest (least negative real part) eigenvalues for low-order holistic models and compare to explicit centered difference approximations on a coarse grid of eight elements on $[0, \pi]$. Figure 13 shows the four largest eigenvalues for the $\mathcal{O}(\gamma^3, \alpha^2)$ (15), $\mathcal{O}(\gamma^4, \alpha^2)$ (16), and $\mathcal{O}(\gamma^5, \alpha^2)$ (17)

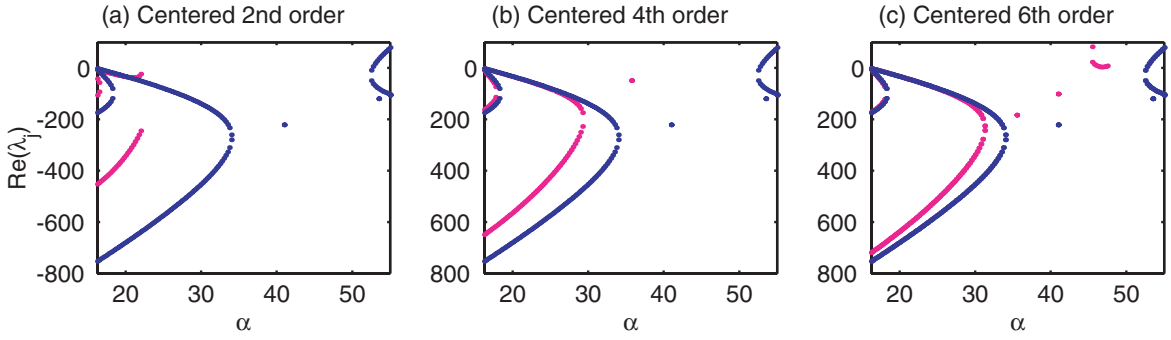


Figure 14. *The four largest eigenvalues along the stable bimodal branch for the* (a) *second-order* (20), (b) *fourth-order* (21), (c) *sixth-order* (22) *centered difference approximations shown in magenta for eight grid points on* $[0, \pi]$. *Accurate eigenvalues are shown in blue.*

holistic models in green and the accurate solution in blue.[6] Figure 13(c) shows that the four largest eigenvalues for the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model closely match the accurate solution over this range of nonlinearity parameter α.

Similarly, Figure 14 shows the four largest eigenvalues for the second-order (20), fourth-order (21), and sixth-order (22) centered difference approximations in magenta on the same coarse grid. The centered difference approximations shown here are of equal stencil width to the corresponding holistic models in Figure 13. Figure 14(a) shows that the second-order centered difference barely approximates the behavior of the stable bimodal branch for $\alpha < 20$. Even the sixth-order centered difference approximation, Figure 14(c), is inferior to the $\mathcal{O}(\gamma^4, \alpha^2)$ holistic model for $\alpha > 30$. This is despite the sixth-order centered difference model having a wider stencil of nine points compared to the seven-point stencil of the $\mathcal{O}(\gamma^4, \alpha^2)$ holistic model. Figures 13 and 14 show that the low-order holistic models are superior to the corresponding centered difference approximations for reproducing the dynamics near the stable bimodal branch.

Compare eigenvalues across the bifurcation diagram. Here we explore a new view of the earlier bifurcation diagrams that additionally depicts the *real part* of the eight largest (least negative) eigenvalues by color. Compare the eigenvalues of the $\mathcal{O}(\gamma^5, \alpha^2)$ (17) holistic model (see Figure 15) on the coarse grid of eight elements on $[0, \pi]$ to accurate ones for the Kuramoto–Sivashinsky PDE (see Figure 16) over the nonlinearity parameter $0 \le \alpha \le 70$. Colors code the magnitude of the real part of the eigenvalues according to the color bar shown; the least negative eigenvalues are plotted above the more negative to give a small band of color for each branch of steady states at each parameter value. Similarly to the bifurcation diagrams shown in section 4, the open squares denote bifurcation points, and the black squares denote Hopf bifurcations. Figure 15, when compared to Figure 16, shows that in addition to reproducing the stability of the accurate Kuramoto–Sivashinsky PDE for $0 \le \alpha \le 70$, as discussed in section 4.2, the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model reproduces well the eigenvalues for most of this range of the nonlinearity parameter α. This accurate modeling of the eigenvalues is evidence of

[6] As in section 4, a sixth-order centered difference approximation with 48 grid points on $[0, \pi]$ provides the accurate reference for solutions.

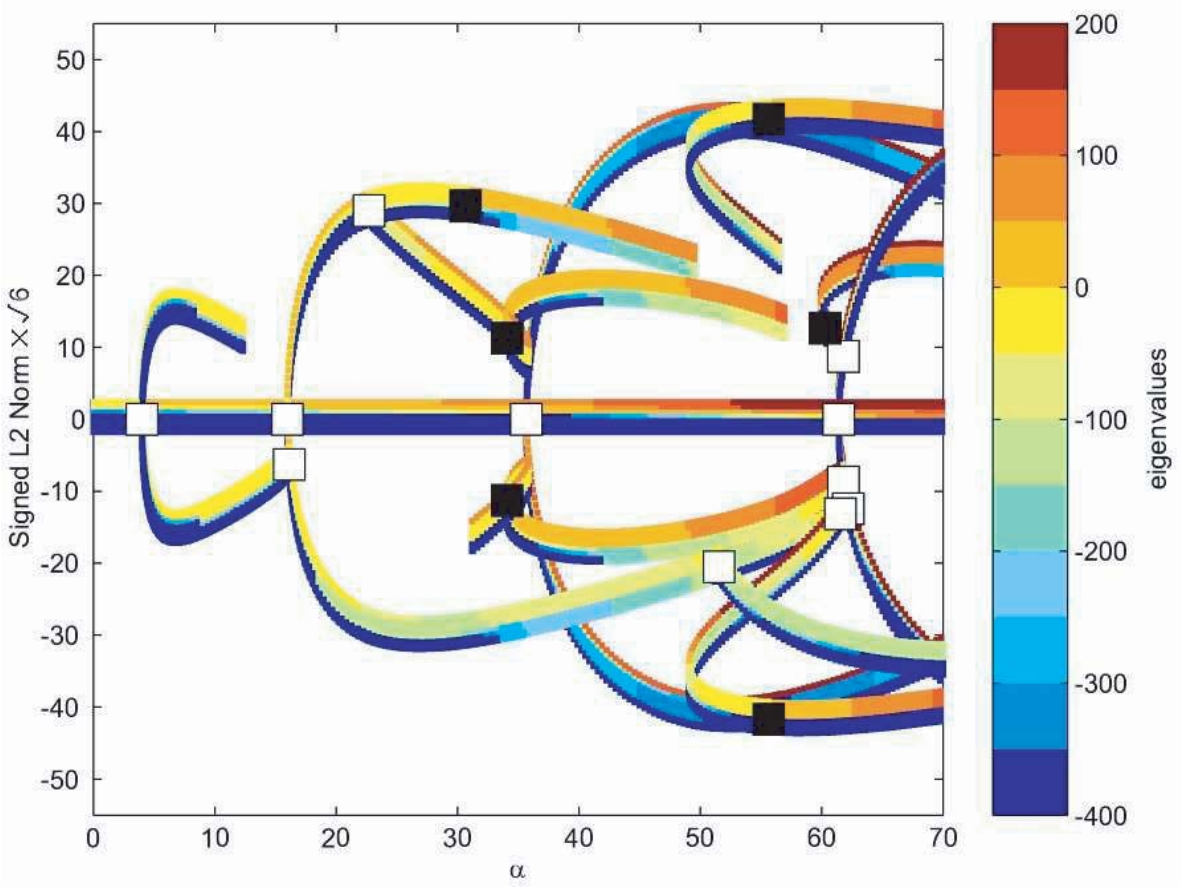


Figure 15. *Bifurcation diagram of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (15) with eight elements and odd symmetry on $[0, \pi]$, depicting the real parts of the eight largest (least negative) eigenvalues color-coded according to the color bar shown.*

accurate modeling of the Kuramoto–Sivashinsky dynamics, at least near the steady states.

5.2. Extend the Hopf bifurcations. Hopf bifurcations give rise to time periodic solutions (limit cycles). We explore the predictions of the various models to see how well they capture these strongly time dependent phenomena.

Here we investigate the bifurcation diagrams obtained by extending the first Hopf bifurcation, at $\alpha = 30.345$, on the positive bimodal branch and the period doubling sequence that ensues. Trajectories in the period doubling sequence are reported and compared by MacKenzie [19]. As before, the holistic models outperform the corresponding centered difference approximations.

Investigate the first Hopf bifurcation HB_1 with a coarse grid of eight elements on $[0, \pi]$. Figure 17 shows the bifurcation diagrams of the low-order holistic models and the accurate bifurcation diagram near the first Hopf bifurcation. The stable limit cycles (light blue) that continue from this bifurcation point undertake a period doubling sequence commencing at a point labeled PD (yellow square). The pair of unstable limit cycles born at PD give rise to the

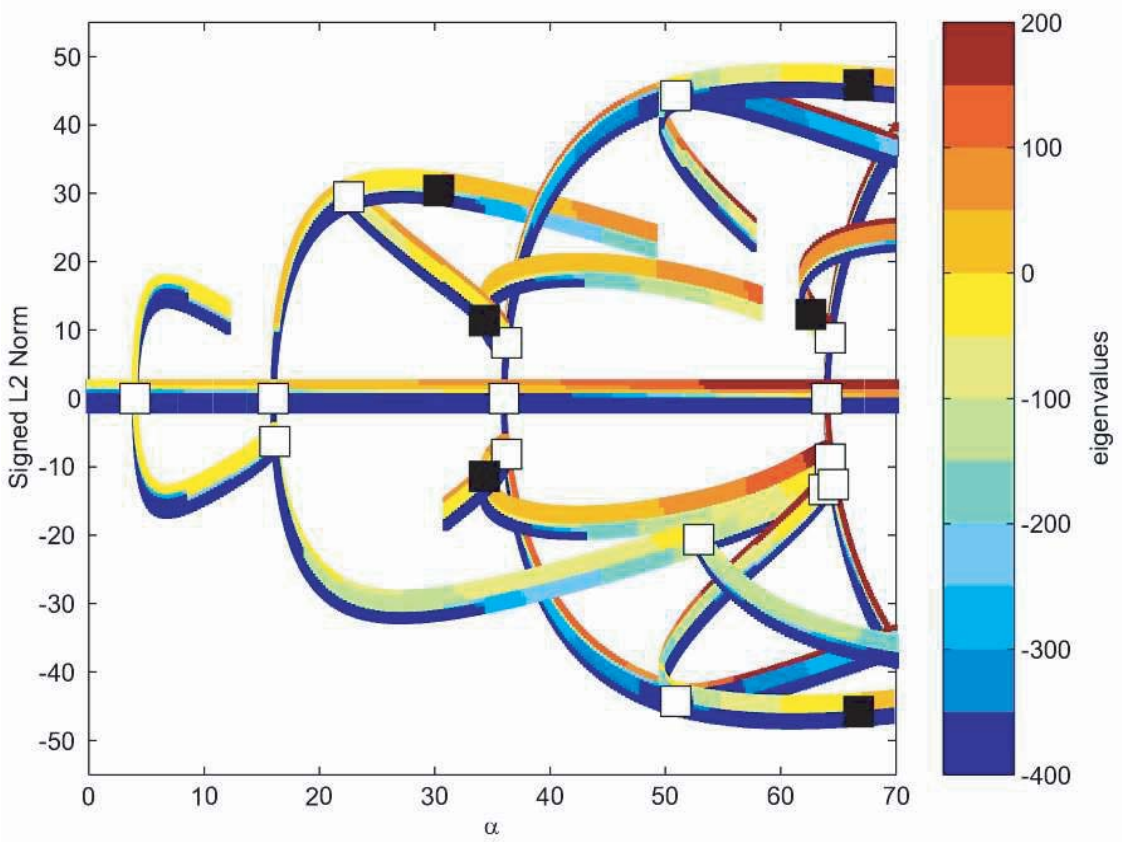


Figure 16. *Bifurcation diagram of the accurate Kuramoto–Sivashinsky* PDE*, depicting the real parts of the eight largest (least negative) eigenvalues, color-coded according to the color bar shown.*

period doubling sequence leading to chaos.

The accurate bifurcation diagram shown is identical to the bifurcation diagram for the same range of α produced by Jolly, Kevrekidis, and Titi [14]. Figure 17(a) shows that even the lowest-order $\mathcal{O}(\gamma^3, \alpha^2)$ (15) holistic model reproduces the first Hopf bifurcation and finds the period doubling point on this coarse grid of eight elements.[7] In comparison, the corresponding second-order centered difference approximation does not even have the first Hopf bifurcation; see Figure 9(b).

Figures 17(b) and (c) show that higher-order holistic models accurately model the first Hopf bifurcation and the resulting stable and unstable limit cycles. The accuracy of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (17) for reproducing these periodic solutions of the Kuramoto–Sivashinsky PDE is remarkable on this coarse grid. Figure 18 shows the corresponding bifurcation diagrams for the fourth-order and sixth-order centered difference approximations with eight grid points on $[0, \pi]$. Compare Figures 18 and 17 to see that the sixth-order cen-

[7]Figure 17(a) displays the bifurcation diagram for $25 \leq \alpha \leq 32$ compared to $30 \leq \alpha \leq 37$ for the other diagrams. Since the first Hopf bifurcation for the $\mathcal{O}(\gamma^3, \alpha^2)$ holistic model occurs at $\alpha = 25.595$ the bifurcation diagram is shifted to contain the important dynamics.

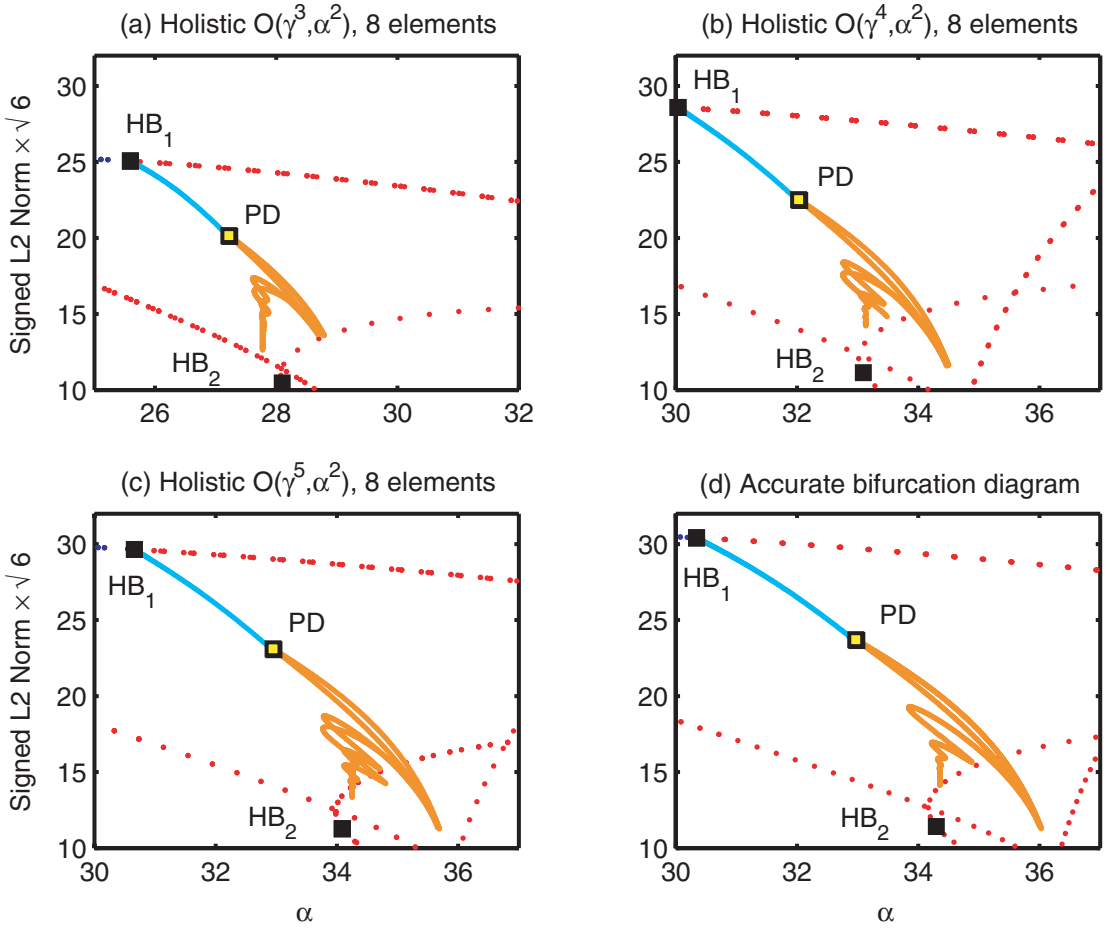


Figure 17. *Bifurcation diagrams near the first Hopf bifurcation for* (a) $\mathcal{O}(\gamma^3, \alpha^2)$ (15), (b) $\mathcal{O}(\gamma^4, \alpha^2)$ (16), (c) $\mathcal{O}(\gamma^5, \alpha^2)$ (17) *holistic models with eight elements on* $[0, \pi]$ *and* (d) *an accurate bifurcation diagram. Stable limit cycles are shown in light blue, and unstable limit cycles are shown in orange.*

tered difference approximation which has a nine-point stencil does not perform as well as the $\mathcal{O}(\gamma^4, \alpha^2)$ holistic model (16), which has a seven-point stencil. Figures 17(b) and (c) show that higher-order holistic models more accurately model the first Hopf bifurcation and the resulting stable and unstable limit cycles. Table 5 shows the parameter values α for the Hopf bifurcations, HB_1, and the initial period doubling point PD: both the $\mathcal{O}(\gamma^4, \alpha^2)$ and $\mathcal{O}(\gamma^5, \alpha^2)$ holistic models are more accurate than the fourth-order and sixth-order centered difference approximations in reproducing the first Hopf bifurcation and the resulting period doubling point.

5.3. Dynamics of periodic patterns without odd symmetry. Consider the Kuramoto–Sivashinsky PDE (1) with solutions that are spatially periodic (31)—we remove the requirement for odd symmetry. Consequently, we now explore traveling wave–like solutions at low nonlinearity α. Also we illustrate the spatio-temporal chaos that occurs at higher α.

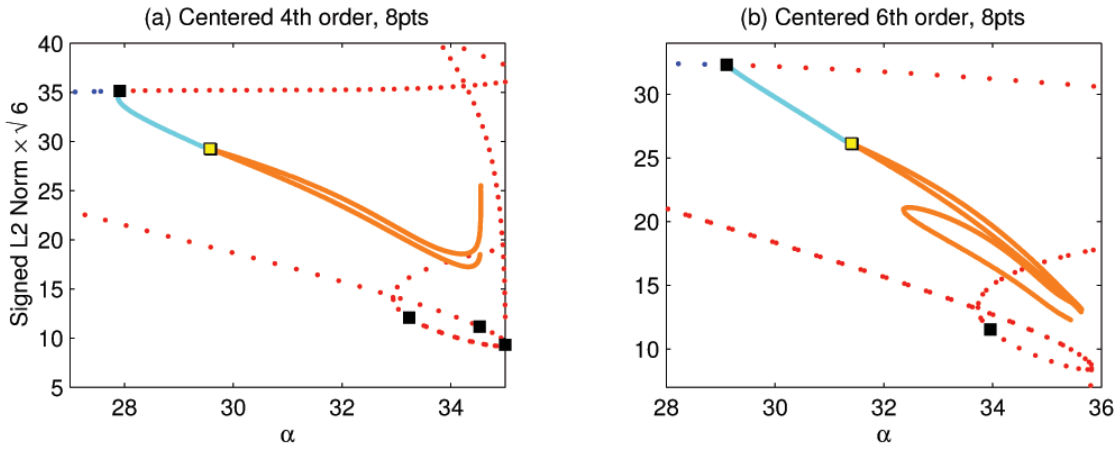


Figure 18. *Bifurcation diagrams near the first Hopf bifurcation for* (a) *fourth-order* (21) *and* (b) *sixth-order* (22) *centered difference approximations with eight grid points on* $[0,\pi]$. *Stable limit cycles are shown in light blue, and unstable limit cycles are shown in orange.*

Table 5

Nonlinearity parameter α *values for the first Hopf bifurcation point* HB_1 *and resulting period doubling point* PD.

Approximation	HB_1	PD
Holistic 8 elements		
$\mathcal{O}(\gamma^3,\alpha^2)$ (15)	25.60	27.22
$\mathcal{O}(\gamma^4,\alpha^2)$ (16)	30.04	32.03
$\mathcal{O}(\gamma^5,\alpha^2)$ (17)	30.66	32.95
Centered 8 points		
2nd-order (20)	—	—
4th-order (21)	27.91	29.57
6th-order (22)	29.11	31.40
Accurate	30.35	32.97

Good performance for holistic models at low nonlinearity α. Consider the holistic models of the Kuramoto–Sivashinsky PDE (1) and (31) for nonlinearity parameter $\alpha = 5$ and $\alpha = 10$ on coarse grids of eight elements on $[0, 2\pi]$; that is, the element size is twice as big as many earlier solutions because the domain is twice as big. Figure 3 shows solutions obtained from the lowest order $\mathcal{O}(\gamma^3,\alpha^2)$ holistic model (15) for $\alpha = 5$ in green, the accurate solution in blue, and the corresponding second-order centered difference approximation (20) with eight points on $[0, 2\pi]$, in magenta. The solutions are shown starting from the half-wave initial condition of $u(x, 0) = |\sin(x/2)|$. Note that the $\mathcal{O}(\gamma^3,\alpha^2)$ holistic model is superior to the second-order centered difference approximation on this coarse grid. In particular, the amplitude of the evolving wave-like solution and the wave speed are more accurately reproduced by the $\mathcal{O}(\gamma^3,\alpha^2)$ holistic model for $\alpha = 5$.

Similarly, Figure 19 shows the solutions at the same times but for larger nonlinearity $\alpha = 10$. For this α the second-order (20) and fourth-order (21) centered difference approximations do not generate a wave-like solution at all. However, the sixth-order centered difference

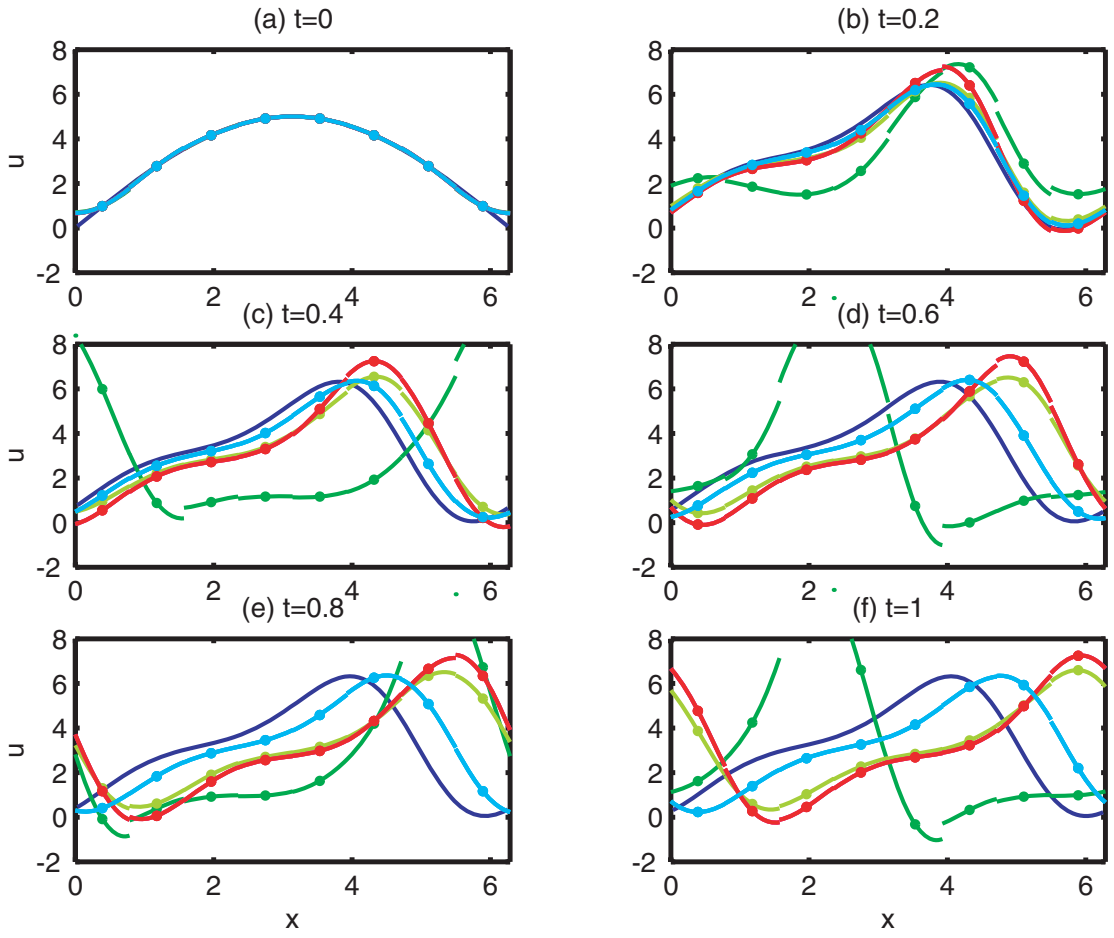


Figure 19. $\alpha = 10$. *Wave-like solutions at times* $t = 0, 0.2, 0.4, 0.6, 0.8, 1$ *for the* $\mathcal{O}(\gamma^3, \alpha^2)$ (15), $\mathcal{O}(\gamma^4, \alpha^2)$ (16), *and* $\mathcal{O}(\gamma^5, \alpha^2)$ (17) *holistic models shown in green, light green, and light blue, respectively, and the sixth-order centered difference approximation* (22) *shown in red on a coarse grid of eight elements on* $[0, 2\pi]$. *The accurate solution is shown in blue.*

approximation (22) does produce the traveling wave–like solution shown in red. The $\mathcal{O}(\gamma^3, \alpha^2)$ holistic model (green) is the least accurate on this coarse grid, but it does reproduce a stable solution on this coarse grid for only a five-point stencil approximation. The $\mathcal{O}(\gamma^4, \alpha^2)$ holistic model (light green) more accurately models the amplitude of the wave compared to the sixth-order centered difference approximation, despite having a smaller stencil width. The $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model is the most accurate at reproducing the evolution of the stable wave-like solution for $\alpha = 10$ on this coarse grid of eight elements.

Good performance for more complex dynamics. For higher values of nonlinearity parameter α for which the Kuramoto–Sivashinsky PDE exhibits more complex dynamics, including spatio-temporal chaos, we compare time averaged power spectra rather than particular traveling waves. Here we investigate the performance of the holistic models on coarse grids for $\alpha = 20$ and 50, using the example of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (17), and compare it with

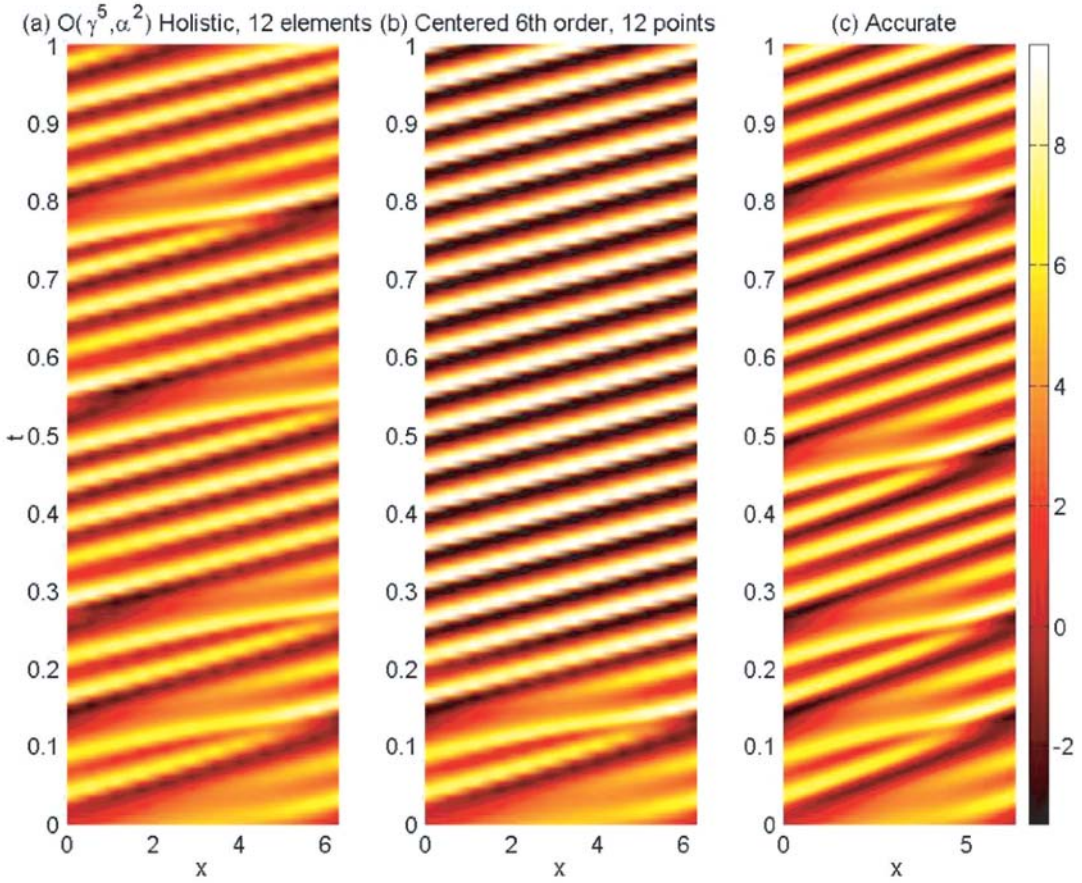


Figure 20. $\alpha = 20$. *Space time plots for* (a) *the* $\mathcal{O}(\gamma^5, \alpha^2)$ *holistic model* (17) *with* 12 *elements on* $[0, 2\pi]$, (b) *sixth-order centered difference approximation* (22) *with* 12 *grid points on* $[0, 2\pi]$, *and* (c) *the accurate solution.*

the sixth-order centered difference approximation (22), as it is of equal stencil width. Further, we also compare the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model on coarse grids to the second-order centered difference approximations of similar accuracy. We find that the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model, but with approximately $1/3$ of the grid points, has comparable accuracy to second-order centered difference approximations.

Figure 20 shows space time plots of (a) the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model (17), (b) the sixth-order centered difference approximation (22), and (c) the accurate solution.[8] The $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model reproduces much of the complex structure of the accurate solution for nonlinearity $\alpha = 20$ with 12 elements. Figure 20(b) shows that the sixth-order centered difference approximation incorrectly finds a periodic solution after approximately $t = 0.2$. Similar behavior occurs for nonlinear parameter $\alpha = 50$ with 24 elements [19]: the holistic discretization

[8]The accurate solutions plotted in this section are computed using a sixth-order centered difference approximation and 256 grid points on the interval $[0, 2\pi]$. This is sufficient grid resolution to capture the important dynamics of the Kuramoto–Sivashinsky PDE for the values of α investigated here.

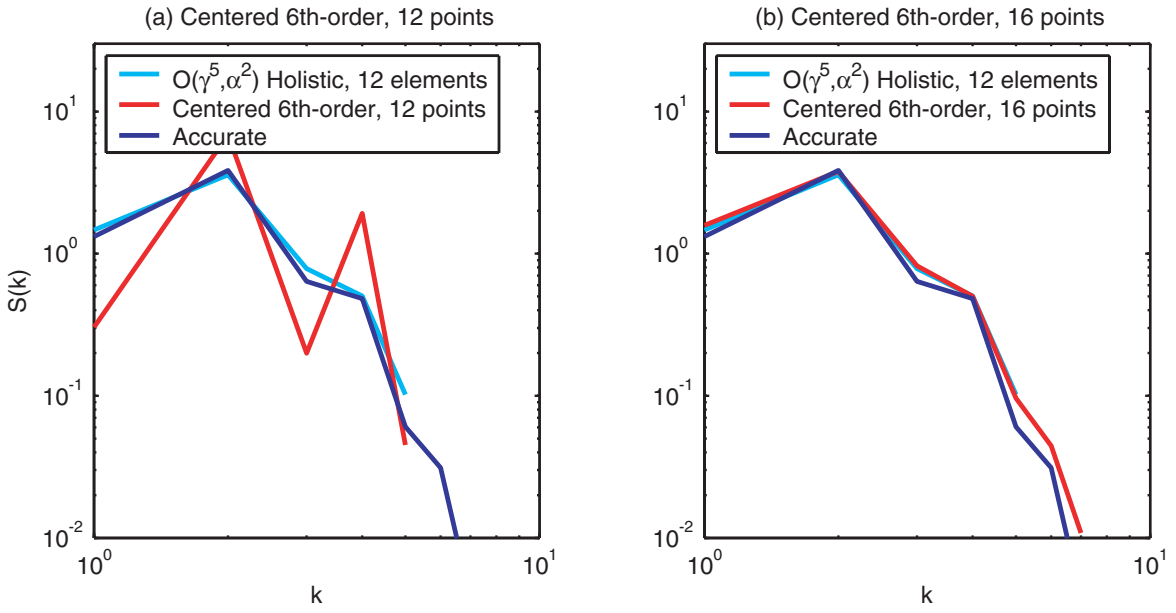


Figure 21. $\alpha = 20$. *Time averaged power spectra for the* $\mathcal{O}(\gamma^5, \alpha^2)$ *holistic model* (17) *with* 12 *elements on* $[0, 2\pi]$ *shown in light blue, and the sixth-order centered difference approximation* (22) *in red for* (a) 12 *grid points on* $[0, 2\pi]$ *and* (b) 16 *grid points on* $[0, 2\pi]$. *The accurate power spectrum is shown in blue.*

accurately models the apparently chaotic dynamics, whereas the centered difference model incorrectly locks onto a periodic traveling wave.

Since the Kuramoto–Sivashinsky PDE at nonlinearity parameter $\alpha = 20$ exhibits more complex time dependent behavior than simple limit cycles, we compare time averaged power spectra, denoted here by $S(k)$ for wavenumber k. Figure 21(a) shows a log-log plot of the time averaged power spectra of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model, the sixth-order centered difference approximation on a coarse grid of 12 elements on $[0, 2\pi]$, and the accurate power spectrum. For this coarse grid of only 12 elements, only five wavenumbers are relevant, as displayed. Note that the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model is superior to the sixth-order centered difference approximation on this coarse grid. Figure 21(b) compares the time averaged power spectrum of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model with 12 elements and the sixth-order centered difference approximation with 16 grid points. The $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model achieves similar accuracy on a coarser grid.

Figures 22(a) and (b) show the power spectra of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model on a coarse grid of 12 elements and the second-order centered difference approximation on the more refined grids of 24 and 36 points, respectively. A refined grid of 36 points is needed to achieve similar accuracy to that for the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model on a coarse grid of 12 elements on $[0, 2\pi]$. That is, through its subgrid scale modeling, the holistic model achieves similar accuracy with one-third the dimensionality.

This investigation of the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model on coarse grids for $\alpha = 20$ and that of MacKenzie [19] for nonlinearity $\alpha = 50$ show that it reproduces accuracy similar to the

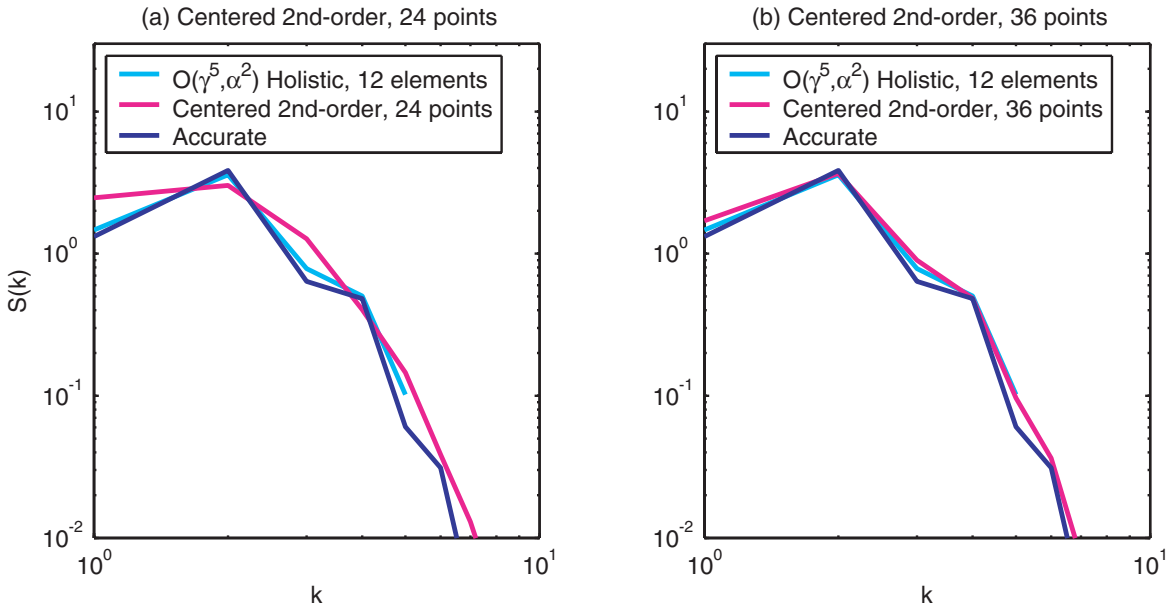


Figure 22. $\alpha = 20$. *Time averaged power spectra for the* $\mathcal{O}(\gamma^5, \alpha^2)$ *holistic model with* 12 *elements on* $[0, 2\pi]$ *shown in light blue, and the second-order centered difference approximation in magenta for* (a) 24 *grid points on* $[0, 2\pi]$ *and* (b) 36 *grid points on* $[0, 2\pi]$. *The accurate spectrum is shown in blue.*

second-order centered difference approximation on a coarse grid of approximately 1/3 the resolution, and accuracy similar to the sixth-order centered difference approximation on grids of approximately 3/4 the resolution. MacKenzie [19] reports that even at $\alpha = 200$ the holistic model qualitatively captures the dynamics of the Kuramoto–Sivashinsky PDE well. This increased accuracy on coarse grids allows larger time steps for explicit time integration schemes, as discussed in section 4.4.

6. Conclusion. Holistic discretization [25] is straightforwardly extended to fourth-order dissipative PDEs through the example of the Kuramoto–Sivashinsky equation [18]. We divide the domain into elements by introducing artificial internal boundary conditions (section 2), which isolate the elements when $\gamma = 0$ but which fully couple the elements to recover the Kuramoto–Sivashinsky dynamics when $\gamma = 1$. Then center manifold theory supports the discretization; see section 2.2. The holistic models listed in section 3 have a dual justification (section 3.3): not only are they supported by center manifold theory for finite element size h, but also the IBCs are specially crafted [26] so that the models are consistent with the Kuramoto–Sivashinsky equation as the grid spacing $h \to 0$.

No formal error bounds currently exist for the holistic method; the difficulty is that the models are based at $\gamma = 0$ but are evaluated at finite $\gamma = 1$. Instead we present a detailed numerical investigation of the holistic models of the steady states (section 4) and time dependent solutions (section 5) of the Kuramoto–Sivashinsky on coarse grids.

We compared, in section 4.4, the accuracy of different approximations in predicting steady states on different grid resolutions. The holistic $\mathcal{O}(\gamma^5, \alpha^2)$ approximation on a grid of eight

elements has accuracy similar to a second-order centered difference approximation on a grid of 16 points. Consequently the holistic model allows a maximum time step that is an order of magnitude larger than that of the explicit centered difference approximation of similar accuracy, while maintaining numerical stability. The accuracy of the holistic approximations to the Kuramoto–Sivashinsky equation on coarse grids and subsequent improved performance justify further application of the holistic method and future investigation of the approach.

Holistic models on coarse grids also model well time dependent phenomena of the Kuramoto–Sivashinsky PDE. In particular, in section 5.1 we saw that the holistic models more accurately model the eigenvalues near the steady states of the first form of the Kuramoto–Sivashinsky PDE compared to explicit centered difference approximations of equal stencil widths. The coarse grid holistic models also more accurately model the first Hopf bifurcation and the resulting period doubling sequence; see section 5.2. Further, in comparison with explicit centered difference models, in section 5.3 we saw good performance for higher values of the nonlinearity parameter α and more accurate predictions of time averaged power spectra: the $\mathcal{O}(\gamma^5, \alpha^2)$ holistic model achieves accuracy similar to that of the second-order and sixth-order centered difference approximations on approximately 1/3 and 3/4 of the grid resolutions, respectively.

This good performance of the holistic models for accurately reproducing both the steady states and the time dependent phenomena of the Kuramoto–Sivashinsky PDE is good evidence that the holistic approach is a powerful method for discretizing dissipative PDEs on coarse grids.

REFERENCES

[1] D. Armbruster, J. Guckenheimer, and P. Holmes, *Kuramoto–Sivashinsky dynamics on the center-unstable manifold*, SIAM J. Appl. Math., 49 (1989), pp. 676–691.

[2] G. Berkooz, P. Holmes, and J. L. Lumley, *The proper orthogonal decomposition in the analysis of turbulence*, in Annual Review of Fluid Mechanics, Vol. 25, Annual Reviews, Palo Alto, CA, 1993, pp. 539–575.

[3] J. Carr, *Applications of Centre Manifold Theory*, Appl. Math. Sci. 35, Springer–Verlag, New York, 1981.

[4] J. Carr and R. G. Muncaster, *The application of centre manifold theory to amplitude expansions.* II. *Infinite dimensional problems*, J. Differential Equations, 50 (1983), pp. 280–288.

[5] M. C. Cross and P. C. Hohenberg, *Pattern formation outside of equilibrium*, Rev. Modern Phys., 65 (1993), pp. 851–1112.

[6] H. Dankowicz, P. Holmes, G. Berkooz, and J. Elezgaray, *Local models of spatio-temporally complex fields*, Phys. D, 90 (1996), pp. 387–407.

[7] E. J. Doedel, R. C. Paffenroth, A. R. Champneys, T. F. Fairgrieve, Yu. A. Kuznetsov, B. Sandstede, and X. Wang, *Auto* 2000*: Continuation and Bifurcation Software for Ordinary Differential Equations (with HomCont)*, technical report, Caltech, Pasadena, CA, 2001.

[8] B. Ermentrout, *XPPAUT* 5.0—*The Differential Equations Tool*, technical report, University of Pittsburgh, 2001, available online at http://www.math.pitt.edu/~bard/bardware/xpp_doc.pdf.

[9] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi, *On the computation of inertial manifolds*, Phys Lett. A, 131 (1988), pp. 433–436.

[10] C. Foias, M. S. Jolly, I. G. Kevrekidis, and E. S. Titi, *On some dissipative fully discrete nonlinear Galerkin schemes for the Kuramoto–Sivashinsky equation*, Phys. Lett. A, 186 (1994), pp. 87–96.

[11] C. Foias, B. Nicolaenko, G. R. Sell, and R. Teman, *Inertial manifolds for the Kuramoto–Sivashinsky equation*, C. R. Acad. Sci. Paris Sér. I Math., 301 (1985), pp. 285–288 (in French).

[12] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and*

Symmetry, Cambridge University Press, Cambridge, UK, 1996.

[13] J. M. Hyman and B. Nicolaenko, *The Kuramoto–Sivashinsky equation: A bridge between PDEs and dynamical systems*, Phys. D, 18 (1986), pp. 113–126.

[14] M. S. Jolly, I. G. Kevrekidis, and E. S. Titi, *Approximate inertial manifolds for the Kuramoto–Sivashinsky equation: Analysis and computations*, Phys. D, 44 (1990), pp. 38–60.

[15] I. G. Kevrekidis, B. Nicolaenko, and J. C. Scovel, *Back in the saddle again: A computer assisted study of the Kuramoto–Sivashinsky equation*, SIAM J. Appl. Math., 50 (1990), pp. 760–790.

[16] Y. Kuramoto, *Diffusion induced chaos in reactions systems*, Progr. Theoret. Phys. Suppl., 64 (1978), pp. 346–367.

[17] S. J. Liao, *The quite general BEM for strongly nonlinear problems*, in Boundary Elements XII, C. A. Brebbia, S. Kim, T. A. Osswald, and H. Power, eds., Computational Mechanics Publications, Southampton, MA, 1995, pp. 67–74.

[18] T. Mackenzie and A. J. Roberts, *Holistic finite differences accurately model the dynamics of the Kuramoto–Sivashinsky equation*, ANZIAM J., 42 (2000), pp. C918–C935.

[19] T. MacKenzie, *Create Accurate Numerical Models of Complex Spatio-temporal Dynamical Systems with Holistic Discretisation*, Ph.D. thesis, University of Southern Queensland, Toowoomba, Australia, 2005.

[20] M. Marion and R. Temam, *Nonlinear Galerkin methods*, SIAM J. Numer. Anal., 26 (1989), pp. 1139–1157.

[21] Y. Pomeau and S. Zaleski, *The Kuramoto–Sivashinsky equation: A caricature of hydrodynamic turbulence?*, in Macroscopic Modelling of Turbulent Flows, Lecture Notes in Phys. 230, Springer, Berlin, 1985, pp. 296–303.

[22] A. J. Roberts, *The utility of an invariant manifold description of the evolution of a dynamical system*, SIAM J. Math. Anal., 20 (1989), pp. 1447–1458.

[23] A. J. Roberts, *Low-Dimensional Modelling of Dynamical Systems*, technical report, 1997, available online at http://arXiv.org/abs/chao-dyn/9705010.

[24] A. J. Roberts, *Low-dimensional modelling of dynamics via computer algebra*, Comput. Phys. Comm., 100 (1997), pp. 215–230.

[25] A. J. Roberts, *Holistic discretisation ensures fidelity to Burgers' equation*, Appl. Numer. Math., 37 (2001), pp. 371–396.

[26] A. J. Roberts, *A holistic finite difference approach models linear dynamics consistently*, Math. Comp., 72 (2002), pp. 247–262.

[27] A. J. Roberts, *Derive boundary conditions for holistic discretisations of Burgers' equation*, in Proc. of 10th Computational Techniques and Applications Conference CTAC-2001, K. Burrage and R. B. Sidje, eds., ANZIAM J., 44 (2003), pp. C664–C686.

[28] A. J. Roberts, *Holistic Discretisation of Dynamical Partial Differential Equations*, technical report, version 1.1, 2003, available online at http://www.sci.usq.edu.au/staff/aroberts/holistic.html.

[29] A. J. Roberts and I. G. Kevrekidis, *Higher order accuracy in the gap-tooth scheme for large-scale dynamics using microscopic simulators*, in Proc. of 12th Computational Techniques and Applications Conference CTAC-2004, R. May and A. J. Roberts, eds., ANZIAM J., 46 (2005), pp. C637–C657.

[30] J. C. Robinson, *The asymptotic completeness of inertial manifolds*, Nonlinearity, 9 (1996), pp. 1325–1340.

[31] C. Scovel, I. G. Kevrekidis, and B. Nicolaenko, *Scaling laws and the prediction of bifurcations in systems modeling pattern formation*, Phys. Lett. A, 130 (1988), pp. 73–80.

[32] G. I. Sivashinsky, *Nonlinear analysis of hydrodynamic instability in laminar flames, Part* I: *Derivation of the basic equations*, Acta Astronaut., 4 (1977), pp. 1176–1206.

[33] A. Vanderbauwhede, *Centre manifolds, normal forms, and elementary bifurcation*, in Dynamics Reported, Wiley, Chichester, UK, 1989, Vol. 2, pp. 89–169.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 3, pp. 403–434

The Moment Map: Nonlinear Dynamics of Density Evolution via a Few Moments*

D. Barkley†, I. G. Kevrekidis‡, and A. M. Stuart†

Abstract. We explore situations in which certain stochastic and high-dimensional deterministic systems behave effectively as low-dimensional dynamical systems. We define and study moment maps, maps on spaces of low-order moments of evolving distributions, as a means of understanding equation-free multiscale algorithms for these systems. The moment map itself is deterministic and attempts to capture the implied probability distribution of the dynamics. By choosing situations where the low-dimensional dynamics can be understood a priori, we evaluate the moment map. Despite requiring the evolution of an ensemble to define the map, this can be an efficient numerical tool, as the map opens up the possibility of bifurcation analyses and other high level tasks being performed on the system. We demonstrate how nonlinearity arises in these maps and how this results in the stabilization of metastable states. Examples are shown for a hierarchy of models, ranging from simple stochastic differential equations to molecular dynamics simulations of a particle in contact with a heat bath.

Key words. equation-free, multiscale, moment map, metastable states

AMS subject classifications. 65C30, 34F05, 60H10, 60H35

DOI. 10.1137/050638667

1. Introduction.

An equation-free framework has recently been developed as a means of computationally analyzing the dynamical behavior of a large class of complex multiscale dynamical systems. The systems may be either stochastic or deterministic with many degrees of freedom and subject to random initial data. The key observation behind the equation-free framework is that in many cases the quantities of interest are averages or low-order moments of evolving distributions, which are smooth in space and time and which evolve effectively as closed low-dimensional systems. In effect, the low-order moments evolve as though they are governed by reduced closed equations, even though the reduced equations are not analytically available. Algorithms for performing scientific computing tasks such as numerical integration or bifurcation and stability analysis of these unavailable reduced equations have been developed (e.g., coarse projective integration, coarse Newton-GMRES; see [40, 12, 25, 26] and references therein). These algorithms are based on traditional continuum numerical analysis, "wrapped" around direct fine-scale simulation. The purpose of this paper is to establish a mathematical framework for understanding the behavior of coarse dynamics and coarse bifurcation methods on problems which exhibit metastable behavior.

*Received by the editors August 23, 2005; accepted for publication (in revised form) by T. Kaper February 27, 2006; published electronically July 17, 2006.
http://www.siam.org/journals/siads/5-3/63866.html

†Mathematics Institute, University of Warwick, Coventry, CV4 7AL United Kingdom (barkley@maths.warwick.ac.uk, stuart@maths.warwick.ac.uk).

‡Department of Chemical Engineering, Program in Applied and Computational Mathematics, and Mathematics, Princeton University, Princeton, NJ 08544 (yannis@Princeton.edu).

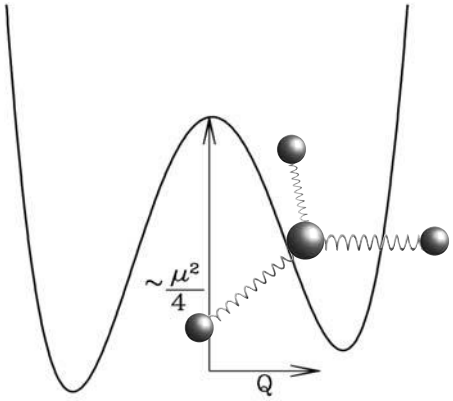


Figure 1. *Model system. A distinguished particle (light gray) sits in an asymmetric double well potential. The particle is coupled via linear springs to N other (bath) particles (of which three are shown as an illustration). The full Hamiltonian system has $N + 1$ degrees of freedom.*

The systems we consider are exemplified by the following model. A particle, called the distinguished particle, with position Q and momentum P, sits in a potential well $V(Q)$. It is coupled via linear springs to a large number N of particles comprising a heat bath; see Figure 1. The potential well considered here is a slightly asymmetric double well. The full system is an $N + 1$ degree-of-freedom Hamiltonian system. A detailed description of the model, including the choice of spring constants, masses, and initial data for the bath particles is given below and in [10, 28, 43]. The important point is that the dynamics of this simple model (and the others that we consider in this paper) is typical of many more complex molecular and stochastic systems in which the state is primarily confined to a few conformations (here defined by the minima of V) with rare switching events between them. This is illustrated in Figure 2 with a typical trajectory and time series for the distinguished particle.

Consider now the dynamics of an ensemble of trajectories for the model system. Figure 3 illustrates the evolution of an ensemble of trajectories all with the same initial conditions $(Q, P) = (0, 3)$ for the distinguished particle, but with different initial data for the bath particles. (The total initial energy of the bath is approximately the same for all realizations; see section 2.) Over a time of order 10 the initial density evolves to a nearly Gaussian density centered near the bottom of the right well, where it remains roughly constant for some time. While not immediately evident, during the initial 10 time units the density is never very far from Gaussian. However, it is evident from the left-hand plot that a small percentage of the realizations are located in the left well at time 10. Over a much longer time scale, $O(10^4)$, the density evolves to the bimodal equilibrium distribution and is hence far from Gaussian. Thus the system exhibits *metastability* with near equilibration within one well dominating over medium time scales, before the system ultimately converges to an equilibrium distribution which sees both wells. The time it takes the system to reach equilibrium clearly depends on the potential barrier height; the time scale of the intermediate time evolution to the well bottom depends only on properties of the particular well.

Our aim is to study the behavior of coarse dynamics and coarse bifurcation methods on problems which exhibit metastable behavior of this type. In section 2 we introduce a hierarchy

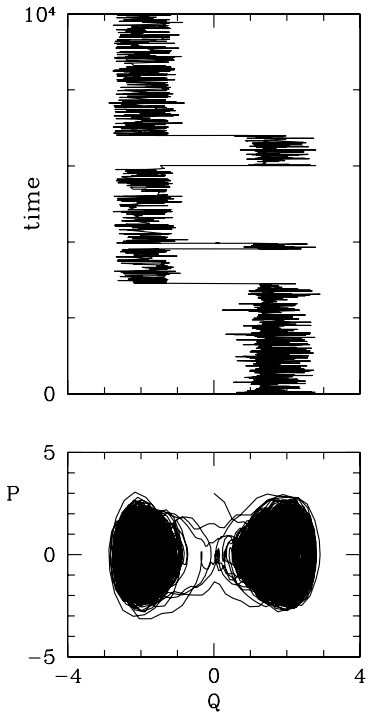


Figure 2. *Typical behavior for the dynamics of the distinguished particle in contact with a bath with $N = 8000$ particles. Most of the time the distinguished particle is located in one of the two potential wells, but occasionally it makes a jump between wells.*

of model problems, all of which exhibit rare transitions between a small number of states, and which we then use throughout as illustrations.

In section 3 we define the discrete-time *moment map* Φ for the first k moments of the ensemble of solutions to a time evolving system. Specifically, Φ will be a deterministic low-dimensional map defined by the short-time evolution of only the low-order moments, in a general setting which applies to systems of both ordinary differential equations (ODEs), with randomness from initial data, and stochastic differential equations (SDEs), with randomness from initial data and Brownian driving noise. Figure 4 illustrates the *first-order moment map* for the heat bath example. The short-term dynamics of the map resemble those of the ensemble, but significantly, the moment map has *stable fixed points* corresponding to means of metastable densities centered in each well. The second-order moment map (not illustrated) additionally captures the widths of the metastable measures. The algorithms we study are based on these maps. The maps, and their fixed points, are good approximations on time scales on which the evolving probability density function is well represented by the number of moments used.

The heart of this paper is section 4, which is devoted to the study of the moment map when applied to a variety of model systems. We use a combination of exact solutions for Gaussian problems, approximate solutions for metastable systems, and numerical experiments. Of central interest is the observation that the moment maps stabilize the metastable states in

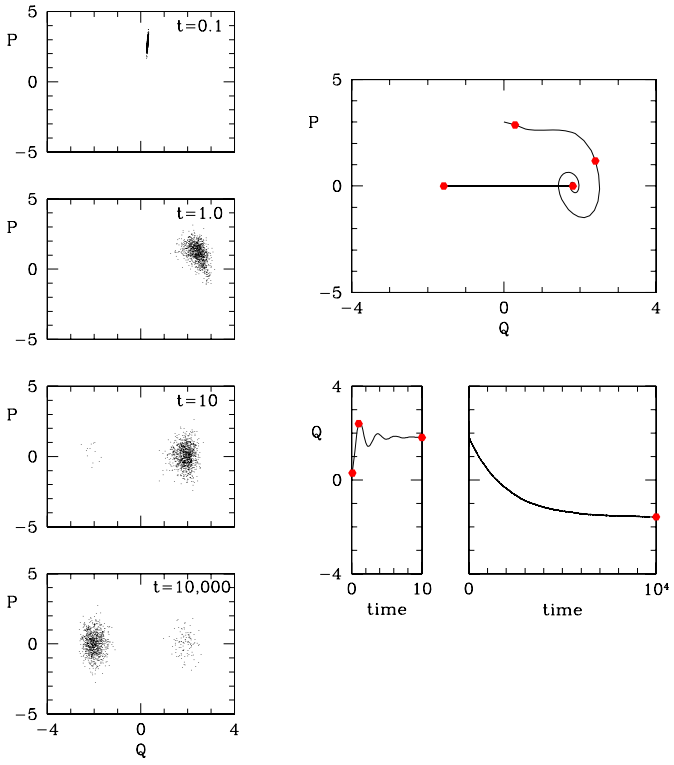


Figure 3. *Evolution of an ensemble of* 10^4 *realizations for the model system. On the left, density plots of position and momentum of the distinguished particle are shown at four times as labeled. For clarity only* 10^3 *realizations are plotted. The initial conditions are* $(Q, P) = (0, 3)$*. The right shows the trajectory and time series for the ensemble expectations. Red points indicate the four times shown on the left. There are* $N = 8000$ *particles in the heat bath.*

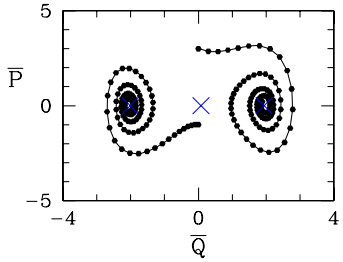


Figure 4. *Dynamics of a moment map for the first-order moments of the position* Q *and momentum* P *of the distinguished particle. The map has three fixed points, two stable foci and one unstable saddle, shown as blue crosses. Two trajectories are shown, one evolving to each of the stable fixed points. The right trajectory has initial condition* $(\bar{Q}, \bar{P}) = (0, 3)$*, corresponding to that in Figure 3. There are* $N = 8000$ *particles in the heat bath.*

the model problems of interest. The moment map is a nonlinear map, defined from the linear flow of probability densities for ODEs and SDEs, with the nonlinearity entering through the process of repeatedly projecting onto the space of moments. This process of *nonlinearization*

creates interesting fixed points that are associated with metastable behavior and are amenable to low-dimensional bifurcation analyses; related issues are addressed in [5, 20]. Section 5 contains our concluding remarks.

2. Model problems. We consider three example systems in this paper. Two are SDEs, and one is the ODE heat bath model described in the introduction and illustrated in Figure 1. A major thrust of this paper is to establish, through computational experimentation, that the moment map stabilizes metastable behavior arising from the slow dynamics between potential wells with large energy barriers; this gives rise to nonlinear phenomena, such as bifurcations, in the moment map. Such phenomena can be illustrated both in SDE models and in the ODE heat bath model. Furthermore, in various parameter limits, the SDEs can be derived as approximations for the heat bath, further justifying their study.

2.1. Heat-bath model. This model problem is defined by the Hamiltonian

$$H(Q,P,q,p) = \frac{1}{2M}P^2 + V(Q) + \sum_{j=1}^{N}\frac{p_j^2}{2m_j} + \sum_{j=1}^{N}\frac{k_j}{2}(q_j - Q)^2, \tag{2.1}$$

where Q and P are the position and momentum of a distinguished particle of unit mass in a potential field $V(\cdot)$. The q_j's and p_j's are the coordinates and momenta of N particles that are referred to as heat bath particles. The jth heat bath particle has mass m_j and interacts with the distinguished particle via a linear spring with stiffness constant k_j. If the distinguished particle were held fixed, it would be the anchor point of N independent oscillators with frequencies $\omega_j = (k_j/m_j)^{1/2}$. The numerical experiments are all conducted with mass $M = 1$.

Initial conditions for the distinguished particle are $Q(0) = Q_0$, $P(0) = P_0$. The initial data for the heat bath particles, $q_j(0) = q_j^0$ and $p_j(0) = p_j^0$, are randomly drawn from a Gibbs distribution with inverse temperature β. The Gibbs measure is conditioned by the (nonrandom) initial data Q_0 and P_0. For fixed Q, P the Hamiltonian (2.1) is quadratic in q, p, and hence the corresponding measure is Gaussian. It is easily verified that

$$\begin{aligned} q_j^0 &= Q_0 + (1/\beta k_j)^{1/2}\xi_j, \\ p_j^0 &= (m_j/\beta)^{1/2}\eta_j, \end{aligned}$$

where $\xi_j, \eta_j \sim \mathcal{N}(0,1)$ are mutually independent sequences of independently and identically distributed (i.i.d.) random variables.

This leaves the specification of the values for k_j and m_j. For our purposes the only important property is that the frequencies $\omega_j = (k_j/m_j)^{1/2}$ cover an increasingly large range in an increasingly dense manner as the number of particles N increases. Hence what we actually specify is the frequencies. These are chosen to be random and uniformly distributed in $[0, N^{1/3}]$,

$$\omega_j = N^{1/3}\nu_j, \quad \nu_j \text{ i.i.d.}, \quad \nu_1 \sim \mathcal{U}[0,1].$$

It is important to note that, in addition to the initial data, the model specification itself contains this random element. We shall be careful to distinguish between the two types of randomness.

From the frequencies the spring constants and masses are given by

$$k_j = \frac{2\alpha^2}{\pi(\alpha^2+\omega_j^2)}\frac{N^{1/3}}{N}, \qquad m_j = \frac{k_j}{\omega_j^2} \tag{2.2}$$

with $\alpha > 0$. See [28] for further details. The parameters α and β are fixed at $\alpha = 100$, $\beta = 2$.

The potential considered in this paper is

$$V(Q) = \frac{Q^4}{4} - \frac{\mu Q^2}{2} + \nu Q, \tag{2.3}$$

where μ and ν are parameters with ν typically small.

2.2. Two-dimensional SDE approximation. For large N and α, the distinguished particle Q in the heat bath model can be approximated by the SDE

$$M\ddot{Q} + \gamma\dot{Q} + V'(Q) = \sqrt{2\gamma/\beta}\,\dot{W}. \tag{2.4}$$

A theorem justifying this approximation can be proved using the techniques of weak convergence, by taking the limit $N \to \infty$ (see [28]) and then $\alpha \to \infty$ (see [34]). In the absence of noise (the zero temperature limit $\beta \to \infty$) this damped Hamiltonian system exhibits decay towards stationary points with zero velocity and positions at the critical points of V. The presence of noise (finite β) then induces transitions between the minima of V, with time scales determined by the well-depths relative to the size of the noise.

2.3. One-dimensional SDE approximation. The stochastic dynamics between potential wells is also present in simple one-dimensional SDEs. A particular instance of such a one-dimensional SDE follows from (2.4) for $M \ll 1$. In the limit $M \to 0$, the solutions of (2.4) converge strongly [33] or weakly [11] to solutions of the SDE

$$\gamma\dot{Q} + V'(Q) = \sqrt{2\gamma/\beta}\dot{W}. \tag{2.5}$$

We will use this problem to illustrate the moment map and its properties on systems exhibiting metastable dynamics within potential wells.

2.4. Remark. In most respects the SDE systems derived above are simpler to treat, and computationally they are far less expensive to simulate, than the full heat-bath model. Therefore, when we later use these models as examples, we will study the models in the opposite order from what has just been presented. We start with the one-dimensional SDE and examine its behavior extensively, and then consider more briefly the two-dimensional SDE and the heat bath system of ODEs.

3. The moment map. The central objects of our study are maps on moments. The basic ingredients are an evolution equation (either an SDE or a system of ODEs), a space of low-order moments, and a measure determined uniquely by low-order moments. We refer to the latter as *lifting*, and the choice of the lifting operator is an essential ingredient in the method. We start with the SDE case, then describe the situation for ODEs. For both the SDE and the ODEs the flow on probability densities is linear. The nonlinearity inherent in the moment map comes from the relationship between the probability density function and its moments. After describing the moment map for the SDE and the ODE, the section concludes with some general remarks.

3.1. The SDE case. Let $x \in C([0,\infty), \mathbb{R}^d)$ solve the following Itô SDE, driven by Brownian motion $W \in C([0,\infty), \mathbb{R}^m)$:

$$\frac{dx}{dt} = f(x) + \sigma(x)\frac{dW}{dt}. \tag{3.1}$$

This includes (2.4) and (2.5) as special cases. We will consider ensembles of solutions of this equation, with ensembles taken over multiple driving noises and random initial data. Let $X_j(t)$ be the jth moment of $x(t)$, with expectation taken with respect to both the driving Brownian motion W and random initial data, the latter being assumed independent of the Brownian motion. Denote the first k moments of x by $X(t) = (X_1(t), \ldots, X_k(t))$.

Let μ be a measure on $\mathbb{R}^d$ determined by exactly k moments, with density $\hat{\rho}(x; X)$. Here X denotes the dependence of the density on the k moments, and we require that the k moments of $\hat{\rho}$ are exactly those given by X.

We now define the map Φ on k moments. Fix a time $\tau \in (0,\infty)$ and set

$$\Phi(X) = X(t = \tau), \tag{3.2}$$

where the initial data $X(0) = x_0$ is distributed with density $\hat{\rho}(x; X)$. (Note that if $\tau = 0$, then Φ is the identity.) We refer to this as the *moment map*. Three choices that need to be made when calculating the moment map are:

(i) the evolution time τ,
(ii) the *number* of moments we choose to use,
(iii) the lifting step: the way we choose to distribute the initial density based on the moments.

We will return to this dependence in more detail below.

The two examples of the measures which will be used throughout this paper are the Dirac measure and the Gaussian measure, uniquely determined by the first moment and the first two moments, respectively. We emphasize that the moment map is a deterministic map.

In practice the moment map must be obtained numerically, through Monte Carlo simulation, for example. However, it is insightful to describe the definition of the map through the Fokker–Planck equation for (3.1). This linear PDE for the probability densities $\rho(x,t)$ propagated by (3.1) is

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (f\rho) + \frac{1}{2}\nabla \cdot \nabla \cdot (\Sigma\rho) := \mathcal{L}^*\rho, \tag{3.3}$$

$$\rho(x,0) = \hat{\rho}(x; X), \tag{3.4}$$

where $\Sigma = \sigma(x)\sigma(x)^T$. Here $\mathcal{L}^*$ is the adjoint of the generator for the process $\mathcal{L}$. We denote the solution by

$$\rho(x,t) = e^{\mathcal{L}^* t}\hat{\rho}(x; X). \tag{3.5}$$

From $\rho(x, t = \tau)$ we can construct $\Phi(X)$ by (3.2). In general the moment map is *nonlinear* because of the nonlinear dependence of $\hat{\rho}(x; X)$ on X and the nonlinearity of the map from $\rho(x,\tau)$ to $X(\tau)$. Thus we have constructed a nonlinear map on $\mathbb{R}^d$ from the linear flow on $\mathbb{R}^d$-valued densities.

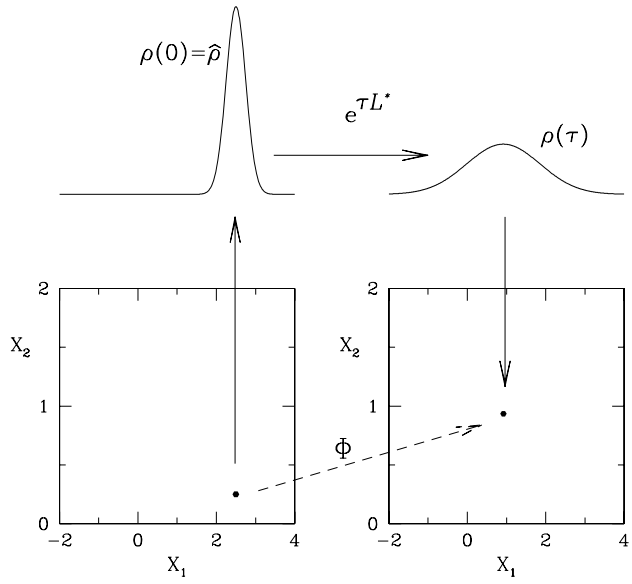


Figure 5. *Sketch illustrating the definition of the moment map. From a point (X_1, X_2) in moment space a density $\hat{\rho}$ is uniquely determined. This initial density is evolved by the system dynamics (e.g., the Fokker–Planck equation). From the resulting density $\rho(\tau)$, τ time units later, a new point in moment space is determined. This point is defined to be the image of (X_1, X_2) under the moment map.*

It is helpful to consider Figure 5, illustrating the definition of the moment map. One should view the map as a composition of three steps: (1) *lifting* from a space of moments (subset of $\mathbb{R}^d$) to a space of probability densities (captured by our choice of $\hat{\rho}(\cdot; X)$), (2) time evolution of the density by the underlying process (a linear flow, given by the map $e^{\mathcal{L}^*\tau}$), (3) projection back to moment space by integrating against the time-evolved measure.

In the case of a single moment, in this paper, we take μ to be a Dirac measure at X, and then $X(\tau) = \mathbb{E}x(\tau)$ can be calculated from

$$\Phi(X) = \int_{\mathbb{R}^d} x e^{\mathcal{L}^*\tau} \delta(x - X) dx.$$

In the case of two moments we have

$$X(\tau) = \{\mathbb{E}x(\tau), \mathbb{E}[x(\tau) - \mathbb{E}x(\tau)][x(\tau) - \mathbb{E}x(\tau)]^T\},$$

and we take μ to be a Gaussian measure with mean and covariance determined by these moments. It is convenient to express the moment map in terms of the mean $\bar{x} \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. We obtain the map

$$\Phi(\bar{x}, \Sigma) = \left\{ \begin{array}{c} \Phi_1(\bar{x}, \Sigma) \\ \Phi_2(\bar{x}, \Sigma) \end{array} \right\}.$$

The functions Φ_i are defined as follows. Since the density $\hat{\rho}$ is determined by the moments we are carrying, in this Gaussian case it is

$$\hat{\rho}(x;\bar{x},\Sigma) = \frac{\exp(-\frac{1}{2}\|\Sigma^{-\frac{1}{2}}(x-\bar{x})\|^2)}{\sqrt{((2\pi)^d \det \Sigma)}}.$$

Then $\Phi_1 : \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^d$ is given by

$$\Phi_1(\bar{x},\Sigma) = \int_{\mathbb{R}^d} x\{e^{\mathcal{L}^*\tau}\hat{\rho}(x;\bar{x},\Sigma)\}dx,$$

and $\Phi_2 : \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is given by

$$\Phi_2(\bar{x},\Sigma) = \int_{\mathbb{R}^d} (x-\Phi_1(\bar{x},\Sigma))(x-\Phi_1(\bar{x},\Sigma))^T\{e^{\mathcal{L}^*\tau}\hat{\rho}(x;\bar{x},\Sigma)\}dx.$$

There is a connection between particle filters and the moment map. In their basic form particle filters represent the desired evolving probability density function as a sum of several delta functions, or Gaussians, not just as one [6]. However, various simplifications are often used in high-dimensional problems, such as the extended Kalman filter and the ensemble Kalman filter, and these methods have some resemblance to the moment map in spirit.

3.2. The ODE case. The moment map can be defined for deterministic problems of the form

$$\frac{dx}{dt} = f(x,y), \tag{3.6}$$

$$\frac{dy}{dt} = g(x,y). \tag{3.7}$$

Here $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^m$, and the randomness is assumed to come entirely from the initial data. In systems characterized by a separation of time scales, it is sometimes the case that one can write an effective reduced model in terms of a subset of (typically slow) variables. Under such appropriate conditions, we might, for example, be interested in finding a map in terms of the first k moments of x alone. Thus the measure μ must be chosen on $\mathbb{R}^d \times \mathbb{R}^m$ so that it is uniquely characterized by X, the first k moments of $x \in \mathbb{R}^d$. It is natural to choose μ to be an invariant measure for the flow, conditioned by knowledge of the first k moments of x; if the flow is Hamiltonian, then a Gibbs measure is often used. We denote the density associated with this measure by $\hat{\rho}(x,y;X)$. This occurs in the heat bath example considered in the introduction, where x represents coordinate and momentum of the distinguished particle, while y represents the heat bath coordinates and momenta.

Rather than the Fokker–Planck equation (3.3) we have the Liouville equation for propagation of probability densities. This is

$$\frac{\partial \rho}{\partial t} = -\nabla_x \cdot (f\rho) - \nabla_y \cdot (g\rho) := \mathcal{L}^*\rho,$$
$$\rho(x,y,0) = \hat{\rho}(x,y;X),$$

and we denote the solution by

$$\rho(x, y, t) = e^{\mathcal{L}^* t}\hat{\rho}(x, y; X).$$

In the case of a Dirac mass we take $\hat{\rho}(x, y; X) = \delta(x - X)\hat{\rho}(y; X)$, and then the map

$$\Phi(X) = \int_{\mathbb{R}^d \times \mathbb{R}^m} x\{e^{\mathcal{L}^* \tau}\delta(x - X)\hat{\rho}(y; X)\}dxdy.$$

Here $\hat{\rho}(y; X)$ is chosen so that $\delta(x - X)\hat{\rho}(y; X)$ is the density of μ conditional on $x = X$. Generalization to Gaussian, and higher moment problems, is also possible.

3.3. General remarks.

- *Notation.* We use X to represent a point in the moment space up to some order k, which will be made explicit for each particular example we consider. In practice, the coordinates used to describe the moment space are dictated by the particular problem. For example, in our case we use the mean and standard deviation as coordinates when considering $k = 2$ and $d = 1$. For $k = 2$ and $d = 2$ we use the two means, the two standard deviations, and the cross correlation as coordinates.
- *Usage.* In discussing moment maps we often do not distinguish between a point X in moment space and the uniquely determined density $\hat{\rho}(\cdot; X)$ based on this point. That is, we sometimes speak of the moment map as mapping X^n to X^{n+1} and sometimes speak of the moment map as mapping density $\hat{\rho}(\cdot; X^n)$ to density $\hat{\rho}(\cdot; X^{n+1})$.
- *Relation to optimal prediction.* The map Φ can be used to generate an approximate vector field by defining

$$F(X) = \frac{\Phi(X) - X}{\tau}. \tag{3.8}$$

  For the ODE case (3.6) and μ a Dirac at X, the limit $\tau \to 0$ coincides with the vector field found by the method of optimal prediction [3]; this is demonstrated in [16]. In general the method of optimal prediction leads to errors which grow linearly in time T [19]. The approach we study here attempts to overcome this error growth by closing the system with a larger number of moments.
- *Previous work.* For simple problems in chemical kinetics, which are modeled by birth-death processes, the equation for the first moment is a closed ODE, in the limit of a large number of independent particles, and the moment map studied here then works well in the Dirac mass case [29, 30, 31]. For more complex problems, such as lattice Boltzmann, a closed effective PDE may sometimes be found, using first and second moments, and again, the moment map works well in this case [40, 12, 41, 9]. In this paper we study examples where no closure is proved to exist, and demonstrate the properties of the moment map. In particular we study the relevance of fixed points of this map to the identification of metastable states. Although no rigorous analysis is presented, the numerical studies show that the moment map has some merit as an approach for elucidating long-term dynamics of large systems, through low-dimensional dynamical systems studies.

- *Lifting.* Initialization of the detailed simulation consistently with coarse-grained observables is the *lifting* step in equation-free computation [40, 12, 25, 38]. This step is obviously not unique, as there exist many ways of initializing a distribution conditioned on a few of its lower-order moments. Our choice of a particular measure depending only on the lower-order moments allows for a systematic initialization of the fine-scale dynamics, a concept that goes back to Ehrenfest (see [8, 17]), and is an important component of our computational approach. In the case of systems with metastability, different effective dynamics will be deduced (different closures will be obtained) depending (a) on the time scale of the observation (the time horizon of the simulation τ with the fine scale solver) and (b) on the nature of the lifting from the moments (the choice of ρ). Over very short times, and initializing with a Dirac delta function, the simulation will effectively sample the local gradient of the well; over medium times, and initializing *within one well*, one will observe equilibration within this well; and over very long times (no matter what the initialization) one will observe the approach to the equilibrium density. If we want to study the system over medium time scales, it is obviously important to use a time τ in the construction of the moment map that is short compared to the escape time between wells, but long enough for the noise (dynamics) to allow the sampling of the features of the well bottom. On a longer time scale it is necessary to incorporate the transition time between wells, as is done, for example, in the method of conditional averaging [36].
- *Computational savings.* The moment map can lead to computational savings in two primary ways:
  (i) the map Φ can be used in finite-dimensional bifurcation and continuation studies;
  (ii) the estimated vector field F can be used to advance the moments over several multiples of the time-step τ.

  In case (i), savings can arise from using accelerated methods, such as Newton iteration in a continuation environment, to find fixed points. In case (ii), savings arise by considering maps of the type

$$X^{n+1} = X^n + l\{\Phi(X^n) - X^n\}$$

  to advance the moments through $l\tau$ time units, using solution of the full problem (3.1) or (3.6) only over τ time units. (If $l = 1$, this simply reduces to the moment map.) The above formula constitutes a *projective forward Euler* explicit coarse integrator. Much more sophisticated integrators, including multistep and implicit ones, can also be used; a rigorous analysis of savings in case (ii) has only recently been initiated [13, 14, 35]. These ideas have been applied to a wide range of problems, both deterministic and stochastic; see, e.g., [15, 12, 29, 30, 31, 41, 38, 9, 18, 27, 22, 37, 4] and other references in [25, 26]. An approach related to case (ii), which can be fully optimized for variance reduction and so forth when explicit time scale separation occurs between the x and y dynamics in (3.8), is outlined in [42]; for a rigorous analysis see [7].

4. Examples. Here we explore a series of examples of moment maps. We start with examples based on the Ornstein–Uhlenbeck (OU) process, for which explicit representation is possible. In these cases the moment maps are linear. We then proceed to the more interesting

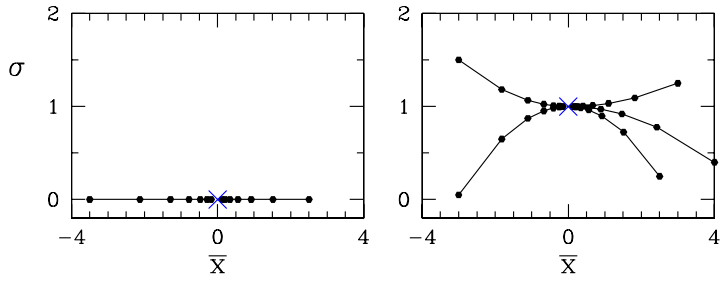


Figure 6. *Phase portraits for first-order (left) and second-order (right) moment maps for the OU process. Trajectories for these linear maps are shown starting from several different initial conditions. Parameters are* $\alpha = 1$, $\lambda = 2$, *and* $\tau = 0.5$. *See also the accompanying animations* (63866_01.gif [714KB], 63866_02.gif [1.4MB], *and* 63866_03.gif [2.5MB]).

nonlinear maps arising from systems with double-well potentials.

4.1. The OU process. Consider the OU process

$$\frac{dx}{dt} = -\alpha x + \sqrt{\lambda}\frac{dW}{dt}. \tag{4.1}$$

This is essentially the simplest example of (3.1) and corresponds to (2.5) in the case of a quadratic potential.

The exact solution of this process is

$$x(t) = e^{-\alpha t}x(0) + \sqrt{\lambda}\int_0^t e^{-\alpha(t-s)}dW(s). \tag{4.2}$$

For a first-order moment map on $\bar{x}$, we take initial data with Dirac measure with density $\hat{\rho}(x;\bar{x}) = \delta(x-\bar{x})$. The map on the first moment $\bar{x}$ is explicitly

$$\Phi(\bar{x}) = \mathbb{E}x(\tau) = \int x\{e^{\mathcal{L}^*\tau}\delta(x-\bar{x})\}dx.$$

From (4.2) we have

$$\Phi(\bar{x}) = e^{-\alpha\tau}\bar{x}. \tag{4.3}$$

The map Φ is linear and has a unique globally attracting fixed point at $\bar{x} = 0$. Figure 6 shows a phase portrait for this simple map.

Before discussing this we consider the second-order moment map with mean $\bar{x}$ and standard deviation σ as coordinates. In this case the initial data has density

$$\hat{\rho}(x;\bar{x},\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}.$$

From (4.2) we then have that the moment map on $(\bar{x},\sigma)$ is given by

$$\Phi_1(\bar{x},\sigma) = e^{-\alpha\tau}\bar{x}, \tag{4.4}$$

$$\Phi_2(\bar{x},\sigma) = \left\{e^{-2\alpha\tau}\sigma^2 + \frac{\lambda}{2\alpha}[1-e^{-2\alpha\tau}]\right\}^{\frac{1}{2}}. \tag{4.5}$$

While not a linear map on $(\bar{x}, \sigma)$, this is linear on $(\bar{x}, \sigma^2)$. Figure 6 shows a phase portrait for map (4.4). The map has the unique fixed point $(\bar{x}, \sigma) = (0, \sqrt{\{\frac{\lambda}{2\alpha}\}})$, which is globally attracting.

The solution to (4.2) is Gaussian if x_0 is Gaussian. Hence the second-order moment map (4.4) gives exact time τ samples of the distribution of the SDE. This unique fixed point of the map characterizes the unique invariant (Gaussian) measure of (4.1). In contrast, the first-order moment map, with μ a Dirac, can only approximate the solution. Although this map does not quantitatively represent the solution, since it contains no information about the width of the measure, it captures the correct dynamics of the first-order moment and shows that probability mass initially far from the origin will be transported towards the origin.

4.2. One-dimensional SDE: The double-well potential. We now present a detailed study of moment maps for the one-dimensional SDE (2.5) with double-well potential. In this section we focus on the dynamics of these maps using numerical simulations. In the next section we analyze the maps, in particular the nonlinearity of the maps, in more detail.

After suitable scaling, SDE (2.5) can be rewritten in the notation of (3.1) as

$$\dot{x} = -V'(x) + \dot{W}. \tag{4.6}$$

We consider the double-well potential

$$V'(x) = x(x^2 - \mu) + \nu. \tag{4.7}$$

For $\nu = 0$ the potential is symmetric about zero, and for ν small this symmetry is weakly broken. The potential has two local minima for $|\nu| < 2(\mu/3)^{3/2}$ and one minimum otherwise.

The first-order (Dirac-based measure) and second-order (Gaussian-based measure) moment maps for this equation are *nonlinear*. In particular, these maps have multiple stable fixed points that undergo bifurcations as the potential (μ or ν) is varied. We are interested in these fixed points and their stable and unstable manifolds as a function of μ for ν fixed. We shall consider two cases, the slightly asymmetric case $\nu = 0.3$ and the symmetric case $\nu = 0$. We resort solely to numerical studies of the moment maps throughout this section. In brief, we use Monte Carlo simulations to evolve densities forward over time interval τ, as in the evolution from $\hat{\rho}$ to ρ in Figure 5. This numerically determines the moment maps. By employing additional techniques, we can compute steady states and bifurcations. The effective fixed point, bifurcation, and continuation calculations require estimates of the Jacobian of the moment map (or its action). This is achieved by using nearby initializations of the moment map (see [29, 30] as well as [40, 25, 23] and the monograph [24] for matrix-free implementations of Newton-GMRES).

We first consider the moment maps for the slightly asymmetric potential with representative values of μ and map time τ. Figure 7 shows phase portraits for both first-order and second-order moment maps. Each map has three fixed points. These are shown together with unstable, and for the second-order map also stable, manifolds of the unstable fixed point.

Figure 8 shows one iteration of both the first- and second-order maps. Consider the first-order map. The initial density $\rho(0) = \hat{\rho}(x; \bar{x})$ is a Dirac delta at $\bar{x}$, here a point slightly to the right of the unstable fixed point near zero. After time $\tau = 0.1$ the mean of the density has

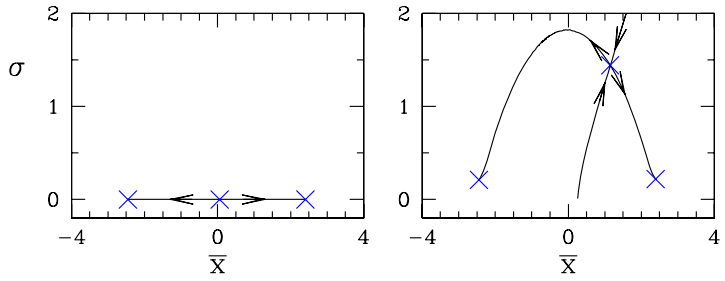


Figure 7. *Phase portraits for first-order (left) and second-order (right) moment maps for the one-variable SDE in the case of a slightly asymmetric potential. Fixed points are indicated by crosses. The stable (for second-order map) and unstable manifolds of the middle fixed point are shown. Note, the stable manifold of the saddle in the second-order map does not pass through the middle fixed point of the first-order map. Parameters are* $\mu = 6$, $\nu = 0.3$, $\tau = 0.1$.

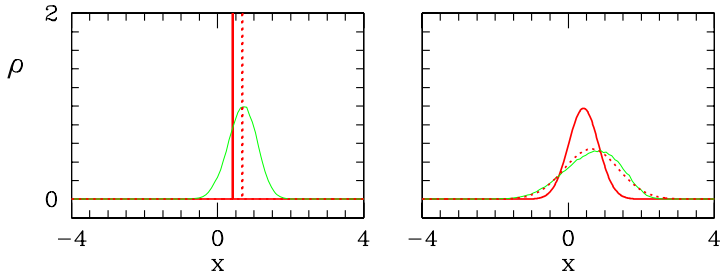


Figure 8. *One iteration of the first and second order maps whose phase portraits are shown in Figure 7. The bold red curve shows* $\rho(0) = \hat{\rho}(x; X^n)$*, the thin green curve shows* $\rho(\tau)$*, and the dash bold red curve shows* $\hat{\rho}(x; X^{n+1})$*. (See Figure 5.)*

moved to the right, and hence the map takes $\bar{x}$ to the right in this case. The density spreads considerably, but over $\tau = 0.1$ it remains nearly symmetric (the mean is indistinguishable from the maximum). The initial density for the next iteration is a Dirac delta displaced to the right.

For the second-order map the initial density $\rho(0) = \hat{\rho}(x; \bar{x})$ is a Gaussian centered at $\bar{x}$ with width σ. Here $(\sigma, \bar{x}) = (0.418, 0.409)$ corresponds to a point on the lower branch of the stable manifold of the saddle fixed point. After $\tau = 0.1$ time units the density has spread, and the mean has moved slightly to the right. The map thus corresponds to substantial increase in σ and small increase in $\bar{x}$. The density $\rho(\tau)$ is slightly non-Gaussian, as can be seen in comparison with the initial (Gaussian) density for the next iteration.

Figure 9 shows all fixed points in Figure 7. In each case we plot the density $\hat{\rho}(\cdot; X)$ corresponding to each fixed point in moment space as well as the density $\rho(\tau)$. The stable fixed points of the second-order map are the metastable measures centered in each well. Specifically, the Gaussian measures $\hat{\rho}(x; \bar{x}, \sigma)$ corresponding to the stable fixed points are indistinguishable from the evolved densities $\rho(\tau)$. As we shall show this is independent of τ over a very large range of τ. Intuitively this is because on any time scale, up to the very long time scale needed

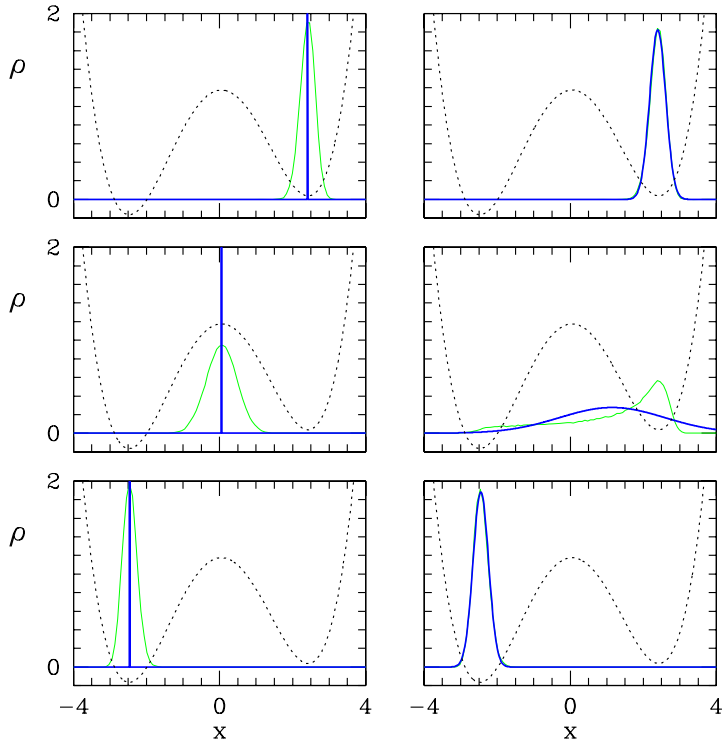


Figure 9. *Densities $\hat{\rho}(\cdot; X)$, shown with bold blue curves, corresponding to fixed points for the first-order (left) and second-order (right) moment maps shown in Figure 7. Thin green curves show the evolved density $\rho(\tau)$ for each case. (For the two stable fixed points of the second-order moment map the evolved density $\rho(\tau)$ is indistinguishable from $\hat{\rho}(\cdot; X)$.) The potential $V(x)$ is shown as a dashed curve.*

to reach the equilibrium distribution, these Gaussian measures are approximately invariant. The stable fixed points of the first-order map are not invariant measures. However, starting from initial condition $\hat{\rho}(x; \bar{x}) = \delta(\bar{x})$, where $\bar{x}$ is indistinguishable from the potential minimum, the density simply fills out (symmetrically), the (locally quadratic) well bottom. Note that $\tau = 0.1$ is close to, but not quite, the time necessary to reach the metastable density starting from the Dirac measure.

The unstable fixed points for the maps are understood as follows. For the first-order map, the unstable fixed point is at the local maximum of the potential. Starting from a Dirac delta $\hat{\rho}(x; \bar{x}) = \delta(\bar{x})$, the density spreads symmetrically, since the maximum is locally quadratic and hence symmetric. Hence after time $\tau = 0.1$ the mean is still at local maximum of the potential. Only for times τ long enough for the density to fill the two wells, and hence have a mean different from the local maximum, would the fixed point be different from the local maximum. For such times the fixed point would approximately be the mean of the equilibrium density. For the second-order map, the saddle fixed point corresponds to a Gaussian $\hat{\rho}(x; \bar{x}, \sigma)$, which is quite broad. The evolved density $\rho(\tau)$ is far from Gaussian; it simply has the property that its first two moment agree with those of the initial Gaussian. The saddle fixed point is quantitatively sensitive to the map time τ (see below). Qualitatively, however, for any value

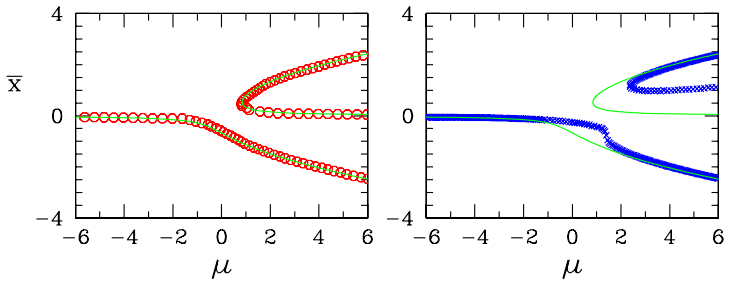


Figure 10. *Bifurcation diagram for first-order (left) and second-order (right) moment maps for the one-variable SDE in the case of a slightly asymmetric potential. Lines show local extrema of the potential. Parameters are* $\nu = 0.3$ *and* $\tau = 0.1$.

of τ the saddle fixed points are broad Gaussians.

The *stable* fixed points of both maps correspond to metastable states (measures) of the underlying process. The metastable states are very nearly Gaussian measures, because the wells are locally quadratic, and they are thus well captured by the low-dimensional moment maps: the densities corresponding to the stable fixed points of the second-order map are virtually indistinguishable from the metastable states. The fixed points of the first-order map capture the means of the metastable distributions. The stable fixed points are insensitive to the value of τ over a large range of τ. (See Figure 12 below.) This lack of sensitivity to τ suggests that these fixed points are meaningful characteristics of the observed dynamics over a range of observation time scales. The unstable (saddle) fixed point for the first-order moment map is also insensitive to the value of τ, but the unstable fixed point for the second-order map is sensitive to τ. This suggests that the fixed points of the first-order moment map provide a useful description of the dynamics (for relatively short times) close to both the saddle and the well bottoms; the second-order map fixed points provide a meaningful description of the effective dynamics close to the bottoms of the two wells, *but not in the neighborhood of the saddle*; this is essentially because a reduced equation in terms of the second-order moments does not appear to successfully close in the neighborhood of the saddle.

We now turn to the behavior of the fixed points as a function of well depth. Figure 10 shows bifurcation diagrams for each moment map as functions of μ, including μ for which the potential has a single well (basically $\mu < 0$). Local extrema of the potential are shown for comparison. The right-most end of each bifurcation diagram ($\mu = 6$) corresponds to the phase portraits just considered. The fixed points for the first-order map follow the extrema of the potential closely for all μ, including near the saddle-node bifurcation. The second-order fixed points do not. In Figure 11 we show fixed points in cases where the potential is far from locally quadratic. One case is $\mu = 0$, and the other is $\mu = 2.4$, very near the value corresponding to the saddle-node bifurcation of the second-order map. In both cases the first-order fixed point is at the potential minimum, while the mean of the second-order fixed point is noticeably different from the minimum, as can also be seen in Figure 10.

We show in Figure 12 how steady states for the second-order moment map are affected by the choice of the map time τ. To understand what the figure shows, it is helpful to first consider the fixed points for $\mu < 0$. Neither $\bar{x}$ nor σ varies significantly with τ, and σ is

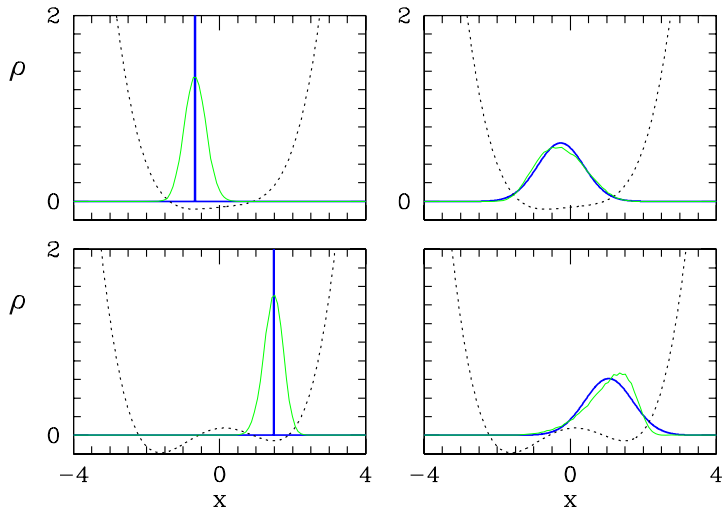


Figure 11. *Densities, shown with bold blue curves, corresponding to fixed points for first-order (left) and second-order (right) moment maps at two values of μ where the nonquadratic aspect of the potential is apparent. Thin green curves show the evolved density $\rho(\tau)$ for each case. The top case is $\mu = 0$. The bottom is $\mu = 2.4$, very near the saddle-node bifurcation for the second-order map in Figure 10. The potential $V(x)$ is shown as a dashed curve.*

not large. While not as easy to see, the stable fixed points for $\mu \gtrsim 2$ behave similarly. The stable fixed points correspond to the uppermost and lowermost branches in the $\bar{x}$-plot and the bottom branch in the σ-plot. Both stable fixed points have approximately the same, relatively small, value of σ. The intermediate, generally unstable, fixed points have larger values of σ; these vary quantitatively and qualitatively with τ. The number of unstable fixed points can change with τ; e.g., at $\mu = 6$, $\tau = 0.1$ there is one saddle fixed point between the two stable fixed points (Figure 7), whereas at $\mu = 6$, $\tau = 1.0$ there are three fixed points between the two stable points (similar to Figure 14 below). The conclusion is that the stable fixed points of the second-order moment map which correspond to stable or metastable measures of the SDE are largely insensitive to the time τ used to define the map. The other fixed points typically correspond to broad distributions and depend quantitatively on τ. This dependence of the fixed points of the moment map on τ suggests that they are not useful features of the effective dynamics—which in turn suggests that the effective behavior *does not usefully close* at the second moment level in the neighborhood of these fixed points.

A further way to quantify the effective closure at a given level is to examine how closely two successive iterates of the moment map agree with one iterate using twice the map time. Specifically, one can consider whether $\Phi_\tau(\Phi_\tau(X)) \simeq \Phi_{2\tau}(X)$ for some particular point X in moment space and some chosen map time τ. If the two are nearly equal, then the system is effectively closed with the given number of moments in the vicinity of point X. We illustrate this in Figure 13 for the second-order moment map. We have computed $||\Phi_\tau(\Phi_\tau(\bar{x}, \sigma)) - \Phi_{2\tau}(\bar{x}, \sigma)||_2$ for a large number of points in moment space with $\tau = 0.1$. The norm of the difference is shown in greyscale. Everywhere in the white region the norm is very small. Starting from any point in the white region, two map iterations over time τ and one iteration

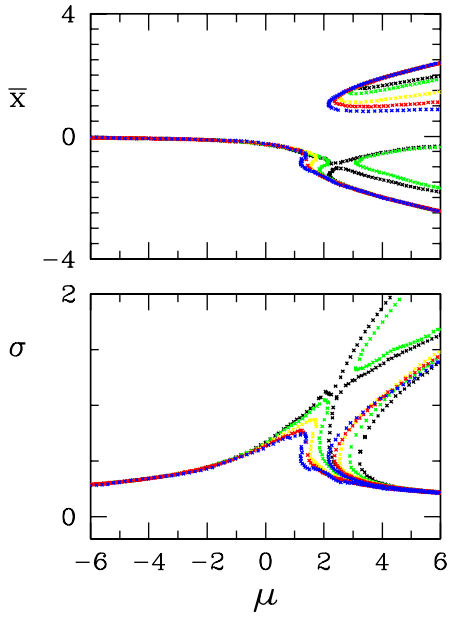


Figure 12. *Bifurcation diagram for a second-order moment map for the one-variable SDE for a variety of values of* τ*:* $\tau = 0.05$ *blue,* $\tau = 0.1$ *red,* $\tau = 0.2$ *yellow,* $\tau = 0.5$ *green,* $\tau = 1.0$ *black. Other parameters are* $\mu = 6$ *and* $\nu = 0.3$*.*

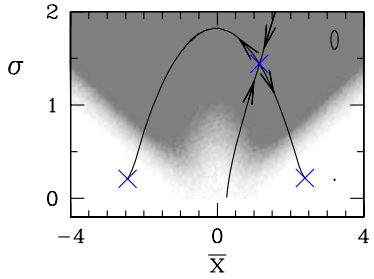


Figure 13. *Effective closure for second-order moment maps for the one-variable SDE in the case of a slightly asymmetric potential. Shown in greyscale is* $||\Phi_\tau(\Phi_\tau(\bar{x},\sigma)) - \Phi_{2\tau}(\bar{x},\sigma)||_2$ *as a function of* $\bar{x},\sigma$*, for* $\tau = 0.1$*. White corresponds to a norm less than* 0.01*, while dark grey corresponds to a norm greater than* 0.1*. For reference, two ellipses are plotted (at* $\bar{x} = 3.2$*): a very small one in the white region corresponds to a neighborhood of size* 0.01*, and one in the grey region corresponds to a neighborhood of size* 0.1*. The phase portrait from Figure 7 is also shown. Parameters are* $\mu = 6$*,* $\nu = 0.3$*,* $\tau = 0.1$*.*

over time 2τ lie within a neighborhood of size shown. As expected, the white region includes the two stable fixed points. In the grey region the norm is quite large. Here two map iterations over time τ and one iteration over time 2τ lie further apart than the neighborhood shown in the grey region. This indicates that is the system does not close in the Gaussian approximation.

Finally we consider the moment maps for the symmetric double-well potential. Figure 14 shows phase portraits, similar to Figure 7 for the asymmetric case, while Figure 15 shows fixed

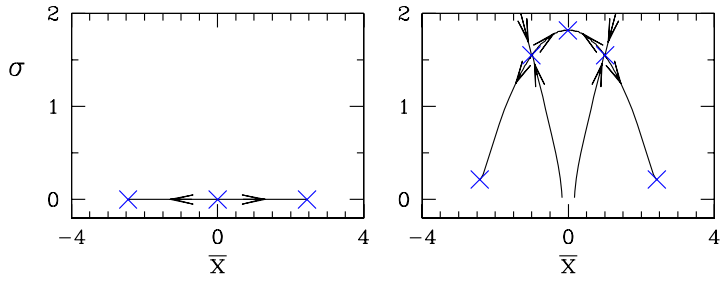


Figure 14. *Phase portrait for first-order (left) and second-order (right) moment maps for the one-variable SDE in the case of a symmetric potential. Fixed points are indicated by crosses. The stable (for second-order map) and unstable manifolds of the saddle fixed point are shown. Parameters are* $\mu = 6$, $\nu = 0$, $\tau = 0.1$.

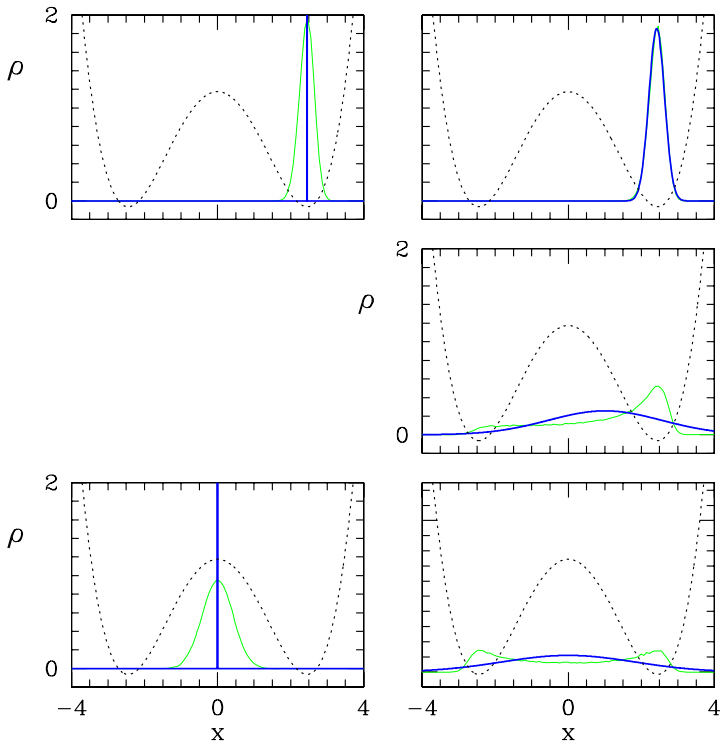


Figure 15. *Fixed points for first-order (left) and second-order (right) moment maps shown in Figure 14. Same conventions as in Figure 9. Top plots show the right stable fixed point. The middle plot shows the right saddle for the second-order map. The bottom plot shows the middle fixed point (saddle for the first-order map and stable fixed point for the second-order map). The other points in Figure 14 are obtained by symmetry.*

points, similar to Figure 9 for the asymmetric case. Figure 16 shows a bifurcation diagram as a function of μ. An important observation from the data presented in the symmetric case is that the moment map produces stable equilibria that are far from metastable states. The Gaussian measures corresponding to these fixed points are very broad.

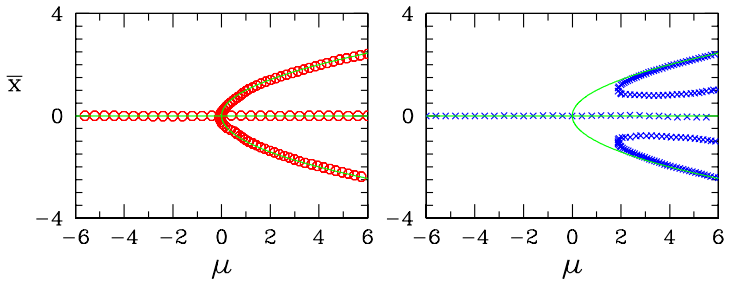


Figure 16. *Bifurcation diagram for first-order (left) and second-order (right) moment maps for the case of the symmetric potential. Lines show local extrema of the potential. Parameters are* $\nu = 0$ *and* $\tau = 0.1$.

4.3. Metastability and the double well potential. We now present some analysis of the nonlinear map. Consider the SDE (4.6). The adjoint of the generator for this process is

$$\mathcal{L}^*\phi(x) = \frac{d}{dx}\{V'(x)\phi(x)\} + \frac{1}{2}\frac{d^2\phi}{dx^2}(x).$$

The equation has a unique invariant density ρ_∞, in the null-space of $\mathcal{L}^*$, given explicitly by

$$\rho_\infty(x) = \mathcal{Z}^{-1}\exp\{-2V(x)\}, \qquad \mathcal{Z} = \int_{\mathbb{R}} \exp\{-2V(x)\}dx. \tag{4.8}$$

The operator $\rho_\infty^{-\frac{1}{2}}\mathcal{L}^*\rho_\infty^{\frac{1}{2}}$ is self-adjoint in the space $L^2(\mathbb{R})$ (see Proposition 2.2 of [21]). We let $\langle\cdot,\cdot\rangle$ denote the weighted $L^2(\mathbb{R})$ inner product

$$\langle\theta,\psi\rangle = \int_{\mathbb{R}} \frac{\theta(x)\psi(x)}{\rho_\infty(x)}dx,$$

and we write the eigenvalue problem

$$\mathcal{L}^*\phi_j(x) = \lambda_j\phi_j(x)$$

with eigenvalues ordered so that

$$0 = \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \cdots.$$

We may choose the normalization

$$\langle\phi_j,\phi_k\rangle = \delta_{ij}.$$

Since $\phi_0(x) = \rho_\infty(x)$, we have

$$\int_{\mathbb{R}} \phi_0(x)dx = 1, \qquad \int_{\mathbb{R}} \phi_j(x)dx = 0, \quad j \geq 1.$$

Now, the solution $\rho(x,t)$ of the Fokker–Planck equation can be expanded in the eigenfunctions ϕ_j as

$$\rho(x,\tau) = \sum_{j=0}^{\infty} a_j(t)\phi_j(x),$$

where

$$a_j(t) = a_j(0)e^{\lambda_j t}.$$

Notice that $a_0(0) = 1$ in all cases, because $\rho(x,0)$ is a probability density function.

Assume that $V(x)$ in (4.6) is a double-well potential with deep wells relative to the noise. Then the analysis and numerical evidence in [2, 21, 32] suggest that it is reasonable to assume that

$$\epsilon := \lambda_1/\lambda_2 \ll 1. \tag{4.9}$$

Thus there exists a spectral gap and, noting that $\lambda_0 = 0$, this suggests that after times τ of order $-1/\lambda_2$ the density $\rho(x,t)$ can be well approximated by only the first two eigenfunctions ϕ_0 and ϕ_1.

Assuming that $V'' \neq 0$ at the two well bottoms, then $\rho_\infty(x)$ is well approximated as the weighted sum of two Gaussians $g_\pm(x)$; this may be verified from (4.8). We may assume that the Gaussians are normalized to be probability densities, and then define their means and standard deviations by

$$\int_{\mathbb{R}} g_\pm(x)dx = 1, \quad \int_{\mathbb{R}} xg_\pm(x)dx = \bar{g}_\pm, \quad \int_{\mathbb{R}} [x-\bar{g}_\pm]^2 g_\pm(x)dx = \bar{\sigma}_\pm^2.$$

We then write

$$\phi_0(x) = \rho_\infty(x) \approx \alpha g_+(x) + (1-\alpha)g_-(x), \tag{4.10}$$

where α determines the relative weight of the two Gaussian contributions. Furthermore, by orthonormality, it may be shown that

$$\phi_1(x) \approx \sqrt{\{\alpha(1-\alpha)\}}[g_+(x) - g_-(x)]. \tag{4.11}$$

We can now use the approximations (4.10), (4.11) to understand the nonlinearity in the moment map for the case of the double-well potential. We focus on the second-order map. In this case $\rho(x,0) = \hat{\rho}(x;\bar{x},\sigma)$ will be Gaussian. For a sufficiently large spectral gap, (4.9), we can choose a map time τ long on the scale of $1/\lambda_2$, while short on the scale of $1/\lambda_1$. Specifically, letting $T = -\lambda_2\tau$, we have

$$\rho(x,\tau) = \phi_0(x) + a_1(0)e^{-\epsilon T}\phi_1(x) + O(e^{-T}).$$

Given small $\epsilon \ll 1$, we can choose T sufficiently large that $e^{-T} \ll 1$ while $\epsilon T \ll 1$. Then letting

$$\beta = a_1(0)\sqrt{\{\alpha(1-\alpha)\}},$$

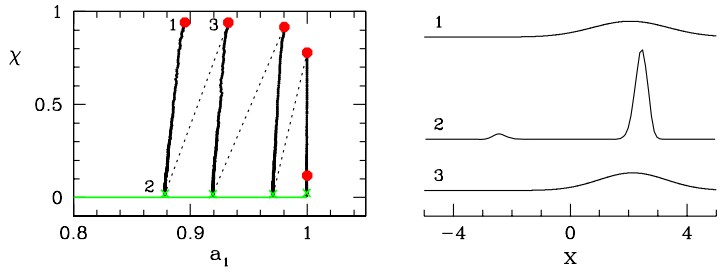


Figure 17. *Iterates of the moment map as seen in a_1, χ coordinates. Red dots show points with Gaussian densities $\hat{\rho}(x; \bar{x}, \sigma)$. The Gaussian densities corresponding to points labeled 1 and 3 are shown to the right. The evolution of the density over time τ is shown by bold curves (actually fine series of bold points) with green crosses indicating the final time. The density corresponding to point 2 is shown to the right. Dashed lines indicate the projections back to Gaussian density, which preserve the mean and standard deviation of the distribution. Parameters are $\mu = 6$, $\nu = 0$, and $\tau = 1.0$.*

we have

$$\begin{aligned} \rho(x,\tau) &\approx (\alpha+\beta)g_+(x) + (1-\alpha-\beta)g_-(x) \\ &= \gamma g_+(x) + (1-\gamma)g_-(x), \end{aligned} \tag{4.12}$$

where $\gamma = \alpha + \beta$.

The key to the dynamics of the moment map is manifest in the formula (4.12). Given any Gaussian with mean and standard deviation $(\bar{x}_0, \sigma_0)$, we project the density $\rho(x, 0) = \hat{\rho}(x; \bar{x}_0, \sigma_0)$ onto the basis $\{\phi_j(x)\}_{j=1}^{\infty}$. Then the evolution over appropriate time τ results in $\rho(x, \tau)$ containing only ϕ_0 and ϕ_1 and is given by (4.12), so that

$$\bar{x}_1 = \gamma \bar{g}_+ + (1-\gamma)\bar{g}_-, \tag{4.13}$$

$$\sigma_1^2 = \beta \bar{\sigma}_+^2 + (1-\beta)\bar{\sigma}_-^2 + (1-\beta)\beta(\bar{g}_+ - \bar{g}_-)^2. \tag{4.14}$$

Since β, γ depend nonlinearly on $(\bar{x}_0, \sigma_0)$, through α and $a_1(0)$, we have a nonlinear map $(\bar{x}_0, \sigma_0) \to (\bar{x}_1, \sigma_1)$. In principle this map can be computed explicitly, though this requires knowing the eigenfunctions ϕ_1 and ϕ_0 accurately.

This analysis can be applied to our computational studies of the double-well potential to gain further insight into the moment map. We employ the symmetric double-well potential considered at the end of section 4.2. In this case the eigenfunctions ϕ_1 and ϕ_0 are accurately approximated by (4.10) and (4.11) with $\alpha = 1/2$. We use $\tau = 1$ in the computations that follow. (We shall see that $\lambda_2 \simeq -5$, so that $\tau\lambda_2 \simeq -5$.)

Figure 17 shows the moment map in the coordinates suggested by the preceding analysis. The corresponding evolution of the density is shown for one iterate of the map. We wish to view the evolution of the density $\rho(x, t)$ in terms of the amplitudes $a_j(t)$ of projections onto the eigenfunctions ϕ_j. We know that $a_0(t) \equiv 1$, so there is no need to show this amplitude. The essential amplitude is a_1. The effect of all the higher amplitudes a_j, $j > 1$, can be summarized

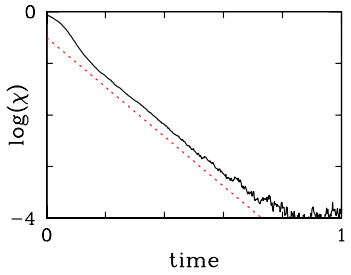


Figure 18. *Exponential decay of higher modes. Dashed curve has slope* −4.8.

by a single scalar χ defined as

$$\chi \equiv \left\| \sum_{j=2}^{\infty} a_j \phi_j \right\|_2 = \|\rho - a_0\phi_0 - a_1\phi_1\|_2,$$

with $\rho = \rho(\cdot, \tau)$.

Consider point 1 in Figure 17. To the right is shown the Gaussian density $\rho(0) = \hat{\rho}(x; \bar{x}_0, \sigma_0)$ determined by a point $(\bar{x}_0, \sigma_0)$ in moment space. This density has a significant projection onto the higher modes ϕ_j, $j > 1$, and χ is significantly greater than zero. From the analysis we expect the density to evolve such that χ decays to zero on a time scale faster than the dynamics of a_1. The thick black curves in Figure 17 show this evolution. There is little change in a_1 as the system evolves toward $\chi = 0$. The decay to $\chi = 0$ is ultimately exponential, as is shown in Figure 18. From this we estimate that $\lambda_2 \simeq -5$ and thus $\exp(\lambda_2 \tau) \simeq 0.007$.

After time $\tau = 1$ the system is at point 2 with corresponding density shown at the right. The resulting density is no longer Gaussian as expected from (4.12). The second-order moment map is obtained by determining $(\bar{x}_1, \sigma_1)$, the mean and standard deviation of density 2. For the next iteration of the map one constructs a Gaussian density with this mean and standard deviation (point 3 in Figure 17) and repeats. The dashed line connecting points 2 and 3 illustrates this. Thus the moment map takes Gaussian $\hat{\rho}(x; \bar{x}_0, \sigma_0)$ determined by $(\bar{x}_0, \sigma_0)$ (point 1) to $\hat{\rho}(x; \bar{x}_1, \sigma_1)$ determined by $(\bar{x}_1, \sigma_1)$ (point 3) where the map can be again be iterated. Note that the evolution from point 1 to point 2 is due to linear flow of the Fokker–Planck equation. Nonlinearity results from the projection of density at point 2 back to $(\bar{x}_1, \sigma_1)$; i.e., going from point 2 to point 3 introduces nonlinearity into the map. Further note that the mean has moved slightly to the right after one iteration of the map. Hence, on the next iteration less of the mass will move into the left well (as in point 2). In this way the map stabilizes the metastable states corresponding to localization of density into a single well.

Now consider exactly the same evolution seen in the moment-map coordinates $(\bar{x}, \sigma)$ in Figure 19. The green curve is that generated by $a_0 = 1$, $-1 \le a_1 \le 1$, $a_j = 0$, $j > 1$. This curve is also shown as green in Figure 17. This is the slow manifold for the system. Starting from point $(\bar{x}_0, \sigma_0)$, point 1, the evolution of $(\bar{x}, \sigma)$ as $\rho(x, t)$ evolves is shown in bold. This is the decay of the modes ϕ_j, $j > 1$. After time τ the system reaches $(\bar{x}_1, \sigma_1)$, point 2. Even

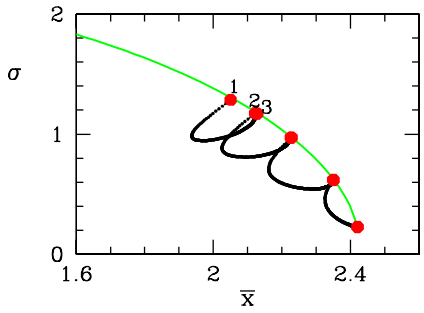


Figure 19. *Iterates of the moment map seen in $(\bar{x}, \sigma)$ coordinates. The points correspond to exactly the same points as in Figure 17. The bold curves (actually a fine series of bold points) show the evolution of $(\bar{x}, \sigma)$ as the density evolves between iterates of the map. The green curve is the slow manifold ($\chi = 0$), also shown in Figure 17.*

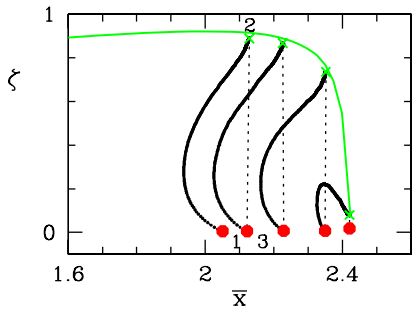


Figure 20. *Iterates of the moment map seen in $(\bar{x}, \zeta)$ coordinates. The points correspond to exactly the same points as in Figures 17 and 19. The bold curves (actually a fine series of bold points) show the evolution of $(\bar{x}, \zeta)$ as the density evolves between iterates of the map. The green curve is the slow manifold.*

though the transient dynamics is such as to move initially away from the slow manifold, the evolution brings the system back, as it must. The projection back to Gaussian density does not change $\bar{x}$ and σ, so points 2 and 3 appear the same in Figure 19.

We know that as the system evolves from point 1 to point 2 the density became non-Gaussian. We show this in Figure 20 by showing the dynamics out of the $(\bar{x}, \sigma)$-plane. Here ζ is given by

$$\zeta \equiv \|\rho - \hat{\rho}(\cdot, \bar{x}, \sigma)\|_2,$$

with $\rho = \rho(\cdot, \tau)$, and is similar in spirit to χ. ζ is a measure of how far the density is from Gaussian, whereas χ is a measure of how far the density is from the slow manifold. However, these have opposite roles and behavior. Gaussian densities correspond to $\zeta = 0$, but since such densities are not in general on the slow manifold they correspond to $\chi \neq 0$.

Starting from point 1 in Figure 20, ζ is necessarily zero since the density (Figure 17 point 1) is Gaussian. As the system evolves toward the slow manifold and becomes non-Gaussian, ζ

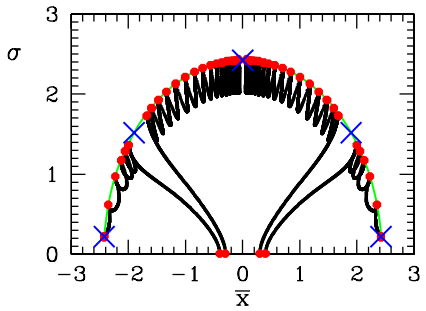


Figure 21. *Iterates of the moment map seen in $(\bar{x}, \sigma)$ coordinates. Same as Figure 19 except over a larger range of coordinates, and same as Figure 14 except that here $\tau = 1$.*

increases (point 2). Projection back to Gaussian moves the system vertically downward to $\zeta = 0$ (point 3).

Now consider the case $a_1 \simeq 1$, near the local minima of the potential wells (we consider only the right well; the left is the same). One sees in Figures 17 and 20 that points on the slow manifold correspond to nearly Gaussian densities. This can be most clearly seen in Figure 20, where the slow manifold falls to near $\zeta = 0$. The end point of the green curve is $a_1 = 1$. In Figure 17 we see that the red points (the Gaussian densities in a coordinates) fall very nearly to $\chi = 0$, meaning that these densities are almost exactly represented by a sum of ϕ_0 and ϕ_1. The small gap between the slow manifold and the Gaussian density (seen in both Figures 17 and 20) reflects the fact that the metastable density is not exactly Gaussian.

In Figure 21 we show the dynamics in $(\bar{x}, \sigma)$ coordinates, showing all the fixed points of the moment map. This is similar to Figure 14(right), except that here $\tau = 1$. The figure shows trajectories starting from four initial conditions. Note that the slow manifold accurately captures the unstable manifold of the saddle fixed points.

4.4. Two-dimensional SDE. In this section we consider moment maps for the two-dimensional SDE (2.4) presented in section 2.2. The dynamics are potentially richer than for the one-dimensional SDE considered up to now. Nevertheless, the stabilization of metastable states is the same as for the one-dimensional SDE.

As before, we examine maps for both first-order and second-order moments. The first-order map can be written as $\Phi(\bar{x}_1, \bar{x}_2)$, where $\bar{x}_1$ and $\bar{x}_2$ are means of Q and P, respectively. The second-order map can be written as $\Phi(\bar{x}_1, \bar{x}_2, \sigma_1, \sigma_2, c)$, where σ_1 and σ_2 are the standard deviations of Q and P, respectively, and c is the cross correlation. Other coordinates could be used for the five-dimensional phase space. We again consider the slightly asymmetric double-well potential (2.3) for parameters similar to those used for the one-dimensional SDE. All results have been obtained numerically from simulations of (2.4) for the case $M = 1$, $\gamma = 1$, $\beta = 2$.

Figure 22 shows typical phase portraits for the first-order and second-order moment maps. The first-order map exhibits the dynamics typical of bistable damped oscillators. Similar to the one-dimensional case, the fixed points are located at, or very close to, $(\bar{x}_c, 0)$, where $\bar{x}_c$ is a local extremum of the potential. The second-order map has five fixed points, similar to

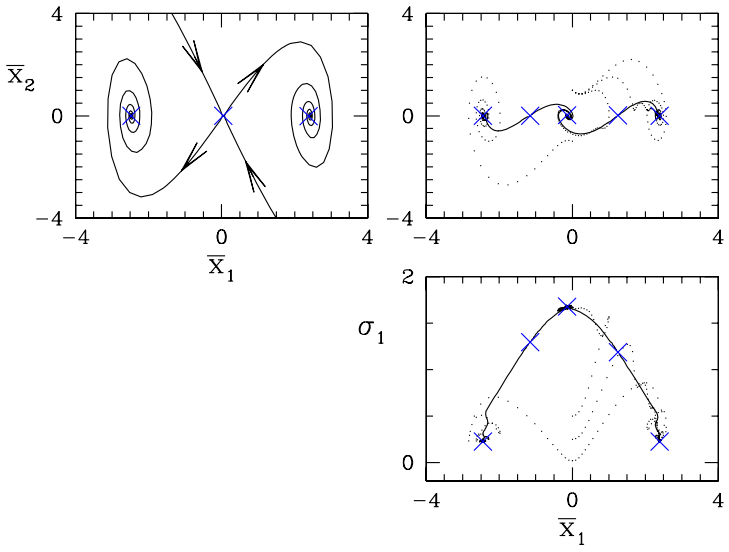


Figure 22. *Phase portrait for first-order (left) and second-order (right) moment maps for the two-variable SDE with a slightly asymmetric double-well potential. Fixed points are indicated by crosses. For the first-order map, stable and unstable manifolds of the saddle fixed point are shown. For the second-order map, unstable manifolds of the saddle fixed point are shown. The stable manifolds are four-dimensional. Points show some representative trajectories. Parameters are* $\mu = 6$, $\nu = 0.3$, $\tau = 0.1$.

those seen in Figure 14, including a stable fixed point with a very broad Gaussian measure. Typically trajectories of interest approach either one or the other of the stable fixed points corresponding to the potential minima. The moment maps for the two-dimensional SDE stabilize metastable states in a very similar way to that found for the one-dimensional SDE.

Figure 23 provides a better view of the dynamics of the second-order moment map and shows how trajectories approach the stable fixed points. A Gaussian measure in two variables can be visualized as an ellipse in the plane specified by five numbers: the center, the semimajor and semiminor axes, and the orientation. Thus for each point in the five-dimensional phase space of the moment map we plot an ellipse in the plane. One can think of each ellipse as corresponding to a level set of a Gaussian density.

Two trajectories are shown, one with the ellipse at every iteration of the map plotted, and the other with only a few representative ellipses plotted. One can see how densities evolve under the map toward stable equilibria. Trajectories from almost all initial condition distributions that are not too broad (i.e., initialized within one of the two metastable wells, as discussed above) evolve to one or the other of the two fixed points, similarly to what is shown in Figure 23.

For completeness we show in Figure 24 bifurcation diagrams for the moment maps for the two-dimensional SDE. These bifurcation diagrams reveal much the same features as for the one-dimensional SDE. Fixed points of the first-order moment map track the potential minima. For the second-order map, stable fixed points corresponding to metastable states exist except in regions near where the potential bifurcates from single- to double-well. Clearly, close to such parameter values, the separation of time scales between equilibration in one well and

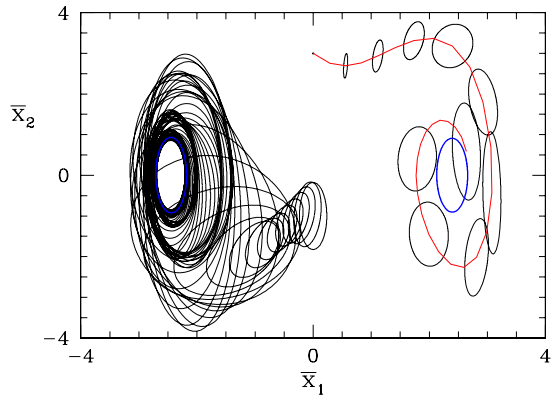


Figure 23. *Dynamics of the second-order moment map for the two-variable SDE. Two trajectories are visualized by plotting ellipses corresponding to points in the five-dimensional phase space. One trajectory starts at* $(\bar{x}_1, \bar{x}_2, \sigma_1, \sigma_2, c) = (0, 3, 0.01, 0.01, 0)$. *Only a few representative ellipses are shown. The curve shows the path of the center of the ellipses. The other trajectory starts at* $(\bar{x}_1, \bar{x}_2, \sigma_1, \sigma_2, c) = (0, -1, 0.1, 0.5, 0)$. *In this case every point on the trajectory is shown. Two stable fixed points of the map are shown as bold blue ellipses. Parameters are* $\mu = 6$, $\nu = 0.3$, $\tau = 0.1$. *See also the accompanying animations* (63866_04.gif [3.8MB] *and* 63866_05.gif [7.5MB]).

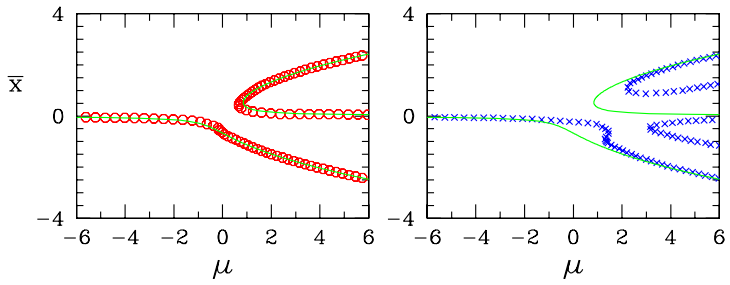


Figure 24. *Bifurcation diagrams for the first-order (left) and second-order (right) moment maps for the two-variable SDE. Parameters are* $\nu = 0.3$, $\tau = 0.1$.

transition to the other is no longer present, and the fixed points that we find depend on the map reporting horizon τ. Fixed points, both stable and unstable, corresponding to broad distributions, separate the stable fixed points. While not shown, we find that the fixed points corresponding to the stabilized metastable states are essentially independent of the time τ over which the map is defined; this suggests that the map is a good effective description in their neighborhood, but not a useful one close to the τ-dependent fixed points. The ability to test the sensitivity of the map dynamics and fixed points to the parameter τ, as well as the ability to use maps of different orders, grants us potential tools with which to "test online" the validity of a given map as an effective model of the detailed dynamics.

4.5. Heat bath. Finally, we consider the dynamics of a particle in a heat bath as described in section 2.1. Moment maps for this example are of the type defined for ODE systems in section 3.2. Nevertheless we expect the moment map for the ODE system to behave very much

like that of the two-dimensional SDE, since the SDE is known to capture the dynamics of the heat-bath model in the limit of a large bath. From a computational viewpoint, in fact, there is not much distinction between the stochastic (SDE) and the deterministic (ODE) cases, since in both cases we numerically compute the moment maps using Monte Carlo simulations to evolve densities over time interval τ.

There are two related new issues, however. The first is that N, the number of particles in the heat bath, is now a parameter which could potentially affect the dynamics of the system. The other related issue is that there is a minimum time interval over which we should define the moment map. For any given bath size N there is a maximum frequency ω_m of the oscillators in the bath, where $\omega_m \simeq N^{1/3}$. We therefore should take the map time τ to be at least of the same order as the $1/\omega_m$. We always use $\tau > 2\pi/\omega_m$.

Other than the preceding two points, the only significant difference between the heat bath and the two-dimensional SDE is that the heat bath requires significantly more computation to evolve densities forward in time. Hence the moment map is much more expensive to compute for the heat bath than for the SDE. For this reason, we have limited our studies to a fixed potential: (2.3) with $\mu = 4$, $\nu = 0.3$. This is the potential used for the simulations shown in the introduction.

Figure 25 shows phase portraits of the moment map for the heat-bath model. The first-order map shows the expected three fixed points at the local extrema of the potential and exhibits the dynamics of bistable damped oscillators. The map has indeed stabilized the metastable states with a well-defined boundary (the stable manifold of the saddle) separating the basins of attraction. The location of the fixed points does not depend in any significant way on the number of particles in the bath or on the map time τ. The stable and unstable manifolds do vary somewhat, primarily with the map time τ, but are always qualitatively as seen in Figure 25.

For the second-order map we show phase portraits in the style of Figure 23 for the two-dimensional SDE system. We show two trajectories for each of three cases. The middle case is for $N = 8000$, the value used for the simulations shown in the introduction. The maximum oscillator frequency in the bath is $\omega_m \simeq 20$, and so we use a map time of $\tau = 0.4 > \pi/10$.

We show a map starting at $(\bar{x}_1, \bar{x}_2, \sigma_1, \sigma_2, c) = (0, 3, 0.01, 0.01, 0)$. This initial condition corresponds exactly to the initial density used Figure 3. One sees the similarity between the trajectory for the moment map and the evolution of the density in Figure 3. However, the "leaking" of mass into the left potential well observed in Figure 3 is prevented by the moment map via the mechanism illustrated in Figure 17. The moment map is nonlinear and has the two stable fixed points where the density is only metastable. Most trajectories evolve to one or the other of these fixed points.

For the same potential we have investigated the effect of N, the number of particles in the heat bath, on the moment map. We find that the map is quite insensitive to N, at least for $N \geq 1000$. For large N one can use a smaller map time τ, and the dynamics do depend weakly on τ. Specifically, the fixed points for the three cases in Figure 25 are very slightly different, though this cannot be seen on the scale of the figure. This difference is not due directly to N but to the value of τ.

5. Conclusions. In this paper we have introduced a mathematical framework intended to outline and clarify some aspects of the coarse-grained approach to analyzing stochastic and

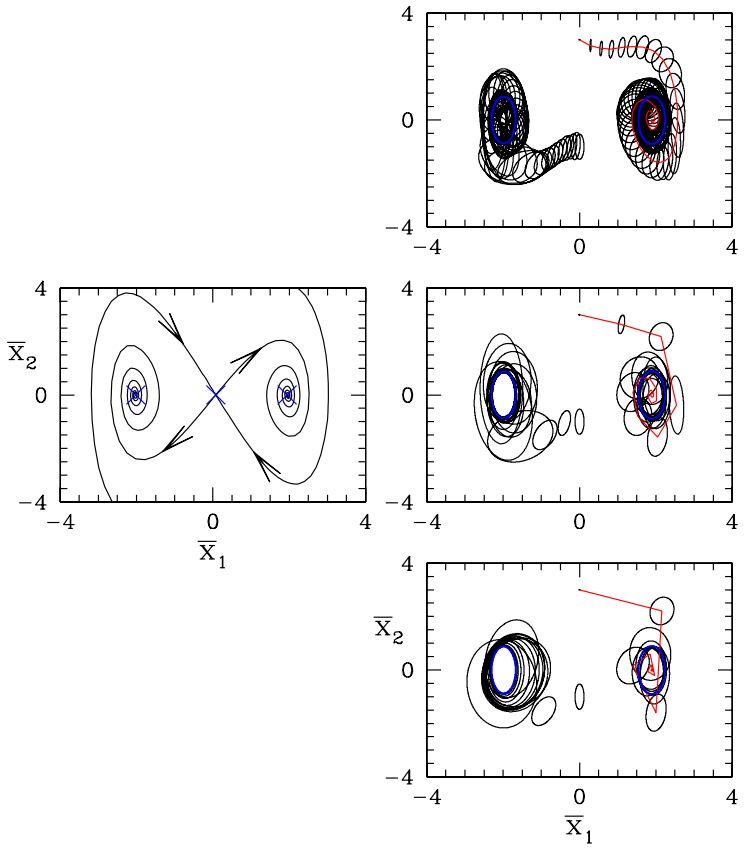


Figure 25. *Phase portraits showing the dynamics of moment maps for the heat bath model. First-order (left) and second-order (right) maps are shown. For the first-order map $N = 8000$ particles is used. For the second-order map the following are used: (top) $N = 256000$, $\tau = 0.1$; (middle) $N = 8000$, $\tau = 0.4$; (bottom) $N = 1000$, $\tau = 0.8$. Other parameters are $\mu = 4$ and $\nu = 0.3$.*

deterministic systems. In particular, we have given a precise definition of the moment map. These are maps on the (low-dimensional) space of low-order moments of probability measures. We have considered these maps both for stochastic systems and for deterministic systems with random initial data. While the underlying evolution of densities in both systems is linear, the moment maps are typically nonlinear. Our main focus has been understanding the origins of this nonlinearity.

In this paper we sought to establish a connection between the dynamics of coarse-grained observables (such as moments of evolving realization ensembles) and the nonlinear dynamics one expects at the deterministic limit. Contemporary estimation techniques would allow us to recover both the deterministic and the stochastic components of an effective *stochastic* model (e.g., [1]). Then, instead of using integral changes of coarse-grained observables, we could directly seek the extrema of an underlying effective potential, or even the extrema of the corresponding equilibrium density (see, e.g., [18, 27, 39]).

We have presented results for a number of model systems. We have first considered the

simple OU process, for which a full analysis is possible. Then, using a combination of numerical studies and analysis, we have considered in most detail a one-dimensional SDE with a double-well potential. This system provides the simplest example of a nonlinear moment map. In particular, this example shows how the moment map can stabilize metastable densities of the underlying linear flow. We have additionally presented numerical results for moment maps computed for two other systems with double-well potentials—a two-dimensional SDE and a deterministic ODE system with many degrees of freedom. Maps for both of these systems show the basic features found for the one-dimensional SDE, namely, nonlinearity and the stabilization of metastable densities.

One of the important issue that arises naturally in this computational framework is the importance of *the observer*. How long does a physical observer have to wait before declaring that an observed quantity is at steady state? This is clearly related to our testing the dependence of the map fixed points on the map reporting horizon. We also saw that the initialization of computational experiments (whether within or outside a well) can be vital in the existence of an effective reduced model (this is related, as we mentioned, to conditional averaging techniques).

REFERENCES

[1] Y. Ait-Sahalia, *Maximum-likelihood estimation of discretely sampled diffusions: A closed-form approximation approach*, Econometrica, 70 (2002), pp. 223–262.

[2] A. Angeletti, C. Castagnari, and F. Zirilli, *Asymptotic eigenvalue degeneracy for a class of one-dimensional Fokker-Planck operators*, J. Math. Phys., 26 (1985), pp. 678–691.

[3] A. Chorin, O. H. Hald, and R. Kupferman, *Optimal prediction and the Mori-Zwanzig representation of irreversible processes*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 2968–2973.

[4] J. Cisternas, C. W. Gear, S. Levin, and I. G. Kevrekidis, *Equation-free modeling of evolving diseases: Coarse-grained computations with individual-based models*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460 (2004), pp. 27621–2779; also available online at http://arXiv.org/abs/nlin.AO/0310011.

[5] P. Deuflhard, W. Huisinga, W. Fischer, and C. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra Appl., 315 (2000), pp. 39–59.

[6] A. Doucet, N. de Freitas, and N. J. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

[7] W. E, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.

[8] P. Ehrenfest, *Collected Scientific Papers*, North-Holland, Amsterdam, 1959.

[9] R. Erban, I. G. Kevrekidis, and H. G. Othmer, *An equation-free computational approach for extracting population-level behavior from individual-based models of biological dispersal*, Phys. D, 215 (2006), pp. 1–24.

[10] G. W. Ford and M. Kac, *On the quantum Langevin equation*, J. Statist. Phys., 46 (1987), pp. 803–810.

[11] C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed., Springer-Verlag, New York, Berlin, 1985.

[12] C. W. Gear, I. G. Kevrekidis, and K. Theodoropoulos, *"Coarse" integration/bifurcation analysis via microscopic simulators: Micro-Galerkin methods*, Comp. Chem. Engrg., 26 (2002), pp. 941–963.

[13] C. W. Gear and I. G. Kevrekidis, *Projective methods for stiff differential equations: Problems with gaps in their eigenvalue spectrum*, SIAM J. Sci. Comput., 24 (2003), pp. 1091–1106.

[14] C. W. Gear and I. G. Kevrekidis, *Telescopic projective integrators for stiff differential equations*, J. Comput. Phys., 187 (2003), pp. 95–109.

[15] C. W. Gear, *Projective Integration Methods for Distributions*, Technical Report NEC TR 2001130, NEC Institute, Princeton, NJ, 2001.

[16] D. Givon, R. Kupferman, and A. M. Stuart, *Extracting macroscopic dynamics: Model problems and algorithms*, Nonlinearity, 17 (2004), pp. R55–R127.

[17] A. N. Gorban, I. V. Karlin, H. C. Oettinger, and L. L. Tatarinova, *Ehrenfest's argument extended to a formalism of nonequilibrium thermodynamics*, Phys. Rev. E (3), 63 (2001), paper 066124.

[18] M. Haataja, D. Srolovitz, and I. G. Kevrekidis, *Apparent hysteresis in a driven system with self-organized drag*, Phys. Rev. Lett., 92 (2004), paper 160603; also available online at http://arXiv.org/abs/cond-mat/0310460.

[19] O. Hald, *Optimal prediction of the Klein-Gordon equation*, Proc. Natl. Acad. Sci. USA, 96 (1999), pp. 4774–4779.

[20] C. Hillermeier, N. Kunstmann, B. Rabus, and P. Tavan, *Topological feature maps with self-organized lateral connections: A population coded, one-layer model of associative memory*, Biol. Cybernet., 72 (1994), pp. 103–117.

[21] W. Huisinga, *Metastability of Markovian Systems*, Ph.D. thesis, Mathematics Department, Free University, Berlin, 2003; available online at http://page.mi.fu-berlin.de/~huisinga/publications/index.html.

[22] G. Hummer and I. G. Kevrekidis, *Coarse molecular dynamics of a peptide fragment: Free energy, kinetics and long time dynamics computations*, J. Chem. Phys., 118 (2003), pp. 10762–10773; also available online at http://arXiv.org/abs/physics/0212108.

[23] C. T. Kelley, I. G. Kevrekidis, and L. Qiao, *Newton-Krylov solvers for time-steppers*, SIAM J. Appl. Dyn. Syst., submitted; also available online at http://arXiv.org/abs/math.DS/0404374.

[24] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, Frontiers Appl. Math. 16, SIAM, Philadelphia, 1995.

[25] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and K. Theodoropoulos, *Equation-free coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level tasks*, Commun. Math. Sci., 1 (2003), pp. 715–762; original version available online at http://arXiv.org/abs/physics/0209043.

[26] I. G. Kevrekidis, C. W. Gear, and G. Hummer, *Equation-free: The computer-assisted analysis of complex, multiscale systems*, AIChE J., 50 (2004), pp. 1346–1354.

[27] D. I. Kopelevich, A. Z. Panagiotopoulos, and I. G. Kevrekidis, *Coarse-grained kinetic computations of rare events: Application to micelle formation*, J. Chem. Phys., 122 (2005), paper 044908; also available online at http://arXiv.org/abs/cond-mat/0407220.

[28] R. Kupferman, A. M. Stuart, J. Terry, and P. Tupper, *Long time behaviour of large mechanical systems with random initial data*, Stoch. Dyn., 2 (2002), pp. 533–562.

[29] A. Makeev, D. Maroudas, and I. G. Kevrekidis, *"Coarse" stability and bifurcation analysis using stochastic simulators: Kinetic Monte Carlo examples*, J. Chem. Phys., 116 (2002), pp. 10083–10091.

[30] A. G. Makeev, D. Maroudas, A. Z. Panagiotopoulos, and I. G. Kevrekidis, *Coarse bifurcation analysis of kinetic Monte Carlo simulations: A lattice gas model with lateral interactions*, J. Chem. Phys., 117 (2002), pp. 8229–8240.

[31] A. G. Makeev and I. G. Kevrekidis, *Equation-free multiscale computations for a lattice-gas model: Coarse-grained bifurcation analysis of the NO+CO reaction on Pt*(100), Chem. Eng. Sci., 59 (2004), pp. 1733–1743.

[32] B. J. Matkowsky and Z. Schuss, *Eigenvalues of the Fokker–Planck operator and the approach to equilibrium for diffusions in potential fields*, SIAM J. Appl. Math., 40 (1981), pp. 242–254.

[33] E. Nelson, *Dynamical Theories of Brownian Motion*; available online at http://www.math.princeton.edu/~nelson/books.html.

[34] G. Pavliotis and A. M. Stuart, *White noise limits for inertial particles in a random field*, Multiscale Model. Simul., 1 (2003), pp. 527–553.

[35] R. Rico-Martinez, C. W. Gear, and I. G. Kevrekidis, *Coarse projective kMC integration: Forward/reverse initial and boundary value problems*, J. Comput. Phys., 196 (2004), pp. 474–489; also available online at http://arXiv.org/abs/nlin.CG/0307016.

[36] C. Schuette, J. Walter, C. Hartmann, and W. Huisinga, *An averaging principle for fast degrees of freedom exhibiting long-term correlations*, Multiscale Model. Simul., 2 (2004), pp. 501–526.

[37] S. Setayeshgar, C. W. Gear, H. G. Othmer, and I. G. Kevrekidis, *Application of coarse integration to bacterial chemotaxis*, Multiscale Model. Simul., 4 (2005), pp. 307–327; also available online at http://arXiv.org/abs/physics/0308040.

[38] C. Siettos, M. D. Graham, and I. G. Kevrekidis, *Coarse Brownian dynamics for nematic liquid crystals: Bifurcation, projective integration and control via stochastic simulation*, J. Chem. Phys., 118 (2003), pp. 10149–10157; also available online at http://arXiv.org/abs/cond-mat/0211455.

[39] S. Sriraman, I. G. Kevrekidis, and G. Hummer, *Coarse nonlinear dynamics of filling-emptying transitions: Water in carbon nanotubes*, Phys. Rev. Lett., 95 (2005), paper 130603; also available online at http://arXiv.org/abs/cond-mat/0503491.

[40] K. Theodoropoulos, Y.-H. Qian, and I. G. Kevrekidis, *"Coarse" stability and bifurcation analysis using timesteppers: A reaction diffusion example*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 9840–9843.

[41] K. Theodoropoulos, K. Sankaranarayanan, S. Sundaresan, and I. G. Kevrekidis, *Coarse bifurcation studies of bubble flow lattice Boltzmann simulations*, Chem. Eng. Sci., 59 (2004), pp. 2357–2362; also available online at http://arXiv.org/abs/nlin.PS/0111040.

[42] E. Vanden Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1 (2003), pp. 377–384.

[43] R. Zwanzig, *Nonlinear generalized Langevin equations*, J. Statist. Phys., 9 (1973), pp. 215–220.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 3, pp. 435–475

Numerical Continuation of Symmetric Periodic Orbits*

Claudia Wulff† and Andreas Schebesch‡

Dedicated to the memory of our dear friend and colleague Karin Gatermann

Abstract. The bifurcation theory and numerics of periodic orbits of general dynamical systems is well developed, and in recent years there has been rapid progress in the development of a bifurcation theory for symmetric dynamical systems. However, there are hardly any results on the numerical computation of those bifurcations yet. In this paper we show how spatio-temporal symmetries of periodic orbits can be exploited numerically. We describe methods for the computation of symmetry breaking bifurcations of periodic orbits for free group actions and show how bifurcations increasing the spatio-temporal symmetry of periodic orbits (including period halving bifurcations and equivariant Hopf bifurcations) can be detected and computed numerically. Our pathfollowing algorithm is based on a multiple shooting algorithm for the numerical computation of periodic orbits via an adaptive Poincaré section and a tangential continuation method with implicit reparametrization.

Key words. numerical continuation, symmetry breaking bifurcations, symmetric periodic orbits

AMS subject classifications. 37G15, 37G40, 37M20, 65P30

DOI. 10.1137/050637170

1. Introduction.

The bifurcation theory and numerics of periodic orbits of general dynamical systems is well developed; see, e.g., [1, 8, 13, 16, 17]. Frequently the considered problems possess certain symmetries. Symmetries change the generic behavior of a dynamical system dramatically, and in recent years there has been rapid progress in the development of a bifurcation theory for periodic orbits of symmetric dynamical systems; see, e.g., [9, 11, 18, 19, 24]. However, there are hardly any results on the numerical computation of those bifurcations yet. Gatermann and Hohmann [10] developed numerical methods for the exploitation of symmetry and the computation of symmetry breaking and symmetry increasing bifurcations of stationary solutions, and implemented those methods in the mixed symbolic numerical code SYMCON. They treat finite symmetry groups. Cliffe, Spence, and Tavener [2] developed numerical methods for the computation of bifurcations of stationary solutions in the case of continuous rotational symmetries.

In this paper we start a systematic theory on numerical bifurcation of symmetric periodic orbits by extending the methods of Gatermann and Hohmann [10] to periodic solutions. We

*Received by the editors July 31, 2005; accepted for publication (in revised form) by J. Guckenheimer April 16, 2006; published electronically July 17, 2006.
http://www.siam.org/journals/siads/5-3/63717.html

†Department of Mathematics and Statistics, University of Surrey, Guildford GU2 7XH, United Kingdom (c.wulff@surrey.ac.uk). This author was partially supported by the Nuffield Foundation and by a European Union Marie Curie fellowship under contract number HPMF-CT-2000-00542.

‡Fachbereich Mathematik und Informatik, Freie Universität Berlin, 14195 Berlin, Germany (schebesc@inf.fu-berlin.de). This author was funded by a European Union Marie Curie studentship at the Marie-Curie Training Site "University of Warwick" during the academic year 2001/2002.

consider a parameter dependent dynamical system

$$\dot{x} = f(x, \lambda), \quad f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n, \tag{1.1}$$

which is equivariant with respect to a finite symmetry group $\Gamma \subset \mathrm{GL}(n)$, i.e.,

$$\gamma f(x, \lambda) = f(\gamma x, \lambda) \quad \forall\, \gamma \in \Gamma,\ x \in \mathbb{R}^n,\ \lambda \in \mathbb{R}.$$

In most parts of this paper we assume that Γ acts freely; i.e., $\gamma x \neq x$ for all $\gamma \neq \mathrm{id}$ and $x \in \mathbb{R}^n$. We numerically continue periodic solutions with respect to the parameter λ, exploiting possible symmetries and computing symmetry breaking and symmetry increasing bifurcations of those periodic orbits. For the computation of periodic orbits we employ the multiple shooting algorithm presented by Deuflhard [5], which we briefly recall in section 2.1. Section 2.2 is concerned with the exploitation of spatial and spatio-temporal symmetries of periodic orbits in the multiple shooting context. In section 2.4 the aspect of continuation is added, and the pathfollowing method of Deuflhard, Fiedler, and Kunkel [7] is used for the continuation of periodic solutions in the single and multiple shooting approach. A different method for the numerical continuation of periodic solutions with symmetry based on Fourier expansions has been presented by Dellnitz [4].

In sections 3 and 4 bifurcations of symmetric periodic orbits are treated. First, symmetry breaking and symmetry increasing bifurcations of periodic orbits are considered (section 3), and then equivariant Hopf bifurcations along periodic orbits (section 4). In these sections numerical techniques of Gatermann and Hohmann [10] are extended from stationary solutions to periodic solutions.

Generic symmetry breaking bifurcations of periodic orbits of free group actions correspond to period doubling bifurcations in the space of group orbits, or equivalently, to period doubling bifurcations of the symmetry reduced Poincaré map [18]. In section 3.4 we show how such symmetry breaking bifurcations can be detected and computed numerically by extending the corresponding methods for nonsymmetric systems; cf. section 3.1.

We have also developed methods for the computation of bifurcations which increase the spatio-temporal symmetry of the periodic solution, in particular an algorithm for the computation of period halving points, which is based on the methods for computing period doubling bifurcation points; see sections 3.2 and 3.5.

In section 4 we derive methods for the detection of equivariant Hopf points along branches of symmetric periodic orbits and present an extended system for the computation of equivariant Hopf points. The main problem is how to deal with multiple Hopf eigenvalues forced by symmetry. The issue of numerically dealing with multiple critical eigenvalues was treated by Cliffe, Spence, and Tavener [2] in the case of continuation of stationary solutions with continuous rotation symmetry.

The numerical methods we present have been implemented in the C code SYMPERCON [20]. In section 5, examples are presented to illustrate the performance of the developed algorithmic tools. In subsection 5.1 we use SYMPERCON, AUTO [8], and CONTENT [17] for the computation of the period doubling cascade of the Lorenz equations to demonstrate the better performance of SYMPERCON. In section 5.2 we compute symmetry breaking bifurcations of periodic orbits of four coupled cells. In section 5.3 we show how oscillations of an electric circuit can be computed efficiently by exploiting their spatio-temporal symmetry.

We partly follow the unpublished manuscripts [20, 22, 23]. For a description of the program SYMPERCON see [22, 20].

2. Continuation of symmetric periodic orbits. In this section we review the multiple shooting method of Deuflhard [5] for the computation of periodic orbits, show how symmetries of periodic orbits can be exploited within the multiple shooting approach, and present a continuation method for symmetric periodic orbits based on the Gauss–Newton method.

2.1. Computation of single periodic orbits—Single shooting method. In this subsection we briefly recollect the algorithm for the computation of periodic orbits of an autonomous ordinary differential equation (ODE)

$$\dot{x} = f(x), \qquad f : \mathbb{R}^n \to \mathbb{R}^n, \tag{2.1}$$

which was introduced in [5].

Let $\Phi_t(\cdot)$ be the flow of (2.1), and let $x(t) = \Phi_t(x^*)$ be a periodic solution of (2.1) of period T^*, i.e., $x(T^*) = x^*$. Then any time shifted solution $x(t+t_0)$, $t_0 \in \mathbb{R}$, is also a periodic solution, because the system (2.1) is autonomous. All these solutions determine the same periodic orbit

$$\mathcal{P} = \mathcal{P}_{x(\cdot)} = \{x(t), t \in \mathbb{R}\}.$$

In order to avoid this nonuniqueness a well-known analytical technique is to fix a Poincaré section $S = S_{x^*}$, which is an $(n-1)$-dimensional affine hyperplane *transversal* to the periodic orbit $\mathcal{P}$ at the point x^*; see, e.g., [16]. Let us use the Poincaré section orthogonal to the orbit

$$S = S_{x^*} = x^* + S'_{x^*}, \quad \text{where} \quad S'_{x^*} = \operatorname{span}(f(x^*))^{\perp}.$$

Then x^* is a fixed point of the Poincaré map (first return map) $\Pi : S \to S$.

Definition 2.1. *We say that a periodic orbit $\mathcal{P}$ with period T^* is* nondegenerate *if*

$$\mathrm{D}_x \Pi(x)|_{x=x^*} - \mathrm{id}$$

is regular for $x^* \in \mathcal{P}$.

In this case x^* is a locally unique fixed point of the Poincaré map Π and a locally unique root of the equation

$$\mathcal{F}(x) := \Pi(x) - x = 0, \quad \text{where} \quad \mathcal{F} : S \to S.$$

Numerically one can either fix the phase by an additional phase condition, as described, e.g., in [1, 16], or solve an underdetermined equation, as in [5]. We follow the latter approach and compute a point $x = x^*$ on the periodic solution together with its period $T = T^*$ by solving the underdetermined equation $F(y) = 0$. Here $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is given by

$$F(x,T) = \Phi_T(x) - x = 0, \quad \text{where} \quad y = (x,T). \tag{2.2}$$

We solve (2.2) by an underdetermined Gauss–Newton method:

$$\begin{aligned} \Delta y^k &= -\mathrm{D}F(y^k)^+ F(y^k), \\ y^{k+1} &= y^k + \Delta y^k, \end{aligned} \tag{2.3}$$

where $DF(y^k)^+$ denotes the Moore–Penrose pseudoinverse of $DF(y^k)$. Remember that for $A \in \mathrm{Mat}(m,n)$, $m \le n$, rank $A = m$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $x = A^+ b$ is defined by

$$Ax = b, \quad x \perp \ker(A),$$

where $\ker(A)$ is the kernel of A. Thus $x = A^+ b$ is the smallest in norm solution of $Ax = b$, and hence the Newton correction Δy^k is the smallest solution of the underdetermined linear system in (2.3).

The Jacobian $DF(x,T)$ of (2.2) in the solution (x^*, T^*) is given by

$$DF(x^*,T^*) = [-\mathrm{id} + D_x\Phi_{T^*}(x^*), f(\Phi_{T^*}(x^*))] = [-\mathrm{id} + D_x\Phi_{T^*}(x^*), f(x^*)]. \tag{2.4}$$

Therefore a kernel vector t^f of $DF(y^*)$ at the solution point $y^* = (x^*, T^*)$ is the tangent $t^f = (f(x^*), 0)$ to the trajectory.

Remark 2.2. This approach can be interpreted as computing periodic orbits in an *adaptive Poincaré section*, which is *approximately orthogonal* to the periodic orbit. Since for the kernel vectors $t^k = (t_x^k, t_T^k)$ of $DF(y^k)$ we have $t^k \to t^f$ as $k \to \infty$, the Gauss–Newton iterate $x^{k+1} = x^k + \Delta x^k$ lies in the adaptive Poincaré section $S_{x^k} = x^k + \mathrm{span}(t_x^k)\perp \approx x^* + \mathrm{span}(f(x^*))^\perp$.

If x^* lies on a nondegenerate periodic orbit, i.e., if $D_x\Pi(x^*) - \mathrm{id}$ is regular, then by (2.4) the Jacobian $DF(x^*,T^*)$ is regular. Since this condition does not depend on the chosen point x^* on $\mathcal{P}$ we get the following convergence result.

Proposition 2.3. *If the periodic orbit $\mathcal{P}$ through $x^* = \Phi_{T^*}(x^*)$ is nondegenerate, then there is a tubular neighborhood U of the periodic orbit $\mathcal{P}$ where there is no other periodic solution with period near T^* and which is such that the Gauss–Newton method (2.3) applied to (2.2) converges for initial data $\hat{x} \in U$ and $\hat{T} \approx T^*$.*

Before we review the extension of this basic shooting method to the multiple shooting context we show how symmetries of periodic orbits can be exploited numerically.

2.2. Symmetries of periodic orbits and how to exploit them. Let $\Gamma \subseteq \mathrm{GL}(n)$ be a finite group, and let f be Γ-equivariant [11], i.e.,

$$f(\gamma x) = \gamma f(x) \quad \forall\, x \in \mathbb{R}^n, \gamma \in \Gamma. \tag{2.5}$$

This condition on the vectorfield (2.1) implies that if $x(t)$ is a solution of the dynamical system (2.1), then $\gamma\, x(t)$ is also a solution. Hence the flow $\Phi_t(\cdot)$ of (2.1) is also Γ-equivariant: $\gamma\Phi_t(x_0) = \Phi_t(\gamma x_0)$ for every $\gamma \in \Gamma$, $x_0 \in \mathbb{R}^n$.

For any $x \in \mathbb{R}^n$ the element $\gamma\, x$ is called *conjugate* to x [11]. An element $\gamma \in \Gamma$ is called a *symmetry* of $x \in \mathbb{R}^n$ if $\gamma x = x$; the set of all symmetries of x (*isotropy subgroup* of x) is given by $K = \Gamma_x = \{\gamma \in \Gamma \mid \gamma x = x\}$. It can be seen easily that the vectorfield f of (2.1) maps the fixed point space of K

$$\mathrm{Fix}(K) = \{x \in \mathbb{R}^n \mid \gamma x = x \;\forall\, \gamma \in K\}$$

into itself. Thus we can restrict the ODE (2.1) to the fixed point space $\mathrm{Fix}(K) \simeq \mathbb{R}^{n_{\mathrm{red}}}$, which has a lower dimension $n_{\mathrm{red}} \le n$. In this way we obtain a symmetry reduced system $f_{\mathrm{red}} : \mathbb{R}^{n_{\mathrm{red}}} \to \mathbb{R}^{n_{\mathrm{red}}}$ which can be computed symbolically (see Gatermann and Hohmann [10]). The symmetry group acting on $\mathrm{Fix}(K)$ is $N(K)/K$, where $N(K)$ is the normalizer of K.

Remark 2.4 (see [11]). If a finite group (or, more generally, a compact group) Γ acts linearly on the phase space $X = \mathbb{R}^n$, i.e., if

$$\gamma x := \vartheta(\gamma)x, \quad \gamma \in \Gamma, x \in X,$$

where $\vartheta : \Gamma \to \mathrm{GL}(n)$ is a homomorphism, then the phase space $X = \mathbb{R}^n$ can be decomposed into a sum $\mathbb{R}^n = X_1 \oplus X_2 \oplus \cdots \oplus X_l$, where the X_i are Γ-invariant vector spaces and cannot be decomposed into smaller Γ-invariant subspaces. Such vectorspaces X_i are called Γ-*irreducible*, and their corresponding reduced group actions $\vartheta_i := \vartheta|_{X_i}$ are called *irreducible representations* of the action of Γ on $\mathbb{R}^n$. If a Γ-irreducible subspace X is also irreducible as a vectorspace over $\mathbb{C}$, then its irreducible representation $\vartheta|_X$ is called an *absolutely irreducible representation*. If it is reducible over $\mathbb{C}$, it is called a *complex irreducible representation*. We will encounter the concept of irreducible representations in the computation of bifurcations; cf. section 4.2.

The *spatial symmetries* K of periodic solutions $x(t)$ are those group elements $\gamma \in \Gamma$ which leave each point on the periodic orbit invariant:

$$K := \Gamma_{x(t)} = \{\gamma \in \Gamma \mid \gamma x(t) = x(t) \; \forall\, t\}.$$

Since the flow Φ_t is Γ-equivariant the set of spatial symmetries K of a periodic solution $x(t)$ does not depend on the time t. In addition to spatial symmetries there are also *spatio-temporal symmetries* which leave the periodic orbit $\mathcal{P} := \mathcal{P}_{x(\cdot)}$ invariant as a whole but not pointwise; i.e., the spatio-temporal symmetries of a periodic orbit $\mathcal{P}$ are given by

$$L := \{\gamma \in \Gamma \mid \gamma\mathcal{P} = \mathcal{P}\}.$$

Each $\gamma \in L$ corresponds to a phase shift $\Theta(\gamma)\, T^*$ of the T^*-periodic solution $x(t)$:

$$\gamma \in L \Rightarrow x(t) = \gamma x(t + \Theta(\gamma)\, T^*), \text{ where } \Theta(\gamma) \in S^1 \simeq \mathbb{R}/\mathbb{Z}. \tag{2.6}$$

So spatio-temporal symmetries come in pairs $(\gamma, \Theta(\gamma)) \in \Gamma \times S^1$. We define an action of the spatio-temporal symmetry group $\Gamma \times S^1$, where $S^1 = \mathbb{R}/\mathbb{Z}$, on T^*-periodic solutions $x(t)$ of (2.1) as follows:

$$((\gamma, \theta)x)(t) := \gamma x(t + \theta T^*), \quad (\gamma, \theta) \in \Gamma \times S^1. \tag{2.7}$$

Note that $\Theta : L \to \mathbb{R}/\mathbb{Z}$ is a group homomorphism with the spatial symmetries K as kernel and that

$$L/K \equiv \mathbb{Z}_\ell, \quad \ell \in \mathbb{N}; \tag{2.8}$$

see [11]. The spatial symmetries of periodic solutions can be exploited by restriction onto the fixed point space $\mathrm{Fix}(K)$, i.e., by using a symmetry reduced system $f_{\mathrm{red}} : \mathrm{Fix}(K) \to \mathrm{Fix}(K)$.

From now on we assume that the spatial symmetry K of the periodic orbit is trivial. Then the spatio-temporal symmetries of the periodic orbit form a finite cyclic group $L = \mathbb{Z}_\ell$. In bifurcation theory the spatio-temporal symmetries of periodic orbits are taken into account by studying the *reduced Poincaré map*. It was first introduced by Fiedler [9] and later used by Lamb and Melbourne [18] and Lamb, Melbourne, and Wulff [19] in order to classify symmetry

breaking bifurcations of periodic orbits; see also section 3. Let $\alpha \in L = \mathbb{Z}_\ell$ be that element in L that corresponds to the smallest possible nonzero phase shift T^*/ℓ:

$$\alpha\, x\left(t + \frac{T^*}{\ell}\right) = x(t) \quad \forall\, t. \tag{2.9}$$

We call this spatio-temporal symmetry the *drift symmetry* of the periodic orbit [25]. For $x^* \in \mathcal{P}$ define the Poincaré section as usual by $S = x^* + \operatorname{span}(f(x^*))^\perp$. Then the *reduced Poincaré map* is defined as

$$\Pi_{\mathrm{red}} = \alpha\hat{\Pi}, \quad \hat{\Pi} : S \to \alpha^{-1}\, S, \tag{2.10}$$

where α is the drift symmetry of the periodic orbit, i.e., satisfies (2.9), and $\hat{\Pi}$ maps $x \in S$ into the point where the positive semiflow through x first hits $\alpha^{-1}\, S$ [9]. The fixed point equation $\Pi_{\mathrm{red}}(x) = x$ or, equivalently, the equation

$$\mathcal{F}(x) = \Pi_{\mathrm{red}}(x) - x = 0, \quad \text{where} \quad \mathcal{F} : S \to S,$$

then determines periodic orbits with spatio-temporal symmetry L. Note that in the case of trivial symmetry $\ell = 1$, $\alpha = \mathrm{id}$, the reduced Poincaré map Π_{red} becomes the standard Poincaré map Π introduced in section 2.1. In order to numerically exploit spatio-temporal symmetries we proceed as in section 2.1: Each point x on a T-periodic orbit with drift symmetry α satisfies the underdetermined equation

$$F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n, \quad F(x, T) = \alpha\Phi_{\frac{T}{\ell}}(x) - x = 0. \tag{2.11}$$

This system is analogous to the corresponding underdetermined system (2.2) in the case of trivial symmetry and reduces to it in the case $\alpha = \mathrm{id}$, $\ell = 1$. It can also be solved by a Gauss–Newton method. Note that it suffices to compute the flow $\Phi_t(\cdot)$ and the Wronskian matrix $\mathrm{D}_x\Phi_t(\cdot)$ only up to time $\frac{T}{\ell}$ instead of T, which is a remarkable reduction of the computational cost in the case of high spatio-temporal symmetry. In a solution point (x^*, T^*) we have

$$\mathrm{D}F(x^*, T^*) = \left[\alpha \mathrm{D}_x\Phi_{\frac{T^*}{\ell}}(x^*) - \mathrm{id}, \frac{1}{\ell} f(\alpha\Phi_{\frac{T^*}{\ell}}(x^*))\right] = \left[\alpha \mathrm{D}_x\Phi_{\frac{T^*}{\ell}}(x^*) - \mathrm{id}, \frac{1}{\ell} f(x^*)\right], \tag{2.12}$$

analogous to the case of trivial symmetry; cf. (2.4). In particular, $\mathrm{D}F(x^*, T^*)$ is closely related to $\mathrm{D}_x\mathcal{F}(x^*) = \mathrm{D}_x\Pi_{\mathrm{red}}(x)|_{x=x^*} - \mathrm{id}$. We extend the definition of nondegeneracy to symmetric periodic orbits as follows.

Definition 2.5. *We say that a symmetric periodic orbit $\mathcal{P}$ with drift symmetry α is* nondegenerate *if $\mathrm{D}\Pi_{\mathrm{red}}(x^*) - \mathrm{id}$ is regular for $x^* \in \mathcal{P}$, where Π_{red} is from (2.10).*

From (2.12) we conclude that $\mathrm{D}F(x^*, T^*)$ in the periodic orbit is regular if and only if the periodic orbit is nondegenerate, and so we get the following proposition, which is analogous to Proposition 2.3.

Proposition 2.6. *If the symmetric periodic orbit through $x^* = \alpha\Phi_{\frac{T^*}{\ell}}(x^*)$ is nondegenerate in the sense of Definition 2.5, then there is a tubular neighborhood U about the periodic orbit $\mathcal{P}$ through x^* with the property that there is no other periodic orbit with symmetry L and period near T^* in U and which is such that the Gauss–Newton method (2.3) applied to (2.11) converges for initial values $\hat{x} \in U$, $\hat{T} \approx T^*$.*

2.3. Multiple shooting approach. In order to numerically compute unstable symmetric periodic solutions we use the just described algorithm in the *multiple* shooting context (cf. [5]): We compute k points on a periodic orbit with spatio-temporal symmetry $L = \mathbb{Z}_\ell$, trivial isotropy, and drift symmetry α by solving the underdetermined equation

$$F(x_1, \dots, x_k, T) = 0, \quad F : \mathbb{R}^N \to \mathbb{R}^M, \tag{2.13}$$

where $N = M + 1 = kn + 1$, $0 = s_1 < \cdots < s_{k+1} = 1$ is a partition of the unit interval, $\Delta s_i = s_{i+1} - s_i$ for $i = 1, \dots k$, and

$$F_i(x_1, \dots, x_k, T) = \begin{cases} \Phi_{\frac{\Delta s_i T}{\ell}}(x_i) - x_{i+1} & \text{for} \quad i = 1, \dots, k-1, \\ \alpha \Phi_{\frac{\Delta s_k T}{\ell}}(x_k) - x_1 & \text{for} \quad i = k. \end{cases} \tag{2.14}$$

The linear systems which arise in the Gauss–Newton method are of the form $Jy = b$, where $y = (x, T) \in \mathbb{R}^{nk+1}$, $x = (x_1, \dots, x_k)$, $b = (b_1, \dots, b_k)$,

$$J = \mathrm{D}F(x,T) = \begin{pmatrix} G_1 & -\mathrm{id} & & & & g_1 \\ & G_2 & -\mathrm{id} & & & g_2 \\ & & \ddots & \ddots & & \vdots \\ & & & G_{k-1} & -\mathrm{id} & g_{k-1} \\ -\mathrm{id} & & & & G_k & g_k \end{pmatrix} = [G, g], \tag{2.15}$$

where G is an (nk, nk)-matrix, g an nk-vector, and

$$\begin{aligned} G_i &= \mathrm{D}_x \Phi_{\frac{\Delta s_i T}{\ell}}(x_i), \quad i = 1, 2, \dots, k-1, \\ G_k &= \alpha \mathrm{D}_x \Phi_{\frac{\Delta s_k T}{\ell}}(x_k), \\ g_i &= \mathrm{D}_T F_i(x,T) = \mathrm{D}_T \Phi_{\frac{\Delta s_i T}{\ell}}(x_i) = \tfrac{\Delta s_i}{\ell} f(\Phi_{\frac{\Delta s_i T}{\ell}}(x_i)), \quad i = 1, \dots, k-1, \\ g_k &= \tfrac{\Delta s_k}{\ell} \alpha f(\Phi_{\frac{\Delta s_k T}{\ell}}(x_k)). \end{aligned}$$

We have

$$Jy = b \;\Leftrightarrow\; [G, g] \begin{pmatrix} x \\ T \end{pmatrix} = b \;\Leftrightarrow\; Gx = b - gT,$$

so we can apply Gaussian block elimination to G to solve these linear systems. This yields the following algorithm:

1. Compute the condensed right-hand side

$$b_c := C(G, b, k) = b_k + G_k b_{k-1} + \cdots + G_k \cdots G_2 b_1.$$

2. Compute the condensed matrix $[G_c, g_c]$ with

$$G_c := G_k \cdots G_1, \qquad g_c := C(G, g, k). \tag{2.16}$$

3. Compute a solution of the condensed system $[G_c - \mathrm{id}, g_c]\binom{x_1}{T} = b_c$, e.g.,

$$\begin{pmatrix} x_1 \\ T \end{pmatrix} = [G_c - \mathrm{id}, g_c]^+ b_c,$$

using QR-decomposition.

4. Compute x via the explicit recursion

$$x_i = G_{i-1}x_{i-1} - b_{i-1} + g_{i-1}T \quad \text{for} \quad i = 2, \dots, k. \tag{2.17}$$

We have now obtained a solution $y = J^- b$, where $y = (x, T)$ and J^- is an outer inverse of J. In order to compute the solution $J^+ b$, where J^+ is the Moore–Penrose pseudoinverse of J, we have to add one more step:

5. Compute the kernel vector $t = (t_x, t_T)$ of J, where $t_x = (t_1, t_2, \dots, t_k)$. Starting from a tangent of the condensed system $\binom{t_1}{t_T}$,

$$[G_c - \mathrm{id}, g_c]\begin{pmatrix} t_1 \\ t_T \end{pmatrix} = 0,$$

we obtain a tangent t of the whole system by

$$t_i = G_{i-1}t_{i-1} + g_{i-1}t_T \quad \text{for} \quad i = 2, \dots, k.$$

In the end we project $y \to y - \frac{\langle y,t\rangle}{\langle t,t\rangle}t$.

An easy computation shows that in a solution point $y^* = (x^*, T^*)$ we have

$$[G_c - \mathrm{id}, g_c] = \left[\alpha \mathrm{D}_x \Phi_{\frac{T^*}{\ell}}(x_1^*) - \mathrm{id}, \frac{1}{\ell} f(x_1^*)\right], \tag{2.18}$$

and so the condensed matrix $E_c := [G_c - \mathrm{id}, g_c]$ equals the Jacobian (2.12) of the single shooting approach in the point x_1^*. The Jacobian J is regular if and only if $[G_c - \mathrm{id}, g_c]$ is regular. Hence we get the following result, which is analogous to Proposition 2.6.

Theorem 2.7. *The Jacobian J of the multiple shooting system (2.13) is regular at the symmetric periodic orbit if and only if the periodic orbit is nondegenerate in the sense of Definition 2.5. In this case the Gauss–Newton method (2.3) applied to (2.13) converges for sufficiently good initial data.*

Remarks 2.8.

(a) The approach for symmetry exploitation in the multiple shooting approach can be transferred to collocation methods (used in AUTO and CONTENT [8, 17]) since collocation can be viewed as a special case of a multiple shooting method where the number of grid points corresponds to the number of multiple shooting points k and the initial value problem solver consists of only one integration step of a collocation method. The advantage of the multiple shooting approach is that it allows the use of adaptive order and stepsize initial value problem solvers for the computation of the flow $\Phi_t(x)$ and the Wronskians $\mathrm{D}_x\Phi_t(x)$, which we use for the evaluation of F and $\mathrm{D}F$, respectively, like the extrapolation codes of Deufhard [6]. These techniques have been implemented in the code SYMPERCON [22, 20]; see section 5.1 for a comparison with AUTO and CONTENT.

(b) Note that in the program packages AUTO and CONTENT [8, 17] a phase condition is used to obtain a unique periodic orbit, whereas we solve an underdetermined equation for the periodic orbit. Since we use the Moore–Penrose pseudoinverse to compute the corrections of the Gauss–Newton method $\Delta y^k = -\mathrm{D}F(y^k)^+ F(y^k)$, we have $\Delta y^k \perp \ker(\mathrm{D}F(y^k))$. Here

$$\ker(\mathrm{D}F(y^k)) \approx \ker(\mathrm{D}F(y^*)) = \mathrm{span}(t^f),$$

where

$$t^f = (f(x_1^*), \ldots, f(x_k^*), 0, 0). \tag{2.19}$$

The condition $\Delta y^k \perp \ker(\mathrm{D}F(y^k))$ is therefore an "adaptive phase condition" which is such that the correction Δy^k is the smallest solution of the equation for the Newton corrections $\mathrm{D}F(y^k)\Delta y^k = -F(y^k)$; cf. Remark 2.2.

(c) Dellnitz [3, 4] computes symmetric periodic orbits by a Galerkin ansatz based on Fourier modes. This method is effective near Hopf bifurcations, since in this situation periodic orbits can be approximated by few Fourier modes.

2.4. Continuation of nondegenerate periodic orbits. In this section we show how the pathfollowing method for stationary solutions described in [7] can be extended to the case of symmetric periodic solutions. We consider the parameter dependent Γ-equivariant dynamical system $\dot{x} = f(x, \lambda)$ from (1.1) again and first look at stationary solutions of (1.1),

$$f(y) = 0, \quad f : \mathbb{R}^{n+1} \to \mathbb{R}^n, \quad y = (x, \lambda). \tag{2.20}$$

If $y^* = (x^*, \lambda^*)$ is a stationary solution and $\mathrm{D}_y f(y)$ is regular at y^*, then (2.20) locally defines a solution branch. We apply the tangential continuation method based on implicit reparametrization presented in [7] to compute this solution branch. By writing $y = (x, \lambda)$ we want to express that the parameter λ does not play any extraordinary role, so that turning points can be treated easily. The pathfollowing algorithm works as follows: If a solution y^* is given, a new guess $\hat{y}$ is computed by setting $\hat{y} = y^* + \epsilon\, t(y^*)$, where $t(y)$ is the normalized kernel vector of $\mathrm{D}_y f(y)$ and hence the continuation tangent, and ϵ is a suitably chosen stepsize. Then an underdetermined Gauss–Newton method as in (2.3) is used for the iteration from the guess $\hat{y}$ back to the solution path. The stepsize control is described in [7]. We now show how to apply this continuation method to symmetric periodic orbits.

2.4.1. Single shooting method. In the case of symmetric periodic solutions we want to compute fixed points of the parameter dependent reduced Poincaré map $\Pi_{\mathrm{red}} : S \times \mathbb{R} \to S$,

$$\Pi_{\mathrm{red}}(x, \lambda) = x \quad \Leftrightarrow \quad \mathcal{F}(x, \lambda) := \Pi_{\mathrm{red}}(x, \lambda) - x = 0, \tag{2.21}$$

or, equivalently, solutions of the parameter dependent nonlinear equation

$$\mathcal{F}(x, \lambda) = 0, \quad \mathcal{F} : S \times \mathbb{R} \to S.$$

We can in principle apply the above described continuation method to this equation. The continuation tangent in a solution point (x^*, λ^*) is simply the kernel vector $t_{\mathcal{F}}^*$ of $\mathrm{D}\mathcal{F}(x^*, \lambda^*)$.

However, in the numerical realization of this idea we want to use the method of adaptive Poincaré sections of sections 2.1 and 2.2; cf. Remark 2.2. So we again introduce the period as a new variable and solve (in the single shooting approach) the underdetermined equation

$$F : \mathbb{R}^{n+2} \to \mathbb{R}^n, \quad F(x, T, \lambda) = \alpha \Phi_{\frac{T}{\ell}}(x, \lambda) - x = 0 \tag{2.22}$$

by a Gauss–Newton procedure. Now the kernel $\ker(\mathrm{D}F)$ of $\mathrm{D}F$ is two-dimensional. In a solution point (x^*, λ^*) one kernel vector of $\mathrm{D}F$ is the tangent to the periodic orbit

$$t^f = (f(x^*, \lambda^*), 0, 0) \in \ker(\mathrm{D}F(x^*, T^*, \lambda^*)).$$

We want to determine the continuation tangent t^* in such a way that it corresponds to the theoretical tangent vector $t^*_{\mathcal{F}}$. First the continuation tangent has to be in the kernel of $\mathrm{D}F$, and second the continuation tangent should lie in the Poincaré section S. Since we choose the Poincaré section orthogonal to the orbit, this leads to the conditions

$$t^* \in \ker(\mathrm{D}F), \quad t^* \perp t^f. \tag{2.23}$$

The Jacobian in the solution point (x^*, T^*, λ^*) is given by

$$\mathrm{D}F(x^*, T^*, \lambda^*) = \left[\alpha \mathrm{D}_x \Phi_{\frac{T^*}{\ell}}(x^*, \lambda^*) - \mathrm{id}, \frac{1}{\ell} f(x^*, \lambda^*), \alpha \mathrm{D}_\lambda \Phi_{\frac{T^*}{\ell}}(x^*, \lambda^*)\right], \tag{2.24}$$

and therefore we get the following proposition, analogous to Proposition 2.6.

Proposition 2.9. *Let x^* lie on a T^*-periodic orbit $\mathcal{P}$ of (1.1) with symmetry $L = \mathbb{Z}_\ell$ and drift symmetry α. Then the Jacobian $\mathrm{D}F(x^*, T^*, \lambda^*)$ is regular if and only if*

$$\mathrm{D}\mathcal{F}(x^*, \lambda^*) = [\mathrm{D}_x \mathcal{F}(x^*, \lambda^*), \mathrm{D}_\lambda \mathcal{F}(x^*, \lambda^*)] = [\mathrm{D}_x \Pi_{\mathrm{red}}(x^*) - \mathrm{id}, \mathrm{D}_\lambda \Pi_{\mathrm{red}}(x^*)] \tag{2.25}$$

is regular. In this case the path of periodic solutions given by (2.21) is locally unique in the following sense: There is a tubular neighborhood U of the periodic orbit $\mathcal{P}$ such that every periodic solution $\hat{x}(t)$ in this neighborhood with period $\hat{T}$ close to T^*, parameter $\hat{\lambda}$ close to λ^*, and drift symmetry α lies on the path of periodic solutions defined by (2.21). Moreover, the Gauss–Newton method (2.3) applied to (2.22) converges to a periodic solution on this path for initial data $\hat{x} \in U$, $\hat{T} \approx T^*$, $\hat{\lambda} \approx \lambda^*$.

Note that for nondegenerate periodic orbits the condition of Proposition 2.9 is always satisfied, but it also holds in a turning point bifurcation; see section 2.5.

2.4.2. Multiple shooting ansatz. In the multiple shooting approach we solve the parameter dependent equation

$$F(x_1, \ldots, x_k, T, \lambda) = 0, \quad F : \mathbb{R}^{nk} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^{nk}, \tag{2.26}$$

where F is as in (2.14). Nearly everything carries over from section 2.3; we just have one more column in the Jacobian consisting of the parameter derivatives

$$P_i = \mathrm{D}_\lambda \Phi_{\frac{\Delta s_i T}{\ell}}(x_i, \lambda), \quad i = 1, \ldots, k-1, \quad P_k = \alpha \mathrm{D}_\lambda \Phi_{\frac{\Delta s_k T}{\ell}}(x_k, \lambda).$$

Therefore, to solve the linear equations $Jy = b$, $y = (x_1, \dots, x_k, T, \lambda)$, we have to compute an additional condensed vector (step 1 in the Gaussian block elimination algorithm), namely

$$p_c := C(P) = P_k + G_k P_{k-1} + \cdots + G_k \cdots G_2 P_1.$$

The condensed matrix in steps 2 and 3 is of the form $[G_c - \mathrm{id}, g_c, p_c]$, the recursion (step 4) has to be modified to

$$x_i = G_{i-1} x_{i-1} - b_{i-1} + g_{i-1} T + P_{i-1} \lambda \quad \text{for} \quad i = 2, \dots, k,$$

and in step 5 we compute an orthonormal basis of the two-dimensional kernel of J and project the preliminary solutions $y = J^- b$ onto the orthogonal complement of this kernel.

As can be seen from the Gaussian block elimination, J has full rank if the condensed matrix $E_c := [G_c - \mathrm{id}, g_c, p_c]$ has full rank. A simple computation shows that the matrix E_c equals the Jacobian of the single shooting approach (2.24). Thus the Gauss–Newton method (2.3) applied to the multiple shooting system of equations (2.26) converges to a solution under the same conditions as the Gauss–Newton method of the single shooting method, namely if the assumptions of Proposition 2.9 are satisfied.

As in the case of the single shooting method, we choose the continuation tangent t^* as the kernel vector of $\mathrm{D}F(y^*)$ (with F from (2.26)), which is orthogonal to t^f from (2.19).

2.5. Turning points. Before we come to the detection and computation of bifurcations of symmetric periodic orbits in sections 3 and 4 we first consider turning points of symmetric periodic orbits. We saw that periodic orbits are solutions of the equation $\mathcal{F}(x, \lambda) = 0$, where $\mathcal{F}$ is as in (2.21). Turning points are characterized by the condition that $\mathrm{D}_x \mathcal{F}(x^*, \lambda^*)$ from (2.25) is singular, but that $\mathrm{D}\mathcal{F}(x^*, \lambda^*)$ has full rank. In this case the solution path $(x(s), \lambda(s))$, of (2.21), $s \in \mathbb{R}$, $x(0) = x^*$, $\lambda(0) = \lambda^*$, satisfies $\lambda'(0) = 0$. Generically $\lambda''(0) \neq 0$ so that the solution path has a turning point in λ. Turning points can be detected by a change of sign of the λ-component t^*_λ of the continuation tangent t^* of the periodic solution, provided that this test is done *after* the tests of other bifurcations. This ordering of the monitoring tests for bifurcations is important, because, as we will see in sections 3.2, 3.5, and 4, a change of sign of the λ-component t^*_λ of the continuation tangent t^* also occurs at flip up bifurcations and Hopf points along periodic orbits.

A turning point between two points $y^{(0)} = (x^{(0)}, T^{(0)}, \lambda^{(0)})$ and $y^{(1)} = (x^{(1)}, T^{(1)}, \lambda^{(1)})$ with continuation tangents $t^{(0)}$ and $t^{(1)}$, respectively, is detected if

$$t^{(0)}_\lambda t^{(1)}_\lambda < 0.$$

The λ-component of the turning point is then computed by Hermite interpolation in exactly the same way as in the case of stationary solutions; see [7]. We construct a cubic polynomial $y(\tau) = (x(\tau), T(\tau), \lambda(\tau))$, where $y : [0, 1] \to \mathbb{R}^N$, over the line $y^{(0)} + \tau(y^{(1)} - y^{(0)})$, $\tau \in [0, 1]$, such that

$$\begin{aligned} y(0) &= y^{(0)}, \qquad y'(0) = \tfrac{\|y^{(1)} - y^{(0)}\|^2}{\langle y^{(1)} - y^{(0)}, t^{(0)} \rangle} t^{(0)}, \\ y(1) &= y^{(1)}, \qquad y'(1) = \tfrac{\|y^{(1)} - y^{(0)}\|^2}{\langle y^{(1)} - y^{(0)}, t^{(1)} \rangle} t^{(1)}, \end{aligned} \tag{2.27}$$

and we solve the quadratic equation $\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}(\tau) = 0$ for its unique root $\hat{\tau} \in [0, 1]$. We then take the value $\hat{y} = y(\hat{\tau})$ of the Hermite polynomial at $\hat{\tau}$ as initial guess for a Gauss–Newton iteration. The periodic solution y^* obtained in this way is accepted as turning point if

$$|t^*_\lambda| < \mathrm{tol},$$

where t^* is the continuation tangent at y^* and tol is the required accuracy of the computation. Otherwise we replace $y^{(0)}$ or $y^{(1)}$ by y^* so that $t^{(0)}_\lambda t^{(1)}_\lambda < 0$ and repeat the procedure.

3. Computation of flip down and flip up bifurcations. In this section we show how generic bifurcations of symmetric periodic orbits to other periodic orbits can be computed numerically. This involves the detection and numerical computation of bifurcation points and the computation of the start off directions for the bifurcating branches. We only have to follow nonconjugate branches and distinguish between two types of symmetry changing bifurcations: There are symmetry increasing bifurcations, which lead to a super group of the symmetry group of the original solution, i.e., the bifurcating solutions possess more symmetry, and there are symmetry breaking bifurcations, which lead to a subgroup of the symmetry group of the original solution.

In this section we treat only bifurcations from periodic orbits to periodic orbits, not Hopf bifurcations (from periodic orbits to stationary solutions)—we will treat these in the next section. Such bifurcations have been classified by Fiedler [9] in the case of cyclic symmetry groups. Bifurcations of periodic orbits in systems with arbitrary finite symmetry group were classified by Lamb and Melbourne [18]; see also [19].

In this paper we assume that the symmetry group of the dynamical system is discrete and that the isotropy K of the periodic orbit is trivial; i.e., we restrict the dynamical system to the corresponding fixed point space $\mathrm{Fix}(K)$. In other words, we do not treat bifurcations to periodic orbits with smaller or bigger isotropy group K. The methods we present extend the techniques of Gatermann and Hohmann [10] for the numerical computation of symmetry changing bifurcations of stationary solutions to the case of periodic solutions. In particular, the symmetry monitoring functions which are used for the detection of symmetry changing bifurcations are related to those used in [10].

Generic bifurcations of periodic orbits with trivial isotropy to other periodic orbits are caused by a period doubling (flip down) or period halving (flip up) bifurcation of the reduced Poincaré map [18, 19]. We start with generic bifurcations without symmetry where the ODE (1.1) is not assumed to be equivariant.

3.1. Detection and computation of period doubling bifurcations. A point (x^*, λ^*) characterizing a periodic solution with period T^* is a period doubling bifurcation point (flip down point) if the Jacobian $\mathrm{D}_x\Pi$ of the Poincaré map has a single eigenvalue -1 in (x^*, λ^*) with eigenvector $v^*_S \in S$ and if this is the only eigenvalue on the complex unit circle; see, e.g., [13, 16]. Let $x(\lambda)$ be the solution branch of $\Pi(x, \lambda) = x$ with $x(\lambda^*) = x^*$. Then we, moreover, require that the path $\mu(\lambda)$ of eigenvalues of $\mathrm{D}_x\Pi(x(\lambda), \lambda)$ with $\mu(\lambda^*) = -1$ satisfy the generic transversality condition

$$\frac{\partial \mu}{\partial \lambda}(\lambda^*) \neq 0. \tag{3.1}$$

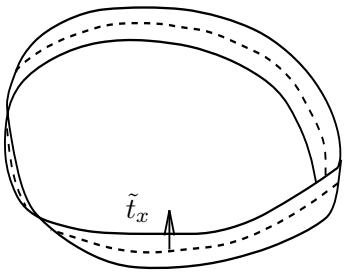


Figure 1. *Period doubling bifurcation in phase space. The bifurcating periodic orbit lies on a Mobius strip. Dashed line: original periodic orbit; solid line: bifurcating periodic orbit with twice the period; $\tilde{t}_x$: x-component of the continuation tangent for the bifurcating branch.*

Under these assumptions (x^*, λ^*) is a pitchfork bifurcation point of the map

$$\tilde{\mathcal{F}}(x, \lambda) = \Pi(\Pi(x, \lambda), \lambda) - x = 0; \tag{3.2}$$

see [13, 16]. The normal form of this bifurcation is

$$\tilde{f}(z, \lambda) = z^3 - \lambda z. \tag{3.3}$$

By a Lyapunov–Schmidt reduction, (3.2) can be reduced to a scalar equation in $z = \langle x - x^*, v_S^* \rangle$. After a suitable change of coordinates this scalar equation takes the form (3.3), up to order 3; see [13, 16]. The vector $\tilde{t}_{\tilde{\mathcal{F}}} = (v_S^*, 0)$ is the tangent vector of the bifurcating branch in (x^*, λ^*). The bifurcating periodic orbits correspond to fixed points of Π^2 and hence have approximately twice the period of the original periodic solution. They lie on a Mobius strip around the original periodic orbit; see Figure 1. The map $\tilde{\mathcal{F}}$ is $\mathbb{Z}_2$-equivariant where the nonlinear $\mathbb{Z}_2$-action is given by the Poincaré map $(x, \lambda) \to (\Pi(x), \lambda)$ (in the normal form (3.3) this $\mathbb{Z}_2$-symmetry becomes $z \to -z$). So a period doubling bifurcation is a $\mathbb{Z}_2$-symmetry breaking bifurcation of equilibria of $\tilde{\mathcal{F}}$. If we consider the T-periodic solutions on the original branch as $\widetilde{T}$-periodic, where $\widetilde{T} := 2T$, the original branch has temporal $\mathbb{Z}_2$-symmetry for the action of the spatio-temporal symmetry group $\Gamma \times \mathbb{R}$ given by (2.7):

$$(\mathrm{id}, \theta)x(t) = x(t + \theta\widetilde{T}) = x(t) \quad \text{for} \quad \theta = 1/2,$$

and the branching solutions are not $\mathbb{Z}_2$-symmetric. Thus we see that even in the case of a trivial symmetry group $\Gamma = \{\mathrm{id}\}$ the period doubling bifurcation corresponds to a $\mathbb{Z}_2$-symmetry breaking bifurcation as the phase shift symmetry (temporal symmetry) of the periodic orbit is broken.

In the following we briefly describe how to numerically detect and compute period doubling bifurcations in nonsymmetric systems. We adapt standard techniques used in the context of collocation methods [1, 8, 17] to our approach for the computation of periodic orbits as solutions of underdetermined systems, as described in section 2. In section 3.5 we show how to adapt these methods to the numerical computation of symmetry breaking bifurcations of symmetric periodic orbits.

3.1.1. Numerical detection of period doubling bifurcations. Period doublings can be detected by a change of sign of

$$d(\lambda) := \det(\mathrm{D}_x \Phi_{T^*}(x^*, \lambda^*) + \mathrm{id}),$$

which occurs due to the transversality condition (3.1). The matrix $\mathrm{D}_x \Phi_{T^*}(x^*, \lambda^*)$ is computed in the single shooting approach to obtain $\mathrm{D}F(x^*, T^*, \lambda^*)$—see (2.4)—and also in the multiple shooting approach in the computation of the condensed matrix G_c—see (2.18) with $\ell = 1$ and $\alpha = \mathrm{id}$.

3.1.2. Computation of period doubling bifurcation points. If a period doubling point has been detected, it can be computed by use of linear interpolation and a Gauss–Newton procedure to iterate back to the solution path: If there is a period doubling point between two consecutively computed periodic solutions $y^{(0)}$ and $y^{(1)}$, then

$$d(\lambda^{(0)}) d(\lambda^{(1)}) < 0.$$

By linear interpolation of the two points $(\lambda^{(0)}, d(\lambda^{(0)}))$ and $(\lambda^{(1)}, d(\lambda^{(1)}))$ we obtain a point $(\hat{\lambda}, 0)$ which gives us an approximation for the parameter value $\hat{\lambda}$ of the bifurcation point. Linear interpolation between the points $y^{(0)}$ and $y^{(1)}$ provides a first guess $\hat{y}$ with parameter $\hat{\lambda}$ for the period doubling point. This guess is then iterated back to a point y^* on the solution path by a Gauss–Newton procedure. If

$$\|y^* - \hat{y}\| \leq \mathrm{tol},$$

y^* is accepted as the period doubling point. If not, then either $y^{(0)}$ or $y^{(1)}$ is replaced by y^*, such that the condition $d(\lambda^{(0)}) d(\lambda^{(1)}) < 0$ is satisfied, and the procedure is repeated.

Doubling the number of multiple shooting points for the bifurcating branch. If we fix the number of multiple shooting points on the bifurcating branch of $2T$-periodic solutions, then after some period doubling bifurcations the multiple shooting method is likely to diverge because the initial number of multiple shooting points will not be appropriate for a periodic solution with a much higher period than the original period. Therefore it is preferable to compute the bifurcating branch of periodic orbits with twice as many multiple shooting points, i.e., to set $\tilde{k} = 2k$ as the number of multiple shooting points of the bifurcating branch. The bifurcation point $\tilde{y} = (\tilde{x}_1, \ldots, \tilde{x}_{\tilde{k}}, \tilde{T}, \tilde{\lambda})$ on the bifurcating branch is given by

$$\tilde{x}_i = x_i^*, \quad \tilde{x}_{i+k} = x_i^* \quad \text{for} \quad i = 1, \ldots, k, \quad \tilde{T} = 2T^*, \quad \tilde{\lambda} = \lambda^*,$$

and the multiple shooting nodes $\tilde{s}_i$, $i = 0, 1, \ldots, \tilde{k}$, of the bifurcating branch are set to

$$\tilde{s}_i = \frac{s_i}{2}, \quad \tilde{s}_{i+k} = \frac{1 + s_i}{2}, \quad i = 1, \ldots, k.$$

3.1.3. Computation of start off directions for the bifurcating branch. After a period doubling bifurcation point has been found, the start off direction for the bifurcating branch has to be computed. The continuation tangent of the original periodic branch is just the usual continuation tangent. The start off direction for the bifurcating branch is computed as follows.

Single shooting ansatz. We first consider the single shooting approach: We want the start off direction for the bifurcating branch of periodic orbits to be orthogonal to the T^*-periodic orbit, so that it lies in the Poincaré section S. In $S \times \mathbb{R}$ the tangent of the bifurcating branch in the bifurcation point should be the vector $\tilde{t}_{\tilde{\mathcal{F}}} = (v_S^*, 0)$, where v_S^* is the eigenvector of $\mathrm{D}_x\Pi(x^*)$ to the eigenvalue -1; see above. It can be computed by projecting the kernel vector v^* of $\mathrm{D}_x\Phi_T(x^*, \lambda^*) + \mathrm{id}$ onto the orthogonal complement of the tangent $f(x^*)$ to the periodic orbit through x^*: Let

$$\tilde{t}_x = v^* - \frac{\langle v^*, f(x^*)\rangle}{\langle f(x^*), f(x^*)\rangle} f(x^*). \tag{3.4}$$

Then we take

$$\tilde{t} = (\tilde{t}_x, \tilde{t}_T, \tilde{t}_\lambda) = (\tilde{t}_x, 0, 0)$$

as the start off direction for the bifurcating periodic solutions. In phase space the bifurcating periodic solutions lie on a Moebius band in the middle of which is the original T^*-periodic solution. The start off direction is tangential to the Moebius band and orthogonal to the original solution (see Figure 1).

Multiple shooting ansatz. As preliminary tangent vector v_1^* of the bifurcating branch for the first multiple shooting point we choose the eigenvector of $G_c + \mathrm{id}$ to -1, where $G_c \approx \mathrm{D}_x\Phi_{T^*}(x^*)$ is the condensed matrix; see (2.16), (2.18). As preliminary tangent start off directions at the multiple shooting points $x_2^*, \ldots, x_k^*$ we take

$$v_j^* = G_j v_{j-1}^*, \quad j = 2, \ldots, k.$$

The first nk components of the start off tangent $\tilde{t}$ of the bifurcating branch in the multiple shooting approach are then obtained by projecting $v^* = (v_1^*, \ldots, v_k^*)$ to the orthogonal complement of the vector $t^f = (f(x_1^*), \ldots, f(x_k^*))$.

Since the number of multiple shooting points is doubled on the bifurcating branch (see section 3.1.2), this gives us only the first half of the x-component

$$\tilde{t}_x = (\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_k, \tilde{t}_{k+1}, \ldots, \tilde{t}_{2k})$$

of the start off tangent $\tilde{t} = (\tilde{t}_x, 0, 0)$ of the bifurcating branch. The whole x-component of $\tilde{t}$ is obtained by copying the first half into the second half and multiplying it by -1, i.e.,

$$\tilde{t}_{k+i} = -\tilde{t}_i, \quad i = 1, \ldots, k. \tag{3.5}$$

As in the single shooting approach, we have $\tilde{t}_T = \tilde{t}_\lambda = 0$.

3.2. Detection and computation of period halving bifurcations. In this section we describe an algorithm for the detection and computation of *period halving bifurcations* (flip up points) along branches of periodic orbits. Again, we assume that the symmetry group Γ of the dynamical system (1.1) is trivial.

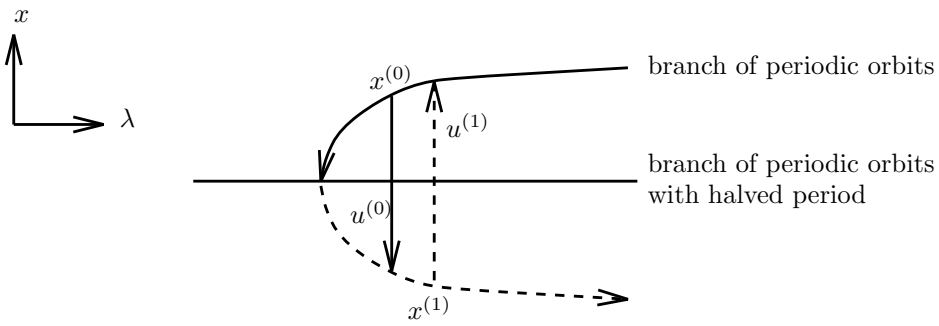


Figure 2. *Detection of period halvings using (3.6). Solid curve: branch of periodic orbits before passing the flip up point; dashed curve: branch of periodic orbits after passing the flip up point. See text for more explanations.*

3.2.1. Detection of period halving bifurcations. Period halvings can be detected as follows: For a solution point $y = (x_1, \dots, x_k, T, \lambda)$ of (2.26) we compute

$$u(y) := \Phi_{\frac{T}{2}}(x_1, \lambda) - x_1 = x_j - x_1.$$

Here one multiple shooting node s_j is set to $s_j = 1/2$ so that no additional initial value problem has to be solved.

If there is a period halving point $y^* = (x^*, T^*, \lambda^*)$ between two consecutively computed periodic solutions $y^{(0)}$ and $y^{(1)}$, the vector u goes through zero. As we saw before, a flip bifurcation corresponds to a pitchfork bifurcation of (3.2). Figure 2 shows the normal form (3.3) of the pitchfork bifurcation and points $x^{(0)}$ and $x^{(1)}$ corresponding to periodic orbits before and after passing the bifurcation. For each parameter value λ, the corresponding points on the solid and dashed curves belong to the same periodic orbit and are conjugate points with respect to the $\mathbb{Z}_2$-symmetry action, which, for (3.2), is given by $x \to \Pi(x, \lambda)$; see above. For the points $x^{(0)}$ and $x^{(1)}$ we denote the difference between $x^{(\nu)}$ and its conjugate $\Pi(x^{(\nu)}, \lambda^{(\nu)})$ by $u^{(\nu)} = \Pi(x^{(\nu)}, \lambda^{(\nu)}) - x^{(\nu)}$, $\nu = 0, 1$. From Figure 2 we see that the vectors $u^{(0)}$ and $u^{(1)}$ are approximately parallel with opposite sign. At the numerically computed solutions $y^{(0)}$ and $y^{(1)}$ the vectors $u(y^{(0)})$ and $u(y^{(1)})$ are good approximations for $u^{(0)}$ and $u^{(1)}$. Therefore a period halving point can be detected by the following condition on the angle between $u(y^{(0)})$ and $u(y^{(1)})$:

$$\frac{\langle u(y^{(0)}), u(y^{(1)})\rangle}{\|u(y^{(0)})\|\|u(y^{(1)})\|} < 0. \tag{3.6}$$

3.2.2. Computation of period halving bifurcation points and start off directions. Assume that a period halving bifurcation has been detected between two consecutively computed periodic solutions $y^{(0)}$ and $y^{(1)}$. We now describe a method for the computation of period halving points which is based on existing methods for detecting period doubling points. The algorithm consists of the following steps:

1. We obtain a first guess for the location of the flip up point by approximating the solution branch by a Hermite interpolating polynomial; cf. Figure 2. We compute

a Hermite polynomial $y(\tau)$ of degree 3 through the two points $y(0) = y^{(0)}$, $y(1) = y^{(1)}$ of the bifurcation diagram, between which a period halving has been detected. This is analogous to the Hermite interpolation used to locate turning points; see section 2.5. By computing the extremum $\lambda(\hat{\tau})$ of the polynomial $\lambda(\tau)$, where $y(\tau) = (x(\tau), T(\tau), \lambda(\tau))$, we obtain a first guess $\hat{y} = y(\hat{\tau}) = (\hat{x}_1, \ldots, \hat{x}_k, \hat{T}, \hat{\lambda})$ for the flip up point. We now halve the period $\hat{T} := \hat{T}/2$. Again, as in the case of period doubling bifurcations, the number of multiple shooting points is adapted with respect to changes of the period. In this case the number of multiple shooting points is $\tilde{k} = j$, where j is such that $s_j = \frac{1}{2}$. Decreasing the number of multiple shooting points at flip up bifurcations obviously speeds up the calculation time.

2. Now we use this first guess as starting value for a flip down point computation. We look for a flip down point in the parameter scope $[\lambda_{\min}, \lambda_{\max}]$, where

$$\lambda_{\min} = \hat{\lambda} - \epsilon, \quad \lambda_{\max} = \hat{\lambda} + \epsilon, \quad \epsilon = \max(\mathrm{tol}, \min(|\lambda^{(0)} - \hat{\lambda}|, |\lambda^{(1)} - \hat{\lambda}|)),$$

and tol is the prescribed accuracy. The period doubling bifurcation point on this branch is then accepted as a period halving point for the original branch.

Remarks 3.1.

(a) In AUTO and CONTENT period halving bifurcations are not properly detected: The programs detect bifurcation points of "unknown type" but do not compute the bifurcating branch of periodic solutions with halved period.

(b) Note that period halving bifurcations cannot be detected by a change of sign of a determinant, in contrast to transcritical and saddle node bifurcations of equilibria. The monitoring function that we use is inspired by the monitoring functions for symmetry increasing bifurcations of equilibria developed by Gatermann and Hohmann [10].

(c) In the computation of bifurcations of equilibria numerical methods based on extended systems are frequently employed (see, e.g., [1, 16]). One could of course also use such a method to locate period halving bifurcations. A period halving point could, for example, be computed by solving the system

$$0 = F(x, T, \lambda, v) := \begin{pmatrix} \Phi_{T/2}(x; \lambda) - x \\ \mathrm{D}_x \Phi_{T/2}(x; \lambda) v + v \\ \|v\|^2 - 1 \end{pmatrix}$$

using a Newton-type method. However, since this requires the approximation of the second derivative of the flow map, it would be more expensive than the method that we suggest.

3.3. Bifurcations of periodic orbits with $\mathbb{Z}_p$-symmetry. In this section we deal with generic symmetry changing but isotropy preserving secondary bifurcations of symmetric periodic orbits. The right-hand side f of the ODE (1.1) is assumed to be Γ-equivariant under a finite group $\Gamma \subset \mathrm{GL}(n)$, as in section 2.2. We assume that the spatial symmetry K of the periodic orbit is trivial (or restrict the dynamics to $\mathrm{Fix}(K)$). This implies that the spatio-temporal symmetry of the periodic orbits is cyclic: $L \simeq \mathbb{Z}_\ell$. We can then, without loss of generality, restrict to the case $\Gamma \simeq \mathbb{Z}_p$ for p a multiple of ℓ; see [18]. We describe how the generic secondary bifurcations of periodic orbits with $\mathbb{Z}_p$-symmetry, which have been classified

by Fiedler [9], can be treated numerically. In this section we deal only with bifurcations of periodic orbits into other periodic orbits (not Hopf bifurcations along branches of periodic orbits).

Let x^* lie on a periodic orbit $\mathcal{P}$ with period T^*, trivial isotropy K, spatio-temporal symmetry $L = \mathbb{Z}_\ell$, and drift symmetry α. Define the Poincaré section as usual by $S = x^* + \operatorname{span}(f(x^*))^\perp$. To examine bifurcations of symmetric periodic orbits the reduced Poincaré map

$$\Pi_{\mathrm{red}} = \alpha\hat{\Pi}, \qquad \hat{\Pi} : S \to \alpha^{-1}S$$

from (2.10) is used, where $\hat{\Pi}$ maps points of S into the points where the positive semiflow through x first hits $\alpha^{-1}\,S$. In the case of trivial isotropy K the relationship between the full Poincaré map Π and the reduced Poincaré map Π_{red} is given by

$$\Pi = \alpha^{-\ell}\Pi_{\mathrm{red}}^{\ell} = \Pi_{\mathrm{red}}^{\ell}. \tag{3.7}$$

Here we used that $\alpha^\ell = \mathrm{id}$. Generic bifurcations of symmetric periodic orbits are bifurcations of the reduced Poincaré map Π_{red}, which arise from an eigenvalue ± 1 of the Jacobian $\mathrm{D}_x\Pi_{\mathrm{red}}$; see [9, 18].

Generic bifurcations of Π_{red} without breaking of the spatial symmetry are turning points and period doublings/halvings (flip down and flip up bifurcations); turning points of Π_{red} lead to turning points of the full Poincaré map Π. They can be detected and computed in the same way as in the case of no symmetry; see section 2.5.

Flip down bifurcations of the reduced Poincaré map Π_{red} lead to pitchfork bifurcations or period doubling bifurcations of Π, depending on whether ℓ is odd or even; see [9]. If ℓ is odd, in particular if the symmetry is trivial (i.e., $\ell = 1$), we have a *flip doubling* (period doubling) bifurcation. If ℓ is even, then a *flip pitchfork* bifurcation takes place, where the period is preserved but the spatio-temporal symmetry halved.

3.3.1. Flip pitchfork bifurcation. First we consider the flip pitchfork bifurcation. Let ℓ be even. Assume that the reduced Poincaré map Π_{red} undergoes a flip down bifurcation. Then the bifurcating solutions $\tilde{x}(s)$, $s \in \mathbb{R}$, $\tilde{x}(0) = x^*$, are fixed points of Π_{red}^2. By (3.7),

$$\Pi(\tilde{x}(s)) = \left(\Pi_{\mathrm{red}}^2\right)^{\ell/2}(\tilde{x}(s)) = \tilde{x}(s),$$

and so the Poincaré map Π undergoes a pitchfork bifurcation. Hence the branching solutions have approximately the same period, but their spatio-temporal symmetry $\tilde{L} = \mathbb{Z}_{\ell/2}$, $\tilde{\ell} = \ell/2$, has been halved. The drift symmetry of the bifurcating periodic orbits is $\tilde{\alpha} = \alpha^2$.

3.3.2. Flip doubling bifurcation. Next we consider the flip doubling bifurcation, i.e., let ℓ be odd. Since $\Pi_{\mathrm{red}}^2(\tilde{x}(s)) = \tilde{x}(s)$ and ℓ is odd the following can be concluded:

$$\Pi^2(\tilde{x}(s)) = \Pi_{\mathrm{red}}^{2\ell}(\tilde{x}(s)) = \tilde{x}(s) \quad \text{and} \quad \Pi(\tilde{x}(s)) = \Pi_{\mathrm{red}}^{\ell}(\tilde{x}(s)) = \Pi_{\mathrm{red}}(\tilde{x}(s)) \neq \tilde{x}(s).$$

So the Poincaré map Π undergoes a period doubling bifurcation without breaking the symmetry $\mathbb{Z}_\ell$: The spatio-temporal symmetry group of the bifurcating branch is $\tilde{L} = \mathbb{Z}_{\tilde{\ell}}$, where $\tilde{\ell} = \ell$, and is generated by the order ℓ element $\tilde{\alpha} = \alpha$. Note that the flip doubling bifurcation reduces to the period doubling bifurcation of nonsymmetric systems (see section 3.1), with $\ell = 1$, $\alpha = \mathrm{id}$, whereas the flip pitchfork bifurcation does not occur for nonsymmetric systems.

3.4. Numerical computation of symmetry breaking bifurcations. Since the generic bifurcations of periodic orbits with underlying symmetry group $\mathbb{Z}_p$ described above are generated by periodic doubling bifurcations of the reduced Poincaré map, they can be treated numerically with the methods for the period doubling bifurcation described in section 3.1.

3.4.1. Detection and computation of flip down bifurcations. Flip down bifurcations are detected by a sign change of

$$d(\lambda) = \det(\alpha \mathrm{D}_x \Phi_{\frac{T^*}{\ell}}(x^*) + \mathrm{id}), \quad \text{where} \quad \alpha \mathrm{D}_x \Phi_{\frac{T^*}{\ell}}(x^*) = G_c,$$

analogously to the case of no symmetry; see section 3.1.1. Once a flip down bifurcation has been detected it can be computed analogously as in the case of nonsymmetric systems; see section 3.1.2.

3.4.2. Initialization of the bifurcating branch. Once a flip down point (x^*, T^*, λ^*) on the original branch has been found, the starting point $\tilde{y} = (\tilde{x}, \tilde{T}, \tilde{\lambda})$ for the bifurcating branch has to be computed. We set $\tilde{T} = T^*$ for a flip pitchfork bifurcation and $\tilde{T} = 2T^*$ otherwise, and $\tilde{\lambda} = \lambda^*$. Since the number of multiple shooting points is doubled, the second half of $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_{2k})$ will be computed by applying the symmetry matrix to the first points:

$$\begin{aligned} \tilde{x}_i &= x_i^* \quad \text{for} \quad i = 1, \ldots, k, \\ \tilde{x}_{i+k} &= \alpha^{-1} x_i^* \quad \text{for} \quad i = 1, \ldots, k. \end{aligned}$$

The tangent vector $\tilde{t} = (\tilde{t}_x, 0, 0)$ of the bifurcating branch is computed in a similar way: The first nk components $(\tilde{t}_1, \ldots, \tilde{t}_k)$ of the x-component $\tilde{t}_x$ of the tangent $\tilde{t}$ of the bifurcating branch are computed as in section 3.1.3, with $\mathrm{D}_x \Phi_{T^*}(x^*)$ replaced by $\alpha \mathrm{D}_x \Phi_{\frac{T^*}{\ell}}(x^*)$. Then the second half of $\tilde{t}_x = (\tilde{t}_1, \ldots, \tilde{t}_{2k})$ is computed by applying the symmetry matrix to the first half and multiplying it by -1:

$$\tilde{t}_{i+k} = -\alpha^{-1} \tilde{t}_i \quad \text{for} \quad i = 1, \ldots, k. \tag{3.8}$$

3.5. Numerical computation of symmetry increasing bifurcations. In this section we extend the algorithms for the detection and computation of period halving points for nonsymmetric systems (section 3.2) to systems with symmetry. The main issue here is the identification of the possible spatio-temporal symmetries of the bifurcating solutions, which are needed for both the detection of bifurcations and the computation of the bifurcating branch.

3.5.1. Detection of flip up bifurcations. As in the nonsymmetric case (see section 3.2), flip up points on a branch of periodic orbits with spatio-temporal symmetry $L = \mathbb{Z}_\ell$, trivial isotropy, and drift symmetry α are detected by the angle condition (3.6)

$$\frac{\langle u(y^{(0)}), u(y^{(1)}) \rangle}{\|u(y^{(1)})\| \|u(y^{(1)})\|} < 0,$$

where $y^{(0)}$ and $y^{(1)}$ are two consecutive points on a branch of periodic solutions and

$$u(x) := \tilde{\alpha} \Phi_{\frac{T}{2\ell}}(x_1, \lambda) - x_1 = \tilde{\alpha} x_j - x_1.$$

Here the multiple shooting node s_j is set to $s_j = \frac{1}{2}$, and $\tilde{\alpha}$ is a group element $\tilde{\alpha}$ which satisfies

$$\tilde{\alpha}^2 = \alpha. \tag{3.9}$$

We now need to classify the possible choices of $\tilde{\alpha}$ and decide whether a *flip up doubling* or a *flip up pitchfork* bifurcation occurs.

Theorem 3.2. *Let i be such that $\alpha = \gamma_p^i$, where γ_p generates the symmetry group $\Gamma = \mathbb{Z}_p$ of (1.1). Similarly, write $\tilde{\alpha} = \gamma_p^{\tilde{i}}$. Then we have the following:*

(a) *Either $\tilde{i} = \frac{i}{2}$ or $\tilde{i} = \frac{i+p}{2}$. Both of these values for $\tilde{i}$ are possible if p and i are even; $\tilde{i} = \frac{i}{2}$ is a possible solution for p odd, i even; and $\tilde{i} = \frac{i+p}{2}$ for p and i odd.*

(b) (i) *If*

$$\ell\tilde{i} = 0 \mod p,$$

then a flip up doubling *bifurcation takes place. The order $\tilde{\ell}$ of the drift symmetry $\tilde{\alpha}$ of the bifurcating branch and its period $\tilde{T}$ in the bifurcation point then satisfy*

$$\tilde{\ell} = \ell, \quad \tilde{T} = \frac{T^*}{2},$$

where T^* is the period of the original periodic orbit at the bifurcation point.

(ii) *If*

$$\ell\tilde{i} \neq 0 \mod p,$$

then a flip up pitchfork *bifurcation takes place. The order $\tilde{\ell}$ of the drift symmetry $\tilde{\alpha}$ of the bifurcating branch and its period $\tilde{T}$ in the bifurcation point then satisfy*

$$\tilde{\ell} = 2\ell, \quad \tilde{T} = T^*.$$

Proof. From (3.9) we get

$$2\tilde{i} = i \mod p,$$

and so

$$\tilde{i} = (i + jp)/2, \quad j \in \mathbb{N}. \tag{3.10}$$

Possible solutions to (3.10) which are different modulo p are

$$\tilde{i} = i/2, \quad \tilde{i} = (i + p)/2.$$

This proves part (a) of the theorem. The rest follows from the definitions of flip pitchfork and flip doubling bifurcations; see sections 3.3.1 and 3.3.2. ∎

Remark 3.3. Consider a periodic orbit with trivial spatio-temporal symmetry, i.e., $\alpha = \mathrm{id}$, $i = 0$. Then, if the order p of the symmetry group $\Gamma = \mathbb{Z}_p$ is odd and the Γ-action is free, every flip up bifurcation is a period halving bifurcation. If p is even, then a flip up bifurcation of Π_{red} can be a period halving bifurcation or a flip up pitchfork bifurcation with $\tilde{\alpha} = \gamma_p^{\frac{p}{2}}$.

3.5.2. Computation of flip up points and start off tangents. Once the spatio-temporal symmetry of the bifurcating periodic orbit is identified the flip up point and the start off directions for the bifurcating branch can be computed in the same way as for nonsymmetric systems; see section 3.2.

4. Computation of equivariant Hopf points. In this section we show how equivariant Hopf points along branches of periodic orbits can be detected and computed. For the sake of simplicity, we first consider the case of a trivial symmetry group $\Gamma = \{\mathrm{id}\}$.

4.1. Hopf bifurcations for nonsymmetric systems. A Hopf bifurcation point is a stationary solution (x^*, λ^*), for which the Jacobian $\mathrm{D}_x f(x^*, \lambda^*)$ has a pair of purely imaginary eigenvalues $\pm\omega^*\mathrm{i}$, $\omega^* \neq 0$. We make the generic assumptions that these eigenvalues are simple and that there is no resonance; i.e., no multiple $\mathrm{i}\omega^* j$, $j \in \mathbb{N}_0 \setminus \{1\}$, is an eigenvalue of $\mathrm{D}_x f(x^*, \lambda^*)$. Denote by $x(\lambda)$ the path of equilibria of (1.1) with parameter λ, so that in particular $x(\lambda^*) = x^*$. Let $\mu(\lambda)$ be the path of eigenvalues of $\mathrm{D}_x f(x(\lambda), \lambda)$ such that $\mu(\lambda^*) = \mathrm{i}\omega^*$ and assume that the generic transversality condition

$$\mathrm{Re}\left(\frac{\partial \mu}{\partial \lambda}(\lambda^*)\right) \neq 0 \tag{4.1}$$

is satisfied. Then (see, e.g., [13, 16]) a unique branch $x(t;\epsilon)$ of periodic solutions emanates from the stationary solution with small amplitude $O(\epsilon)$ and period $T(\epsilon) \approx T(0) = 2\pi/\omega^*$. This surface of periodic solutions is tangential to the real eigenspace N_{ω^*} of $\pm\omega^*\mathrm{i}$, i.e., $\mathrm{D}_\epsilon x(t;0) \in N_{\omega^*}$, and generically agrees (after a smooth coordinate change) to second order with a paraboloid $\lambda - \lambda^* = C(z_1^2 + z_2^2)$; see, e.g., [13, 16]. We can consider the Hopf point (x^*, λ^*) as an S^1-invariant $2\pi/\omega^*$-periodic solution with respect to the action (2.7) of the temporal symmetry group on the periodic solutions $x(t)$ of (1.1):

$$x^*(t) \equiv x^* \quad \forall\, t \implies (\mathrm{id}, \theta)\, x^*(t) = x^*(t) \quad \forall\, t.$$

Hence the Hopf bifurcation is an S^1-symmetry breaking bifurcation.

4.1.1. Detection of Hopf points along branches of periodic orbits. If a pathfollowing algorithm for periodic solutions runs through a Hopf point, the continuation direction changes its sign and the same path of periodic orbits is computed again. Therefore Hopf points which occur during the pathfollowing of periodic orbits should be detected so that the pathfollowing routine can be stopped at the Hopf point. If there is a Hopf point between two consecutively computed periodic orbits $y^{(0)}$ and $y^{(1)}$, where $y = (x, T, \lambda)$, $x = (x_1, \ldots, x_k)$, then the vectors $f(x_i)$, which are the infinitesimal generators of the S^1-symmetry in the point x_i, go through zero. Thus $f(x_i)$ is a symmetry monitoring function in this case. If the angle between the vectors $f(x_i)$ of two consecutively computed periodic orbits is greater than 90 degrees, i.e., if for some $i \in \{1, \ldots, k\}$

$$\frac{\langle f(x_i^{(0)}), f(x_i^{(1)})\rangle}{\|f(x_i^{(0)})\|\|f(x_i^{(1)})\|} < 0,$$

then there is a Hopf point on the branch of periodic orbits between the $y^{(0)}$ and $y^{(1)}$; cf. Figure 3.

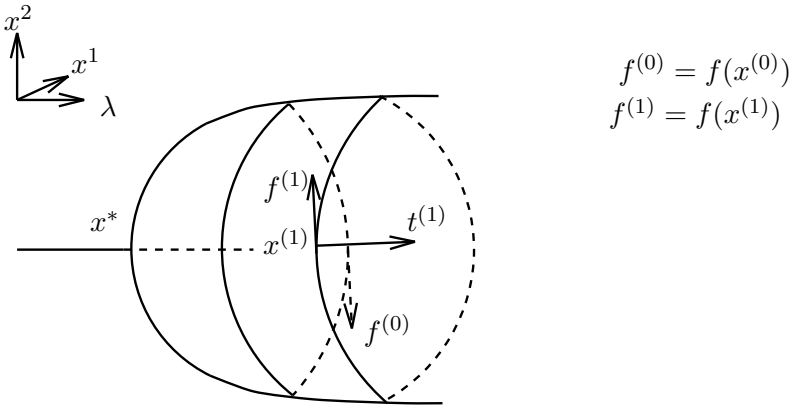


Figure 3. *Detection of Hopf points between two periodic orbits* $y^{(0)} = (x^{(0)}, T^{(0)}, \lambda^{(0)})$ *and* $y^{(1)} = (x^{(1)}, T^{(1)}, \lambda^{(1)})$ *(single shooting method). Here* $t^{(1)}$ *is the continuation tangent at the point* $y^{(1)}$.

Remark 4.1. Note that in the program packages AUTO [8] and CONTENT [17], the numerical part of which is based on AUTO, Hopf points along periodic orbits are not detected properly. They are detected as general cycle branching points, but when switching to the bifurcating branch of stationary solutions, an error occurs.

4.1.2. Computation of Hopf bifurcations of nonsymmetric systems. If a Hopf point along a path of periodic orbits is detected, it can be computed by an extended system [12, 14]; see also the review in [1]. We use a slightly different form of extended system which is underdetermined and does not fix the phase of the Hopf eigenvector to be computed. We present this extended system briefly in this subsection.

Let x^* be a Hopf point, i.e., an equilibrium $f(x^*, \lambda^*) = 0$, the Jacobian $\mathrm{D}_x f(x^*, \lambda^*)$ of which has a pair of purely imaginary eigenvalues $\pm \mathrm{i}\omega^*$. Hence

$$\mathrm{D}_x f(x^*, \lambda^*)(v^* + \mathrm{i}w^*) = \mathrm{i}\omega^*(v^* + \mathrm{i}w^*).$$

So we get

$$\mathrm{D}_x f(x^*, \lambda^*)v^* = -\omega^* w^*, \quad \mathrm{D}_x f(x^*, \lambda^*)w^* = \omega^* v^*, \quad \|w^*\|^2 + \|v^*\|^2 = 1.$$

Define $F : \mathbb{R}^{3n+2} \to \mathbb{R}^{3n+1}$, where

$$F(x, \lambda, v, w, \omega) = \begin{pmatrix} f(x, \lambda) \\ \mathrm{D}_x f(x, \lambda)v + \omega w \\ \mathrm{D}_x f(x, \lambda)w - \omega v \\ \langle v, v\rangle + \langle w, w\rangle - 1 \end{pmatrix}. \tag{4.2}$$

Then $F = 0$ yields the Hopf point and its imaginary eigenvalue and corresponding eigenvector. Moreover we have the following result.

Theorem 4.2. *If the eigenvalue* $\pm \mathrm{i}\omega^*$ *is simple, if* $\mathrm{D}_x f(x^*, \lambda^*)$ *is invertible, and if the transversality condition* (4.1) *holds, then the Gauss–Newton method applied to* (4.2) *converges for sufficiently good initial data.*

The proof of this theorem is basically contained in [1, 12, 14]. As shown there, the transversality condition $\operatorname{Re}\mu'(\lambda^*) \neq 0$ implies that any kernel vector $t = (t_x, t_\lambda, t_v, t_w, t_\omega)$ of the derivative $\mathrm{D}F(x^*, \lambda^*, v^*, w^*, \omega^*)$ of F in the Hopf point satisfies $t_\omega = 0$, $t_\lambda = 0$, and $t_x = 0$, and hence (t_v, t_w) satisfies the equations

$$0 = \mathrm{D}_x f(x^*, \lambda^*)t_v + \omega^* t_w, \tag{4.3}$$

$$0 = \mathrm{D}_x f(x^*, \lambda^*)t_w - \omega^* t_v, \tag{4.4}$$

$$0 = 2\langle v^*, t_v\rangle + 2\langle w^*, t_w\rangle. \tag{4.5}$$

Therefore $t_v + \mathrm{i}t_w$ is a Hopf eigenvector. Equation (4.5) and the fact that the Hopf eigenvalue $\mathrm{i}\omega^*$ is a simple eigenvalue of $\mathrm{D}_x f(x^*, \lambda^*)$ imply that the kernel of $\mathrm{D}F$ in the Hopf point is one-dimensional. Hence $\mathrm{D}F$ has full rank in the Hopf point, and the Gauss–Newton method applied to (4.2) converges for sufficiently good initial data. As we will see later (see sections 4.2.1 and 5.2), symmetry often enforces multiple Hopf eigenvalues, so that the extended system (4.2) has to be modified in the case of equivariant Hopf points.

Initial guess for the Gauss–Newton iteration. An initial guess for a Hopf point detected between two periodic orbits $y^{(0)} = (x^{(0)}, T^{(0)}, \lambda^{(0)})$ and $y^{(1)} = (x^{(1)}, T^{(1)}, \lambda^{(1)})$, where $x = (x_1, \ldots, x_k)$, can be obtained by Hermite interpolation $y(\tau)$ between those points over the line $y^{(0)} + \tau(y^{(1)} - y^{(0)})$, $\tau \in [0, 1]$, such that $y(0) = y^{(0)}$, $y(1) = y^{(1)}$, and by computing the point $\hat{y} = y(\hat{\tau}) = (x(\hat{\tau}), T(\hat{\tau}), \lambda(\hat{\tau}))$ such that $\lambda'(\hat{\tau}) = 0$, analogously to the computation of initial guesses for a turning point; see section 2.5. We then set

$$\hat{x} := x_1(\hat{\tau}),$$

define an approximation for the Hopf frequency $\hat{\omega}$ as

$$\hat{\omega} = \frac{2\pi}{\hat{T}}, \quad \text{where } \hat{T} = T(\hat{\tau}),$$

and define an approximation for the parameter value of the Hopf point as

$$\hat{\lambda} = \lambda(\hat{\tau}).$$

We moreover define initial guesses $\hat{v} + \mathrm{i}\hat{w}$ for the Hopf eigenvector as

$$\hat{v} = c\frac{\mathrm{d}}{\mathrm{d}\tau}x_1(\hat{\tau}), \quad \hat{w} = -\frac{1}{\hat{\omega}}\mathrm{D}_x f(\hat{x}, \hat{\lambda})\hat{v}$$

with c such that

$$\|\hat{v}\|^2 + \|\hat{w}\|^2 = 1.$$

The point $(\hat{x}, \hat{\lambda}, \hat{v}, \hat{w}, \hat{\omega})$ is then used as an initial guess for the Newton iteration applied to (4.2).

4.2. Detection and computation of equivariant Hopf points. In this section we extend the methods for the computation of Hopf points of nonsymmetric systems from section 4.1 to systems with symmetry. The main issue here is how to deal with multiple Hopf eigenvalues forced by symmetry which lead to convergence failure of the extended system (4.2) for the computation of Hopf points of nonsymmetric systems.

4.2.1. Equivariant Hopf points. We are starting from a Γ-invariant stationary solution (x^*, λ^*), i.e., $\Gamma_{x^*} = \Gamma$. We assume that the Jacobian $\mathrm{D}_x f(x^*, \lambda^*)$ has a pair $\pm\omega^*\mathrm{i}$ of purely imaginary eigenvalues, $\omega^* \neq 0$, and that there are no resonances; i.e., $\pm j\mathrm{i}\omega^*$, $j = 0, 2, 3, 4, \ldots$, is not an eigenvalue of $\mathrm{D}_x f(^*, \lambda^*)$. As before, let N_{ω^*} be the real eigenspace of $\mathrm{D}_x f(x^*, \lambda^*)$.

Lemma 4.3 (see [11]). *If $\gamma \in \Gamma_x$, then $\gamma \mathrm{D}_x f(x, \lambda) = \mathrm{D}_x f(x, \lambda)\gamma$. Moreover, every eigenspace of $\mathrm{D}_x f(x, \lambda)$ is Γ_x-invariant.*

Proof. The first statement follows from the Γ-equivariance of f and the γ-invariance of x. For the second statement let u be a complex eigenvector of $A = \mathrm{D}_x f(x, \lambda)$ to the eigenvalue μ. Since $\gamma A = A\gamma$ we have $A\gamma u = \gamma A u = \gamma\mu u$, so that γu is an eigenvector of A to the eigenvalue μ as well. ∎

As a consequence, $\mathrm{D}_x f(x^*, \lambda^*)$ is Γ-equivariant and N_{ω^*} is Γ-invariant. Hence N_{ω^*} can be decomposed into irreducible Γ-invariant subspaces; see Remark 2.4.

Definition 4.4 (see [11]). *We call an eigenvalue μ of a Γ-equivariant matrix A a Γ-simple eigenvalue of A if the real eigenspace N of A to the eigenvalue μ is irreducible.*

We make the generic assumption that $\mathrm{i}\omega^*$ is a Γ-simple eigenvalue of $\mathrm{D}_x f(x^*, \lambda^*)$. This means that $\mathrm{i}\omega^*$ has the lowest multiplicity allowed by the symmetry group Γ.

Since $\mathrm{D}_x f(x^*, \lambda^*)$ is invertible, by the implicit function theorem applied to $\mathrm{Fix}(\Gamma) = \mathbb{R}^{n_{\mathrm{red}}}$ there is a path $x(\lambda)$ of Γ-invariant equilibria of (1.1) with $x(\lambda^*) = x^*$. As in the case of the standard Hopf bifurcation, we assume that the transversality condition (4.1) holds for the path $\mu(\lambda)$ of the pair of eigenvalues of $\mathrm{D}_x f(x(\lambda), \lambda)$ with $\mu(\lambda^*) = \mathrm{i}\omega^*$.

We define the operation of $\Gamma \times S^1$ on a T-periodic solution $x(t)$ as in (2.7)

$$(\gamma, \theta)x(t) = \gamma x(t + \theta T) \quad \text{for} \quad (\gamma, \theta) \in \Gamma \times S^1,$$

and the operation of $\Gamma \times S^1$ on the real eigenspace N_{ω^*} of $\pm\omega^*\mathrm{i}$ of $\mathrm{D}_x f(x^*, \lambda^*)$ by

$$(\gamma, \theta)u = \gamma e^{\theta \mathrm{D}_x f(x^*, \lambda^*) T^*} u, \quad (\gamma, \theta) \in \Gamma \times S^1, \quad u \in N_{\omega^*}, \tag{4.6}$$

where $T^* = \frac{2\pi}{\omega^*}$.

Theorem 4.5 (equivariant Hopf theorem [11]). *Let the above conditions be satisfied. If then for a subgroup $L \subset \Gamma \times S^1$ the fixed point space*

$$N^L_{\omega^*} := \{u \in N_{\omega^*} : (\gamma, \theta)u = u \; \forall \, (\gamma, \theta) \in L\} \tag{4.7}$$

satisfies the condition

$$\dim N^L_{\omega^*} = 2, \tag{4.8}$$

then there is a unique branch $x(t; \epsilon)$ of periodic solutions with amplitude $O(\epsilon)$ bifurcating from (x^*, λ^*) with $\mathrm{D}_\epsilon x(t; 0) \in N^L_{\omega^*}$, with parameter $\lambda(\epsilon)$ such that $\lambda(0) = \lambda^*$, with minimal periods $T(\epsilon)$ such that $T(0) = 2\pi/\omega^*$, and with L as spatio-temporal symmetry group.

As in the nonsymmetric case, the bifurcating periodic orbits generically lie on a paraboloid; see Figure 3.

Remark 4.6. The equivariant Hopf theorem provides the spatio-temporal symmetries L of the bifurcating periodic orbits and the planes $N^L_{\omega^*}$ from which the start off directions for the emanating periodic orbits can be chosen: We define the phase space for the bifurcating

periodic orbit with symmetry L as $\mathrm{Fix}(K)$, where K, the group of spatial symmetries of the periodic orbit, is the kernel of the homomorphism Θ of the periodic orbit (see (2.6)), and thus

$$K = \{\gamma \in \Gamma, \quad (\gamma, 0) \in L\}.$$

We compute the integer ℓ such that $L/K \simeq \mathbb{Z}_\ell$ (see (2.8)), and the drift symmetry α of the periodic orbit by determining the element (α, θ^*) in L with the smallest nonzero phase shift $\theta^* = \frac{1}{\ell}$; cf. (2.9). We identify the Hopf point (x^*, λ^*) with a periodic orbit which has period $T^* = \frac{2\pi}{\omega^*}$, multiple shooting points $x_i^* = x^*$ for $i = 1, \dots k$, and the parameter λ^*. Similarly as in the nonsymmetric case (cf. [1]), we then define the continuation tangent $t^* = (t_x^*, t_T^*, t_\lambda^*)$ at this Hopf periodic orbit as follows:

$$t_T^* = 0, \quad t_\lambda^* = 0, \quad t_x^* = (t_1^*, \dots, t_k^*),$$

where

$$t_i^* = \cos(s_i T^*/\ell)v^* + \sin(s_i T^*/\ell)w^*, \quad i = 1, \dots, k.$$

Here $v^* + \mathrm{i}w^*$ is the eigenvector to the purely imaginary eigenvalue $\omega^*\mathrm{i}$ of $\mathrm{D}_x f(x^*, \lambda^*)$ which is determined by the condition $v^*, w^* \in N_{\omega^*}^L$. See sections 5.2 and 5.3 for applications.

4.2.2. Detection of equivariant Hopf points. Equivariant Hopf points along branches of periodic orbits are detected in the same way as Hopf points of nonsymmetric systems; see section 4.1.1.

4.2.3. Computation of equivariant Hopf bifurcations. As mentioned in section 4.1.2, the extended system (4.2) for the computation of Hopf points of nonsymmetric systems converges only if the Hopf eigenvalue $\mathrm{i}\omega^*$ is simple. In the case of symmetric dynamical systems this assumption can be satisfied only if the corresponding irreducible representation is one-dimensional. In general the symmetry might enforce multiple eigenvalues (see section 5.2 for an example). Therefore the numerical method for computing Hopf points from section 4.1.2 has to be modified in the case of symmetric dynamical systems. In this section we present an efficient algorithm for computing equivariant Hopf points, which applies to Hopf points that satisfy the conditions of the equivariant Hopf theorem.

Assume that an equivariant Hopf point (x^*, λ^*) with Hopf eigenvalue $\pm\mathrm{i}\omega^*$, $\omega^* > 0$, has been detected numerically along a branch of periodic orbits of the Γ-equivariant ODE (1.1) with drift symmetry α of order ℓ and, for simplicity, trivial isotropy K (restrict the dynamics to $\mathrm{Fix}(K)$ and replace Γ by $N(K)/K$ otherwise). As before we denote by $L \simeq \mathbb{Z}_\ell$ the spatio-temporal symmetry of the branch of periodic orbits. Then the Hopf point x^* is L-invariant: $L \subseteq \Gamma_{x^*}$. We make the assumptions of the equivariant Hopf Theorem 4.5, replacing Γ by Γ_{x^*}, and denote the plane tangent to the branch of periodic orbits at the Hopf point by $N_{\omega^*}^L$, as in (4.7).

We will now formulate an algorithm for the computation of the equivariant Hopf point (x^*, λ^*) along with the Hopf frequency ω^* and the corresponding Hopf eigenvector $v^* + \mathrm{i}w^*$ satisfying $v^*, w^* \in N_{\omega^*}^L$.

Note that the condition of the equivariant Hopf Theorem 4.5, that $N^L_{\omega^*}$ is two-dimensional, is equivalent to requiring that there be only one eigenvector $v^* + \mathrm{i}w^*$ of $\mathrm{D}_x f(x^*, \lambda^*)$ to the eigenvalue $\pm\mathrm{i}\omega^*$ satisfying

$$v^* + \mathrm{i}w^* = \alpha \mathrm{e}^{\frac{2\pi}{\omega^*\ell}\mathrm{D}_x f(x^*,\lambda^*)}(v^* + \mathrm{i}w^*). \tag{4.9}$$

Solving the nonlinear equation (4.9) numerically using an extended system would involve the computation of the exponential $\exp(\frac{2\pi}{\omega\ell}\mathrm{D}_x f(x^*, \lambda^*))$ of $\mathrm{D}_x f(x^*, \lambda^*)$, which is in general expensive. An extended system involving $\mathrm{D}_x f(x^*, \lambda^*)$ rather than its exponential, like the system (4.2) in the case of nonsymmetric systems, is therefore preferable. To derive such a system we use the following approach. Note that

$$\mathrm{D}_x f(x^*, \lambda^*)(v^* + \mathrm{i}w^*) = \mathrm{i}\omega^*(v^* + \mathrm{i}w^*),$$

and so

$$\exp\left(\frac{2\pi}{\omega^*\ell}\mathrm{D}_x f(x^*, \lambda^*)\right)(v^* + \mathrm{i}w^*) = \mathrm{e}^{\frac{2\pi\mathrm{i}}{\ell}}(v^* + \mathrm{i}w^*),$$

and hence

$$\alpha(v^* + \mathrm{i}w^*) = \mathrm{e}^{-\frac{2\pi\mathrm{i}}{\ell}}(v^* + \mathrm{i}w^*). \tag{4.10}$$

Thus $v^* + \mathrm{i}w^*$ lies in the the complex eigenspace of α to the eigenvalue $\mathrm{e}^{-2\pi\mathrm{i}/\ell}$, which we denote by $X^c_\ell \subset X^c = \mathbb{C}^n$, and v^* and w^* lie in the real eigenspace of α to the eigenvalue $e^{-2\pi\mathrm{i}/\ell}$, which we denote by X_ℓ. So X_ℓ is the L-invariant subspace of $X = \mathbb{R}^n$, where α, the generator of $L \simeq \mathbb{Z}_\ell$, acts as a rotation by $-2\pi/\ell$.

The following lemma is crucial for the numerical computation of equivariant Hopf points.

Lemma 4.7. *Let the assumptions of the equivariant Hopf Theorem 4.5 hold. If $v + \mathrm{i}w \in X^c_\ell$ is an eigenvector of $\mathrm{D}_x f(x^*, \lambda^*)$ to the eigenvalue $\mathrm{i}\omega^*$, then $v + \mathrm{i}w = c(v^* + \mathrm{i}w^*)$ for some $c \in \mathbb{C}$, where $v^* + \mathrm{i}w^*$ is a Hopf eigenvector with $v^*, w^* \in N^L_{\omega^*}$.*

Proof. The vector $v + \mathrm{i}w$ satisfies (4.9), and by the assumption of the equivariant Hopf Theorem 4.5 there is only one such eigenvector (over $\mathbb{C}$), namely $v^* + \mathrm{i}w^*$. This proves the lemma. ∎

Due to this lemma, we can solve (4.9) as follows: We first compute the space X^ℓ_c. Then we compute the L-invariant Hopf point together with a Hopf eigenvector which lies in X_ℓ by an extended system.

First step of the algorithm. We compute an orthonormal basis of X^c_ℓ and store it as row vectors of a matrix E_ℓ. We assume that $\ell > 1$, i.e., that the branch of periodic orbits along which a Hopf bifurcation has been detected has nontrivial spatio-temporal symmetry $L = \mathbb{Z}_\ell$. We consider the following two cases.

Case 1: $\ell = 2$. We compute the kernel $X_2 = \ker(B_2)$ of $B_2 = \alpha + \mathrm{id}_n$ and store an orthonormal basis of it in the row vectors of the (d, n)-matrix E_2. Here $d = \dim X_2$, and X_2 is the L-invariant subspace of $\mathbb{R}^n$ where the action of the symmetry group $L = \mathbb{Z}_2$ generated by α is given by $\alpha|_{X_2} = -1$. Hence,

$$E_2^T E_2 = \mathrm{id}\,|_{X_2}. \tag{4.11}$$

Case 2: $\ell > 2$. In this case we compute the null space of the $(2n, 2n)$-matrix

$$B_\ell := \begin{pmatrix} \alpha - \cos(2\pi/\ell)\,\mathrm{id}_n & -\sin(2\pi/\ell)\,\mathrm{id}_n \\ \sin(2\pi/\ell)\,\mathrm{id}_n & \alpha - \cos(2\pi/\ell)\,\mathrm{id}_n \end{pmatrix}. \tag{4.12}$$

We store an orthonormal basis of $\ker(B_\ell)$ in the row vectors of the matrix $E_\ell = [E_V, E_W] \in \mathrm{Mat}(d, 2n)$, where $E_V, E_W \in \mathrm{Mat}(d, n)$.

Before we continue with the description of the algorithm we present the following lemma.

Lemma 4.8. *Let $\ell > 2$. Then the following hold true:*

(a) *If $(v, w) \in \ker B_\ell$, $v, w \in \mathbb{R}^n$, we have $v + \mathrm{i}w \in X_\ell^c$.*

(b) *If $(v, w) \in \ker B_\ell$, then so is $(-w, v)$. In particular, $\ker B_\ell$ has even dimension $d = 2\hat{d}$, $\hat{d} \in \mathbb{N}$, the eigenvalue $\mathrm{e}^{-2\pi\mathrm{i}/\ell}$ of α has geometric multiplicity $\hat{d}$, and $d = \dim X_\ell$.*

(c) *For any $x \in \mathbb{R}^d$ we have $E_V^T x + \mathrm{i}E_W^T x \in X_\ell^c$.*

(d) *Both E_V^T and E_W^T have X_ℓ as range.*

Proof.

(a) By definition, $v + \mathrm{i}w \in X_\ell^c$ if and only if $v + \mathrm{i}w$ satisfies (4.10). Taking real and imaginary parts of the left- and right-hand sides of (4.10), we see that $v + \mathrm{i}w$ satisfies (4.10) if and only if $(v, w) \in \ker B_\ell$.

(b) follows from (a) and the fact that $(v, w) \simeq v + \mathrm{i}w$ and $\mathrm{i}(v + \mathrm{i}w) \sim (-w, v)$ are linearly independent over $\mathbb{R}$.

(c) follows from the definition of E_ℓ.

(d) follows from (a) and (b). ∎

Second step of the algorithm. Again we consider two cases: $\ell = 2$, $\ell > 2$.

Case 1: $\ell = 2$. We solve an extended system

$$F(x_{\mathrm{red}}, \lambda, v_{\mathrm{red}}, w_{\mathrm{red}}, \omega) = 0,$$

similarly to (4.2), where now

$$v_{\mathrm{red}}, w_{\mathrm{red}} \in X_\ell \simeq \mathbb{R}^d, \quad x_{\mathrm{red}} \in \mathrm{Fix}(L) \simeq \mathbb{R}^{n_{\mathrm{red}}}$$

and

$$F : \mathbb{R}^{n_{\mathrm{red}}+2d+2} \to \mathbb{R}^{n_{\mathrm{red}}+2d+1}.$$

Let $f_{\mathrm{red}} := f|_{\mathrm{Fix}(L)}$ and let $Q : \mathbb{R}^n \to \mathbb{R}^{n_{\mathrm{red}}}$ be the matrix which contains an orthonormal basis of $\mathrm{Fix}(L)$, $\dim \mathrm{Fix}(L) = n_{\mathrm{red}}$, as row vectors. Then we define F as

$$F(x_{\mathrm{red}}, \lambda, v_{\mathrm{red}}, w_{\mathrm{red}}, \omega) = \begin{pmatrix} f_{\mathrm{red}}(x_{\mathrm{red}}, \lambda) \\ E_2 \mathrm{D}_x f(x, \lambda)|_{x=Q^T x_{\mathrm{red}}} E_2^T v_{\mathrm{red}} + \omega w_{\mathrm{red}} \\ -\omega v_{\mathrm{red}} + E_2 \mathrm{D}_x f(x, \lambda)|_{x=Q^T x_{\mathrm{red}}} E_2^T w_{\mathrm{red}} \\ \langle v_{\mathrm{red}}, v_{\mathrm{red}} \rangle + \langle w_{\mathrm{red}}, w_{\mathrm{red}} \rangle - 1 \end{pmatrix}. \tag{4.13}$$

Case 2: $\ell > 2$. In this case the real part v^* of the Hopf eigenvector $v^* + \mathrm{i}w^*$ and the knowledge of the drift symmetry α of the branch of periodic orbits along which a Hopf

bifurcation was detected determine the Hopf eigenvector uniquely since $v^*+\mathrm{i}w^*$ satisfies (4.10) and so

$$w^* = \frac{1}{\sin(2\pi/\ell)}\left(\alpha v^* - \cos\left(\frac{2\pi}{\ell}\right)v^*\right).$$

Therefore only the real part of the Hopf eigenvector, the Hopf point, its parameter, and the Hopf frequency need to be computed by an extended system. We define $F : \mathbb{R}^{n_{\mathrm{red}}+d+2} \to \mathbb{R}^{n_{\mathrm{red}}+d+1}$ by

$$F(x_{\mathrm{red}}, \lambda, v_{\mathrm{red}}, \omega) = \begin{pmatrix} f_{\mathrm{red}}(x_{\mathrm{red}}, \lambda) \\ E_V(\mathrm{D}_x f(x,\lambda)|_{x=Q^T x_{\mathrm{red}}} E_V^T v_{\mathrm{red}} + \omega E_W^T v_{\mathrm{red}}) \\ \langle v_{\mathrm{red}}, v_{\mathrm{red}}\rangle - 1 \end{pmatrix}. \tag{4.14}$$

Proposition 4.9.

(a) *If $\ell = 2$ and $(x^*_{\mathrm{red}}, \lambda^*, v^*_{\mathrm{red}}, w^*_{\mathrm{red}}, \omega^*)$ is a solution to $F = 0$ as defined in (4.13), then $x^* = Q^T x^*_{\mathrm{red}}$ is an equivariant Hopf point with Hopf frequency ω^*. Moreover, a Hopf eigenvector $v^* + \mathrm{i}w^*$ with $v^*, w^* \in N^L_{\omega^*}$ is given by $v^* = E_2^T v^*_{\mathrm{red}}$, $w^* = E_2^T w^*_{\mathrm{red}}$.*

(b) *Similarly, if $\ell > 2$ and $(x^*_{\mathrm{red}}, \lambda^*, v^*_{\mathrm{red}}, \omega^*)$ is a solution to $F = 0$ as defined in (4.14), then $x^* = Q^T x^*_{\mathrm{red}}$ is an equivariant Hopf point with Hopf frequency ω^*. Moreover, a Hopf eigenvector $v^* + \mathrm{i}w^*$ with $v^*, w^* \in N^L_{\omega^*}$ is given by $v^* = E_V^T v^*_{\mathrm{red}}$, $w^* = E_W^T v^*_{\mathrm{red}}$.*

For the proof we need the following lemma.

Lemma 4.10. *Let A be an (n,n)-matrix which is equivariant with respect to a linear $\mathbb{Z}_\ell$-action on $\mathbb{R}^n$. Let α generate $\mathbb{Z}_\ell$ and define X_ℓ as before. Then X_ℓ is A-invariant.*

Proof. Let $v \in X_\ell$. Then there is some $w \in X_\ell$ such that $v + \mathrm{i}w$ is an eigenvector of α to the eigenvalue $\exp(-2\pi\mathrm{i}/\ell)$. By the $\mathbb{Z}_\ell$-equivariance of A, also $A(v + \mathrm{i}w)$ is an eigenvector of α to the eigenvalue $\exp(-2\pi\mathrm{i}/\ell)$, and so both Av and Aw lie in X_ℓ. ∎

Note that X_ℓ is an *isotypic component* of the $\mathbb{Z}_\ell$-action on $\mathbb{R}^n$, and that generally isotypic components for a linear action of a group Γ are invariant under Γ-equivariant matrices [11].

Proof of Proposition 4.9.

(a) *Case $\ell = 2$.* The first equation $f_{\mathrm{red}}(x, \lambda) = 0$ of $F = 0$ implies that x^*_{red} is an equilibrium of $f_{\mathrm{red}}(\cdot, \lambda^*)$. Hence $x^* = Q^T x^*_{\mathrm{red}}$ is an L-invariant equilibrium of $f(\cdot, \lambda^*)$. From the other equations in $F = 0$ we conclude that $v^*_{\mathrm{red}} + \mathrm{i}w^*_{\mathrm{red}}$ is an eigenvector of $E_2 \mathrm{D}_x f(x^*, \lambda^*) E_2^T$ to the eigenvalue $\mathrm{i}\omega^*$. Since $x^* \in \mathrm{Fix}(L)$ the derivative $\mathrm{D}_x f(x^*, \lambda^*)$ is L-equivariant by Lemma 4.3 and therefore, by Lemma 4.10, maps X_2 into itself. Let $v^* = E_2^T v^*_{\mathrm{red}}$, $w^* = E_2^T w^*_{\mathrm{red}}$. Because of (4.11), $v^* + \mathrm{i}w^* \in X_2^c$ is an eigenvector of $\mathrm{D}_x f(x^*, \lambda^*)$ to the eigenvalue $\mathrm{i}\omega^*$. Hence $(x^* = Q^T x^*_{\mathrm{red}}, \lambda^*)$ is a Hopf point. Lemma 4.7 now implies that $v^*, w^* \in N^L_{\omega^*}$.

(b) *Case $\ell > 2$.* As in the case $\ell = 2$, the first equation of $F = 0$ implies that $x^* = Q^T x^*_{\mathrm{red}}$ is an L-invariant equilibrium of $f(\cdot, \lambda^*)$. Let $v^* = E_V^T v^*_{\mathrm{red}}$ and $w^* = E_W^T w^*_{\mathrm{red}}$. From Lemma 4.8(c) we conclude that $v^* + \mathrm{i}w^* \in X_\ell^c$. Since $\mathrm{D}_x f(Q^T x_{\mathrm{red}}(\lambda), \lambda)$ is L-equivariant by Lemma 4.3 and hence maps X_ℓ into itself by Lemma 4.10, and since $E_V|_{X_\ell}$ is an isomorphism by Lemma 4.8(d), the other equations in $F = 0$ imply that

$$\mathrm{D}_x f(x^*, \lambda^*)v^* + \omega^* w^* = 0, \tag{4.15}$$

and so

$$\mathrm{Re}(\mathrm{D}_x f(x^*, \lambda^*)(v^* + \mathrm{i}w^*) - \mathrm{i}\omega^*(v^* + \mathrm{i}w^*)) = 0.$$

Multiplying (4.15) by α and using the L-equivariance of $\mathrm{D}_x f(x^*, \lambda^*)$, we obtain

$$\mathrm{D}_x f(x^*, \lambda^*)\alpha v^* + \omega^* \alpha w^* = 0.$$

From the fact that $v^* + \mathrm{i}w^* \in X_\ell^c$ we conclude

$$\mathrm{Re}\left(\mathrm{e}^{-2\pi \mathrm{i}/\ell}(\mathrm{D}_x f(x^*, \lambda^*)(v^* + \mathrm{i}w^*) - \mathrm{i}\omega^*(v^* + \mathrm{i}w^*))\right) = 0.$$

Since $\mathrm{e}^{2\pi \mathrm{i}/\ell} \notin \mathbb{R}$ this implies that

$$\mathrm{D}_x f(x^*, \lambda^*)(v^* + \mathrm{i}w^*) = \mathrm{i}\omega^*(v^* + \mathrm{i}w^*)$$

and therefore (x^*, λ^*) is a Hopf point and $v^* + \mathrm{i}w^* \in X_\ell^c$ a Hopf eigenvector. By Lemma 4.7 we then get $v^*, w^* \in N^L_{\omega^*}$. ∎

Analogously to Theorem 4.2 we have the following.

Theorem 4.11. *If the assumptions of the equivariant Hopf Theorem 4.5 hold and if the initial guess is good enough, then the Gauss–Newton method applied to (4.13) for $\ell = 2$ and applied to (4.14) for $\ell > 2$ converges.*

Proof. As in the nonsymmetric case (Theorem 4.2; see [1, 12, 14]) we show that $\mathrm{D}F$, with F from (4.13) for $\ell = 2$ and from (4.14) for $\ell > 2$, has full rank in the Hopf point. As before we consider the cases $\ell = 2$ and $\ell > 2$ separately.

Case $\ell = 2$. The assumptions of the equivariant Hopf theorem imply that $\mathrm{i}\omega^*$ is a simple eigenvalue of $\mathrm{D}_x f(x^*, \lambda^*)|_{X_2}$ and hence, due to (4.11), also of $E_2 \mathrm{D}_x f(x^*, \lambda^*) E_2^T$. The proof that $\mathrm{D}F$ has full rank in the Hopf point is therefore very similar to the nonsymmetric case: The only difference is that we require $v, w \in X_2$ and $x \in \mathrm{Fix}(L)$. We omit the details.

Case $\ell > 2$. In this case we have

$$\begin{aligned}
&\mathrm{D}_{(x_{\mathrm{red}}, \lambda, v_{\mathrm{red}}, \omega)} F(x_{\mathrm{red}}, \lambda, v_{\mathrm{red}}, \omega) \\
&\quad = \begin{pmatrix}
\mathrm{D}_{x_{\mathrm{red}}} f_{\mathrm{red}}(x_{\mathrm{red}}, \lambda) & \mathrm{D}_\lambda f_{\mathrm{red}}(x_{\mathrm{red}}, \lambda) & 0 & 0 \\
E_V \mathrm{D}_x^2 f E_V^T v_{\mathrm{red}} Q^T & E_V \mathrm{D}_x \mathrm{D}_\lambda f E_V^T v_{\mathrm{red}} & E_V(\mathrm{D}_x f E_V^T + \omega E_W^T) & E_V E_W^T v_{\mathrm{red}} \\
0 & 0 & 2 v_{\mathrm{red}}^T & 0
\end{pmatrix},
\end{aligned}$$

where f is short for $f(Q^T x_{\mathrm{red}}, \lambda)$. Let $t = (t_{x_{\mathrm{red}}}, t_\lambda, t_{v_{\mathrm{red}}}, t_\omega)$ be a kernel vector of $\mathrm{D}F$ in the equivariant Hopf point:

$$\begin{aligned}
0 &= \mathrm{D}_{x_{\mathrm{red}}} f_{\mathrm{red}}(x^*_{\mathrm{red}}, \lambda^*) t_{x_{\mathrm{red}}} + \mathrm{D}_\lambda f_{\mathrm{red}}(x^*_{\mathrm{red}}, \lambda^*) t_\lambda, \\
0 &= E_V \mathrm{D}_x^2 f E_V^T v^*_{\mathrm{red}} Q^T t_{x_{\mathrm{red}}} + E_V \mathrm{D}_x \mathrm{D}_\lambda f E_V^T v^*_{\mathrm{red}} t_\lambda \\
&\quad + E_V(\mathrm{D}_x f E_V^T + \omega^* E_W^T) t_{v_{\mathrm{red}}} + E_V E_W^T v^*_{\mathrm{red}} t_\omega, \\
0 &= 2\langle v^*_{\mathrm{red}}, t_{v_{\mathrm{red}}} \rangle.
\end{aligned} \tag{4.16}$$

We need to show that $\ker(\mathrm{D}F)$ is one-dimensional. Similarly, as in the proof of convergence for the nonsymmetric case, Theorem 4.2 (see, e.g., [1, 12, 14]), we conclude from the first equation of (4.16) that

$$(t_{x_{\mathrm{red}}}, t_\lambda) = t_\lambda (x'_{\mathrm{red}}(\lambda^*), 1).$$

Since

$$\mathrm{D}_x^2 f(x^*,\lambda^*)x'(\lambda^*) + \mathrm{D}_x\mathrm{D}_\lambda f(x^*,\lambda^*) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)\Big|_{\lambda=\lambda^*},$$

where $x(\lambda) = Q^T x_{\mathrm{red}}(\lambda)$, we deduce from the second equation of (4.16) that

(4.17)
$$E_V\left(t_\lambda \frac{\mathrm{d}}{\mathrm{d}\lambda}\mathrm{D}_x f(x(\lambda),\lambda)\Big|_{\lambda=\lambda^*} E_V^T v_{\mathrm{red}}^* + (\mathrm{D}_x f(x^*,\lambda^*)E_V^T + \omega^* E_W^T)t_{v_{\mathrm{red}}} + t_\omega E_W^T v_{\mathrm{red}}^*\right) = 0.$$

Since $E_V^T v_{\mathrm{red}}^*, E_V^T t_v, E_W^T v_{\mathrm{red}}^*, E_W^T t_v \in X_\ell$ and $E_V|_{X_\ell}$ is an isomorphism by Lemma 4.8(d) and since $\mathrm{D}_x f(x(\lambda),\lambda)$ maps X_ℓ into itself by Lemmata 4.3 and 4.10, (4.17) implies that

$$t_\lambda \frac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)\Big|_{\lambda=\lambda^*} E_V^T v_{\mathrm{red}}^* + \mathrm{D}_x f(x^*,\lambda^*)E_V^T t_v + \omega^* E_W^T t_v + t_\omega E_W^T v_{\mathrm{red}}^* = 0.$$

Let $t_v = E_V^T t_{v_{\mathrm{red}}}$ and $t_w = E_W^T t_{v_{\mathrm{red}}}$. By Lemma 4.8(c) we have $t_v + \mathrm{i}t_w \in X_\ell^c$, which we will need later on. Denote $v^* = E_V^T v_{\mathrm{red}}^*$, $w^* = E_W^T v_{\mathrm{red}}^*$ such that, by Proposition 4.9(b), $v^* + \mathrm{i}w^*$ is a Hopf eigenvector. Then we get

$$\begin{aligned}
0 &= t_\lambda \frac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)\Big|_{\lambda=\lambda^*} v^* + \mathrm{D}_x f(x^*,\lambda^*)t_v + \omega^* t_w + t_\omega w^* \\
&= \mathrm{Re}\left(t_\lambda \frac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)\Big|_{\lambda=\lambda^*} (v^* + \mathrm{i}w^*) + \mathrm{D}_x f(x^*,\lambda^*)(t_v + \mathrm{i}t_w)\right) \\
&\quad - \mathrm{Re}\left(\mathrm{i}\omega^*(t_v + \mathrm{i}t_w) + \mathrm{i}t_\omega(v^* + \mathrm{i}w^*)\right).
\end{aligned}$$

By Lemma 4.8(c), $v^* + \mathrm{i}w^* \in X_\ell^c$. Therefore

$$\begin{aligned}
0 &= \alpha t_\lambda \tfrac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)|_{\lambda=\lambda^*} v^* + \alpha \mathrm{D}_x f(x^*,\lambda^*)t_v + \omega^* \alpha t_w + t_\omega \alpha w^* \\
&= t_\lambda \tfrac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)|_{\lambda=\lambda^*} \alpha v^* + \mathrm{D}_x f(x^*,\lambda^*)\alpha t_v + \omega^* \alpha t_w + t_\omega \alpha w^* \\
&= \mathrm{Re}\left(\mathrm{e}^{-2\pi\mathrm{i}/\ell}(t_\lambda \tfrac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)|_{\lambda=\lambda^*}(v^* + \mathrm{i}w^*) + \mathrm{D}_x f(x^*,\lambda^*)(t_v + \mathrm{i}t_w))\right) \\
&\quad - \mathrm{Re}\left(\mathrm{e}^{-2\pi\mathrm{i}/\ell}(\mathrm{i}\omega^*(t_v + \mathrm{i}t_w) - \mathrm{i}t_\omega(v^* + \mathrm{i}w^*))\right),
\end{aligned}$$

where we used the L-equivariance of $\mathrm{D}_x f(x(\lambda),\lambda)$ (Lemma 4.3) in the second line. Since $\mathrm{e}^{-2\pi\mathrm{i}/\ell} \notin \mathbb{R}$ for $\ell > 2$ these last two equations imply that

(4.18)
$$\begin{aligned}
0 &= t_\lambda \tfrac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathrm{D}_x f(x(\lambda),\lambda)|_{\lambda=\lambda^*}(v^* + \mathrm{i}w^*) \\
&\quad + \mathrm{D}_x f(x^*,\lambda^*)(t_v + \mathrm{i}t_w) - \mathrm{i}\omega^*(t_v + \mathrm{i}t_w) - \mathrm{i}t_\omega(v^* + \mathrm{i}w^*).
\end{aligned}$$

Let $v(\lambda) + \mathrm{i}w(\lambda)$, $v(\lambda^*) = v^*$, $w(\lambda^*) = w^*$, $v(\lambda), w(\lambda) \in X_\ell$, be the path of eigenvectors of $\mathrm{D}_x f(x(\lambda),\lambda)$ to the eigenvalues $\mu(\lambda)$ with $\|v(\lambda)\| = 1$. Let $u^* = v^* + \mathrm{i}w^*$, and let $u_L^* \in X_\ell^c$ be the uniquely determined left eigenvector of $\mathrm{D}_x f(x^*,\lambda^*)$ to the eigenvalue $\mathrm{i}\omega^*$ with $\langle u_L^*, u^*\rangle = 1$. As in the nonsymmetric case (see [1, 14, 12]), (4.18) then implies that

$$t_\lambda u_L^* \frac{\mathrm{d}}{\mathrm{d}\lambda}\mathrm{D}_x f(x(\lambda),\lambda)_{\lambda=\lambda^*} u^* = \mathrm{i}t_\omega.$$

This combined with

$$\left\langle u_L^*, \frac{\mathrm{d}}{\mathrm{d}\lambda} \mathrm{D}_x f(x(\lambda), \lambda)_{\lambda=\lambda^*} u^* \right\rangle = \mu'(\lambda^*)$$

and the assumption that $\operatorname{Re}\mu'(\lambda^*) \neq 0$ gives $t_\lambda = 0$. Hence $t_x = 0$ and $t_\omega = 0$. As in the proof of Theorem 4.2, (4.18) reduces to

$$\mathrm{D}_x f(x^*, \lambda^*)(t_v + \mathrm{i}t_w) - \mathrm{i}\omega^*(t_v + \mathrm{i}t_w) = 0$$

so that $t_v + \mathrm{i}t_w$ is a Hopf eigenvector. Since $t_v + \mathrm{i}t_w \in X_\ell^c$ we deduce from Lemma 4.7 that $t_v + \mathrm{i}t_w = c(v^* + \mathrm{i}w^*)$ for some $c \in \mathbb{C}$. By the definition of v^*, w^* and t_v, t_w this is equivalent to

$$E_V^T t_{v_{\mathrm{red}}} + \mathrm{i}E_W^T t_{v_{\mathrm{red}}} = c(E_V^T v_{\mathrm{red}}^* + \mathrm{i}E_W^T v_{\mathrm{red}}^*). \tag{4.19}$$

By Lemma 4.8 the map $E_V E_V^T \in \mathrm{Mat}(d)$ is invertible. Therefore we conclude from (4.19) that $t_{v_{\mathrm{red}}} \in \operatorname{span}(v_{\mathrm{red}}^*, w_{\mathrm{red}}^*)$, where $w_{\mathrm{red}}^* := (E_V E_V^T)^{-1} E_V E_W^T v_{\mathrm{red}}^*$. From the last row of $\mathrm{D}F$ we get the additional condition $\langle v_{\mathrm{red}}^*, t_{v_{\mathrm{red}}} \rangle = 0$ on $t_{v_{\mathrm{red}}}$. Hence the kernel of $\mathrm{D}F$ is one-dimensional and $\mathrm{D}F$ has full rank in the Hopf point. ■

Initial guess for the Gauss–Newton iteration. As in the nonsymmetric case (see section 4.1.2), we use Hermite interpolation between two consecutive periodic orbits $y^{(0)}$ and $y^{(1)}$, between which a Hopf bifurcation was detected, and determine the point $\hat{\tau}$ on the interpolating polynomial $y(\tau) = (x(\tau), T(\tau), \lambda(\tau))$ with $\lambda'(\hat{\tau}) = 0$. This way we obtain an initial approximation $\hat{x} = x(\hat{\tau})$ for the Hopf point, an initial approximation $\hat{\lambda} = \lambda(\hat{\tau})$ for its parameter, and an initial approximation $\hat{\omega} = \frac{2\pi}{\hat{T}}$ for its frequency, where $\hat{T} = T(\hat{\tau})$. If $\ell = 2$, we let

$$\hat{v} = c\, E_2 \frac{\mathrm{d}}{\mathrm{d}\tau} x_1(\hat{\tau}), \quad \hat{w} = -c \frac{1}{\hat{\omega}} \mathrm{D}_x f(\hat{x}, \hat{\lambda}) \frac{\mathrm{d}}{\mathrm{d}\tau} x_1(\hat{\tau}),$$

where c is such that $\langle \hat{v}, \hat{v} \rangle + \langle \hat{w}, \hat{w} \rangle = 1$, and take the point $(\hat{x}_{\mathrm{red}} = Q\hat{x}, \hat{\lambda}, \hat{v}, \hat{w}, \hat{\omega})$ as initial guess for the Gauss–Newton iteration applied to (4.13). If $\ell > 2$, we define

$$\hat{v} = c E_V \frac{\mathrm{d}}{\mathrm{d}\tau} x_1(\hat{\tau})$$

with $c \in \mathbb{R}$ such that $\langle \hat{v}, \hat{v} \rangle = 1$ and use $(\hat{x}_{\mathrm{red}}, \hat{\lambda}, \hat{v}, \hat{\omega})$ as initial guess for the Gauss–Newton iteration applied to (4.14).

5. Applications.

In this section we illustrate with some examples the numerical methods for the continuation of symmetric periodic orbits which we presented in the preceding sections and implemented in the code SYMPERCON [20].

5.1. The Lorenz model—Comparison with AUTO and CONTENT.

There are many programs for the numerical continuation of periodic orbits of nonsymmetric systems. Two of the most well known and widely used programs are AUTO [8] and CONTENT [17], the numerical part of which is based on AUTO. In both programs collocation is used to find

periodic solutions. This approach is equivalent to the multiple shooting ansatz with the multiple shooting points being the grid points and the IVP solver having only one step.

One of the key new features of SYMPERCON compared to those program packages is the exploitation of symmetries of periodic orbits and the computation of symmetry breaking and symmetry increasing bifurcations, and we present some applications of these methods in the next sections. In this section we give numerical evidence that the program package SYMPERCON is also competitive when applied to continuation of nonsymmetric periodic orbits. We compare the programs AUTO, CONTENT, and SYMPERCON using an example which is taken from the CONTENT Tutorial (ver 1.4) [17] "ODEs: Lorenz system: Continuation of limit cycles and branch switching." Starting with the Lorenz system

$$\begin{aligned} x_1' &= -\sigma x_1 + \sigma x_2, \\ x_2' &= -x_1 x_3 + \lambda x_1 - x_2, \\ x_3' &= x_1 x_2 - b x_3, \end{aligned}$$

where

$$\sigma = 10, \qquad b = 8/3,$$

all programs find a periodic solution for the initial guess

$$(x^*, T^*, \lambda^*) = ((16.2, 57.4, 250.8), 0.411, 312)$$

from [17]. We continue this limit cycle with respect to the parameter λ with required relative error $\mathrm{tol} = 10^{-5}$. The Lorenz system has the $\mathbb{Z}_2$-symmetry $(x_1, x_2, x_3) \to (-x_1, -x_2, x_3)$, and the above periodic orbit is $\mathbb{Z}_2$-symmetric. All programs detect a bifurcation of this periodic orbit at $\lambda = 312.97$. This is a symmetry breaking flip pitchfork bifurcation, so the bifurcating periodic orbits are nonsymmetric. The bifurcating branch of nonsymmetric periodic orbits, which we take as the primary branch in this comparison, subsequently undergoes a period doubling cascade, and we compared AUTO, CONTENT, and SYMPERCON in their performance computing this bifurcation cascade.

While SYMPERCON automatically doubles the number of multiple shooting points at each flip bifurcation up to a given maximum, in AUTO and CONTENT the entered number of grid points has to be set by the user. Since this number of grid points has to be increased at period doubling bifurcations—otherwise the Newton method fails to converge—we doubled the number manually. However, after the fourth period doubling we did not manage to configure CONTENT in such a way that it would find any more period doubling points. With AUTO we found a fifth period doubling bifurcation. In comparison, SYMPERCON found periodic orbits of 2^6 times the original period, with period doubling bifurcations starting from the primary branch at parameters $\lambda = 229.41$, $\lambda = 218.21$, $\lambda = 215.97$, $\lambda = 215.49$, $\lambda = 215.39$, and $\lambda = 215.37$ (see Figure 4), and (when run with the compiler gcc under linux) even a seventh period doubling point at $\lambda = 215.36$. All programs were very sensitive to changes of the parameters of the computation (like the number of initial grid points m, the initial continuation steplength, etc.) and the choice of C compiler used.

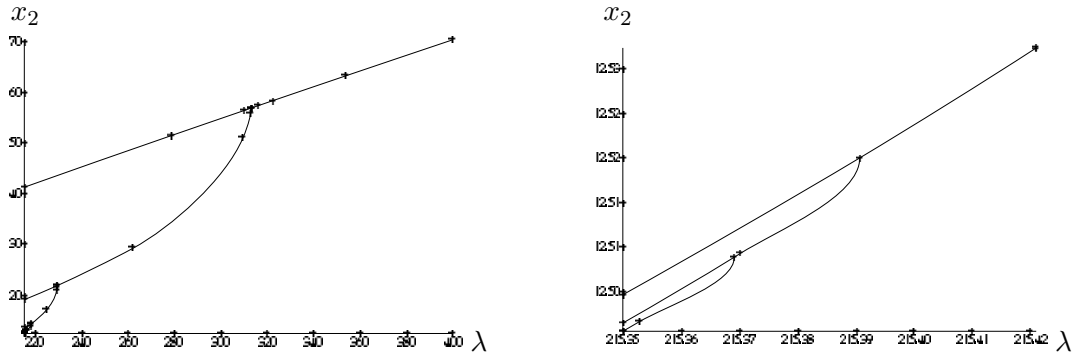


Figure 4. *Feigenbaum cascade in the Lorenz model as computed by SYMPERCON. The right figure is a zoom of the lower-left corner of the left figure.*

5.2. Symmetry breaking bifurcations in coupled cells. In this section we apply our methods to compute equivariant Hopf points and symmetry breaking flip pitchfork bifurcations in the Brusselator model of coupled cells, with parameters as in [3]. We consider four identical cells in which the same reaction takes place and which are coupled by diffusion. The equations for the 4-cell Brusselator are the following ($j = 1, 3, 5, 7$, $x_i := x_{i-8}$ for $i > 8$):

$$\begin{aligned} x_j' &= A - (B+1)x_j + x_j^2 x_{j+1} + \lambda(-3x_j + x_{j+2} + x_{j+4} + x_{j+6})/1000, \\ x_{j+1}' &= Bx_j - x_j^2 x_{j+1} + \lambda(-3x_{j+1} + x_{j+3} + x_{j+5} + x_{j+7})/1000, \end{aligned}$$

where $A = 2.0$, $B = 5.9$. The problem is invariant with respect to permutations of the cells, and thus the equations are S_4-equivariant. We denote the elements of $\Gamma = S_4$ by γ_{ijkl}, e.g.,

$$\gamma_{2314} = \begin{pmatrix} 0 & 0 & \mathrm{id} & 0 \\ \mathrm{id} & 0 & 0 & 0 \\ 0 & \mathrm{id} & 0 & 0 \\ 0 & 0 & 0 & \mathrm{id} \end{pmatrix}, \quad \mathrm{id} \in \mathrm{Mat}(2,2).$$

From [3] we took the following equivariant Hopf point:

$$(x^*, \lambda^*) = (2.0, 2.95, 2.0, 2.95, 2.0, 2.95, 2.0, 2.95, 20.45).$$

The corresponding imaginary eigenvalue $\omega^*\mathrm{i}$ has the value $\omega^* = 0.62058$ and belongs to an absolutely irreducible three-dimensional representation; cf. Remark 2.4. The real eigenspace N_{ω^*} of $\omega^*\mathrm{i}$ is of the form $N_{\omega^*} = V \oplus W$ with

$$\begin{aligned} V &= \{(v_1, 0, v_2, 0, v_3, 0, -v_1 - v_2 - v_3, 0) \mid v_1, v_2, v_3 \in \mathbb{R}\}, \\ W &= \{(0, w_1, 0, w_2, 0, w_3, 0, -w_1 - w_2 - w_3) \mid w_1, w_2, w_3 \in \mathbb{R}\}. \end{aligned}$$

Let e_j^V, e_j^W be the jth unit vectors of V and W, respectively. Then the vectors $e_j^V + \mathrm{i}e_j^W$, $j = 1, 2, 3$, are eigenvectors to the Hopf eigenvalues $\mathrm{i}\omega^*$:

$$\mathrm{D}_x f(x^*, \lambda^*)(e_j^V + \mathrm{i}e_j^W) = \mathrm{i}\omega^*(e_j^V + \mathrm{i}e_j^W), \quad j = 1, \ldots, 3. \tag{5.1}$$

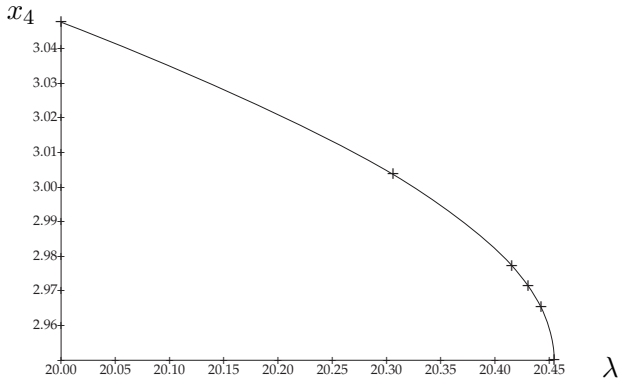


Figure 5. *Equivariant Hopf bifurcation of $\mathbb{Z}_4$-symmetric periodic orbits of the Brusselator.*

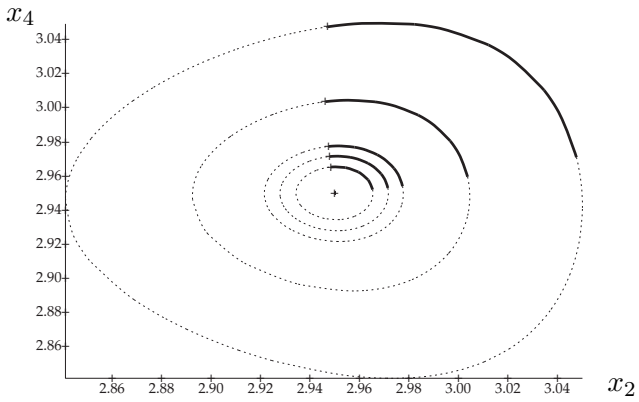


Figure 6. *$\mathbb{Z}_4$-symmetric periodic orbits near an equivariant Hopf point of the Brusselator. The bold parts of the trajectories are computed numerically.*

Using the equivariant Hopf theorem (Theorem 4.5), we can now compute the symmetries and initial planes of the emanating periodic solutions, as in [3]. As an example we consider periodic orbits with spatio-temporal symmetry group

$$L = \left\{ \mathrm{id}, \left(\gamma_{3142}, \frac{3}{4} \right), \left(\gamma_{4321}, \frac{1}{2} \right), \left(\gamma_{2413}, \frac{1}{4} \right) \right\},$$

which means

$$L = \mathbb{Z}_4, \qquad K = \{\mathrm{id}\}.$$

In order to compute the starting plane $N_{\omega^*}^L$ we need to know how S^1 acts on N_{ω^*}. Let $u = (v, w) \in N_{\omega^*} = V \oplus W$. Due to (5.1) and (4.6) we have for $\theta \in S^1 \simeq \mathbb{R}/\mathbb{Z}$

$$(\mathrm{id}, \theta)u = \begin{pmatrix} \cos(2\pi\theta)v - \sin(2\pi\theta)w \\ \sin(2\pi\theta)v + \cos(2\pi\theta)w \end{pmatrix}.$$

By the equivariant Hopf Theorem 4.5 the starting plane $N_{\omega^*}^L$ is determined by the condition

$$N_{\omega^*}^L = \{u \in N_{\omega^*} \mid (\gamma, \theta)u = u \; \forall \, (\gamma, \theta) \in L\} = \left\{ u \in N_{\omega^*} \mid \left(\gamma_{2413}, \frac{\pi}{2} \right) u = u \right\}.$$

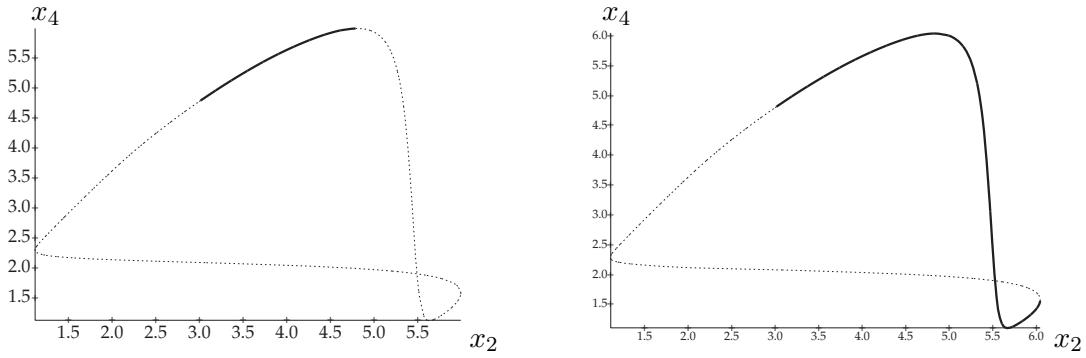


Figure 7. *Symmetric periodic orbits of the 4-cell Brusselator before and after a flip pitchfork bifurcation from $\mathbb{Z}_4$-symmetry to $\mathbb{Z}_2$-symmetry. The bold parts of the trajectories are computed numerically.*

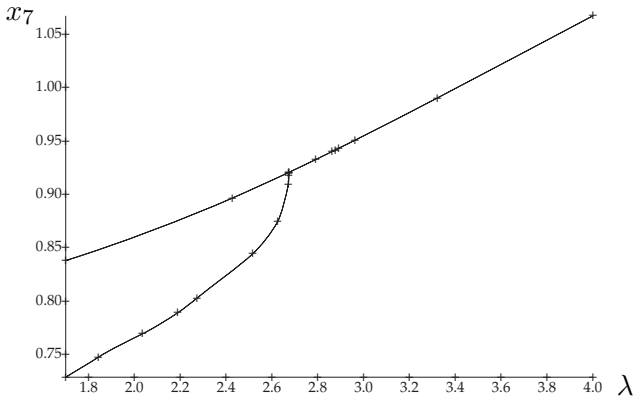


Figure 8. *Flip pitchfork bifurcation from $\mathbb{Z}_4$- to $\mathbb{Z}_2$-symmetry in the 4-cell Brusselator.*

Here we used that $L = \mathbb{Z}_4$ is generated by γ_{2413}. After a short computation we obtain

$$N^L_{\omega^*} = \{(u_1, u_2, -u_2, u_1, u_2, -u_1, -u_1, -u_2) \mid u_1, u_2 \in \mathbb{R}\}.$$

The $\mathbb{Z}_4$-symmetric periodic orbits bifurcating from the equivariant Hopf point are shown in Figures 5 and 6. These solutions have already been computed in [3] by a Galerkin method based on Fourier modes.

Using the methods of section 3, SYMPERCON finds that this branch of $\mathbb{Z}_4$-symmetric periodic orbits undergoes a flip pitchfork bifurcation at $\lambda^* = 2.6738$. A point x^* on the periodic orbit at the symmetry breaking bifurcation point is given by

$$x^* = (1.7041, 3.0132, 0.86523, 4.8152, 3.6590, 1.5591, 0.92057, 6.0440),$$

and the period of this periodic solution is $T^* = 5.0410$. The $\mathbb{Z}_4$-periodic solution close to the symmetry breaking bifurcation (at the parameter $\lambda = 2.7904$) and a $\mathbb{Z}_2$-symmetric periodic orbit after the bifurcation (at the parameter $\lambda = 2.6734$) can be seen in Figure 7; the bifurcation diagram is shown in Figure 8.

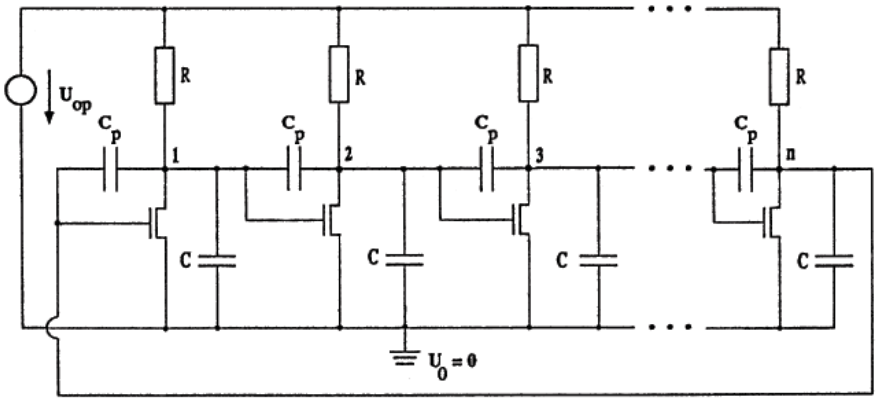


Figure 9. *Circuit of ring oscillator.*

5.3. Electronic ring oscillator. In this section we simulate an electrical circuit consisting of n MOSFET-inverters, taken from Kampowsky, Rentrop, and Schmidt [15]; see Figure 9. Such a circuit can be modeled in the following way: Let U_i be the voltage at the ith node; then using Kirchhoff's law, we get the differential equations

$$f, \tilde{f} : \mathbb{R}^n \to \mathbb{R}^n, \quad \dot{U} = f(U), \quad f(U) = -C^{-1}\tilde{f}(U)$$

for the vector $U = (U_1, \ldots, U_n)$ of the voltages. Here C is the capacity matrix

$$C = \begin{pmatrix} 2C_p + C & -C_p & & -C_p \\ -C_p & 2C_p + C & -C_p & \\ & \ddots & \ddots & \\ -C_p & & -C_p & 2C_p + C \end{pmatrix},$$

and the function f is given by

$$\tilde{f}(U) = \begin{pmatrix} 1/R(U_1 - U_{\text{op}}) + g(U_n, U_1, U_0) \\ 1/R(U_2 - U_{\text{op}}) + g(U_1, U_2, U_0) \\ \vdots \\ 1/R(U_n - U_{\text{op}}) + g(U_{n-1}, U_n, U_0) \end{pmatrix},$$

where

$$g(U_G, U_D, U_S) = K \max\{(U_G - U_S - U_T), 0\}^2 - K \max\{(U_G - U_D - U_T), 0\}^2.$$

The constants $K, U_T, U_0, U_{\text{op}}, R$ and the capacities C and C_p are technical parameters, which have (after appropriate scaling) the values $K = 0.2$, $U_T = 1$, $U_0 = 0$, $U_{\text{op}} = 5$, $R = 5$, $C = 0.21$, $C_p = 5 \cdot 10^{-3}$; see [15]. From [15] we took the initial values of a periodic solution, namely $U_1 = 4.2087$, $U_2 = 0.917$, $U_3 = 2.6534$, $U_4 = 4.2762$, $U_5 = 0.6928$, $T = 6.298$. This periodic solution is shown in Figure 10. Furthermore there is a stationary solution given by

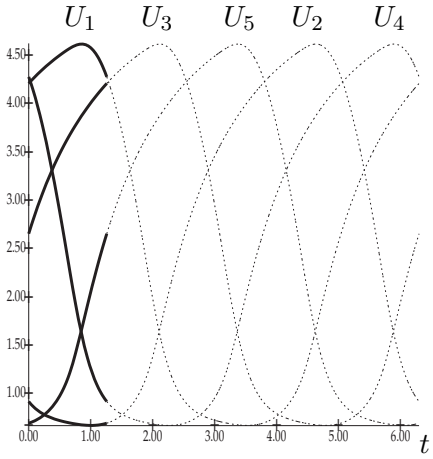


Figure 10. *Periodic solution for* 5 *inverters.*

$U_i(t) \equiv 2.56155$, $i = 1, \ldots, 5$. Our aim is now the numerical simulation of oscillations of large electrical circuits, i.e., the computation of periodic orbits for large n.

The function $f : \mathbb{R}^n \to \mathbb{R}^n$ is $\mathbb{Z}_n$-equivariant where the generating element γ_n of

$$\mathbb{Z}_n = \{\mathrm{id}, \gamma_n, \ldots, \gamma_n^{n-1}\}$$

is acting on $\mathbb{R}^n$ by shifting the components to the right:

$$\gamma_n(x_1, \ldots, x_n) = (x_2, \ldots, x_n, x_1).$$

This representation of $\mathbb{Z}_n$ is called the *regular representation* of $\mathbb{Z}_n$; see [21]. Using SYMPERCON, we checked that the above periodic solution (with $n = 5$) has spatio-temporal symmetry $L = \mathbb{Z}_5$. Since the stationary solution is $\mathbb{Z}_n$-invariant the periodic solution might have branched from the stationary solution via an equivariant Hopf bifurcation. So we introduce U_{op} as a continuation parameter and examine the stationary $\mathbb{Z}_n$-invariant solutions $U = (u, \ldots, u) \in \mathbb{R}^n$, $u \in \mathbb{R}$, which are given by the equation

$$\begin{aligned} &1/R(u - U_{\mathrm{op}}) + g(u, u, U_0) = 0 \\ \iff &1/R(u - U_{\mathrm{op}}) + K(\max(u - U_T, 0))^2 = 0. \end{aligned}$$

In $U = (u, \ldots, u)$ the Jacobian $\mathrm{D}_U f(U)$ is $\mathbb{Z}_n$-symmetric:

$$\gamma\, \mathrm{D}_U f(U) = \mathrm{D}_U f(U)\, \gamma \quad \forall\, \gamma \in \Gamma = \mathbb{Z}_n. \tag{5.2}$$

From elementary representation theory it is known that in a regular representation each irreducible representation ϑ_i is contained n_i times, where n_i is the dimension of the representation ϑ_i; see [21]. Finite cyclic groups $\mathbb{Z}_n$ have only two-dimensional complex irreducible representations (as defined in Remark 2.4). Over $\mathbb{C}$ they are given by

$$\vartheta_j(\gamma_n) = e^{\frac{2\pi \mathrm{i} j}{n}}, \quad j = 0, \ldots, n-1; \tag{5.3}$$

see [21]. Let X_j denote the Γ-irreducible subspace of $\mathbb{R}^n$ belonging to ϑ_j. Then X_j, $j = 0, \dots, n-1$, is spanned by the real part v_j and imaginary part w_j of the vector

$$v_j + \mathrm{i}w_j = (1, e^{\frac{2\pi \mathrm{i} j}{n}}, e^{\frac{2\pi \mathrm{i} 2j}{n}}, \dots, e^{\frac{2\pi \mathrm{i}(n-1)j}{n}}).$$

Using (5.2), we conclude that the Jacobian $\mathrm{D}_U f(U)$ can be diagonalized over $\mathbb{C}$. Here $v_j + \mathrm{i}w_j$ is an eigenvector to the eigenvalue

$$\lambda_j = \frac{1/R + 2K \max(u - U_T, 0) e^{\frac{2\pi \mathrm{i}(n-1)j}{n}}}{C + (2 - 2\cos\frac{2\pi j}{n}) C_p}, \tag{5.4}$$

as can be computed easily. Setting $\mathrm{Re}(\lambda_j) = 0$, we get simple expressions for the stationary solution $u = u^{(j)}$ and the corresponding parameter $U_{\mathrm{op}}^{(j)}$, for which an equivariant Hopf bifurcation can occur, and also for the Hopf frequency ω_j. Using the equivariant Hopf Theorem 4.5, we conclude that every bifurcating periodic solution has the symmetry $L = \mathbb{Z}_n$. However, we also have to determine the homomorphism Θ of the bifurcating periodic orbits and thereby their isotropy K and drift symmetry α. For this we have to know the operation of $\Gamma \times S^1$ on the Hopf eigenspace $N_{\omega^*} = X_j$; see (4.6). We compute that $\Gamma \times S^1$ operates on the irreducible subspace $X_j = \mathrm{span}(v_j, w_j)$ in the following way:

$$(\gamma, \theta)(v_j + \mathrm{i}w_j) = \vartheta_j(\gamma) e^{2\pi \mathrm{i}\theta}(v_j + \mathrm{i}w_j), \tag{5.5}$$

where $\gamma \in \Gamma$, $\theta \in S^1 \equiv \mathbb{R}/\mathbb{Z}$. We can characterize the homomorphism Θ_j of the bifurcating periodic orbits tangential to $N_{\omega^*} = X_j$ by the homomorphism $\Theta_j(\gamma_n)$ applied to the generating element γ_n of $\mathbb{Z}_n$. From (5.5) and (5.3) we get

$$\Theta_j(\gamma_n) = -j/n.$$

Let r be the greatest common divisor of n and j and $n_{\mathrm{red}} := n/r$. Then K is generated by $\gamma_n^{n_{\mathrm{red}}}$, and the fixed point space $\mathrm{Fix}(K)$ contains all vectors in $\mathbb{R}^n$ which consist of r equal sections of length n_{red}. Therefore n_{red} is the dimension of the reduced system.

After restriction onto the fixed point space $\mathrm{Fix}(K) = \mathbb{R}^{n_{\mathrm{red}}}$ we obtain $L = \mathbb{Z}_{n_{\mathrm{red}}}$, $K = \{\mathrm{id}\}$. Then $\gamma_{n_{\mathrm{red}}} \equiv \gamma_n K$ is a generating element of $\mathbb{Z}_{n_{\mathrm{red}}}$. The group $\mathbb{Z}_{n_{\mathrm{red}}}$ also acts on $\mathbb{R}^{n_{\mathrm{red}}}$ by cyclically permuting the components of every $x \in \mathbb{R}^{n_{\mathrm{red}}}$. The representation $\vartheta_{j_{\mathrm{red}}}$ with $j_{\mathrm{red}} := j/r$ is the irreducible representation of $\mathbb{Z}_{n_{\mathrm{red}}}$ on $\mathbb{R}^{n_{\mathrm{red}}}$, which corresponds to ϑ_j. Thus,

$$N_{\omega^*}^{\mathrm{red}} = \langle V_{j_{\mathrm{red}}}, W_{j_{\mathrm{red}}} \rangle \subset \mathbb{R}^{n_{\mathrm{red}}}$$

is the initial plane for the periodic solutions in the reduced coordinates. Finally we have to determine the element $\alpha \in \mathbb{Z}_{n_{\mathrm{red}}}$ with the smallest phase shift $\Theta_{\mathrm{red}}(\alpha)$. Let $\alpha = \gamma_{n_{\mathrm{red}}}^k$, where k is an integer between 1 and $n_{\mathrm{red}} - 1$. Then α is given by the equation $\Theta_{\mathrm{red}}(\alpha) = 1/n_{\mathrm{red}}$, i.e.,

$$\vartheta_{j_{\mathrm{red}}}(\alpha) e^{2\pi \mathrm{i}/n_{\mathrm{red}}} = 1 \;\Leftrightarrow\; j_{\mathrm{red}}\, k \equiv 1 \pmod{n_{\mathrm{red}}}.$$

By the method just described we can compute initial values for periodic solutions for arbitrary n, and by exploiting symmetry we can reduce the amount of work effectively: If the considered periodic solution has large spatial symmetry, then the ODE is reduced to a system with small

Table 1

Amount of work for the computation of symmetric periodic orbits with trivial isotropy of the electrical circuit depending on the number n of inverters.

	With exploitation of symmetry	Without exploitation of symmetry
Flow $\Phi_t(x)$	$O(n)$	$N_{\text{step}} * O(n)$
Wronskian matrix $\mathrm{D}_x\Phi_t(x)$	$O(n)$	$N_{\text{step}} * O(n^2)$
Linear algebra	$O(n)$	$O(n^3)$

dimension n_{red}. For example, in the case $n = 1000$, there is a periodic solution with $K = \mathbb{Z}_{200}$, $n_{\text{red}} = 5$, which corresponds to the above computed solution with $n = 5$.

If the periodic solution does not have any spatial symmetry, but only spatio-temporal symmetry, then only the integration interval can be reduced by exploitation of symmetry. Let N_{step} be the number of integration steps for the full integration interval $[0, T]$. If the symmetry is not exploited, $N_{\text{step}} * O(n)$ multiplications and divisions are needed for the computation of the flow; exploiting the symmetry, we can reduce the number of integration steps from N_{step} to N_{step}/n, since the integration interval is only $[0, T/n]$. Hence for large n the number of time steps is $O(1)$, and so the number of multiplications and divisions for the computation of the flow is $O(n)$ when symmetry is exploited. Since the derivatives $\mathrm{D}_x f(x)$ are tridiagonal matrices and hence contain only $O(n)$ nonzero components, $N_{\text{step}} * O(n^2)$ multiplications are needed for the computation of the Wronskian matrices $\mathrm{D}_x\Phi_t(x)$ by integration of the variational equation if the symmetry is not exploited; if the symmetry is exploited, then for large n only a few steps of integration are necessary for the computation of the Wronskian matrices. Therefore the Wronskian matrices are sparse and contain only nonzero entries close to the diagonal. Hence by exploitation of symmetry the amount of work for the computation of the Wronskian reduces to $O(n)$. For the solution of the linear equations which arise in the Gauss–Newton method, $O(n^3)$ multiplications are needed if the symmetry is not exploited, and the amount of work for the linear algebra reduces to $O(n)$ due to the sparse structure of the Wronskians if the symmetry is exploited. Altogether, we obtain a remarkable reduction of the computational cost by exploitation of symmetry; cf. Table 1.

Conclusion and future directions. In this paper we have presented efficient algorithms for the computation of generic symmetry changing but isotropy preserving bifurcations of periodic orbits in systems where the symmetry group is *discrete*. These bifurcations were analyzed by Golubitsky, Stewart, and Schaeffer (equivariant Hopf bifurcation; see, e.g., [11]) and Fiedler (flip doubling and flip pitchfork bifurcations; see [9]). General symmetry breaking bifurcations of periodic orbits in systems with discrete symmetry group were classified by Lamb and Melbourne [18]; see also [19] and [24] for the case of continuous symmetry groups. In future work we will extend our numerical results to compute symmetry changing bifurcations of periodic orbits of arbitrary Lie group actions. In the paper [26] we extend the methods presented here to Hamiltonian systems with continuous symmetries, building on the persistence results of [25].

Acknowledgments. This paper is dedicated to the memory of Karin Gatermann who died on New Year's Day 2005. She pioneered the use of symmetry methods in numerical bifurcation theory and introduced one of us (CW) to this topic back in 1991. Karin gave us lots of valuable

advice and encouragement and was a dear friend.

CW thanks the Freie Universität Berlin, particularly the research group of Bernold Fiedler, for their hospitality during visits when parts of this paper were written. We both thank Mark Roberts for many helpful discussions and for inviting us to the Warwick Symposium on Geometric Mechanics and Symmetry in 2001/2002, during which this project was started. Finally we acknowledge the EU-TMR network MASIE for funding our participation in various summer schools, workshops, and conferences.

REFERENCES

[1] W.-J. Beyn, A. Champneys, E. Doedel, W. Govaerts, Y. A. Kuznetsov, and B. Sandstede, *Numerical continuation, and computation of normal forms*, in Handbook of Dynamical Systems, Vol. 2, B. Fiedler, ed., Elsevier–North Holland, Amsterdam, 2002, Chap. 4, pp. 149–219.

[2] K. A. Cliffe, A. Spence, and S. J. Tavener, O(2)*-symmetry breaking bifurcation: With application to the flow past a sphere in a pipe*, Internat. J. Numer. Methods Fluids, 32 (2000), pp. 175–200.

[3] M. Dellnitz, *A computational method and pathfollowing for periodic solutions with symmetry*, in Continuation and Bifurcations: Numerical Techniques and Applications, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 313, D. Roose, B. de Dier, and A. Spence, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990, pp. 153–167.

[4] M. Dellnitz, *Computational bifurcation of periodic solutions in systems with symmetry*, IMA J. Numer. Anal., 12 (1992), pp. 429–455.

[5] P. Deuflhard, *Computation of periodic solutions of nonlinear ODE's*, BIT, 24 (1984), pp. 456–466.

[6] P. Deuflhard, *Recent progress in extrapolation methods for ordinary differential equations*, SIAM Rev., 27 (1985), pp. 505–535.

[7] P. Deuflhard, B. Fiedler, and P. Kunkel, *Efficient numerical pathfollowing beyond critical points*, SIAM J. Numer. Anal., 24 (1987), pp. 912–927.

[8] E. J. Doedel, R. C. Paffenroth, A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. E. Oldeman, B. Sandstede, and X.-J. Wang, *AUTO*2000*: Continuation and Bifurcation Software for Ordinary Differential Equations*, technical report, Department of Computer Science, Concordia University, Montreal, QC, 2000, available from http://indy.cs.concordia.ca/auto/.

[9] B. Fiedler, *Global Bifurcation of Periodic Solutions with Symmetry*, Lecture Notes in Math., Springer-Verlag, Berlin, Heidelberg, New York, 1988.

[10] K. Gatermann and A. Hohmann, *Symbolic exploitation of symmetry in numerical pathfollowing*, IMPACT Comput. Sci. Engrg., 3 (1991), pp. 330–365.

[11] M. Golubitsky, I. Stewart, and D. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. 2, Springer-Verlag, Berlin, Heidelberg, New York, 1988.

[12] A. Griewank and G. W. Reddien, *The calculation of Hopf points by a direct method*, IMA J. Numer. Anal., 3 (1983), pp. 295–303.

[13] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Appl. Math. Sci. 42, Springer-Verlag, New York, Heidelberg, Berlin, 1990.

[14] A. D. Jepson, *Part* II: *Numerical Hopf Bifurcation*, Ph.D. thesis, Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA, 1981.

[15] W. Kampowsky, P. Rentrop, and W. Schmidt, *Classification and numerical simulation of electric circuits*, Surveys Math. Indust., 2 (1992), pp. 23–65.

[16] Y. Kuznetsov, *Elements of Applied Bifurcation Theory*, Appl. Math. Sci. 112, Springer, New York, 1995.

[17] Y. A. Kuznetsov and V. V. Levitin, *CONTENT: A Multiplatform Environment for Continuation and Bifurcation Analysis of Dynamical Systems*, Centrum voor Wiskunde en Informatica, 1997, available from ftp://ftp.cwi.nl/pub/CONTENT.

[18] J. S. W. Lamb and I. Melbourne, *Bifurcation from discrete rotating waves*, Arch. Ration. Mech. Anal., 149 (1999), pp. 229–270.

[19] J. Lamb, I. Melbourne, and C. Wulff, *General bifurcations from periodic solutions with spatiotemporal symmetry, including mode interactions and resonances*, J. Differential Equations, 191 (2003), pp. 377–407.

[20] A. Schebesch, *SYMPERCON—A Package for the Numerical Continuation of Symmetric Periodic Orbits*, Diplomarbeit, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2004; the current version of SYMPERCON can be obtained online from http://www.maths.surrey.ac.uk/personal/st/C.Wulff/.

[21] J.-P. Serre, *Linear Representations of Finite Groups*, Springer-Verlag, New York, Heidelberg, Berlin, 1977.

[22] C. Wulff, *Numerische Pfadverfolgung von periodischen Lösungen mit Symmetrie*, Diplomarbeit, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1993.

[23] C. Wulff, A. Hohmann, and P. Deuflhard, *Numerical Continuation of Periodic Orbits with Symmetry*, Preprint SC 94-12, Konrad-Zuse-Zentrum, Berlin, 1994.

[24] C. Wulff, J. Lamb, and I. Melbourne, *Bifurcation from relative periodic solutions*, Ergodic Theory Dynam. Systems, 21 (2001), pp. 605–635.

[25] C. Wulff, *Persistence of Hamiltonian relative periodic orbits*, J. Geom. Phys., 48 (2003), pp. 309–338.

[26] C. Wulff and A. Schebesch, *Numerical Continuation of Hamiltonian Relative Periodic Orbits*, preprint, Department of Mathematics, University of Surrey, Guildford, Surrey, UK.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 5, No. 3, pp. 476–507

Symmetry and Automated Branch Following for a Semilinear Elliptic PDE on a Fractal Region*

John M. Neuberger†, Nándor Sieben†, and James W. Swift†

Abstract. We apply the gradient Newton–Galerkin algorithm (GNGA) of Neuberger and Swift to find solutions to a semilinear elliptic Dirichlet problem on the region whose boundary is the Koch snowflake. In a recent paper, we described an accurate and efficient method for generating a basis of eigenfunctions of the Laplacian on this region. In that work, we used the symmetry of the snowflake region to analyze and postprocess the basis, rendering it suitable for input to the GNGA. The GNGA uses Newton's method on the eigenfunction expansion coefficients to find solutions to the semilinear problem. This article introduces the *bifurcation digraph*, an extension of the lattice of isotropy subgroups. For our example, the bifurcation digraph shows the 23 possible symmetry types of solutions to the PDE, and the 59 generic symmetry-breaking bifurcations among these symmetry types. Our numerical code uses continuation methods and follows branches created at symmetry-breaking bifurcations, and so the human user does not need to supply initial guesses for Newton's method. Starting from the known trivial solution, the code automatically finds at least one solution with each of the symmetry types that we predict can exist. Such computationally intensive investigations necessitated the writing of automated branch following code, whereby symmetry information was used to reduce the number of computations per GNGA execution and to make intelligent branch following decisions at bifurcation points.

Key words. snowflake, symmetry, bifurcation, semilinear elliptic PDE, GNGA

AMS subject classifications. 20C35, 35P10, 65N25

DOI. 10.1137/050640485

1. Introduction.

We seek numerical solutions to the semilinear elliptic boundary value problem

$$\begin{aligned} \Delta u + f_\lambda(u) &= 0 \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

where Δ is the Laplacian operator, $\Omega \subset \mathbb{R}^2$ is the region whose boundary $\partial\Omega$ is the Koch snowflake, $u : \Omega \to \mathbb{R}$ is the unknown function, and $f_\lambda : \mathbb{R} \to \mathbb{R}$ is a one-parameter family of odd functions. For convenience, we refer to Ω as the *Koch snowflake region*. This article is one of the first to consider a nonlinear PDE on a region with fractal boundary. In this paper, we choose the nonlinearity to be

$$f_\lambda(u) = \lambda u + u^3 \tag{2}$$

*Received by the editors September 16, 2005; accepted for publication (in revised form) by W. Beyn June 22, 2006; published electronically September 21, 2006. This research was partially supported by NSF grant DMS-0074326.
http://www.siam.org/journals/siads/5-3/64048.html
†Department of Mathematics and Statistics, Northern Arizona University, P.O. Box 5717, Flagstaff, AZ 86011-5717 (John.Neuberger@nau.edu, Nandor.Sieben@nau.edu, Jim.Swift@nau.edu).

and treat $\lambda \in \mathbb{R}$ as the bifurcation parameter. When the parameter is fixed, we will sometimes use f in place of f_λ. Using this convention, note that $\lambda = f'(0)$.

This paper exploits the hexagonal symmetry of the Koch snowflake region and the fact that f is odd. Our nonlinear code would work with any region with hexagonal symmetry and any odd "superlinear" function f (see [4]), and with minor modification for other classes of nonlinearities as well. We chose to work with odd f primarily because of the rich symmetry structure. The explicit shape of Ω represents a considerable technological challenge for the computation of the eigenfunctions [16, 27], which are required as input to the nonlinear code.

It is well known that the eigenvalues of the Laplacian under this boundary condition satisfy

$$0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \to \infty, \tag{3}$$

and that the corresponding eigenfunctions $\{\psi_j\}_{j\in\mathbb{N}}$ can be chosen to be an orthogonal basis for the Sobolev space $H = H_0^1(\Omega) = W_0^{1,2}(\Omega)$, and an orthonormal basis for the larger Hilbert space $L^2 = L^2(\Omega)$. The inner products are

$$\langle u, v\rangle_H = \int_\Omega \nabla u \cdot \nabla v \, dx \quad \text{and} \quad \langle u, v\rangle_2 = \int_\Omega u \, v \, dx,$$

respectively (see [1, 9, 15, 17]). Theorem 8.37 and subsequent remarks in [9] imply that the eigenfunctions are in $C^\infty(\Omega)$. In [17], properties of the gradients of eigenfunctions near boundary points are explored in light of the lack of regularity of $\partial\Omega$.

Using the gradient Newton–Galerkin algorithm (GNGA; see [26]), we seek approximate solutions $u = \sum_{j=1}^M a_j\psi_j$ to (1) by applying Newton's method to the eigenfunction expansion coefficients of the gradient $\nabla J(u)$ of a nonlinear functional J whose critical points are the desired solutions. The definition of J, the required variational equations, a description of the GNGA, and a brief history of the problem are the subjects of section 2.

The GNGA requires as input a basis spanning a sufficiently large but finite dimensional subspace $B_M = \text{span}\{\psi_1, \ldots, \psi_M\}$, corresponding to the first M eigenvalues $\{\lambda_j\}_{j=1}^M$. As described in [27], a grid G_N of N carefully placed points is used to approximate the eigenfunctions. These are the same grid points used for the numerical integrations required by Newton's method. Section 3 briefly describes the process we use for generating the eigenfunctions.

Section 4 concerns the effects of symmetry on automated branch following. The symmetry theory for linear operators found in [27] is summarized, and then the extensions required for nonlinear operators are described. Symmetry-breaking bifurcations are analyzed in a way that allows an automated system to follow the branches created at the bifurcations. As we develop the theory, we present specific examples applying the general theory to equation (1) on the snowflake region. In particular, we find that there are 23 different symmetry types of solutions to (1) and 59 generic symmetry-breaking bifurcations. The symmetry types and bifurcations among them are summarized in a *bifurcation digraph*, which generalizes the well-known lattice of isotropy subgroups (see [10]). As far as we know, the bifurcation digraph is a new way to organize the information about the symmetry-breaking bifurcations.

Section 5 describes how understanding the symmetry allows remarkable increases in the efficiency of the GNGA. Section 6 describes the automated branch following. We use repeated

executions of the GNGA or a slightly modified algorithm (parameter-modified GNGA) to follow solution branches of (1), (2). The GNGA uses Newton's method, which is known to work well if it has a good initial approximation. The main shortcoming of Newton's method is that it works poorly without a good initial approximation. We avoid this problem by starting with the trivial solution ($u = 0$). The symmetry-breaking bifurcations of the trivial solution are found by the algorithm, and the primary branches are started. The program follows the branches by continuation methods, and then follows the new branches created at symmetry-breaking bifurcations. To follow an existing branch, we vary λ slightly between executions. To start new solution branches created at bifurcation points, we treat λ as a variable and fix one of the null eigenfunctions of the Hessian evaluated at the bifurcation point. The symmetry analysis tells us which null eigenfunction to use. In this way solutions with all 23 symmetry types are found automatically, starting from $u = 0$, without having to guess any approximations for Newton's method.

In our experiments, many bifurcation diagrams were generated by applying the techniques mentioned above. A selection of these diagrams is provided in section 7, along with contour plots of solutions to (1) corresponding to each of the 23 symmetry types predicted to exist. We include evidence of the convergence of our algorithm as the number of modes M and grid points N increase.

Many extensions to our work are possible, including enforcing different boundary conditions on the same region, solving similar semilinear equations on other fractal regions, and applying the methodology to partial difference equations (PdE) on graphs [25]. Section 8 discusses some of these possible extensions. In particular, we are in the process of rewriting the suite of programs. We plan to be able to solve larger problems using a parallel environment. We will be able to solve problems with larger symmetry groups by automating the extensive group theoretic calculations. This concluding section also has a discussion of the convergence of the GNGA.

2. GNGA.

We now present the variational machinery for studying (1) and follow with a brief description of the general GNGA. Section 6 contains more details of the implementation of the algorithm for our specific problem. Let $F_\lambda(u) = \int_0^u f_\lambda(s)\,ds$ for all $u \in \mathbb{R}$ define the primitive of f_λ. We then define the action functional $J : \mathbb{R} \times H \to \mathbb{R}$ by

$$J(\lambda, u) = \int_\Omega \left\{ \tfrac{1}{2}|\nabla u|^2 - F_\lambda(u) \right\} dx. \tag{4}$$

We will sometimes use $J : H \to \mathbb{R}$ to denote $J(\lambda, \cdot)$. The class of nonlinearities f found in [2, 4, 5, 25, 28] implies that J is well defined and of class C^2 on H. The choice (2) we make in this paper belongs to that class. Critical points of J are by definition weak solutions of (1) (see, for example, [4, 28, 9]), and clearly classical solutions are critical points. The usual "bootstrap" argument of repeatedly applying Theorem 8.10 of [9] can be used in our case. Specifically, H_0^k is embedded in L^q for all $q \geq 2$ when the space dimension n is 2, regardless of the regularity of $\partial\Omega$ (due to the zero Dirichlet boundary condition; see [1]). Hence $u \in H^k$ implies $f(u) \in H^k$ as well. As a result, if u is a critical point, then $u \in C^\infty(\Omega) \cap C(\bar{\Omega})$; hence a classical solution. If one considered boundary conditions, space dimensions, and nonlinear terms other than the choices made in this paper, it could happen that critical points would

be weak, not classical, solutions. Regardless, our approximations lie in $B_M \subset C^\infty$. Here, the existence proofs for positive, negative, and sign-changing exactly once solutions from [4, 28] immediately give at least three nontrivial (classical) solutions for our specific superlinear boundary value problem; appealing to symmetry implies the existence of even more solutions (see, for example, [25]).

The choice of H for the domain is crucial to the analysis of the PDE (see [4, 24] and references therein), as well as for understanding the theoretical basis of effective steepest descent algorithms (see [7, 22, 23], for example). We will work in the coefficient space $\mathbb{R}^M \cong B_M$. The *coefficient vector of* $u \in B_M$ is the vector $a \in \mathbb{R}^M$ satisfying $u = \sum_{j=1}^M a_j \psi_j$. Using the corresponding eigenvalues (3) and integrating by parts, the quantities of interest are

$$g_j = J'(u)(\psi_j) = \int_\Omega \{\nabla u \cdot \nabla \psi_j - f(u)\,\psi_j\} = a_j \lambda_j - \int_\Omega f(u)\,\psi_j, \quad \text{and} \tag{5}$$

$$h_{jk} = J''(u)(\psi_j, \psi_k) = \int_\Omega \{\nabla \psi_j \cdot \nabla \psi_k - f'(u)\,\psi_j\,\psi_k\} = \lambda_j \delta_{jk} - \int_\Omega f'(u)\,\psi_j\,\psi_k, \tag{6}$$

where δ_{jk} is the Kronecker delta function. Note that there is no need for numerical differentiation when forming gradient and Hessian coefficient vectors and matrices in implementing Algorithm 2.1; this information is encoded in the eigenfunctions.

The vector $g \in \mathbb{R}^M$ and the $M \times M$ matrix h represent suitable projections of the L^2 gradient and Hessian of J, restricted to the subspace B_M, where all such quantities are defined. For example, for $u = \sum_{j=1}^M a_j \psi_j$, $v = \sum_{j=1}^M b_j \psi_j$, and $w = \sum_{j=1}^M c_j \psi_j$, we have

$$P_{B_M} \nabla_2 J(u) = \sum_{j=1}^M g_j \psi_j, \quad J'(u)(v) = g \cdot b, \quad \text{and} \quad J''(u)(v,w) = hb \cdot c = b \cdot hc.$$

We can identify g with the approximation $P_{B_M} \nabla_2 J(u)$ of $\nabla_2 J(u) = \Delta u + f(u)$, which is defined for $u \in B_M$. The solution χ to the M-dimensional linear system $h\chi = g$ is then identified with the (suitably projected) search direction $(D_2^2 J(u))^{-1} \nabla_2 J(u)$, which not only is defined for $u \in B_M$, but also is there equal to $(D_H^2 J(u))^{-1} \nabla_H J(u)$. We use the least squares solution of $h\chi = g$. In practice, the algorithm works even near bifurcation points where the Hessian is not invertible.

The heart of our code is Newton's method in the space of eigenfunction coefficients.

Algorithm 2.1 (GNGA).

1. Choose initial coefficients $a = \{a_j\}_{j=1}^M$, and set $u = \sum a_j \psi_j$.
2. Loop
   (a) Calculate the gradient vector $g = \{J'(u)(\psi_j)\}_{j=1}^M$ from (5).
   (b) Calculate the Hessian matrix $h = \{J''(u)(\psi_j, \psi_k)\}_{j,\,k=1}^M$ from (6).
   (c) Exit loop if $||g||$ is sufficiently small.
   (d) Solve $h\chi = g$ for the Newton search direction $\chi \in \mathbb{R}^M$.
   (e) Replace $a \leftarrow a - \chi$ and update $u = \sum a_j \psi_j$.
3. Calculate $\mathrm{sig}(h)$ and J for the approximate solution.

If Newton's method converges, then we expect that u approximates a solution to the PDE (1), provided that M is sufficiently large and the eigenfunctions and numerical integrations are sufficiently accurate. See section 8.

Our estimate for the Morse index (MI) of the critical point of J is the signature of h, denoted $\mathrm{sig}(h)$, which is defined as the number of negative eigenvalues of h. This measures the number of linearly independent directions away from u in which J decreases quadratically.

The basic Algorithm 2.1 is modified to take advantage of the symmetry of our problem. The M integrations required in step (a) and the $M(M+1)/2$ integrations in step (b) are reduced to fewer integrations if the initial guess has nontrivial symmetry.

We often use a "parameter-modified" version of the GNGA (pmGNGA). In this modification, λ is treated as an unknown variable, and one of the M coefficients a_k is fixed. Along a given branch, symmetry generally forces many coefficients to be zero. When a bifurcation point is located by observing a change in MI, we can predict the symmetry of the bifurcating branches using the symmetry of the null eigenfunctions of the Hessian. By forcing a small nonzero component in the direction of a null eigenfunction (orthogonal to the old branch's smaller invariant subspace), we can assure that the pmGNGA will not converge to a solution lying on the old branch. Another benefit of the pmGNGA is that it can handle a curve bifurcating to the right as well as one bifurcating to the left. In our system, the branches that bifurcate to the right have saddle node bifurcations where they turn around and go to the left. The pmGNGA can follow such branches, while the normal GNGA cannot.

The implementation of pmGNGA is not difficult. The M equations are still

$$g_i = J'(u)(\psi_i) = 0,$$

but the M unknowns are

$$\tilde{a} = (a_1, \ldots, a_{k-1}, \lambda, a_{k+1}, \ldots, a_M),$$

and the value of one coefficient, a_k, is fixed. Consequently, we replace the Hessian matrix h with a new matrix $\tilde{h}$, where the kth column is set to $\partial g_i/\partial\lambda = -a_i$:

$$\tilde{h}_{ij} = \begin{cases} h_{ij} & \text{if } j \neq k, \\ -a_i & \text{if } j = k. \end{cases}$$

The search direction $\tilde{\chi}$ is the solution to the system $\tilde{h}\tilde{\chi} = g$. The pmGNGA step is

$$\tilde{a} \leftarrow \tilde{a} - \tilde{\chi},$$

and then u and λ are updated. After Newton's method converges, the kth column of the original h_{ij} is calculated, and the MI of the solution, $\mathrm{sig}(h)$, is computed.

We conclude this section with a very brief history of the analytical and numerical aspects of the research into (1), given our type of nonlinearity f. Our introduction to this general subject was [4], where a sign-changing existence result was proven. This theorem is extended in [5]; we indicate briefly in section 7 where this so-called CCN solution (Castro–Cossio–Neuberger; see [4, 5]) can be found on our bifurcation diagrams. The article [7] was our first success in using symmetry to find higher MI solutions. The GNGA was developed in [26], wherein a much more detailed description of the variational structure and numerical implementation can be found. The first implementation of the GNGA for regions where the eigenfunctions are not known in closed form is in [12], where the region is a Bunimovich stadium. The article [24] provides a historical overview of the authors' experimental results using variants of the mountain pass algorithm (MPA, MMPA, HLA) and the GNGA, as well as recent analytical results and a list of open problems; the references found therein are extensive.

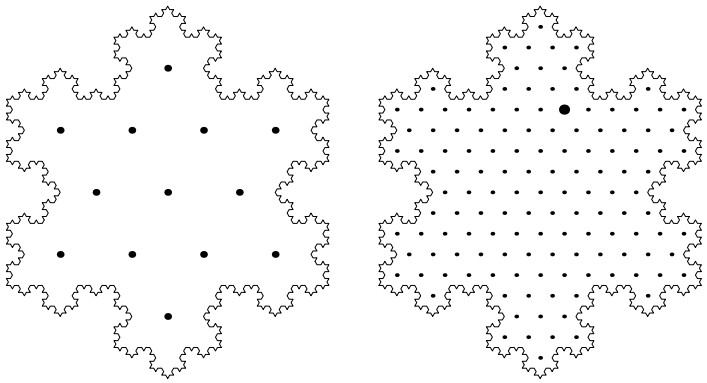


Figure 1. *The Koch snowflake region Ω with the grids G_{13} and G_{133} at levels $\ell = 2$ and 3, respectively. A generic grid point (which is not on any line of reflection symmetry) is indicated in the larger grid.*

3. The basis of eigenfunctions. In [27], we describe theoretical and computational results that lead to the generation of a basis of eigenfunctions solving

$$\Delta u + \lambda u = 0 \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega. \tag{7}$$

That paper details the grid technique and symmetry analysis that accompanied the effort; we briefly summarize those results in this section.

The Koch snowflake is a well-known fractal, with Hausdorff dimension $\log_3 4$. Following Lapidus et al. [16], we take our snowflake to be inscribed in a circle of radius $\frac{\sqrt{3}}{3}$ centered about the origin. We use a triangular grid G_N of N points to approximate the snowflake region. Then, we identify $u : G_N \to \mathbb{R}$ with $u \in \mathbb{R}^N$, that is,

$$u(x_i) = u_i \tag{8}$$

at grid points $x_i \in G_N$. Our paper [27] differs from [16] in that we use a different placement of the grid points and a different method of enforcing the boundary condition, resulting in more accurate eigenvalue estimates with fewer points. Figure 1 depicts the level 2 and 3 grids in the family of grids used in [27] to compute eigenfunctions; we used the first M eigenfunctions computed at levels 4, 5, and 6 in our nonlinear experiments. The number of grid points at level ℓ is $N = (9^\ell - 4^\ell)/5$, and the spacing between grid points is $h = 2/3^\ell$. We computed the eigenvalues and eigenfunctions for (7) using the Arnoldi package (ARPACK; see [19]) and the following approximation to the Laplacian with zero-Dirichlet boundary conditions:

$$-\Delta u(x) \approx \frac{2}{3h^2}\left((12 - \text{number of neighbors})\, u(x) - \sum\{\text{neighboring values of } u\}\right). \tag{9}$$

The ARPACK is based upon an algorithmic variant of the Arnoldi process called the implicitly restarted Arnoldi method and is ideally suited for finding the eigenpairs of the large sparse matrices associated with the discretization of the Laplacian.

4. Symmetry: The lattice of isotropy subgroups and the bifurcation digraph. This section describes equivariant bifurcation theory as it applies to the branching of solutions to (1); see [6, 10, 11, 18]. We are able to describe the expected symmetry types of solutions to (1), as traditionally arranged in a lattice of isotropy subgroups. We introduce the *bifurcation digraph*, a refinement of the lattice, which shows every possible generic bifurcation from one symmetry type to another as a directed edge which is labeled with information about the bifurcation. The bifurcation digraph is of interest in its own right and summarizes the essential information required by our automated branch following code. In this project, GAP (groups, algorithms, and programming; see [8]) was used solely to verify the symmetry analysis we did by hand. In our continuing projects GAP is a useful tool since it can perform the tedious calculations and write the information in a format that can be read by the branch following code. Matthews [21] has used GAP to do similar calculations. We apply this methodology to the snowflake domain being considered in this paper. The analysis shows that solutions fall into 23 symmetry types, and that there are 59 types of generic symmetry-breaking bifurcations.

Group actions and the lattice of isotropy subgroups. Let Γ be a finite group and V be a real vector space. A *representation* of Γ is a homomorphism $\alpha : \Gamma \to GL(V)$. Where convenient, we identify $GL(V)$ with the set of invertible matrices with real coefficients. Every representation α corresponds to a unique *group action* of Γ on V by the rule $\gamma \cdot v := \alpha(\gamma)(v)$ for all $\gamma \in \Gamma$ and $v \in V$. We will usually use the action rather than the representation. The *group orbit* of v is $\Gamma \cdot v = \{\gamma \cdot v \mid \gamma \in \Gamma\}$.

Example 4.1. Let

$$\mathbb{D}_6 := \langle \rho, \sigma \mid \rho^6 = \sigma^2 = 1,\, \rho\,\sigma = \sigma\rho^5 \rangle$$

be the dihedral group with 12 elements. It is convenient to define $\tau = \rho^3\sigma$. It follows that $\sigma\tau = \tau\sigma = \rho^3$. The group $\mathbb{D}_6$ is the symmetry of a regular hexagon, and of the Koch snowflake region Ω. The standard $\mathbb{D}_6$ action on the plane is given by

$$\begin{aligned} \rho \cdot (x, y) &= \left(\tfrac{1}{2}x + \tfrac{\sqrt{3}}{2}y, -\tfrac{\sqrt{3}}{2}x + \tfrac{1}{2}y\right), \\ \sigma \cdot (x, y) &= (-x, y), \\ \tau \cdot (x, y) &= (x, -y). \end{aligned} \tag{10}$$

In this action, ρ is a rotation by 60°, σ is a reflection across the y-axis, and τ is a reflection across the x-axis. These group actions are depicted in Figure 13, near the end of the paper.

We will denote subgroups of $\mathbb{D}_6$ by listing the generators. While any given subgroup of $\mathbb{D}_6$ can be defined using only ρ and σ, we find it geometrically descriptive to use τ in certain cases. For example, we prefer $\langle \rho^2, \tau \rangle$ to the equivalent $\langle \rho^2, \rho\sigma \rangle$. In order to make relationships among subgroups intuitive, we often include τ when its membership is implied by the other generators (see, for example, Figure 2 below).

The standard $\mathbb{D}_6$ group action (10) is not the only action we consider. For a function $u \in L^2(\Omega)$ and group element $\gamma \in \mathbb{D}_6$, we define $(\gamma \cdot u)(x) = u(\gamma^{-1} \cdot x)$. In this paper, a vector u defined by $u_i = u(x_i)$, for a given grid $G_N = \{x_i\}_{i=1}^N$, is a discrete approximation of a function on Ω. The $\mathbb{D}_6$ group action on $u \in \mathbb{R}^N$ is a permutation of the components:

$(\gamma \cdot u)_i = u(\gamma^{-1} \cdot x_i)$. Given a function $u \in L^2(\Omega)$ or $\mathbb{R}^N$, the group orbit $\mathbb{D}_6 \cdot u$ consists of functions obtained from u by a reflection or rotation.

Example 4.2. The group $\mathbb{D}_6 \times \mathbb{Z}_2$, where $\mathbb{Z}_2 = \{1, -1\}$, acts on $L^2(\Omega)$ in a natural way. For all $(\gamma, z) \in \mathbb{D}_6 \times \mathbb{Z}_2$, define

$$(\gamma, z) \cdot u = z(\gamma \cdot u).$$

We will denote $(\gamma, 1) \in \mathbb{D}_6 \times \mathbb{Z}_2$ by γ and $(\gamma, -1) \in \mathbb{D}_6 \times \mathbb{Z}_2$ by $-\gamma$. With this natural notation, $(-\gamma) \cdot u = -(\gamma \cdot u)$, which we call simply $-\gamma \cdot u$.

Let us recall some facts about group actions, following [6, 10, 11]. The *isotropy subgroup* or *stabilizer of* $v \in V$ *in* Γ is

$$\mathrm{Stab}(v, \Gamma) := \{\gamma \in \Gamma \mid \gamma \cdot v = v\}.$$

The isotropy subgroup measures the symmetry of v, and is sometimes called the little group of v, or Γ_v. If the group Γ is understood, we may simply write $\mathrm{Stab}(v)$ in place of $\mathrm{Stab}(v, \Gamma)$. The *stabilizer of a subset* $W \subseteq V$ *in* Γ is $\mathrm{Stab}(W, \Gamma) := \{\gamma \in \Gamma \mid \gamma \cdot W = W\}$. This must be distinguished from the *point stabilizer of a subset*

$$\mathrm{pStab}(W, \Gamma) := \{\gamma \in \Gamma \mid \gamma \cdot v = v \text{ for all } v \in W\} = \bigcap \{\mathrm{Stab}(v, \Gamma) \mid v \in W\}.$$

Another commonly used notation is Γ_W for the stabilizer and $\Gamma_{(W)}$ for the point stabilizer. Note that $\mathrm{pStab}(W, \Gamma)$ is always normal in $\mathrm{Stab}(W, \Gamma)$, and the *effective symmetry group acting on* W is $\mathrm{Stab}(W, \Gamma)/\mathrm{pStab}(W, \Gamma)$, which acts faithfully on W.

If Σ is a subgroup of Γ, then the *fixed point subspace of* Σ *in* V is

$$\mathrm{Fix}(\Sigma, V) := \{v \in V \mid \ \gamma \cdot v = v \text{ for all } \gamma \in \Sigma\}.$$

Another notation for the fixed point subspace is V_Σ. We write $\mathrm{Fix}(\Sigma)$ when V is understood.

An *isotropy subgroup* of the Γ action on V is the stabilizer of some point $v \in V$. For some group actions, not every subgroup of Γ is an isotropy subgroup.

Example 4.3. Consider the $\mathbb{D}_6$ action on the plane $\mathbb{R}^2$ described in (10). It is well known that $\langle \rho \rangle$ is not an isotropy subgroup of this action.

Now consider the $\mathbb{D}_6$ action on the function space $L^2(\Omega)$. We give a standard argument that every subgroup of $\mathbb{D}_6$ is an isotropy subgroup. Start with a function u^* that is zero everywhere except for a small region, and suppose that the region is distinct from each of its nontrivial images under the $\mathbb{D}_6$ action. Then for any subgroup $\Sigma \leq \mathbb{D}_6$, the *average of the function* u^* *over* Σ, defined as

$$P_\Sigma(u^*) = \frac{1}{|\Sigma|} \sum_{\gamma \in \Sigma} \gamma \cdot u^*, \tag{11}$$

has isotropy subgroup Σ. Therefore every subgroup of the $\mathbb{D}_6$ action on $L^2(\Omega)$ is an isotropy subgroup. The average over the group is an example of a Haar operator, and $P_\Sigma : V \to \mathrm{Fix}(\Sigma, V)$ is an orthogonal projection operator [36].

Similarly, every subgroup of $\mathbb{D}_6$ is an isotropy subgroup of the $\mathbb{D}_6$ action on $\mathbb{R}^N$, the space of functions on the grid G_N, provided $\ell \geq 3$. This follows from averaging the function that is

1 at a generic lattice point and 0 elsewhere. Recall that a *generic point* is one whose isotropy subgroup is trivial. Figure 1 shows that the level two grid G_{13} does not have a generic point, while the level three grid G_{133} does. Thus, the space of functions on G_{133} has the same isotropy subgroups as $L^2(\Omega)$, but a much smaller space has this same property. Start with any generic point $x_1 \in \Omega$. Then $\mathbb{D}_6$ acts on the space of functions on the 12 points $\mathbb{D}_6 \cdot x_1$. This $\mathbb{D}_6$ action on $\mathbb{R}^{12}$ has the same structure of isotropy subgroups as the $\mathbb{D}_6$ action on $L^2(\Omega)$, and is the $\mathbb{D}_6$ action used in our GAP calculations. The corresponding 12-dimensional representation is the well-known *regular representation* of $\mathbb{D}_6$ (see [29, 31, 34]).

The symmetry of functions is described by two related concepts. A function $q : V \to \mathbb{R}$ is Γ-*invariant* if $q(\gamma \cdot v) = q(v)$ for all $\gamma \in \Gamma$ and all $v \in V$. Similarly, an operator $T : V \to V$ is Γ-*equivariant* if $T(\gamma \cdot v) = \gamma \cdot T(v)$ for all $\gamma \in \Gamma$ and all $v \in V$.

Example 4.4. The energy functional J defined in (4) is $\mathbb{D}_6 \times \mathbb{Z}_2$-invariant. The nonlinear PDE (1) can be written as $(\Delta + f)(u) = 0$, where $\Delta + f$ is a $\mathbb{D}_6 \times \mathbb{Z}_2$-equivariant operator. (There are subtleties concerning the domain and range of Δ. See [6, 7] for a careful treatment of the function spaces.) In particular, $\Delta + f$ is $\mathbb{D}_6$-equivariant because the snowflake region Ω has $\mathbb{D}_6$ symmetry, and $(\Delta + f)(-u) = -(\Delta + f)(u)$, since f is odd. As a consequence, if u is a solution to (1), then so is every element in its group orbit $(\mathbb{D}_6 \times \mathbb{Z}_2) \cdot u$.

The isotropy subgroups and fixed point subspaces are important because of the following simple yet powerful results; see [6, 10, 11].

Proposition 4.5. *Suppose that* Γ *acts linearly on* V, $T : V \to V$ *is* Γ*-equivariant, and* Σ *is an isotropy subgroup of* Γ.

(a) *If* $v \in \mathrm{Fix}(\Sigma)$, *then* $T(v) \in \mathrm{Fix}(\Sigma)$. *Thus,* $T|_{\mathrm{Fix}(\Sigma)} : \mathrm{Fix}(\Sigma) \to \mathrm{Fix}(\Sigma)$ *is defined.*

(b) $\mathrm{Stab}(\mathrm{Fix}(\Sigma)) = N_\Gamma(\Sigma)$, *the normalizer of* Σ *in* Γ, *and* $\mathrm{pStab}(\mathrm{Fix}(\Sigma)) = \Sigma$.

(c) $T|_{\mathrm{Fix}(\Sigma)}$ *is* $N_\Gamma(\Sigma)$*-equivariant.*

(d) $T|_{\mathrm{Fix}(\Sigma)}$ *is* $N_\Gamma(\Sigma)/\Sigma$*-equivariant, and* $N_\Gamma(\Sigma)/\Sigma$ *acts faithfully on* $\mathrm{Fix}(\Sigma)$.

If Σ is a subgroup of Γ, the normalizer of Σ in Γ is defined to be $N_\Gamma(\Sigma) := \{\gamma \in \Gamma \mid \gamma\Sigma = \Sigma\gamma\}$, which is the largest subgroup of Γ for which Σ is a normal subgroup. The presence of the normalizer in Proposition 4.5(b) is interesting, since the normalizer is a property of the abstract groups and is independent of the group action.

Example 4.6. As a consequence of Proposition 4.5, we can solve the PDE (1), written as $(\Delta+f)(u) = 0$, by restricting u to functions in $\mathrm{Fix}(\Sigma, L^2(\Omega))$. This leads to a simpler problem since the function space $\mathrm{Fix}(\Sigma, L^2(\Omega))$ is simpler than $L^2(\Omega)$. An example of this is in Costa, Ding, and Neuberger [7]. The techniques of that paper, applied to our problem, would find sign-changing solutions with MI 2 within the space $\mathrm{Fix}(\mathbb{D}_6, L^2(\Omega))$. This space consists of all functions which are unchanged under all of the rotations and reflections of the snowflake region.

Proposition 4.5 also applies to the GNGA, since the Newton method iteration mapping is $\mathbb{D}_6 \times \mathbb{Z}_2$-equivariant. If the initial guess is in a particular fixed point subspace, all the iterates will be in that fixed point subspace. This fact can be used to speed numerical calculations, as described in section 5.

Two subgroups Σ_1, Σ_2 of Γ are conjugate ($\Sigma_1 \sim \Sigma_2$) if $\Sigma_1 = \gamma\Sigma_2\gamma^{-1}$ for some $\gamma \in \Gamma$. The *symmetry type* of $v \in V$ for the Γ action is the conjugacy class of $\mathrm{Stab}(v, \Gamma)$. Note that $\mathrm{Stab}(\gamma \cdot v) = \gamma\,\mathrm{Stab}(v)\gamma^{-1}$. Thus, every element of a group orbit $\Gamma \cdot v$ has the same symmetry type.

Let $\mathcal{S} = \{S_i\}$ denote the set of all symmetry types of a Γ action on V. The set $\mathcal{S}$ has a natural partial order, with $S_i \leq S_j$ if there exists $\Sigma_i \in S_i$ and $\Sigma_j \in S_j$ such that $\Sigma_i \leq \Sigma_j$. The partially ordered set $(\mathcal{S}, \leq)$ is called the *lattice of isotropy subgroups* of the Γ action on V [10]. The *diagram* of the lattice of isotropy subgroups is a directed graph with vertices S_i and arrows $S_i \leftarrow S_j$ if and only if $S_i \lneq S_j$ and there is no symmetry type between S_i and S_j.

Example 4.7. The symmetry type of a solution u to our PDE (1) for the $\mathbb{D}_6 \times \mathbb{Z}_2$ action is the conjugacy class of $\mathrm{Stab}(u, \mathbb{D}_6 \times \mathbb{Z}_2)$; we refer to this as the symmetry type of u, without reference to $\mathbb{D}_6 \times \mathbb{Z}_2$. The discussion of $\mathbb{D}_6$ acting on $L^2(\Omega)$ in Example 4.3 can easily be extended to $\mathbb{D}_6 \times \mathbb{Z}_2$ acting on $L^2(\Omega)$. Note that if $-1 \in \Sigma \leq \mathbb{D}_6 \times \mathbb{Z}_2$, then the average of any function over Σ is $u = 0$. Therefore the only isotropy subgroup of $\mathbb{D}_6 \times \mathbb{Z}_2$ which contains -1 is $\mathbb{D}_6 \times \mathbb{Z}_2$ itself. On the other hand, the argument in Example 4.3 shows that any subgroup of $\mathbb{D}_6 \times \mathbb{Z}_2$ which does not contain -1 is an isotropy subgroup. Therefore, $\Sigma \leq \mathbb{D}_6 \times \mathbb{Z}_2$ is an isotropy subgroup of this group action if and only if $\Sigma = \mathbb{D}_6 \times \mathbb{Z}_2$ or $-1 \notin \Sigma$.

This result allowed us to compute the isotropy subgroups by hand. We verified our calculations using GAP. There are exactly 23 conjugacy classes of isotropy subgroups for the $\mathbb{D}_6 \times \mathbb{Z}_2$ action on $L^2(\Omega)$, shown in condensed form in Figure 2. Thus, a solution to the PDE (1) has one of 23 different symmetry types.

Irreducible representations and the isotypic decomposition. In order to understand the symmetry-breaking bifurcations we need to first understand irreducible representations and the isotypic decomposition of a group action. The information about the irreducible representations is summarized in character tables [29, 31, 32, 34]. For our purposes, irreducible representations over the field $\mathbb{R}$ are required; see [6, 10, 11]. The *irreducible representations of* Γ are homomorphisms from Γ to the space of $d_j \times d_j$ real matrices $\gamma \mapsto \alpha^{(j)}(\gamma)$, such that no proper subspace of $\mathbb{R}^{d_j}$ is invariant under $\alpha^{(j)}(\gamma)$ for all $\gamma \in \Gamma$. The *dimension of the irreducible representation* $\alpha^{(j)}$ is d_j. We call $W \subseteq V$ a Γ-*invariant subspace of* V if $\Gamma \cdot W \subseteq W$. An *irreducible subspace of* V is an invariant subspace with no proper invariant subspaces. Every irreducible subspace of the Γ action on V corresponds to a unique (up to similarity) irreducible representation of Γ. The dimension of the irreducible subspace is the same as the dimension of the corresponding irreducible representation.

For each irreducible representation $\alpha^{(j)}$ of Γ, the *isotypic component* of V for the Γ action, denoted by $V_\Gamma^{(j)}$, is defined to be the direct sum of all of the irreducible subspaces corresponding to the fixed $\alpha^{(j)}$ [6, 10, 11, 27]. The *isotypic decomposition* of V is then

$$V = \bigoplus_j V_\Gamma^{(j)}. \tag{12}$$

It is possible that one or more of the isotypic components is the single point at the origin. These can be left out of the isotypic decomposition. A description of the isotypic components in terms of projection operators is given in [27].

For any group Γ, we denote the trivial representation by $\alpha^{(1)}$. That is, $\alpha^{(1)}(\gamma) = 1$ for all $\gamma \in \Gamma$. Thus, if Γ is an isotropy subgroup of a Γ_0 action on V, then

$$V_\Gamma^{(1)} = \mathrm{Fix}(\Gamma, V).$$

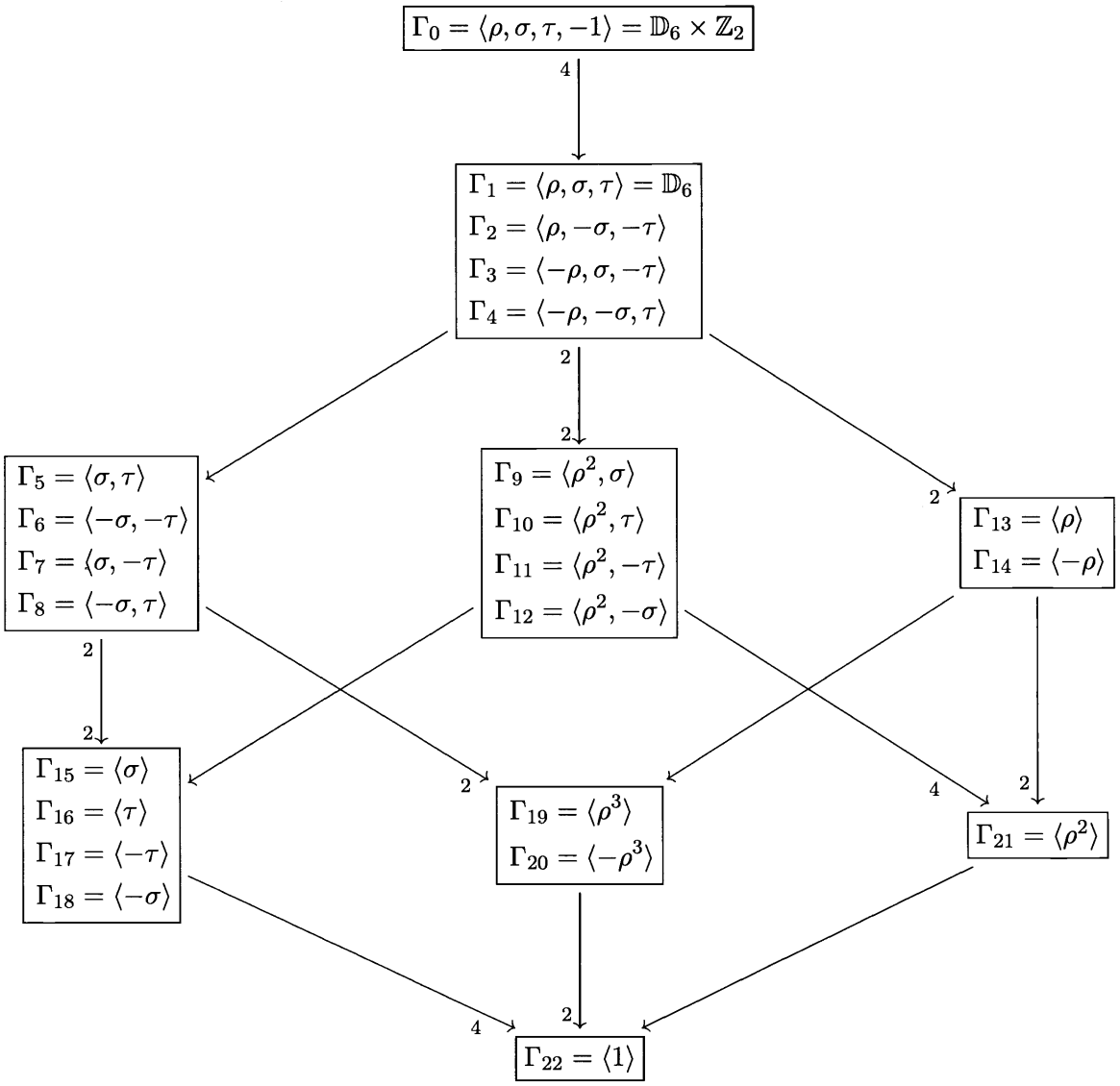


Figure 2. *The condensed diagram of the isotropy lattice (see [10]) for the $\mathbb{D}_6 \times \mathbb{Z}_2$ action on $L^2(\Omega)$. The vertices of this diagram are the symmetry types (equivalence classes of isotropy subgroups). We follow the convention [6, 10, 11] that one element Γ_i of each symmetry type $S_i = [\Gamma_i]$ is listed. The representatives Γ_i have the property that $\Gamma_i \leq \Gamma_j$ if and only if $S_i \leq S_j$. Contour plots of solutions to PDE (1) with each of the 23 symmetry types are given in Figures 13 and 14. The diagram of the isotropy lattice is condensed as in [32]. The small numbers on the edges give the number of connections emanating from each symmetry type in a box. A missing small number means 1. For example, the two arrows representing $[\Gamma_{21}] \leq [\Gamma_{13}]$ and $[\Gamma_{21}] \leq [\Gamma_{14}]$ in the full diagram are collapsed to a single arrow in the condensed diagram. For Γ_0 through Γ_4, the τ generator is redundant since $\tau = \rho^3\sigma$, but its presence makes the subgroups manifest. For example, $\Gamma_2 = \langle \rho, -\sigma, -\tau \rangle = \langle \rho, -\sigma \rangle$, but the three generators make it clear that $\langle -\sigma, -\tau \rangle \leq \langle \rho, -\sigma, -\tau \rangle$.*

Example 4.8. Let us consider the $\mathbb{D}_6 = \langle \rho, \sigma, \tau \rangle$ action on $L^2(\Omega)$. We need to consider the six irreducible representations of $\mathbb{D}_6$, which are listed in [27], to find the isotypic decomposition of $L^2(\Omega)$. Since these isotypic components are central to our problem, we drop the $\mathbb{D}_6$ and define $V^{(j)} := V^{(j)}_{\mathbb{D}_6}$, $j = 1, 2, \ldots, 6$, as follows:

$$
\begin{aligned}
V^{(1)} &= \{u \in L^2(\Omega) \mid \rho \cdot u = u,\ \sigma \cdot u = u,\ \tau \cdot u = u\},\\
V^{(2)} &= \{u \in L^2(\Omega) \mid \rho \cdot u = u,\ \sigma \cdot u = -u,\ \tau \cdot u = -u\},\\
V^{(3)} &= \{u \in L^2(\Omega) \mid \rho \cdot u = -u,\ \sigma \cdot u = u,\ \tau \cdot u = -u\},\\
V^{(4)} &= \{u \in L^2(\Omega) \mid \rho \cdot u = -u,\ \sigma \cdot u = -u,\ \tau \cdot u = u\},\\
V^{(5)} &= \{u \in L^2(\Omega) \mid \rho^3 \cdot u = u,\ u + \rho^2 \cdot u + \rho^4 \cdot u = 0\},\\
V^{(6)} &= \{u \in L^2(\Omega) \mid \rho^3 \cdot u = -u,\ u + \rho^2 \cdot u + \rho^4 \cdot u = 0\}.
\end{aligned} \tag{13}
$$

Example 4.9. The isotypic decomposition of $\Gamma_{13} = \langle \rho \rangle \cong \mathbb{Z}_6$ illustrates some features of *real* representation theory. The irreducible representations of $\mathbb{Z}_6$ over $\mathbb{C}$ are all one-dimensional. They are $\alpha^{(j)}(\rho) = (e^{i\pi/3})^{j-1}$ for $j = 1, 2, \ldots, 6$. Over the field $\mathbb{R}$, however, the one-dimensional irreducible representations of $\mathbb{Z}_6$ are given by

$$
\alpha^{(1)}(\rho) = 1, \qquad \alpha^{(2)}(\rho) = -1, \tag{14}
$$

and the two-dimensional irreducible representations of $\mathbb{Z}_6$, up to similarity transformations, are given by

$$
\alpha^{(3)}(\rho) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \qquad \alpha^{(4)}(\rho) = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}. \tag{15}
$$

Note that $\alpha^{(3)}(\rho)$ is matrix for a rotation by 120°, and $\alpha^{(4)}(\rho)$ is a 60° rotation matrix.

An irreducible representation over $\mathbb{R}$ is called *absolutely irreducible* if it is also irreducible over $\mathbb{C}$. For example, all of the irreducible representations of $\mathbb{D}_6$ listed in [27] are absolutely irreducible, as are the one-dimensional irreducible representations of $\mathbb{Z}_6$ in (14). On the other hand, the two-dimensional irreducible representations of $\mathbb{Z}_6$ in (15) are *not* absolutely irreducible.

The four isotypic components of the $\langle \rho \rangle$ action on $L^2(\Omega)$ are

$$
V^{(1)}_{\langle \rho \rangle} = \{u \in L^2(\Omega) \mid \rho \cdot u = u\} = V^{(1)} \oplus V^{(2)},
$$

$$
V^{(2)}_{\langle \rho \rangle} = \{u \in L^2(\Omega) \mid \rho \cdot u = -u\} = V^{(3)} \oplus V^{(4)},
$$

$$
V^{(3)}_{\langle \rho \rangle} = V^{(5)}, \quad \text{and } V^{(4)}_{\langle \rho \rangle} = V^{(6)}.
$$

If we had used the complex irreducible representations, some of the corresponding isotypic components would contain complex-valued functions. It is more natural to use real irreducible representations and consider only real-valued functions. The price we pay is that most of the

representation theory found in books, and built into GAP, is done for complex irreducible representations.

The isotypic decomposition for each of the 23 isotropy subgroups, Γ_i, of $\mathbb{D}_6 \times \mathbb{Z}_2$ can be written as a direct sum of some subset of the eight spaces $V^{(j)}$, for $j = 1, \ldots, 4$, and $V_1^{(j)}$ and $V_2^{(j)}$ for $j = 5, 6$ defined in (13) and [27]. The C++ program can easily check whether a function is in any of the isotypic components $V_{\Gamma_i}^{(j)}$ of B_M for each of the Γ_i, $i = 0, 1, \ldots, 22$, actions.

Symmetry-breaking bifurcations. The fact that there are 23 *possible* symmetry types of solutions to the PDE (1) raises the question, do solutions with each of these symmetry types exist? Clearly the trivial solution $u = 0$, with symmetry type S_0, exists. Our procedure for finding approximate solutions with each of these symmetry types is to start with the trivial solution and recursively follow solution branches created at symmetry-breaking bifurcations.

Let us start by abstracting the PDE defined by (1), which depends on the real parameter λ. Let V be an inner product space, and $J : \mathbb{R} \times V \to \mathbb{R}$ be a family of Γ_0-invariant functions that depends on a parameter λ. That is, $J(\lambda, \gamma \cdot u) = J(\lambda, u)$ for all $\gamma \in \Gamma_0$ and $u \in V$. It is understood that Γ_0 is the *largest* known group for which J is invariant; of course J is also invariant under any subgroup of Γ_0. We will use Γ, or Γ_i, to refer to an isotropy subgroup of the "full" group Γ_0. Consider the steady-state bifurcation problem $g(\lambda, u) = 0$, where $g(\lambda, u) = \nabla J(\lambda, u)$. Throughout this paper, the gradient ∇ acts on the u component. The solutions to $g(\lambda, u) = 0$ are critical points of J, and so we use the terms "solution" and "critical point" interchangeably. Note that $g : \mathbb{R} \times V \to V$ is a family of Γ_0-equivariant gradient operators on V. That is, $g(\lambda, \gamma \cdot u) = \gamma \cdot g(\lambda, u)$. For our PDE, $\Gamma_0 = \mathbb{D}_6 \times \mathbb{Z}_2$. In the numerical implementation, $V = \mathbb{R}^M \cong B_M$ and g is defined in (5).

We define a *branch of solutions* to be a connected component of $\{(\lambda, u) \in \mathbb{R} \times L^2(\Omega) \mid g(\lambda, u) = 0, \text{Stab}(u) = \Gamma\}$, where Γ is called the isotropy subgroup, or symmetry, of the branch. A branch of solutions B_1 has a *symmetry-breaking bifurcation* at the *bifurcation point* $(\lambda^*, u^*) \in B_1$ if a branch of solutions, B_2, with a different symmetry, has (λ^*, u^*) as a limit point but $(\lambda^*, u^*) \notin B_2$. We say that branch B_2 is *created* at this bifurcation, and often refer to B_1 as the *mother branch* and B_2 as the *daughter branch*. The symmetry of the daughter branch is always a proper subgroup of the symmetry of the mother branch. That is, the daughter has less symmetry than the mother.

The main tool for finding bifurcation points is the Hessian of the energy functional, h. If (λ^*, u^*) is a bifurcation point, then $h(\lambda^*, u^*)$ is not invertible, since otherwise the implicit function theorem would guarantee the existence of a unique local solution branch. The MI of a critical point (λ, u) is defined to be the number of negative eigenvalues of $h(\lambda, u) = D^2 J(\lambda, u)$, provided that no eigenvalue is 0. The Hessian is symmetric, so all of its eigenvalues are real. The MI on a branch of solutions typically changes at a bifurcation point.

Example 4.10. The trivial solution to (1), (2) is $u = 0$, and the *trivial branch* is $\{(\lambda, 0) \mid \lambda \in \mathbb{R}\}$. Since $h(\lambda, 0)(v) = \Delta v + \lambda v$, the bifurcation points of the trivial branch are $(\lambda_i, 0)$, where λ_i, $i \in \mathbb{N}$, are the eigenvalues (3). If $\lambda_i < \lambda < \lambda_{i+1}$, then the MI of the trivial solution $(\lambda, 0)$ is i. The ith *primary branch* is created at the bifurcation point $(\lambda_i, 0)$ on the trivial branch. In cases with double eigenvalues there are two branches created at the same point in our problem. For example, the second and third primary branches are created at $\lambda_2 = \lambda_3$.

Near $(\lambda_i, 0)$, the solutions on the ith primary branch are approximately some constant times the ith eigenfunction of the Laplacian, ψ_i.

We define a *degenerate critical point*, or a *degenerate solution*, to be a point (λ^*, u^*) that satisfies $g(\lambda^*, u^*) = 0$ and $\det h(\lambda^*, u^*) = 0$. Thus, every bifurcation point is a degenerate critical point. Some degenerate critical points are not bifurcation points. For example, when a branch folds over and is not monotonic in λ, the *fold point* is degenerate but is not a bifurcation point as we have defined it. (Note that we avoid the term "saddle-node bifurcation" since there is really no bifurcation.)

Let us develop some notation to talk about bifurcations. Suppose that (λ^*, u^*) is an isolated degenerate critical point of a Γ_0-equivariant system $g(\lambda, u) = 0$. Let $\Gamma = \mathrm{Stab}(u^*, \Gamma_0)$, and define $L := h(\lambda^*, u^*)$. Note that Γ, not Γ_0, is important as far as the bifurcation of (λ^*, u^*) is concerned. Let E be the null space of the Γ-equivariant operator L. We call E the *center eigenspace*. Let Γ' be the point stabilizer of E. The definitions are repeated here for reference:

$$\Gamma := \mathrm{Stab}(u^*, \Gamma_0), \quad L := h(\lambda^*, u^*), \quad E := N(L), \quad \Gamma' := \mathrm{pStab}(E, \Gamma). \tag{16}$$

If $e \in E$, then $L(e) = 0$ by definition. For any $\gamma \in \Gamma$, $\gamma \cdot e \in E$ since the Γ-equivariance of L implies that $L(\gamma \cdot e) = \gamma \cdot L(e) = 0$. Hence,

$$\mathrm{Stab}(E, \Gamma) = \Gamma.$$

Note that $\mathrm{Stab}(E, \Gamma)/\mathrm{pStab}(E, \Gamma) = \Gamma/\Gamma'$ acts faithfully on E. In the usual case where (λ^*, u^*) is a bifurcation point, not just a degenerate critical point, we say that Γ/Γ' is the *symmetry group of the bifurcation*, or that (λ^*, u^*) undergoes a *bifurcation with* Γ/Γ' *symmetry*.

In the notation of (16), L sends each of the isotypic components $V_\Gamma^{(j)}$ to itself [27, 31, 34]. Barring "accidental degeneracy," the center eigenspace E is a Γ-irreducible subspace. Thus, E is typically a subspace of exactly one isotypic component $V_\Gamma^{(j)}$, and $\dim(E)$ is the dimension d_j of the corresponding irreducible representation, $\alpha^{(j)}$. Furthermore, the point stabilizer of E is the kernel of $\alpha^{(j)}$ and can be computed without knowing E. In summary, at a generic bifurcation point there is some irreducible representation $\alpha^{(j)}$ of Γ such that

$$E \text{ is } \Gamma\text{-irreducible}, \quad E \subseteq V_\Gamma^{(j)}, \quad \dim(E) = \Delta MI = d_j, \quad \Gamma' = \{\gamma \in \Gamma \mid \alpha^{(j)}(\gamma) = I\}.$$

Accidental degeneracy is discussed in [27, 31, 34]. We did not encounter any accidental degeneracy in our numerical investigation of (1), (2), so we will not discuss it further here.

We finally have the background to describe the bifurcations which occur in equivariant systems. The goal is to predict what solutions will be created at each of the symmetry-breaking bifurcations, and to know what vectors in E to use to start these branches using the pmGNGA. While such a prediction is impossible for some complicated groups, we can determine how to follow all of the bifurcating branches in the system (1), (2). We follow the treatment and notation of [10, 11]. At a symmetry-breaking bifurcation we can translate (λ^*, u^*) to the origin, and we could, in principle, do an equivariant Liapunov–Schmidt reduction or center manifold reduction to obtain reduced bifurcation equations $\tilde{g} : \mathbb{R} \times E \to E$, where $\tilde{g}(0, 0) = 0$, $D\tilde{g}(0, 0) = 0$, and $\tilde{g}$ is $\Gamma := \mathrm{Stab}(u^*)$-equivariant. It is important to realize that we do not actually need to perform the Liapunov–Schmidt reduction.

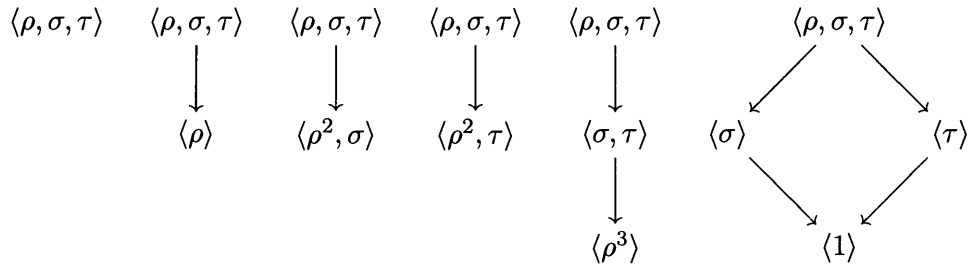


Figure 3. *Diagrams of the six isotropy lattices for the actions of $\mathbb{D}_6 = \langle \rho, \sigma, \tau \rangle$ on each of the six isotypic components $V^{(j)}$ of the $\mathbb{D}_6$ action on $L^2(\Omega)$. This describes the six possibilities (barring accidental degeneracy) for the $\mathbb{D}_6$ action on the center eigenspace E at a degenerate critical point.*

The most powerful tool for understanding symmetry-breaking bifurcations is the equivariant branching lemma. Recall that absolutely irreducible representations were defined in Example 4.9. See [6, 10, 11] for a thorough discussion of the equivariant branching lemma, including further references.

Theorem 4.11 (equivariant branching lemma (EBL)). *Suppose that Γ acts absolutely irreducibly on the space E, and let $\tilde{g} : \mathbb{R} \times E \to E$ be Γ-equivariant. Assume that Γ acts nontrivially, so that $\tilde{g}(\lambda, 0) = 0$. Since Γ acts absolutely irreducibly, $D\tilde{g}(\lambda, 0) = c(\lambda) I_d$ for some function $c : \mathbb{R} \to \mathbb{R}$, where I_d is the identity matrix of size $d = \dim(E)$. Assume that $c(0) = 0$ and $c'(0) \neq 0$. Let Σ be an isotropy subgroup of the Γ action on E with $\dim \mathrm{Fix}(\Sigma, E) = 1$. Then there are at least two solution branches of $\tilde{g}(\lambda, u) = 0$ with isotropy subgroup Σ created at $(0, 0)$.*

The EBL, combined with Liapunov–Schmidt theory, implies that there are at least two solution branches of the full problem $g(\lambda, u) = 0$ with isotropy subgroup Σ created at the bifurcation point (λ^*, u^*). We call these newly created branches *EBL branches* since their existence can be predicted by the EBL. Other branches created at a bifurcation are called *non-EBL branches*.

Following [6, 10, 11], we define a *maximal isotropy subgroup* of a Γ action on V to be an isotropy subgroup $\Sigma \neq \Gamma$ with the property that if Θ is an isotropy subgroup such that $\Sigma \leq \Theta$, then $\Theta = \Sigma$ or $\Theta = \Gamma$. In other words, a maximal isotropy subgroup is a maximal proper isotropy subgroup. If $\dim(\mathrm{Fix}(\Sigma, E)) = 1$, then Σ is a maximal isotropy subgroup of the Γ action on E. The converse, however, is not true.

In gradient systems, for example the PDE (1), more can be said. If Σ is any maximal isotropy subgroup of the Γ action on E, then there is typically a solution branch created at the bifurcation with isotropy subgroup Σ. If $\dim \mathrm{Fix}(\Sigma, E) \geq 2$, the branch created is an example of a non-EBL branch. See [30] for a discussion of bifurcations in gradient systems.

By Proposition 4.5, the effective symmetry group of $\tilde{g}$, restricted to $\mathrm{Fix}(\Sigma, E)$, is $N_\Gamma(\Sigma)/\Sigma$. This effective symmetry group determines how solutions with symmetry Σ bifurcate.

Example 4.12. Consider a degenerate critical point with isotropy subgroup $\Gamma_1 = \mathbb{D}_6 = \langle \rho, \sigma, \tau \rangle$. Barring accidental degeneracy, the center eigenspace E is a subspace of one of the six isotypic components of the $\mathbb{D}_6$ action on $L^2(\Omega)$ described in Example 4.8. Figure 3 shows the lattice of isotropy subgroups for $\mathbb{D}_6$ acting on each of these six isotypic components

$V^{(j)}$. These six cases can be distinguished by determining which isotypic component an arbitrary eigenfunction in E belongs to. We shall go through each of these six cases and describe the resulting bifurcation. Recall that $\Gamma = \Gamma_1 = \mathbb{D}_6$ for each of these six cases, and $\Gamma' = \mathrm{pStab}(E, \Gamma)$.

$$\begin{aligned}
&E \subseteq V^{(1)} \Rightarrow \Gamma' = \Gamma_1 = \langle \rho, \sigma, \tau \rangle, && \dim E = 1,\ \Gamma/\Gamma' \cong \langle 1 \rangle,\\
&E \subseteq V^{(2)} \Rightarrow \Gamma' = \Gamma_{13} = \langle \rho \rangle, && \dim E = 1,\ \Gamma/\Gamma' \cong \mathbb{Z}_2,\\
&E \subseteq V^{(3)} \Rightarrow \Gamma' = \Gamma_9 = \langle \rho^2, \sigma \rangle, && \dim E = 1,\ \Gamma/\Gamma' \cong \mathbb{Z}_2,\\
&E \subseteq V^{(4)} \Rightarrow \Gamma' = \Gamma_{10} = \langle \rho^2, \tau \rangle, && \dim E = 1,\ \Gamma/\Gamma' \cong \mathbb{Z}_2,\\
&E \subseteq V^{(5)} \Rightarrow \Gamma' = \Gamma_{19} = \langle \rho^3 \rangle, && \dim E = 2,\ \Gamma/\Gamma' \cong \mathbb{D}_3,\\
&E \subseteq V^{(6)} \Rightarrow \Gamma' = \Gamma_{22} = \langle 1 \rangle, && \dim E = 2,\ \Gamma/\Gamma' \cong \mathbb{D}_6.
\end{aligned}$$

The first case, $E \subseteq V^{(1)} = \mathrm{Fix}(\Gamma_1, L^2(\Omega))$, does not lead to a symmetry-breaking bifurcation. The $\mathbb{D}_6$ action on E is trivial, so the EBL does not apply. The degenerate critical point (u^*, λ^*) is typically a fold point (or saddle-node), not a bifurcation point. In the neighborhood of the fold point there is only one solution branch, with isotropy subgroup Γ_1, and the branch lies to one side of $\lambda = \lambda^*$ or the other.

The next three cases, with $\Gamma/\Gamma' \cong \mathbb{Z}_2$ symmetry, are called *pitchfork bifurcations*. Clearly, the only maximal isotropy subgroup is Γ' in each case, and the EBL applies. The effective symmetry group acting on E is $\mathbb{Z}_2$, so there are two conjugate solution branches created at the bifurcation. In the branch following code we follow one of these branches using the pmGNGA starting with any eigenvector $e \in E$.

The next case, with $E \subseteq V^{(5)}$, is a bifurcation with $\mathbb{D}_3$ symmetry. The maximal isotropy subgroup $\Gamma_5 = \langle \sigma, \tau \rangle$ satisfies

$$\dim \mathrm{Fix}(\Gamma_5, E) = 1 \quad \text{and} \quad N_{\Gamma_1}(\Gamma_5)/\Gamma_5 = \langle 1 \rangle.$$

Our branch following code uses a projection operator to find an eigenvector $e \in E$ with $\mathrm{Stab}(e, \Gamma_1) = \Gamma_5$. The pmGNGA using this eigenvector e will follow one of the solution branches created at the bifurcation, and the pmGNGA using the negative eigenvector $-e$ will find a branch that is not conjugate to the first. Bifurcations with $\mathbb{D}_3$ symmetry are typically transcritical, and two $\mathbb{D}_3$-orbits of branches are created at the bifurcation [10, 11].

The last case, with $E \subseteq V^{(6)}$, is a bifurcation with $\mathbb{D}_6$ symmetry. There are two maximal symmetry types, the conjugacy classes of Γ_{15} and Γ_{16}. A calculation shows that

$$\dim \mathrm{Fix}(\Gamma_{15}, E) = \dim \mathrm{Fix}(\Gamma_{16}, E) = 1 \quad \text{and} \quad N_{\Gamma_1}(\Gamma_{15})/\Gamma_{15} = N_{\Gamma_1}(\Gamma_{16})/\Gamma_{16} = \mathbb{Z}_2.$$

To follow one branch from each of the group orbits of solution branches created at this bifurcation, it suffices to use the pmGNGA twice, with the eigenvectors $e_1, e_2 \in E$, where $\mathrm{Stab}(e_1, \Gamma_1) = \Gamma_{15}$ and $\mathrm{Stab}(e_2, \Gamma_1) = \Gamma_{16}$. It is well known that these EBL-branches are typically the only branches created at a bifurcation with $\mathbb{D}_6$ symmetry [10, 11].

Example 4.13. Consider a degenerate critical point with isotropy subgroup $\Gamma_{13} = \langle \rho \rangle \cong \mathbb{Z}_6$. Barring accidental degeneracy, the center eigenspace E is a subspace of one of the four isotypic

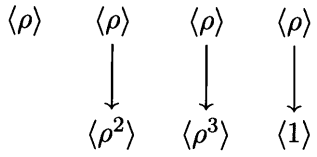


Figure 4. *The diagrams of the four isotropy lattices for the actions of $\Gamma_{13} = \langle\rho\rangle$ on each of the four isotypic components $V^{(j)}_{\langle\rho\rangle}$ of the Γ_{13} action on $L^2(\Omega)$. This describes the four possibilities (barring accidental degeneracy) for the Γ_{13} action on the center eigenspace E at a degenerate critical point.*

components $V^{(j)}_{\langle\rho\rangle}$ defined in Example 4.9. Figure 4 shows the lattice of isotropy subgroups for Γ_{13} acting on each of these four isotypic components. Recall that $\Gamma = \Gamma_{13} = \langle\rho\rangle$ for each of these cases, and the minimal isotropy subgroup is $\Gamma' = \mathrm{pStab}(E, \Gamma)$. We shall go through each of the four cases, and describe the resulting bifurcation:

$$
\begin{aligned}
E \subseteq V^{(1)}_{\langle\rho\rangle} = V^{(1)} \oplus V^{(2)} \quad &\Rightarrow \quad \Gamma' = \Gamma_{13} = \langle\rho\rangle, \qquad \dim E = 1, \quad \Gamma/\Gamma' \cong \langle 1\rangle,\\
E \subseteq V^{(2)}_{\langle\rho\rangle} = V^{(3)} \oplus V^{(4)} \quad &\Rightarrow \quad \Gamma' = \Gamma_{21} = \langle\rho^2\rangle, \qquad \dim E = 1, \quad \Gamma/\Gamma' \cong \mathbb{Z}_2,\\
E \subseteq V^{(3)}_{\langle\rho\rangle} = V^{(5)} \quad &\Rightarrow \quad \Gamma' = \Gamma_{19} = \langle\rho^3\rangle, \qquad \dim E = 2, \quad \Gamma/\Gamma' \cong \mathbb{Z}_3,\\
E \subseteq V^{(4)}_{\langle\rho\rangle} = V^{(6)} \quad &\Rightarrow \quad \Gamma' = \Gamma_{22} = \langle 1\rangle, \qquad \dim E = 2, \quad \Gamma/\Gamma' \cong \mathbb{Z}_6.
\end{aligned}
$$

The first two cases are analogous to the first two cases in Example 4.12. When $\Gamma/\Gamma' \cong \langle 1\rangle$ there is a fold point, but no symmetry-breaking bifurcation. There is a pitchfork bifurcation when $\Gamma/\Gamma' \cong \mathbb{Z}_2$. The next two cases are interesting because Γ_{13} does not act absolutely irreducibly on E, and the EBL does not apply. In both cases Γ' is a maximal isotropy subgroup.

In the third case, where $E \subseteq V^{(3)}_{\langle\rho\rangle} = V^{(5)}$, every eigenfunction in the two-dimensional E has isotropy subgroup Γ_{19}. Since we have a gradient system, we know that solution branches with isotropy subgroup Γ_{19} are created at this bifurcation with $\mathbb{Z}_3$ symmetry. The bifurcation is well understood, and it looks like a bifurcation with $\mathbb{D}_3$ symmetry, except that the "angle" of the bifurcating solutions in the E plane is arbitrary. This means that trial and error is needed, in general, to find eigenfunctions in E for which the pmGNGA will converge. If a branch is found for a starting eigenfunction e, then the eigenfunction $-e$ is used to find the other solution branch.

In the fourth case, where $E \subseteq V^{(4)}_{\langle\rho\rangle} = V^{(6)}$, every eigenfunction in E has the same isotropy subgroup: $\Gamma_{22} = \langle 1\rangle$. Gradient bifurcations with $\mathbb{Z}_6$ symmetry look like bifurcations with $\mathbb{D}_6$ symmetry, except that the angle in the E plane is arbitrary. Again, trial and error is needed to find starting eigenfunctions for which the pmGNGA converges.

The bifurcation digraph. A calculation similar to those summarized in Examples 4.12 and 4.13 was done for each of the isotropy subgroups of the $\mathbb{D}_6 \times \mathbb{Z}_2$ action on $L^2(\Omega)$. The calculations were done by hand and verified with GAP. There are 59 generic symmetry-breaking bifurcations, one for each isotypic component $V^{(j)}_{\Gamma_i}$ on which Γ_i acts nontrivially. The amount of information is overwhelming, so we display the essential results in what we call a bifurcation digraph.

Definition 4.14. *The* bifurcation digraph *of the Γ_0 action on a real vector space V is a directed graph with labeled arrows. The vertices are the symmetry types (equivalence classes of isotropy subgroups). Given $\Sigma \leq \Gamma$, two isotropy subgroups of the Γ_0 action on V, we draw an arrow from $[\Gamma]$ to $[\Sigma]$ if and only if Σ is a maximal isotropy subgroup of the Γ action on some isotypic component $V_\Gamma^{(j)}$ of V. Each arrow has the label Γ/Γ', where Γ' is the kernel of the Γ action on $V_\Gamma^{(j)}$. Furthermore, each arrow is either solid, dashed, or dotted. The arrow is*

$$\begin{aligned} \textit{solid if } & \dim \operatorname{Fix}(\Sigma, E) = 1 \textit{ and } N_\Gamma(\Sigma)/\Sigma = \mathbb{Z}_2, \\ \textit{dashed if } & \dim \operatorname{Fix}(\Sigma, E) = 1 \textit{ and } N_\Gamma(\Sigma)/\Sigma = \langle 1 \rangle, \textit{ and} \\ \textit{dotted if } & \dim \operatorname{Fix}(\Sigma, E) \geq 2, \end{aligned}$$

where E is any irreducible subspace contained in $V_\Gamma^{(j)}$.

Note that if $\dim \operatorname{Fix}(\Sigma, E) = 1$, then $N_\Gamma(\Sigma)/\Sigma$ is either $\mathbb{Z}_2$ or $\langle 1 \rangle$, since these are the only linear group actions on $E \cong \mathbb{R}^1$. Thus, the three arrow types (solid, dashed, and dotted) exhaust all possibilities.

Theorem 4.15. *For a given Γ_0 action on V, every arrow in the diagram of the isotropy lattice is an arrow in the bifurcation digraph.*

Proof. Suppose that $[\Gamma] \to [\Sigma]$ is an arrow in the diagram of the isotropy lattice. Then some $\Sigma^* \in [\Sigma]$ is a maximal isotropy subgroup of the Γ action on V. Choose $u^* \in V$ such that $\operatorname{Stab}(u^*, \Gamma) = \Sigma^*$. Such a u^* exists since Σ^* is an isotropy subgroup. Now consider the isotypic decomposition $\{V_\Gamma^{(j)}\}_{j \in J}$ of V. We can write $u^* = \sum_{j \in J} u^{(j)}$, where $u^{(j)} \in V_\Gamma^{(j)}$ are uniquely determined. Let γ be any element of Σ^*. Then $\gamma \cdot u^* = \sum_{j \in J} \gamma \cdot u^{(j)} = u^*$. Since each of the components $V_\Gamma^{(j)}$ is Γ-invariant, $\gamma \cdot u^{(j)} = u^{(j)}$ for each $j \in J$. Thus $\Sigma^* \leq \operatorname{Stab}(u^{(j)}, \Gamma)$ for each $j \in J$. Either $\operatorname{Stab}(u^{(j)}, \Gamma) = \Gamma$ or $\operatorname{Stab}(u^{(j)}, \Gamma) = \Sigma^*$, since Σ^* is a maximal isotropy subgroup of the Γ action on V. If $\operatorname{Stab}(u^{(j)}, \Gamma) = \Gamma$ for all $j \in J$, then $\operatorname{Stab}(u^*, \Gamma) = \Gamma$. But $\operatorname{Stab}(u^*, \Gamma) \neq \Gamma$, so $\operatorname{Stab}(u^{(j)}, \Gamma) = \Sigma^*$ for some $j \in J$, and Σ^* is a maximal isotropy subgroup of the Γ action on this component $V_\Gamma^{(j)}$ of V. Therefore the bifurcation digraph has an arrow from $[\Gamma]$ to $[\Sigma^*] = [\Sigma]$. ∎

Theorem 4.15 says that the bifurcation digraph is an extension of the diagram of the isotropy lattice. The bifurcation digraph has more arrows, in general. As with the lattice of isotropy subgroups, we usually draw a single element Γ of the equivalence class $[\Gamma]$ for each vertex of the bifurcation digraph.

An arrow from Γ to Σ in the bifurcation digraph indicates that a Γ_0-equivariant gradient system $g(\lambda, u) = 0$ can have a generic symmetry-breaking bifurcation where a mother branch with isotropy subgroup Γ creates a daughter branch with isotropy subgroup Σ. The symmetry group of the bifurcation is Γ/Γ', and the center eigenspace at the bifurcation point is the Γ-irreducible space E. The information encoded in the label and arrow type is used by the heuristics of our branch following algorithm. A solid arrow indicates that every e in the one-dimensional space $\operatorname{Fix}(\Sigma, E)$ satisfies $\gamma \cdot e = -e$ for some $\gamma \in \Sigma$. Thus, there is typically a pitchfork bifurcation in the space $\operatorname{Fix}(\Sigma, E)$. A dashed arrow indicates that $\gamma \cdot e = e$ for all $e \in \operatorname{Fix}(\Sigma, E)$ and $\gamma \in \Sigma$. Thus, the daughter branches bifurcating in the directions e and $-e$ are not conjugate. A dotted arrow indicates that the EBL does not apply to this bifurcation.

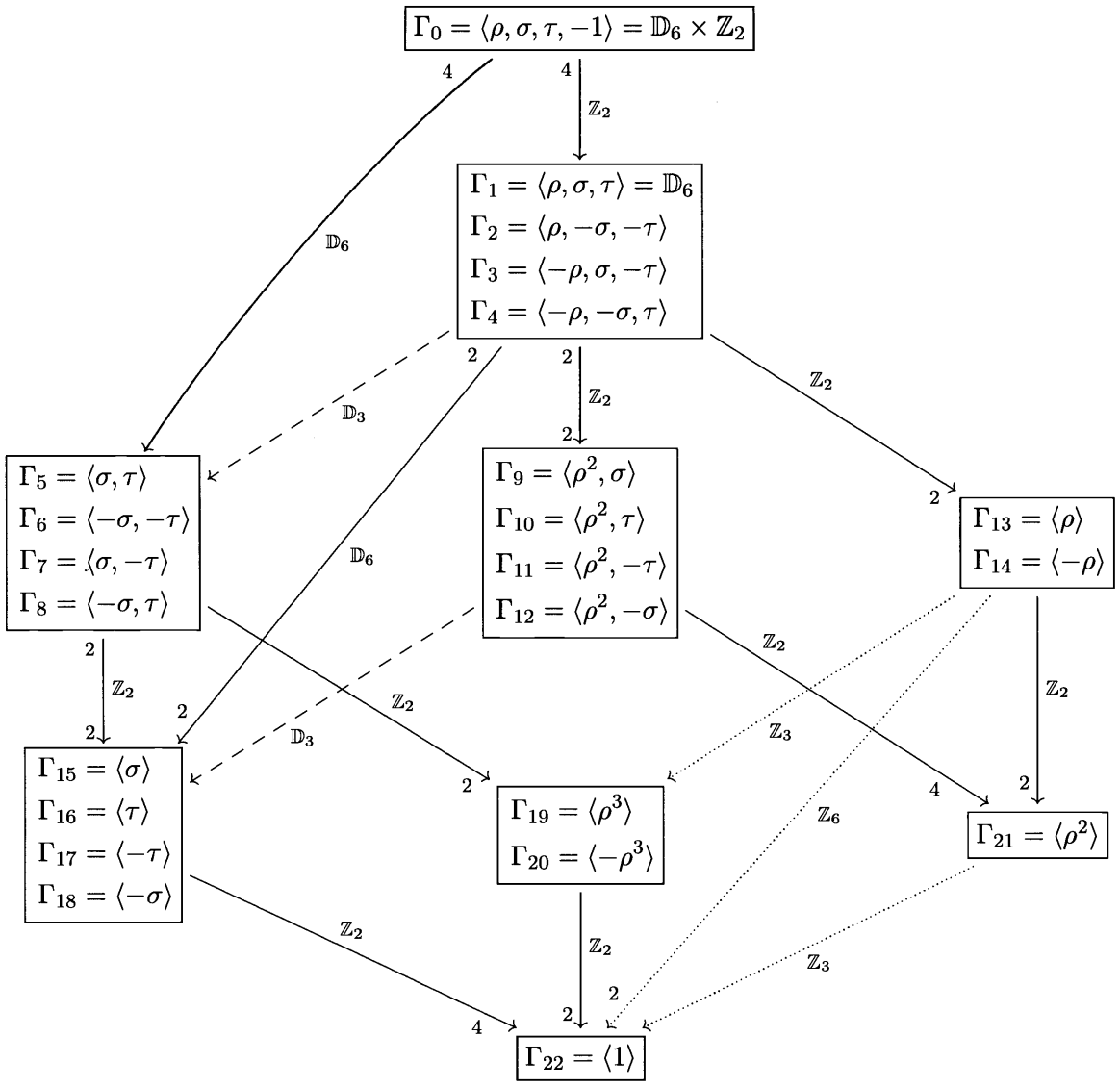


Figure 5. *The bifurcation digraph for the $\mathbb{D}_6 \times \mathbb{Z}_2$ action on $L^2(\Omega)$ extends the diagram of the isotropy lattice. The digraph shown is condensed as in Figure 2. The arrows indicate generic symmetry-breaking bifurcations. The MI of the mother branch changes by 1 at bifurcations with $\mathbb{Z}_2$ symmetry, and it changes by 2 at all other bifurcations shown here.*

As mentioned above, branching of solutions corresponding to a dotted arrow is generic in gradient systems [30, 10].

A condensed bifurcation digraph for the $\mathbb{D}_6 \times \mathbb{Z}_2$ action on $L^2(\Omega)$ is shown in Figure 5. The calculations for the directed edges coming from Γ_1 and Γ_{13} are described in Examples 4.12 and 4.13, respectively. The digraph has 65 directed edges, but there are only five possibilities for the symmetry group of the bifurcation: $\Gamma/\Gamma' = \mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_6$, $\mathbb{D}_3$, or $\mathbb{D}_6$. The symmetry-breaking bifurcation with each of these symmetries is well understood [10, 11], and each is

described briefly in Example 4.12 or 4.13. This digraph is of great help in writing an automated code for branch following.

In our problem the label Γ/Γ' and arrow type are sufficient to characterize the bifurcation completely. For more complicated groups, the label may need to contain more information about the action of Γ on E. For example the label $\Gamma/\Gamma' = \mathbb{S}_4$ would be ambiguous, since $\mathbb{S}_4$ has two faithful irreducible representations with different lattices of isotropy subgroups.

5. Symmetry and computational efficiency. Several modifications of Algorithm 2.1 take advantage of symmetry to speed up the calculations. The symmetry forces many of the components of the gradient and Hessian to be zero. We identified these zero components and avoided doing the time-consuming numerical integrations to compute them. At the start of the C++ program, the isotropy subgroup, Γ_i, of the initial guess is computed. Recall that there are M modes in the Galerkin space B_M, so $\dim(B_M) = M$. Define $M_i := \dim(\mathrm{Fix}(\Gamma_i, B_M))$. We chose the representatives Γ_i within each conjugacy class so that $\mathrm{Fix}(\Gamma_i, B_M)$ is a coordinate subspace of B_M. Thus, $M - M_i$ components of the gradient $g(\lambda, u)$ are zero if $\mathrm{Fix}(u) = \Gamma_i$. The numerical integrations in (5) are done only for the M_i potentially nonzero components of g. Similarly, $M_i(M_i + 1)/2$ rather than $M(M + 1)/2$ numerical integrations are needed to compute the part of the Hessian matrix h needed by the GNGA algorithm: The numerical integrations in (6) are done only if ψ_j and ψ_k are both in $\mathrm{Fix}(\Gamma_i, B_M)$. The system $h\chi = g$ for the Newton step χ reduces to a system of M_i equations in M_i unknowns. After Newton's method converges to a solution, the full Hessian needs to be calculated in order to compute the MI. Here, too, we can take advantage of the symmetry: Since h is Γ_i-equivariant, $h_{jk} = 0$ if ψ_j and ψ_k are in different isotypic components $V_{\Gamma_i}^{(j)}$ of B_M.

As an example, consider the execution time for approximating a solution with Γ_1 symmetry using $M = 300$ modes and a level $\ell = 5$ grid on a 1GHz PC. Our C++ code uses only $M_1 = 30$ modes, and takes about 1.5 seconds per Newton step, compared to 44 seconds when the symmetry speedup is not implemented.

6. Automated branch following. The branch following code is a complex collection of about a dozen Perl scripts, Mathematica and Gnuplot scripts, and a C++ program. These programs write and call each other fully automatically and communicate through output files, pipes, and command line arguments. A complete bifurcation diagram can be produced by a single call to the main Perl script.

Two choices for the function of u plotted against λ are shown in Figure 6. In most bifurcation diagrams we plot approximate solutions u evaluated at a generic point $(2/27, 4\sqrt{3}/27)$ (the big dot in Figure 1) versus the parameter λ; other choices for the vertical axis such as $J(u)$ or $\|u\|_\infty$ lead to less visible separation of branches. Two conjugate solutions can have different values at the generic point, but since our program follows only one branch in each group orbit this does not cause a problem.

The C++ program implements the GNGA algorithm. Its input is a vector of coefficients $a \in \mathbb{R}^M$ for an initial guess in Newton's method, an interval for λ, a stepsize for λ, and several other parameters such as the grid level. It finds solutions on a single branch of the bifurcation diagram. Every solution is written as a single line in an output file. This line contains all the

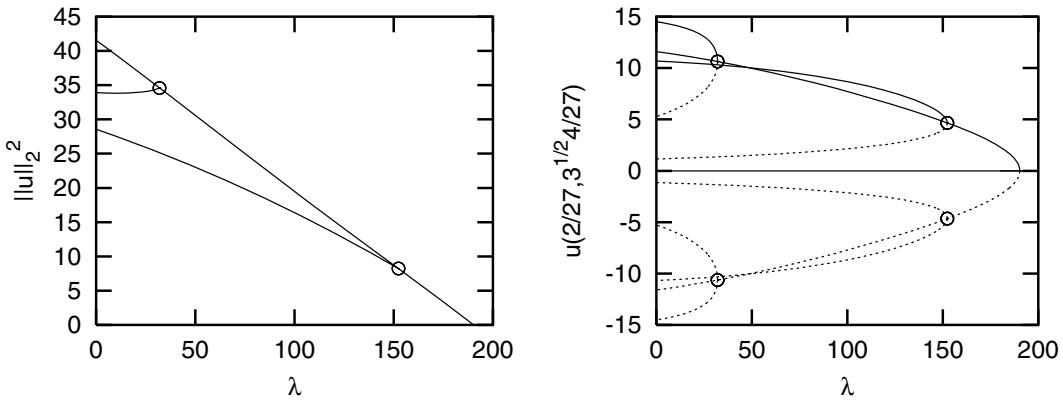


Figure 6. *Bifurcation diagrams of the sixth primary branch (which bifurcates from λ_6), showing $||u||_2^2$ and $u(2/27, 4\sqrt{3}/27)$ as functions of λ. Since $||u||_2^2$ is a $\mathbb{D}_6 \times \mathbb{Z}_2$-invariant function of u, each group orbit of solution branches is shown as one curve on the left. The disadvantage of plotting $||u||_2^2$ is that the curves in many bifurcation diagrams are not well separated. The point $(2/27, 4\sqrt{3}/27)$ is not on any of the reflection axes of the snowflake region. There are two primary branches with symmetry S_1, four secondary branches with symmetry S_9, and four secondary branches with symmetry S_{10}. Our choice for the bifurcation diagrams in this paper combines the advantages of both views: $u(2/27, 4\sqrt{3}/27)$ is plotted as a function of λ for exactly one branch (the solid lines) from each group orbit. Unless indicated otherwise, all figures were produced with level $\ell = 5$ and $M = 300$ modes.*

information about the solution and can be used to write an input file for a subsequent call to the same C++ program.

The C++ program finds one branch (referred to as the main branch) and a short segment of each of the daughter branches created at bifurcations of the main branch. The coefficients approximating the first solution on the branch are supplied to the C++ program. Newton's method is used to find this first solution, then λ is incremented, and the next solution is found. The program attempts to follow the main branch all the way to the final λ, usually 0. Heuristics are used to double or halve the λ stepsize when needed, keeping the stepsize in the interval from the initial stepsize (input to the C++ program) to 1/32 of the initial stepsize. For example, the stepsize is halved if Newton's method does not converge, if it converges to a solution with the wrong symmetry, or if more than one bifurcation is detected in one λ step.

The MI is computed at each λ value on the main branch. When the MI changes, a subroutine is called to handle the bifurcation before the main branch is continued. If the MI changes from m_1 to m_2, we define $m = \max\{m_1, m_2\}$. Then the bifurcation point is approximated by using the secant method to set the mth eigenvalue of the Hessian $h(u)$ to zero as a function of λ. The GNGA is needed at each step of the secant method to compute $u = u(\lambda)$. We find that the GNGA works well even though we are approximating a solution for which the Hessian is singular.

After the bifurcation point is approximated, a short segment of each bifurcating branch is computed, and one output file is written for each branch, using Algorithm 6.1. If the EBL holds, then we know exactly which critical eigenvector to use for each branch.

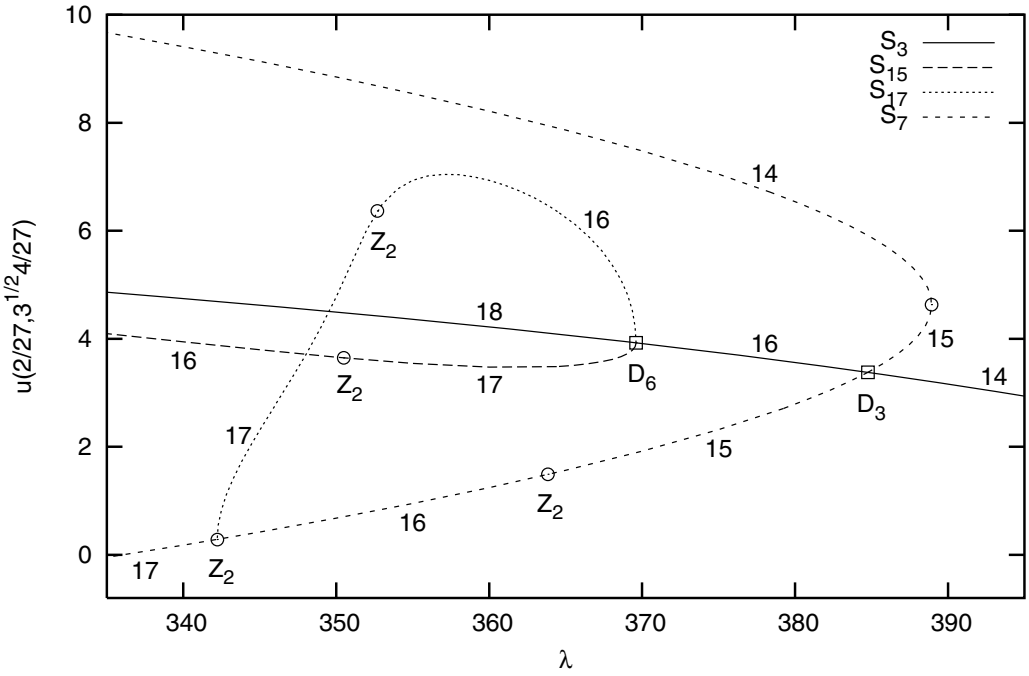


Figure 7. *A partial bifurcation diagram of the 14th primary branch showing a $\mathbb{D}_6$, a $\mathbb{D}_3$, and several $\mathbb{Z}_2$ bifurcations. At the $\mathbb{D}_6$ bifurcation, 12 branches in two different group orbits are born. In accordance with Figure 6, only two branches are followed and shown on this bifurcation diagram. An animation showing the followed branch with symmetry type S_{15} is shown in 64048_01.gif [152KB], and an animation of the followed branch with symmetry type S_{17} is in 64048_02.gif [246KB]. Note that this branch with S_{17} symmetry "dies" at a bifurcation with $\mathbb{Z}_2$ symmetry, showing that we cannot always make a consistent distinction between secondary and tertiary branches. At the $\mathbb{D}_3$ bifurcation, six branches in two different group orbits are born. As before, only two branches are followed. An animation showing the "upper" branch with symmetry type S_7, through the bifurcation point and continuing to the "lower" branch with symmetry type S_7 is shown in 64048_03.gif [178KB]. For clarity, the branches bifurcating from three of the $\mathbb{Z}_2$ bifurcations are not shown. The numbers next to a branch indicate the MI of the solution; the MI changes by 2 at a square, and by 1 at a circle.*

Algorithm 6.1 (`follow_branch`).

1. Input: bifurcation point (λ, a), one critical eigenvector $e \in \mathbb{R}^M$, and stepsize $\Delta\lambda < 0$. Output: A file is written for one daughter branch.
2. Write (λ, a) to output file. Set $t = 0.1$. Set $\lambda_b = \lambda$.
3. Compute index k so that $|e_k| \geq |e_i|$ for all $i \in \{1, \ldots, M\}$.
4. Repeat until $\lambda_b - \lambda < \Delta\lambda$, or $t < 0.1/32$, or some maximum number of points have been written to the file.
   (a) Do the pmGNGA with initial guess $(\lambda, a + t\,e)$, fixing coefficient k.
   (b) If Newton's method converges, replace (λ, a) by the solution found and write this point to the file; else $t \leftarrow t/2$.

Note that the pmGNGA can follow a branch that bifurcates to the right or the left. Those that bifurcate to the right usually turn over in a saddle-node "bifurcation" that does not offer any difficulty for the pmGNGA. Figures 7 and 8 show several examples of bifurcations.

The EBL does not hold at bifurcations with $\mathbb{Z}_3$ and $\mathbb{Z}_6$ symmetry in our problem, since

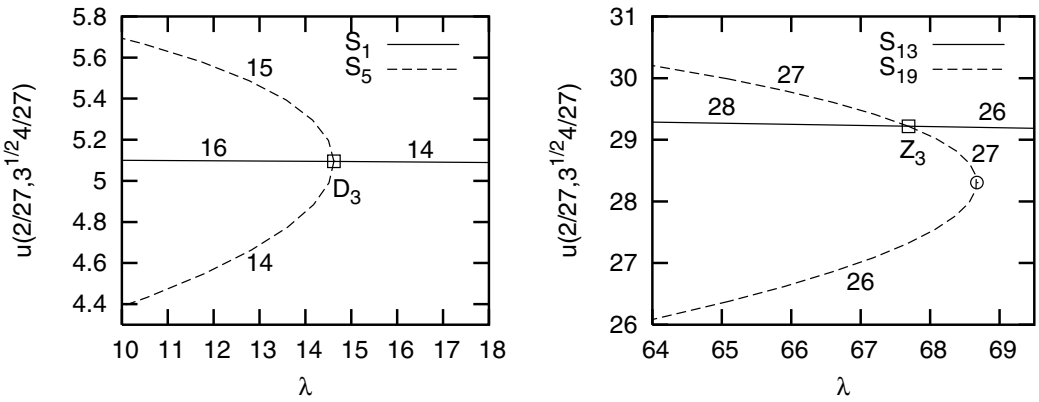


Figure 8. *The $\mathbb{D}_3$ bifurcation of the 13th primary branch is on the left. This is the only observed $\mathbb{D}_3$ bifurcation that is not transcritical. An animation of the upper branch with symmetry type S_5, through the bifurcation point and continuing with the lower branch, is shown in* 64048_04.gif [118KB]. *A $\mathbb{Z}_3$ bifurcation of a daughter of the 24th primary branch is shown on the right. The branches created at this bifurcation are not described by the EBL. An animation of the branches with symmetry type S_{19} is shown in* 64048_05.gif [101KB].

the two-dimensional center eigenspace does not have a one-dimensional subspace with more symmetry. Figure 8 shows one of the few bifurcations with $\mathbb{Z}_3$ symmetry that we observed. By good fortune, the branches with symmetry type S_{19} were successfully followed using the same eigenvectors one would choose for a bifurcation with $\mathbb{D}_3$ symmetry. A better method for following bifurcating solutions that are not predicted by the EBL would be to use the pmGNGA with random (normalized) eigenvectors in E repeatedly until it appears that all equivalence classes of solutions have been found.

The branch following code is called recursively by a main Perl script. Initially, the C++ program follows the trivial branch on a given λ range. This results in an output file for the trivial branch and another output file for each bifurcating primary branch. Then the short parts of the primary branches are followed with more calls to the C++ program. Any bifurcating branch results in a new output file, and the Perl script makes another call to the C++ program to continue that branch. The main Perl script's most important job is bookkeeping. It saves the output files with distinct names, and calls the branch following code to continue each of the new branches. The process stops when all the branches are fully followed within the given λ range.

In this way, a complete bifurcation diagram is produced by a single invocation of the main Perl script. There is no need to guess initial conditions for input to Newton's method, since the trivial solution is known exactly ($a = 0$) and all the other solutions are followed automatically.

The main Perl script calls several other smaller scripts. For example, there is a script which extracts solutions from output files and feeds them to the branch following code as input. Another script creates Gnuplot scripts on the fly to generate bifurcation diagrams. Perl scripts are used to automatically number and store the output files and create human readable reports about them.

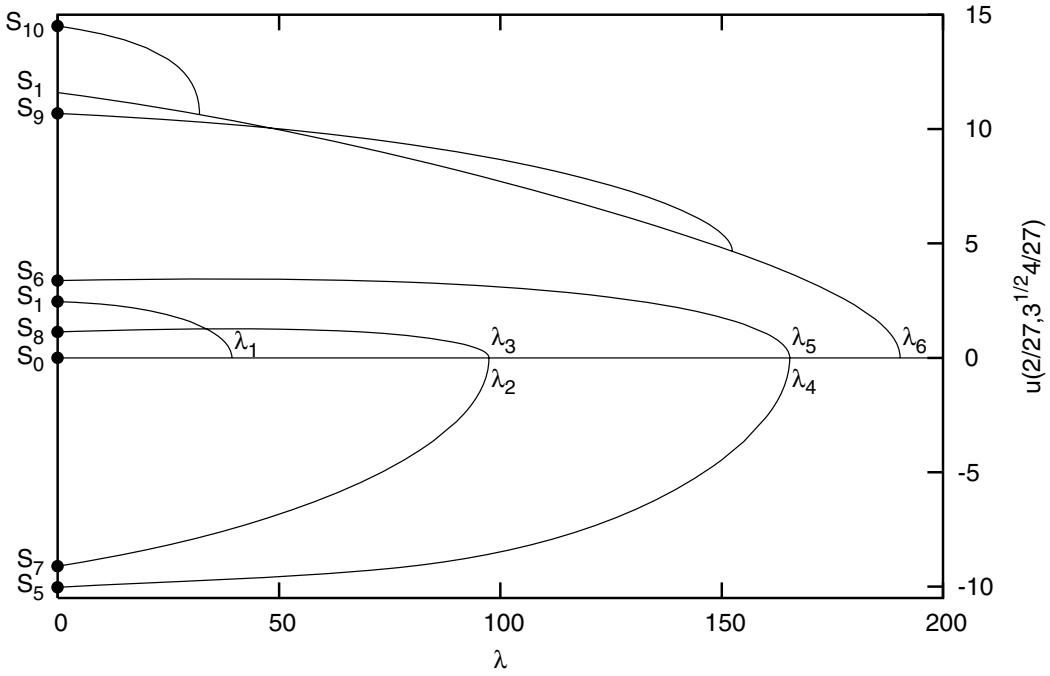


Figure 9. *The complete bifurcation diagram for the first six primary branches bifurcating from the trivial branch. The second branch, with symmetry S_7, contains the CCN solution. The dots at $\lambda = 0$ in Figures 9–12 correspond to solutions depicted in Figures 13 and 14. We used the level 5 grid with 300 modes in creating all bifurcation diagrams. In Figure 15 convergence data for the solution of symmetry type S_{10} at $\lambda = 0$ is provided.*

7. Numerical results. Our goal was to find solutions to (1), (2) at $\lambda = 0$ with each of the 23 symmetry types. The 24th primary branch is the first one with symmetry type S_2, so we followed the first 24 primary branches. With level $\ell = 5$ and $M = 300$ modes, which gave our most accurate results, this found solutions with all symmetry types except S_{11} and S_{14}. We then searched the first 100 primary branches, following only solutions with symmetry above S_{11} and S_{14} on the bifurcation digraph (Figure 5). In this way we found solutions with all 23 symmetry types. The bifurcation diagrams which lead to these solutions are shown in Figures 9–12. We chose one solution at $\lambda = 0$ with each symmetry type by taking the one descended from the lowest primary branch. These choices are indicated by dots in Figures 9–12, and the corresponding contour diagrams of the solutions are shown in Figures 13 and 14. The contour diagrams use white for $u > 0$ and black for $u < 0$, and gray indicates $u = 0$. Equally spaced contours are drawn along with dots for local extrema. Details about the technique for generating these contour diagrams are found in [27].

We ran our experiments using a range of modes and levels in order to observe convergence and qualitative stability of the implementation of our algorithm. At level $\ell = 5$ we have computed 300 eigenfunctions, so $M \leq 300$ is possible. At level $\ell = 6$ we computed only 100 eigenfunctions. Due to our limited computational resources, using more than 100 modes on level 6 was not practical.

As an indication of the convergence, consider the bifurcation diagram in Figure 9. The

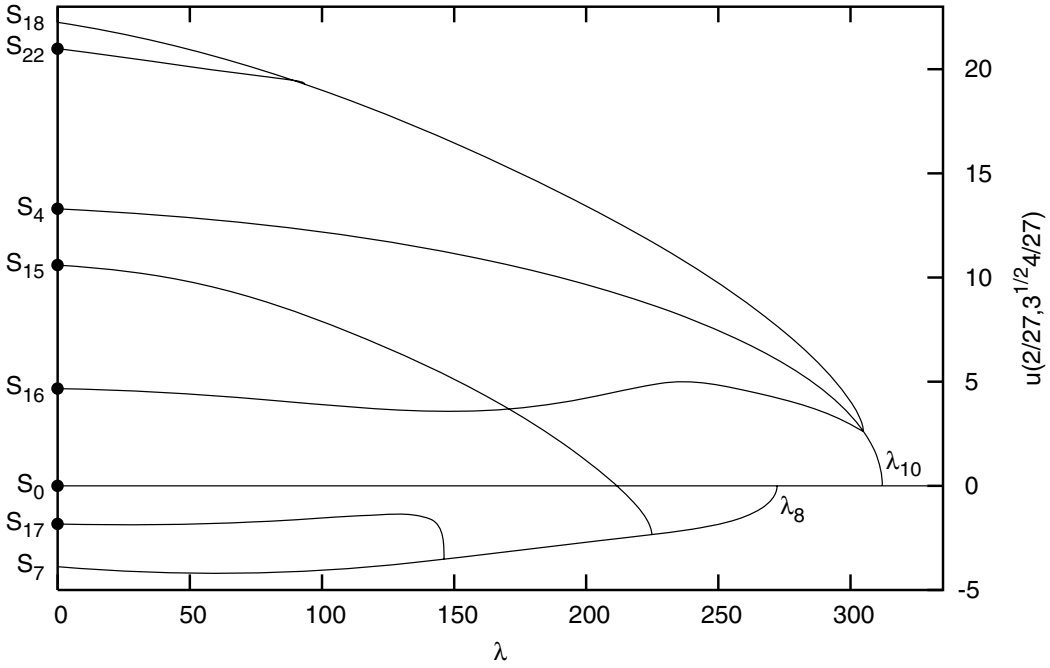


Figure 10. *A partial bifurcation diagram showing some of the solutions bifurcating from the 8th and 10th primary branches. Again, the dots at $\lambda = 0$ indicate solutions shown in Figures 13 and 14. The contour plots as a function of λ are animated for the branches ending with the dots indicating symmetry types S_{15} (64048_06.gif [126KB]), S_{17} (64048_07.gif [161KB]), S_{16} (64048_08.gif [115KB]), and S_{22} (64048_09.gif [147KB]).*

diagram looks qualitatively the same for any choice of ℓ and M that we use. The position of the bifurcation point creating the S_{10} solution (near $\lambda = 30$) changes slightly, according to the following table:

	$\ell = 4$	$\ell = 5$	$\ell = 6$
$M = 100$	35.3931	34.9814	34.9252
$M = 200$	32.1131	32.2964	
$M = 300$		32.0518	

The level 5 and 6 approximations with $M = 100$ modes are very close, but increasing the mode number has more of an effect. This indicates that the results with $(\ell, M) = (5, 300)$ are more accurate than those with $(6, 100)$. Figure 15 shows how $u(2/27, 4\sqrt{3}/27)$ varies with mode number and ℓ for the solution with S_{10} symmetry at $\lambda = 0$ shown in Figures 9 and 13. The horizontal segments of the graphs correspond to the addition of modes with zero coefficients for this solution. Based on this and other similar convergence results, we chose to use level 5 with 300 modes in most of our numerical experiments.

8. Conclusions. We are currently working on a more general program for recursive branch following in symmetric systems. Starting with any graph, the analogue to (1) is the partial *difference* equation (PdE) $Lu + f(u) = 0$ [25], where L is the well-known discrete Laplacian on that graph and u is a real-valued function on the vertices. Discretizing a PDE as we have

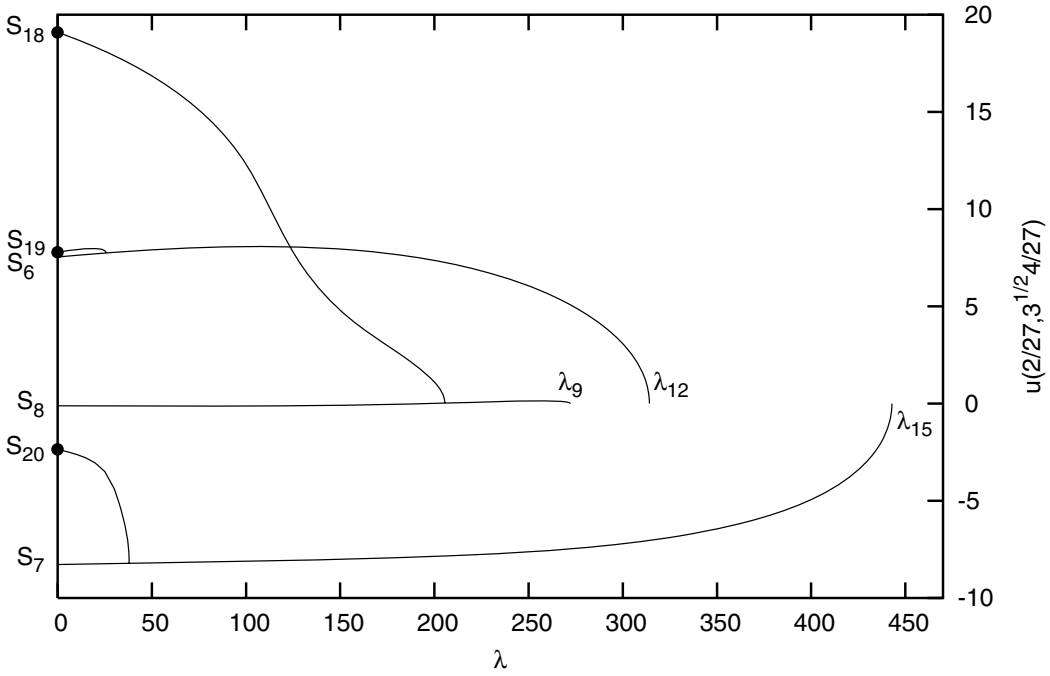


Figure 11. *A partial bifurcation diagram providing three additional symmetry types. For clarity, the trivial branch is not shown in this and the next figure.*

done in this paper leads to a PdE on a graph with a large number of vertices. The grid points are the vertices of the graph, and the edges of the graph connect nearest neighbor grid points. Starting with an arbitrary graph, our new suite of programs will analyze the symmetry of the graph and compute the bifurcation diagrams for the PdE on the graph.

The programs we describe in the current paper will work with other superlinear odd f and other regions with hexagonal symmetry. The nonlinearity f needs to be superlinear since our program assumes that the branches eventually "go to the left." Our general program will not have this restriction; the GNGA and pmGNGA will be replaced by a single method of branch following that is able to go through fold points and has no prejudice about the parameter increasing or decreasing. This new method of branch following has already been successfully implemented in [33]. We hope to write the new code so that a cluster of computers can be used in parallel, with each computer following a single branch at one time, under the control of a central PERL script.

With minor modifications, our program would analyze the PDE (1) even when f is not odd. The appropriate bifurcation digraph for $\mathbb{D}_6$ acting on $L^2(\Omega)$ is a subgraph of the digraph in Figure 5, and so the bifurcating branches would be followed properly unless the symmetry of the mother solution is incorrectly identified. The Perl scripts which start with the trivial branch would have to be modified, since $u = 0$ is not a solution when f is not odd (unless $f(0) = 0$). If $f(0) = 0$, the trivial branch exists, but its bifurcations are not properly described by the bifurcation digraph in Figure 5, and some special code would be needed to handle these bifurcations.

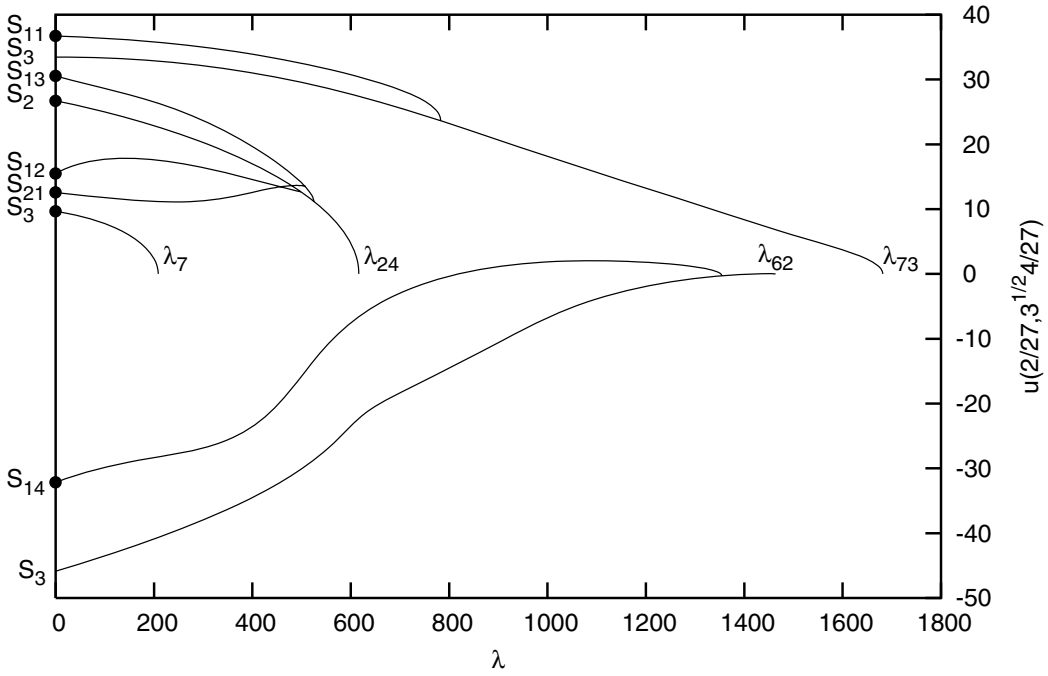


Figure 12. *A partial bifurcation diagram containing solutions of the seven remaining symmetry types. Primary branch* 24 *is the first branch with symmetry type* S_2*. The symmetry types* S_{14} *and* S_{11} *were found by searching the first* 100 *primary branches, following only those branches which can lead to solutions with the desired symmetry. These two solutions are included for completeness, but their existence for the PDE would have to be confirmed with more modes and a higher-level approximation of the eigenfunctions.*

It is valid to ask the question "does the GNGA converge" (as implemented in this current research). While we do not have a complete proof affirming the positive of this conjecture, many references contain relevant theorems. The GNGA is an implementation of Newton's method, which indeed converges under standard assumptions. In [14], one finds the classical fixed point iteration proof that Newton's method in $\mathbb{R}^N$ converges when the initial guess is sufficiently close to a nondegenerate zero of the object function. This proof applies almost without change to the infinite dimensional case. Also addressed in [14] are algorithms where the object function and/or its derivative are only approximated; this would apply to our implementation due to numerical integration errors, as well as owing to our imperfect knowledge of the eigenfunctions and corresponding eigenvalues. While not discussed exactly in the cited literature, elementary fixed point arguments indicate that the restriction of our object function ∇J to sufficiently large subspaces B_M will still result in convergent iterations. It would be worthwhile to string these type of results together in order to obtain a "best possible" GNGA convergence theorem. Monograph [13] gives an easy introduction into some of the details of implementing Newton's method to solve nonlinear problems. Further, in the spirit of [7] and [35], by the invariance of the Newton map, any convergence result should hold in fixed point subspaces corresponding to a given symmetry type. The articles [20, 35] and others by those authors discuss the convergence of algorithms similar to the GNGA, at times also considering symmetry restrictions. Finally, the well-known book [3] contains relevant

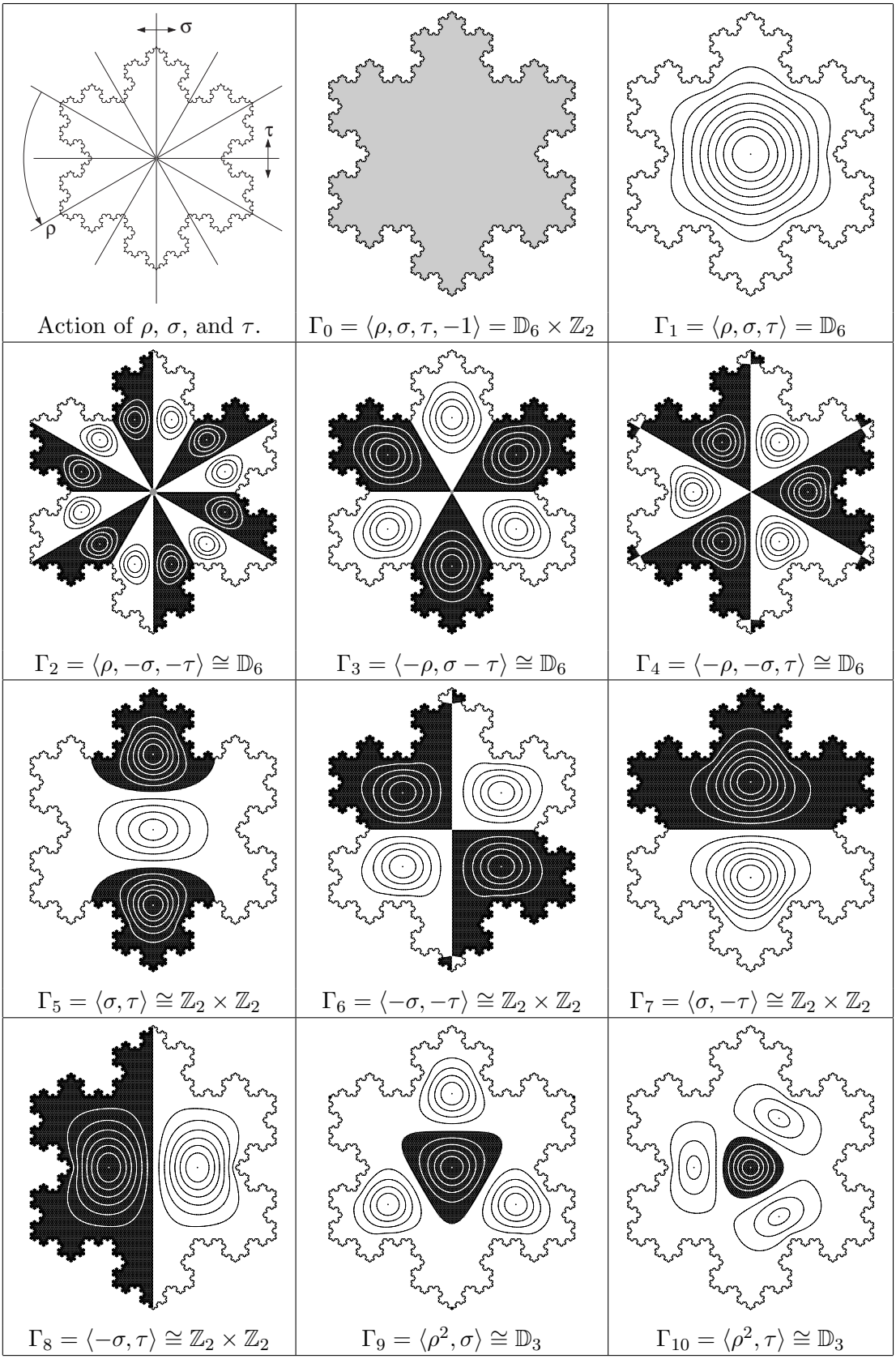


Figure 13. *The action of the generators of $\mathbb{D}_6$ on the plane, along with contour plots of solutions with symmetry types $S_0, \ldots, S_{10}$ at $\lambda = 0$. Recall that $S_i = [\Gamma_i]$.*

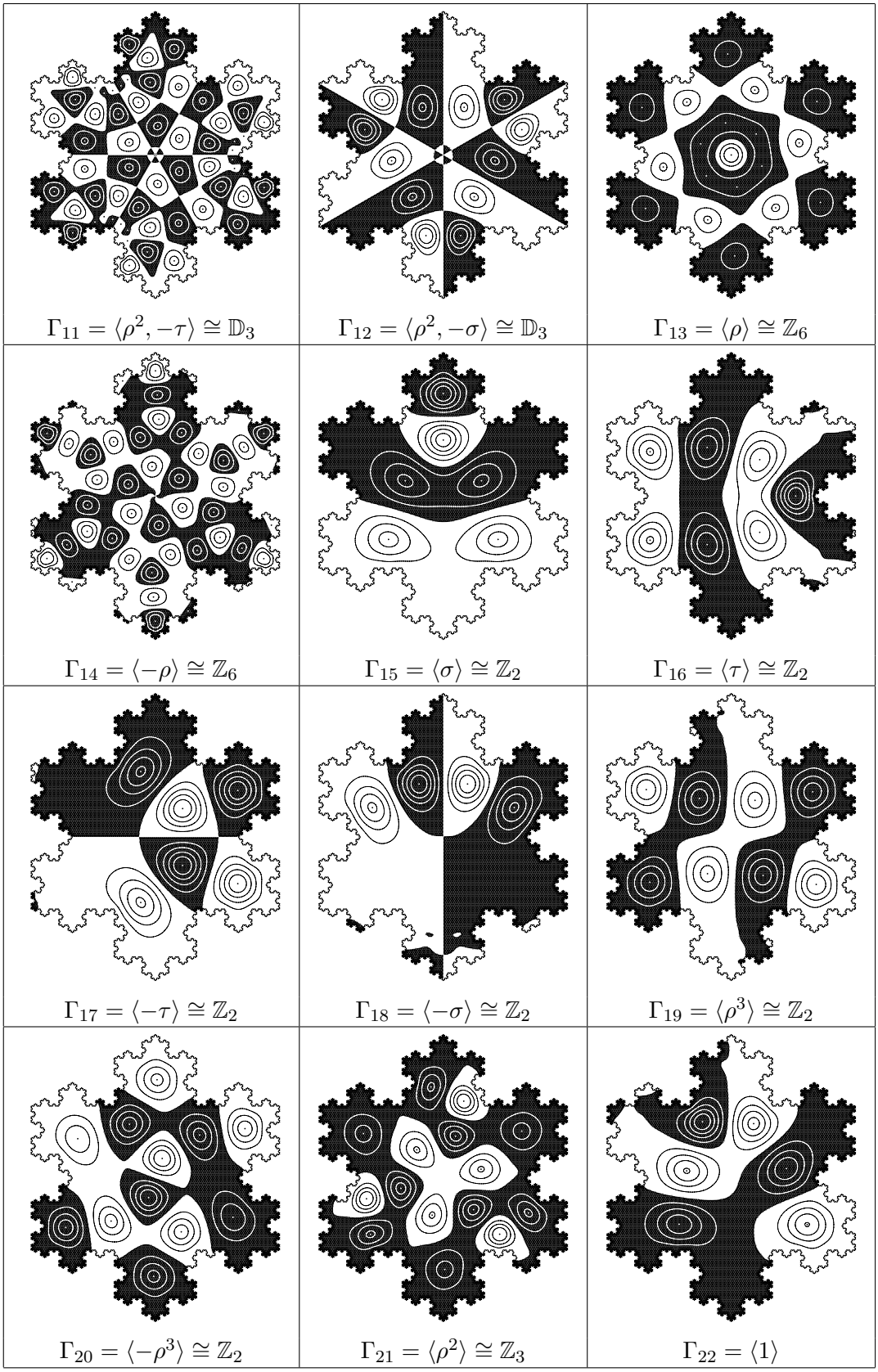


Figure 14. *Contour plots of solutions with symmetry types $S_{11}, \ldots, S_{22}$ at $\lambda = 0$.*

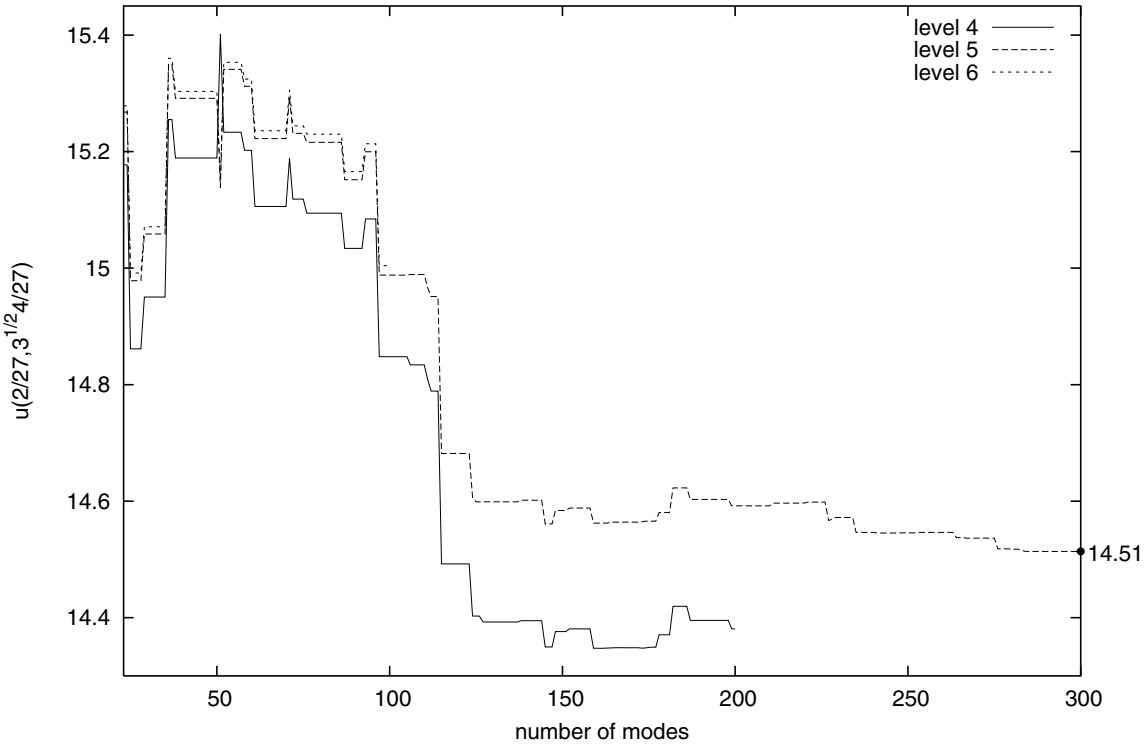


Figure 15. *A plot of* $u(2/27, 4\sqrt{3}/27)$ *as a function of the number of modes for the lowest energy solution at* $\lambda = 0$ *with symmetry type* S_{10}*. The point at* $M = 300$ *matches the point labeled* S_{10} *in Figure 9.*

convergence results for Newton and approximate Newton iterative fixed point algorithms.

In summary, we have written a suite of programs that automatically computes the bifurcation diagram of the PDE (1), (2). The program finds solutions with each of the 23 symmetry types by following solution branches which are connected to the trivial branch by a sequence of symmetry-breaking bifurcations. A thorough understanding of the possible symmetry-breaking bifurcations is required for this task. We introduced the bifurcation digraph, which summarizes the results of the necessary symmetry calculations. For the group $\mathbb{D}_6 \times \mathbb{Z}_2$, these calculations were done by hand and verified by the GAP computer program [8, 21]. In the future, we plan to implement automated branch following in systems where the symmetry group is so complicated that GAP is necessary.

REFERENCES

[1] R. A. Adams, *Sobolev Spaces*, Pure Appl. Math. 65, Academic Press, New York, London, 1975.
[2] A. Ambrosetti and P. H. Rabinowitz, *Dual variational methods in critical point theory and applications*, J. Funct. Anal., 14 (1973), pp. 349–381.
[3] M. S. Berger, *Nonlinearity and Functional Analysis. Lectures on Nonlinear Problems in Mathematical Analysis*, Pure Appl. Math., Academic Press, New York, London, 1977.
[4] A. Castro, J. Cossio, and J. M. Neuberger, *Sign-changing solutions for a superlinear Dirichlet*

problem, Rocky Mountain J. Math., 27 (1997), pp. 1041–1053.

[5] A. Castro, P. Drabek, and J. M. Neuberger, *A sign-changing solution for a superlinear Dirichlet problem,* II, in Proceedings of the Fifth Mississippi State Conference on Differential Equations and Computational Simulations, Electron. J. Differ. Equ. Conf. 10, Southwest Texas State University, San Marcos, TX, 2003, pp. 101–107.

[6] P. Chossat and R. Lauterbach, *Methods in Equivariant Bifurcations and Dynamical Systems*, Adv. Ser. Nonlinear Dynam. 15, World Scientific, River Edge, NJ, 2000.

[7] D. Costa, Z. Ding, and J. M. Neuberger, *A numerical investigation of sign-changing solutions to superlinear elliptic equations on symmetric domains*, J. Comput. Appl. Math., 131 (2001), pp. 299–319.

[8] GAP Group, *GAP—Groups, Algorithms, and Programming*, 2002, http://www.gap-system.org.

[9] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics Math., Springer-Verlag, Berlin, 2001.

[10] M. Golubitsky, I. Stewart, and D. G. Schaefer, *Singularities and Groups in Bifurcation Theory*, Vol. 2, Appl. Math. Sci. 69, Springer-Verlag, New York, 1988.

[11] M. Golubitsky and I. Stewart, *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space*, Progr. Math. 200, Birkhäuser-Verlag, Basel, 2002.

[12] J. Hineman and J. M. Neuberger, *GNGA for general regions: Semilinear elliptic PDE and crossing eigenvalues*, Commun. Nonlinear Sci. Numer. Simul., (2005), to appear.

[13] C. T. Kelley, *Solving Nonlinear Equations with Newton's Method*, Fundam. Algorithms 1, SIAM, Philadelphia, 2003.

[14] C. T. Kelley, *Iterative Methods for Optimization*, Frontiers Appl. Math. 18, SIAM, Philadelphia, 1999.

[15] M. Lapidus, *Fractal drum, inverse spectral problems for elliptic operators and a partial resolution of the Weyl–Berry conjecture*, Trans. Amer. Math. Soc., 325 (1991), pp. 465–529.

[16] M. L. Lapidus, J. W. Neuberger, R. L. Renka, and C. A. Griffith, *Snowflake harmonics and computer graphics: Numerical computation of spectra on fractal drums*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 6 (1996), pp. 1185–1210.

[17] M. Lapidus and M. Pang, *Eigenfunctions of the Koch snowflake domain*, Comm. Math. Phys., 172 (1995), pp. 359–376.

[18] R. Lauterbach and S. Maier, *Symmetry-breaking at non-positive solutions of semilinear elliptic equations*, Arch. Ration. Mech. Anal., 126 (1994), pp. 299–331.

[19] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, Software Environ. Tools 6, SIAM, Philadelphia, 1998.

[20] Y. Li and J. Zhou, *Convergence results of a local minimax method for finding multiple critical points*, SIAM J. Sci. Comput., 24 (2002), pp. 865–885.

[21] P. C. Matthews, *Automated symmetry-breaking calculations*, LMS J. Comput. Math., 7 (2004), pp. 101–119.

[22] J. W. Neuberger, *Sobolev Gradients and Differential Equations*, Lecture Notes in Math. 1670, Springer-Verlag, Berlin, 1997.

[23] J. M. Neuberger, *A numerical method for finding sign-changing solutions of superlinear Dirichlet problems*, Nonlinear World, 4 (1997), pp. 73–83.

[24] J. M. Neuberger, *GNGA: Recent progress and open problems for semilinear elliptic PDE*, in Variational Methods: Open Problems, Recent Progress, and Numerical Algorithms, Contemp. Math. 357, AMS, Providence, RI, 2004, pp. 201–237.

[25] J. M. Neuberger, *Nonlinear elliptic partial difference equations on graphs*, J. Exper. Math., 15 (2005), pp. 91–107.

[26] J. M. Neuberger and J. W. Swift, *Newton's method and Morse index for semilinear elliptic PDEs*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 11 (2001), pp. 801–820.

[27] J. M. Neuberger, N. Sieben, and J. W. Swift, *Computing eigenfunctions on the Koch snowflake: A new grid and symmetry*, J. Comput. Appl. Math., 191 (2005), pp. 126–142.

[28] P. Rabinowitz, *Minimax Methods in Critical Point Theory with Applications to Differential Equations*, CBMS Reg. Conf. Ser. Math. 65, AMS, Providence, RI, 1986.

[29] W. R. Scott, *Group Theory*, Prentice–Hall, Englewood Cliffs, NJ, 1964.

[30] J. Smoller and A. G. Wasserman, *Symmetry-breaking for positive solutions of semilinear elliptic equations*, Arch. Ration. Mech. Anal., 95 (1986), pp. 217–225.
[31] S. Sternberg, *Group Theory and Physics*, Cambridge University Press, Cambridge, UK, 1994.
[32] A. D. Thomas and G. V. Wood, *Group Tables*, Shiva Math. Ser. 2, Shiva Publishing, Nantwich, UK, 1980; distributed by Birkhäuser Boston, Cambridge, MA.
[33] T. J. Thompson, *Estimating Solutions for the Ginzburg-Landau Superconductivity Model in Thin Disks*, M.S. thesis, Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ, 2005.
[34] M. Tinkham, *Group Theory and Quantum Mechanics*, McGraw–Hill, New York, Toronto, London, 1964.
[35] Z.-Q. Wang and J. Zhou, *A local minimax-Newton method for finding multiple saddle points with symmetries*, SIAM J. Numer. Anal., 42 (2004), pp. 1745–1759.
[36] Z.-Q. Wang and J. Zhou, *An efficient and stable method for computing saddle points with symmetries*, SIAM J. Numer. Anal., 43 (2005), pp. 891–907.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 3, pp. 508–527

Stability of Coupled Map Networks with Delays*

Fatihcan M. Atay† and Özkan Karabacak‡

Abstract. We consider networks of coupled scalar maps, with weighted connections which may include a time delay, and study the stability of equilibria with respect to the delays and connection structure. We prove that the largest eigenvalue of the graph Laplacian determines the effect of the connection topology on stability. The stability region in the parameter plane shrinks with increasing values of the largest eigenvalue, or of the time delay of the same parity. In particular, all bipartite graphs have an identical stability region, regardless of the delay or graph size, which is also the smallest stability region among those of all graphs. Furthermore, for certain parameter ranges, unstable (and possibly chaotic) maps can be stabilized via diffusive coupling with an odd time delay, provided that the network does not have a nontrivial and connected bipartite component. On the other hand, stabilization is not possible for even values of the delay or for bipartite networks.

Key words. network, delay, stability, synchronization, chaos, Laplacian

AMS subject classifications. 39A11, 37E05, 94C15

DOI. 10.1137/060652531

1. Introduction.

Coupled map networks were introduced in the 1980s as models of various physical phenomena [1, 2]. Since then, they have become one of the prototype systems for studying collective behavior, such as chaotic synchronization and cluster formation [3]. The interaction between the nodes in such networks is often modeled by a diffusion operator, which intuitively should favor homogenous (i.e., synchronized) behavior of the whole network. Nevertheless, in view of a chaotic system's sensitive dependence on initial conditions, it is a nontrivial finding of the early 1990s that coupled chaotic systems can indeed synchronize [4]. The synchronization of identical chaotic units is rather well understood, and the relevant conditions can be expressed in terms of the largest Lyapunov exponent of the individual maps and the spectrum of the diffusion operator [5]. Because of the diffusive nature of coupling, the synchronized solution is identical to the behavior of individual units in isolation. Thus, a network of chaotic maps will itself be chaotic when synchronized. Partly because of this, the equilibrium solutions of these networks did not receive much attention, although spatially homogeneous equilibria are a special form of synchrony. On the other hand, if one takes into account the time delays in the information transmission between the units, the synchronized dynamics is no longer the same as the isolated behavior [6]. In particular, delays can induce stability in coupled identical limit cycle oscillators [7, 8], leading to the phenomenon

*Received by the editors February 20, 2006; accepted for publication (in revised form) by B. Krauskopf April 28, 2006; published electronically September 26, 2006.
http://www.siam.org/journals/siads/5-3/65253.html
†Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, D-04103 Leipzig, Germany (atay@member.ams.org).
‡Istanbul Technical University, Faculty of Electrical and Electronic Engineering, Electronics Engineering Department, Maslak, TR-34469, Istanbul, Turkey (karabacak@ehb.itu.edu.tr).

of amplitude death, which was known earlier for nonidentical oscillators [9, 10]. This suggests the possibility that a synchronized network of chaotic maps might exhibit a stable equilibrium solution if delays are introduced into the model.

In this paper we present a stability analysis of networks of diffusively coupled scalar maps, both with and without time delays. We give necessary and sufficient conditions for stability in terms of the parameters of the individual map and the coupling function, the coupling delay, and the connectivity operator. We show that the role of the coupling topology is characterized solely by the largest eigenvalue of the coupling operator, which provides an order relation for comparing graphs with respect to their stability properties. More precisely, the larger the largest eigenvalue, the smaller the stability region in the parameter space. A similar result has been proved for continuous-time systems near Hopf bifurcation [11]. In the discrete-time setting of this paper, we are able to dispense with the assumption of Hopf bifurcation and extend the result to arbitrary maps. Nevertheless, discrete time has its own peculiarities and introduces some important differences in the stability picture, depending on the parity of the delay. Specifically, we prove that an unstable fixed point of the map can be stable for the network, provided that the delay is odd and that the network has no nontrivial bipartite components, whereas such stabilization is never possible with even (or zero) delays. Moreover, for delays of the same parity, the stability region becomes monotonically smaller as the delay increases, and we calculate the limiting profile of the stability region.

We consider coupled systems of the form

$$x_i(t+1) = f(x_i(t)) + \frac{1}{d_i}\sum_{j=1}^{n} a_{ij} g(x_i(t), x_j(t-\tau)), \quad i = 1, \dots, n. \tag{1}$$

Here $x_i(t) \in \mathbb{R}$ denotes the state of the ith node at time $t \in \mathbb{Z}$, and $\tau \in \mathbb{Z}^+$ is the signal transmission delay between the nodes. ($\mathbb{Z}^+$ denotes the nonnegative integers.) The function $f : \mathbb{R} \to \mathbb{R}$ describes the individual dynamics of the units, and $g : \mathbb{R}^2 \to \mathbb{R}$ is the interaction between a pair of units. We assume that g satisfies the general diffusion condition

$$g(x, x) = 0 \quad \forall x \in \mathbb{R}. \tag{2}$$

The numbers a_{ij} determine the connection structure of the network. In the simplest case, a_{ij} takes on binary values, depending on whether or not there is a connection between the nodes i and j. In other words, $a_{ij} = 1$ if nodes i and j are neighbors, and $a_{ij} = 0$ otherwise, with the stipulation $a_{ij} = a_{ji}$ and $a_{ii} = 0$ for all i, j. (More generally, one might have weighted connections where $a_{ij} \in \mathbb{R}^+$.) The number of neighbors of the ith node is denoted by $d_i = \sum_j a_{ij}$. Disregarding the trivial case of isolated nodes, it is assumed that $d_i > 0$ for all i.

Our interest in this paper is the stability of spatially uniform equilibrium solutions of (1). By virtue of the diffusion condition (2), $\mathbf{x}^* := (x^*, \dots, x^*) \in \mathbb{R}^n$ is a spatially uniform equilibrium solution if and only if $x^* \in \mathbb{R}$ is a fixed point of f. The main question is how the stability of $\mathbf{x}^*$ for the coupled system is related to the stability of x^* for the isolated map. Suppose that $f'(x^*)$ as well as the partial derivatives $D_i g$ of g at (x^*, x^*) exist, and denote

$$b = f'(x^*) \quad \text{and} \quad c = D_2 g(x^*, x^*) = -D_1 g(x^*, x^*),$$

where the last equality is a consequence of (2). Then the linear variational equation about $\mathbf{x}^* = (x^*, \dots, x^*)$ is

$$u_i(t+1) = bu_i(t) + \frac{c}{d_i}\sum_{j=1}^{n} a_{ij}[u_j(t-\tau) - u_i(t)], \quad i = 1, \dots n, \tag{3}$$

with $u_i = x_i - x^*$. The asymptotic stability of the zero solution of the linear equation (3) is equivalent to the exponential stability of $\mathbf{x}^*$ in the nonlinear equation (1) (see, e.g., [12]).

In the following sections we present a detailed analysis of the stability of (3). Section 2 derives necessary and sufficient conditions for stability. Section 3 studies the relationship between stability and the delays. Section 4 treats the role of the network topology. In section 5 we look at the implications of the results for the nonlinear system (1) and also mention some extensions to high-dimensional maps. Section 6 concludes the work

2. Stability criteria. We study the stability of the linear system (3). Using the fact that $d_i = \sum_j a_{ij}$, (3) can be written in the vector form

$$\mathbf{u}(t+1) = (b-c)\mathbf{u}(t) + c\mathbf{D}^{-1}\mathbf{A}\mathbf{u}(t-\tau), \tag{4}$$

where $\mathbf{u} = (u_1, \dots, u_n)$, $\mathbf{A} = [a_{ij}]$, and $\mathbf{D} = \mathrm{diag}\{d_1, \dots, d_n\}$ is a diagonal matrix. Thus, the characteristic values s corresponding to (3) are given by the solutions of

$$\det(s^{\tau+1}\mathbf{I} - s^{\tau}(b-c)\mathbf{I} - c\mathbf{D}^{-1}\mathbf{A}) = 0. \tag{5}$$

Consequently, the zero solution of (3) is asymptotically stable if and only if all roots s of (5) lie inside the unit circle.

The dynamics of the system (3) is intimately related to the underlying connection structure. Therefore, we shall make use of some graph-theoretical ideas in the analysis. We identify the indices i with the vertices of a graph $\mathcal{G}$, whose adjacency matrix is $\mathbf{A} = [a_{ij}]$, and the vertex degrees are given by d_i. We assume that $\mathcal{G}$ is a simple and nontrivial graph; i.e., it contains at least one edge and no self-connections. The matrix $\mathbf{D}^{-1}\mathbf{A}$ appearing in (4) depends only on the connection structure of the graph $\mathcal{G}$. It is related to a (normalized) Laplacian operator $\mathbf{L}$, which encapsulates the diffusive nature of the coupling. In matrix form, $\mathbf{L}$ is given by

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}. \tag{6}$$

Although $\mathbf{L}$ is in general not symmetric, it can be shown to be a self-adjoint operator with respect to a certain inner product. We list some relevant spectral properties of $\mathbf{L}$ in the next lemma (see, e.g., [13, 5, 11]). Recall that $\mathcal{G}$ is called a *complete graph* if every vertex is connected to every other vertex, and is called *bipartite* if its vertex set can be divided into two parts V_1 and V_2 such that every edge has one end in V_1 and one in V_2.

Lemma 1. *Let $\mathcal{G}$ be a graph on n vertices, $n \geq 2$. Then $\mathbf{L}$ is a self-adjoint and positive semidefinite operator; thus its eigenvalues λ_k are real and nonnegative, and its eigenvectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ form a complete orthogonal basis for $\mathbb{R}^n$. Its smallest eigenvalue is zero and*

corresponds to the eigenvector $(1, 1, \ldots, 1)$. *If* $\mathcal{G}$ *has no isolated vertices, then the largest eigenvalue* $\lambda_{\max}$ *of* $\mathbf{L}$ *satisfies*

$$\frac{n}{n-1} \leq \lambda_{\max} \leq 2. \tag{7}$$

Furthermore, $\lambda_{\max} = n/(n-1)$ *if and only if* $\mathcal{G}$ *is a complete graph of* n *vertices, and* $\lambda_{\max} = 2$ *if and only if a connected component of* $\mathcal{G}$ *is bipartite and nontrivial.*

The lemma implies that for large complete graphs the largest eigenvalue $\lambda_{\max}$ is close to one. We remark that, if self-connections are included for the nodes, then $\lambda_{\max}$ for complete graphs is exactly one. At the other extreme are bipartite graphs, for which $\lambda_{\max} = 2$. This is a richer set of graphs, which includes regular lattices in arbitrary dimensions, cycles with an even number of vertices, and all trees. We shall occasionally refer to these two extreme cases in the analysis. The set of eigenvalues of $\mathbf{L}$ is often referred to as the spectrum of the corresponding graph. We note that the spectrum of a graph is the union of the spectra of its connected components.

The system (1) can be generalized in a straightforward way to take into account different connection strengths between pairs of nodes. In this setting, a_{ij} are arbitrary nonnegative numbers denoting the strength of the connection between the ith and jth nodes, where a zero value for a_{ij} indicates that i and j are not connected. The graph $\mathcal{G}$ is now a weighted graph, where the vertex degrees $d_i = \sum_j a_{ij}$ may have noninteger values. We still assume $a_{ij} = a_{ji}$, $a_{ii} = 0$, and $d_i > 0$. The Laplacian is defined by (6), and the spectral properties given in Lemma 1, and in particular (7), remain valid for this more general case [13].

Now using (6) and Lemma 1, one has

$$\mathbf{D}^{-1}\mathbf{A}\mathbf{v}_k = (\mathbf{I} - \mathbf{L})\mathbf{v}_k = (1 - \lambda_k)\mathbf{v}_k, \quad k = 1, \ldots, n,$$

where $\mathbf{v}_k$ and λ_k are the eigenvectors and eigenvalues, respectively, of the Laplacian $\mathbf{L}$. Hence, in the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, $\mathbf{D}^{-1}\mathbf{A}$ can be written as a diagonal matrix, and the characteristic equation (5) can be expressed as the product

$$\prod_{k=1}^{n} p_k(s) = 0, \tag{8}$$

where

$$p_k(s) := s^{\tau+1} - (b - c)s^{\tau} - c(1 - \lambda_k). \tag{9}$$

We conclude that the zero solution of (3) is asymptotically stable if and only if all roots of (9) lie inside the unit circle for all $k = 1, \ldots, n$.

We first give a sufficient condition for stability that is independent of the delay τ.

Theorem 1. *The zero solution of* (3) *is asymptotically stable for all* $\tau \in \mathbb{Z}^+$ *if*

$$|b - c| + |c| < 1, \tag{10}$$

or equivalently,

$$|b| < 1 \quad \textit{and} \quad |b - 2c| < 1. \tag{11}$$

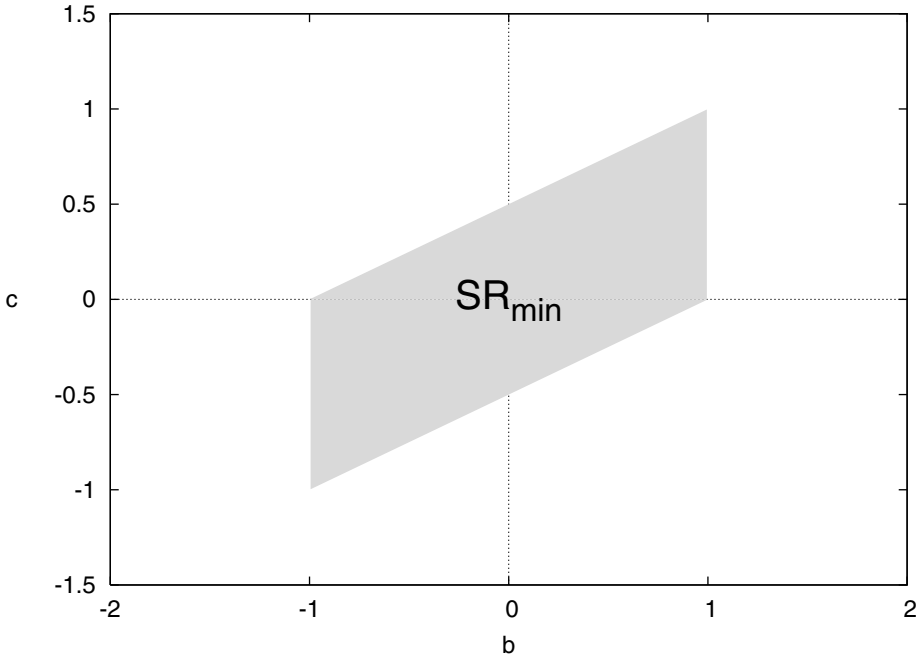


Figure 1. *Minimal stability region.*

Proof. For $\tau > 0$, a result due to Clark (Lemma 6 in the appendix) implies that the roots of (9) lie inside the unit circle, provided

$$|b - c| + |c(1 - \lambda_k)| < 1. \tag{12}$$

For $\tau = 0$, the characteristic roots are directly solved from (9) as $s = (b - c) + c(1 - \lambda_k)$; so one again has stability if (12) is satisfied. However, (10) implies (12) since $\lambda_k \in [0, 2]$ by Lemma 1, and so (10) is a sufficient condition for stability. The equivalence of (10) and (11) is proved in Lemma 3 in the appendix. ∎

In the parameter space (b, c), we call the set

$$\mathrm{SR}_{\min} = \{(b, c) \in \mathbb{R}^2 : |b - c| + |c| < 1\} \tag{13}$$

$$= \{(b, c) \in \mathbb{R}^2 : |b| < 1 \text{ and } |b - 2c| < 1\} \tag{14}$$

the *minimal stability region*, since by Theorem 1 it is necessarily included in the stability region for any network and any choice of time delay. We shall prove later that $\mathrm{SR}_{\min}$ is actually the exact stability region for certain graphs. Figure 1 depicts the shape of the minimal stability region. The bounding lines are

$$c = \frac{b \pm 1}{2} \quad \text{and} \quad b = \pm 1, \tag{15}$$

corresponding to the inequalities in (14).

The precise conditions for stability are given by the next theorem.

Theorem 2. *Let τ be a positive integer. The zero solution of (3) is asymptotically stable if and only if one of the following hold for both $\lambda = 0$ and $\lambda = \lambda_{\max}$ (i.e., for the smallest and the largest eigenvalues of the Laplacian):*

(i) *τ is odd and*

$$|b-c| - 1 < -c(1-\lambda) < \sqrt{(b-c)^2 + 1 - 2\,|b-c|\cos\Phi}. \tag{16}$$

(ii) *τ is even,*

$$|b - c\lambda| < 1, \quad \textit{and} \tag{17}$$

$$|c| < \sqrt{(b-c)^2 + 1 - 2\,|b-c|\cos\Phi}, \tag{18}$$

where Φ is the unique number satisfying

$$\frac{\sin\left((\tau+1)\Phi\right)}{\sin\left(\tau\Phi\right)} = |b-c|\,, \qquad \Phi \in \left(0, \frac{\pi}{\tau+1}\right). \tag{19}$$

For $\tau = 0$, the zero solution is asymptotically stable if and only if (17) holds for $\lambda = 0$ and $\lambda = \lambda_{\max}$.

We will make repeated use of the following simple lemma, whose proof is omitted.

Lemma 2. *Let $p < q$ and $\lambda_1 \le \lambda_2$ be real numbers, and $h : \mathbb{R} \to \mathbb{R}$ a monotone[1] function. Then the inequality*

$$p < h(\lambda) < q$$

is satisfied for all $\lambda \in [\lambda_1, \lambda_2]$ if and only if it is satisfied for both $\lambda = \lambda_1$ and $\lambda = \lambda_2$.

Proof of Theorem 2. We first prove sufficiency. Assume that τ is odd and (16) holds for $\lambda = 0$ and $\lambda = \lambda_{\max}$. Then,

$$|b-c| - 1 < -c, \tag{20}$$

$$\text{and } |b-c| - 1 < -c(1-\lambda_{\max}). \tag{21}$$

If $c \ge 0$, then (20) implies $|b-c| - 1 < 0$, whereas if $c < 0$, then (21) implies $|b-c| - 1 < 0$. In either case we have

$$|b-c| < 1 < \frac{\tau+1}{\tau}\,. \tag{22}$$

On the other hand, Lemma 2 implies that (16) holds for all eigenvalues $\lambda_k \in [0, \lambda_{\max}]$ of the Laplacian. We now apply a result of Kuruklis (Lemma 7 in the appendix) with $a_1 = (b-c)$ and $a_2 = -c(1-\lambda_k)$ and the inequalities (16) and (22), proving that all roots of the polynomials

[1]We say $h : \mathbb{R} \to \mathbb{R}$ is monotone increasing (resp., decreasing) if $h(x_1) \le h(x_2)$ whenever x_1 is less than (resp., greater than) x_2. A *monotone* function is one that is either monotone increasing or monotone decreasing on $\mathbb{R}$.

(9) are within the unit circle for $k = 1, \dots, n$. Next assume that τ is even, and (17) and (18) hold. Since (17) is equivalent to

$$-1 < b - c\lambda < 1$$

and holds for $\lambda = 0$ and $\lambda = \lambda_{\max}$, it holds for all $\lambda \in [0, \lambda_{\max}]$ by Lemma 2. In particular, it holds for $\lambda = 1$, which implies (22). Furthermore, $|c(1-\lambda)| \le |c|$ for all $\lambda \in [0, \lambda_{\max}] \subset [0, 2]$. Again, an application of Lemma 7 proves stability.

To prove necessity, assume that all roots of the polynomials (9) are within the unit circle for $k = 1, \dots, n$. If τ is odd, Lemma 7 implies (16) for any $\lambda = \lambda_k$. If τ is even, we use Lemma 7 again to conclude

$$|b - c\lambda_k| < 1,$$

and

$$|c(1-\lambda_k)| < \sqrt{(b-c)^2 + 1 - 2|b-c|\cos\Phi}$$

for all eigenvalues λ_k of the Laplacian, in particular for the zero and the largest eigenvalues, which imply (17)–(18). On the other hand, when $\tau = 0$, the roots of (9) can be explicitly solved as $s = b - c\lambda_k$. Thus, the zero solution of (3) is asymptotically stable if and only if (17) holds for $\lambda = \lambda_k$, $k = 1, \dots, n$. By Lemma 2, it is necessary and sufficient that this inequality hold for $\lambda = 0$ and $\lambda = \lambda_{\max}$. Finally, the uniqueness of the solution Φ of (19) is proved in Lemma 5 in the appendix. ■

Theorem 2 gives precise conditions for the stability of the coupled system, which will be explored in more detail in the remainder of the paper. The first important observation is that the network topology enters the stability criteria only through the largest eigenvalue $\lambda_{\max}$ of the Laplacian. The other parameters affecting stability are the time delay τ and the derivatives $b = f'(x^*)$ and $c = D_2 g(x^*, x^*)$. For a given $\lambda_{\max}$ and τ, we denote the stability region in the (b, c) parameter plane as

$$\mathrm{SR}_{\tau,\lambda_{\max}} = \{(b,c) \in \mathbb{R}^2 : \text{the zero solution of (3) is asymptotically stable}\}. \tag{23}$$

In addition to the shape of $\mathrm{SR}_{\tau,\lambda_{\max}}$ in the (b, c) plane, it is of interest how it depends on the time delay and the network structure, as encapsulated by the quantities τ and $\lambda_{\max}$, respectively. We give a detailed study of this problem in the following sections.

3. Delays and stability.

In this section we investigate how the stability region $\mathrm{SR}_{\tau,\lambda_{\max}}$ defined by (23) depends on the time delay τ. In Theorem 2, the delay affects the stability conditions (16)–(18) implicitly through the quantity Φ defined in (19). In the special case $\tau = 1$ and the limiting case $\tau \to \infty$, Φ can be explicitly solved and the bounds of the stability region can be analytically determined. Furthermore, the case $\tau = 0$ follows directly from (9), as stated in Theorem 2. For the remaining cases Φ is easy to solve numerically, for instance by the Newton–Raphson method. Figure 2 shows Φ as a function of $|b - c|$ for various values of τ. In the following, we obtain detailed information on how the stability of the system depends on the time delay.

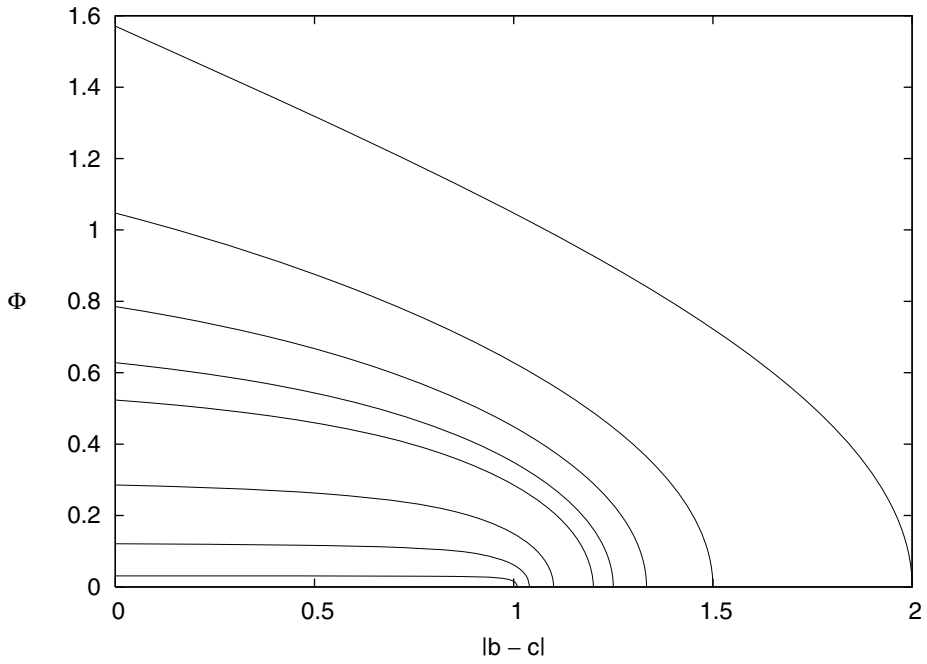


Figure 2. *The solution* Φ *of* (19)*, plotted for* τ *equal to (from top to bottom)* $1, 2, 3, 4, 5, 10, 25$*, and* 100*.*

3.1. A monotonicity property. We start by proving a monotone dependence on the delay and the limiting behavior for large delays.

Theorem 3. *Let* $\lambda_{\max} \in [1, 2]$*, and let* τ_1 *and* τ_2 *be positive integers which are both odd or both even. If* $\tau_2 > \tau_1$*, then*

$$\mathrm{SR}_{\tau_2,\lambda_{\max}} \subset \mathrm{SR}_{\tau_1,\lambda_{\max}}. \tag{24}$$

Furthermore, if $\{\tau_j : j = 1, 2, \dots\}$ *is a sequence of even (or odd) integers tending to infinity, then*

$$\lim_{j\to\infty} \mathrm{SR}_{\tau_j,\lambda_{\max}} \subset \mathrm{cl}(\mathrm{SR}_{\min}), \tag{25}$$

where cl *denotes closure in* $\mathbb{R}^2$*.*

Proof. Suppose that τ_1 and τ_2 are both odd or both even, with $\tau_1 < \tau_2$. We fix a point $(b, c) \in \mathrm{SR}_{\tau_2,\lambda_{\max}}$. We will prove that $(b, c) \in \mathrm{SR}_{\tau_1,\lambda_{\max}}$. To show this, let $\Phi = \Phi(\tau)$ be the solution of (19) belonging to the interval $(0, \pi/(\tau + 1))$. By Lemma 5 in the appendix, Φ is a decreasing function of τ, and so $\cos\Phi$ is increasing in τ. Therefore, the radicands in (16) and (18) are decreasing functions of τ, which implies that (16) or (18) is satisfied for $\Phi(\tau_2)$ whenever it is satisfied for $\Phi(\tau_1)$. This proves (24).

Now let $\{\tau_j\}$ be a sequence of even integers tending to ∞. By the first part of the theorem, the limit $S = \lim_{j\to\infty} \mathrm{SR}_{\tau_j,\lambda_{\max}}$ exists and equals

$$S = \bigcap_{j=1}^{\infty} \mathrm{SR}_{\tau_j,\lambda_{\max}}.$$

Hence, if $(b, c) \in S$, then b and c satisfy (17) for $\lambda = 0$ and $\lambda = \lambda_{\max}$ and (18) for all j. Now, (17) is linear in λ, so it also holds for $\lambda = 1$ by Lemma 2, which implies

$$|b - c| - 1 < 0. \tag{26}$$

On the other hand, since $\Phi_j \triangleq \Phi(\tau_j) \in (0, \pi/(\tau_j + 1))$, we have $\cos \Phi_j \to 1$, and using this in (18) gives

$$|c| \leq ||b - c| - 1|. \tag{27}$$

Combining (27) with (26) yields

$$|c| \leq 1 - |b - c|, \tag{28}$$

which describes the closure of the minimal stability region $\mathrm{SR}_{\min}$ given by (13). Hence, $S \subset \mathrm{cl}(\mathrm{SR}_{\min})$, proving (25). A similar argument works also for the case when $\{\tau_j\}$ are odd integers. In this case b and c satisfy (16) as $\cos \Phi \to 1$; that is,

$$|b - c| - 1 < -c(1 - \lambda) \leq ||b - c| - 1|$$

for $\lambda = 0$ and $\lambda = \lambda_{\max}$, and therefore also for $\lambda = 1$. Evaluating at $\lambda = 1$ gives (26) as above. Evaluating at $\lambda = 0$ and using (26) yields

$$|b - c| - 1 < -c \leq 1 - |b - c|, \tag{29}$$

which implies (28) and proves (25). ■

By Theorem 1, the set $\mathrm{SR}_{\min}$ is contained in every stability region $\mathrm{SR}_{\tau,\lambda_{\max}}$. Hence, Theorem 24 implies that the stability region $\mathrm{SR}_{\tau,\lambda_{\max}}$ approaches $\mathrm{SR}_{\min}$ for large delays. Nevertheless, there are some important qualitative differences resulting from the parity of the delay, as we show next.

3.2. Zero delay and even delays. By Theorem 2, the stability region for $\tau = 0$ is given by

$$\mathrm{SR}_{0,\lambda_{\max}} = \{(b, c) : |b| < 1 \text{ and } |b - c\lambda_{\max}| < 1\}. \tag{30}$$

Figure 3 depicts the region $\mathrm{SR}_{0,\lambda_{\max}}$, together with the minimal stability region $\mathrm{SR}_{\min}$. The boundary curves are given by

$$c = \frac{b \pm 1}{\lambda_{\max}} \quad \text{and} \quad b = \pm 1. \tag{31}$$

Note that if $\lambda_{\max}^A > \lambda_{\max}^B$, then

$$\mathrm{SR}_{0,\lambda_{\max}^A} \subsetneq \mathrm{SR}_{0,\lambda_{\max}^B}.$$

Furthermore, comparison of (31) and (15) shows that the stability region of a bipartite graph ($\mathrm{SR}_{0,2}$) coincides with the minimal stability region $\mathrm{SR}_{\min}$. We shall generalize these observations to the case of positive delays in section 4.

The stability regions for other even delays are similar in shape, although they get smaller with increasing τ. Figure 3 shows the stability regions for $\tau = 0$ and $\tau = 2$. In all cases, the stability region is a subset of the strip $|b| < 1$, as follows from Theorem 2 by evaluating (17) at $\lambda = 0$. In other words, an unstable fixed point of the isolated map f cannot be stabilized by any coupling strength or topology when the delay is even.

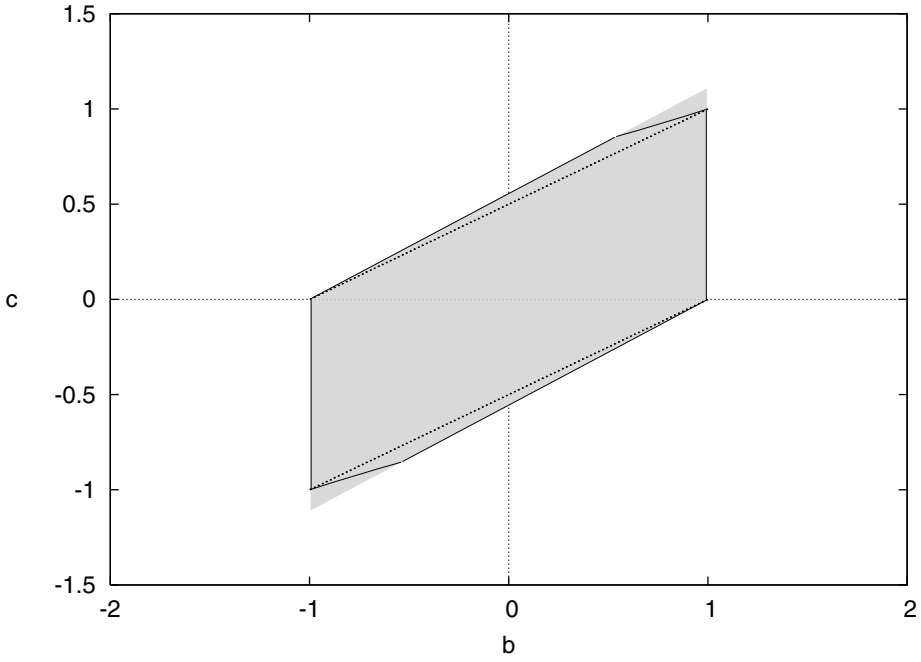


Figure 3. *Stability regions for even delays: $\tau = 0$ (gray area) and $\tau = 2$ (inside solid boundaries), plotted for $\lambda_{\max} = 1.8$. The dotted lines delineate the minimal stability region.*

3.3. Odd delays. When $\tau = 1$, the equation (19) for Φ yields $\cos\Phi = |b - c|/2$. Substituting into (16), we obtain

$$|b - c| - 1 < -c(1 - \lambda) < 1 \tag{32}$$

for $\lambda = 0, \lambda_{\max}$. A straightforward calculation involves these four inequalities in different regions according to the signs of c and $b - c$. The resulting stability region is depicted in Figure 4. For comparison, the stability boundaries for the undelayed case are also shown in the figure. The stability changes introduced by the delay can be summarized as follows. For $c \geq 0$ one easily obtains $|b - c| - 1 < -c$ from (32), which coincides with the upper half of the minimal stability region (13),

$$0 < c < \frac{b+1}{2}, \quad |b| < 1.$$

For $c < 0$, three different domains can be identified: in the region

$$(1 - \lambda_{\max}) < b < 1, \qquad \frac{b-1}{\lambda_{\max}} < c < 0$$

the stability is unchanged from the undelayed case; for

$$-1 < b < (1 - \lambda_{\max}), \quad -1 < c < 0,$$

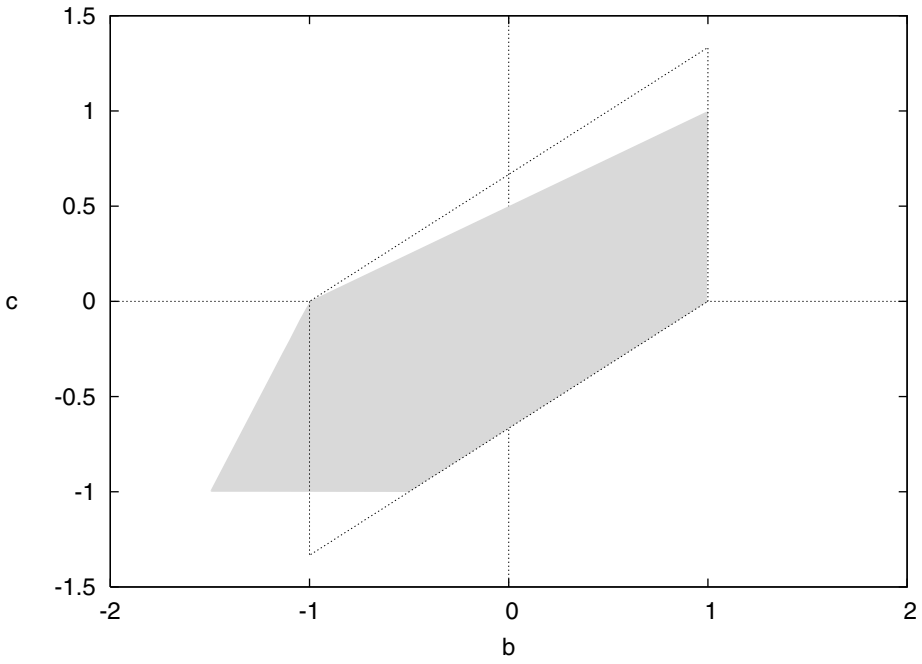


Figure 4. *Stability regions for* $\tau = 1$ *(gray area) and* $\tau = 0$ *(inside dotted lines). The value of* $\lambda_{\max}$ *is* 1.5.

it is worsened; and for

$$(\lambda_{\max} - 3) < b < -1, \qquad -1 < c < \frac{b+1}{(2-\lambda_{\max})} \tag{33}$$

it is improved. The last possibility is particularly interesting, as it extends the stability region beyond the strip $|b| < 1$. In other words, an unstable fixed point of f can be stabilized in a network with $\tau = 1$, provided that (33) holds.

The stability region is similar for other odd values of the delay. Figure 5 depicts a close-up view of the region where stability extends into the domain $b < -1$ by the virtue of delays. In accordance with Theorem 3, the stability region gets smaller as the delay gets larger; however, it still extends into $b < -1$. As already mentioned, the stability region is necessarily confined to the strip $|b| < 1$ for even values of the delay. Hence, stabilization of unstable maps by coupling is a feature of odd delays only.

4. Stability and network topology. We now turn to the role of the connection topology on stability. We have already seen that the stability depends on the graph topology only through the value of the largest eigenvalue $\lambda_{\max}$ of the graph Laplacian. We now prove a monotonicity property; namely, the smaller the value of $\lambda_{\max}$, the larger the stability region on the (b, c)-plane for the same value of τ.

Theorem 4. *Let* $1 \le \lambda_A \le \lambda_B \le 2$. *Then* $\mathrm{SR}_{\tau,\lambda_B} \subset \mathrm{SR}_{\tau,\lambda_A}$.

Proof. Let $(b, c) \in \mathrm{SR}_{\tau,\lambda_B}$. Then by Theorem 2, b and c satisfy either (16) or (17)–(18), depending on whether τ is odd or even, for both $\lambda = 0$ and $\lambda = \lambda_B$. Because these inequalities

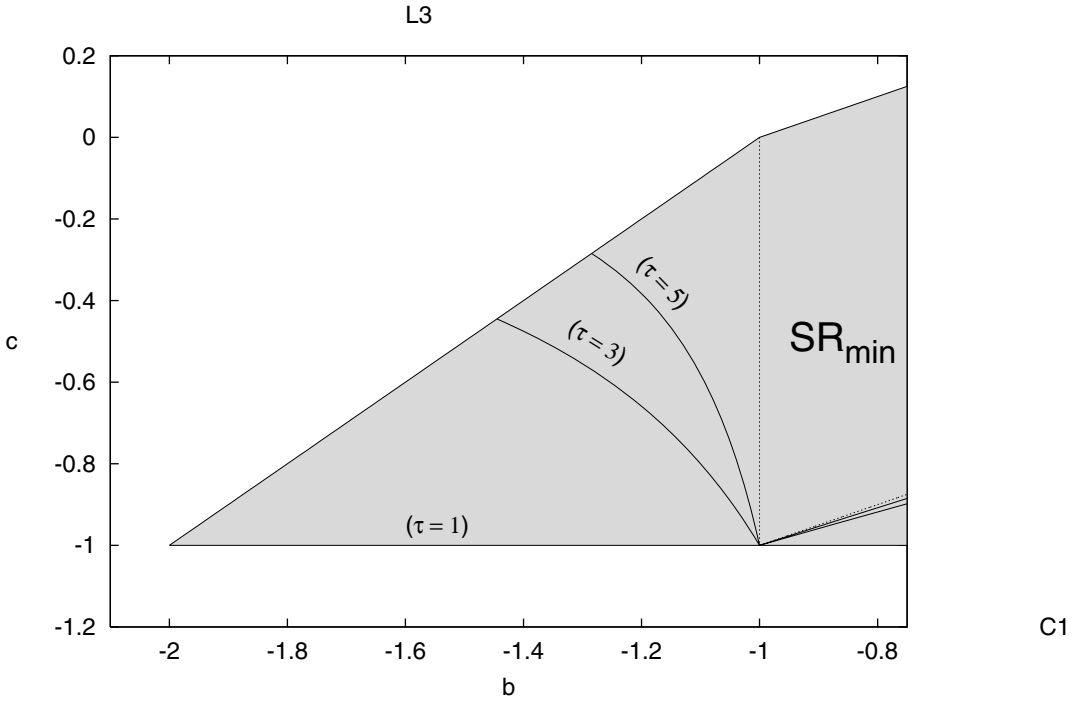


Figure 5. *Part of the stability region for odd delays, plotted for* $\lambda_{\max} = 1$. *As the delay increases, the stability region approaches* $\mathrm{SR}_{\min}$.

are linear in λ, they also hold for $\lambda = \lambda_A$, by Lemma 2. Since the conditions (16) or (17)–(18) are also sufficient for stability, $(b, c) \in \mathrm{SR}_{\tau,\lambda_A}$. ■

Theorem 4 provides an ordering relation for graphs with respect to their stability properties. Thus, if $\mathcal{G}_A$ and $\mathcal{G}_B$ are two graphs such that $\lambda_A \leq \lambda_B$, where λ_A and λ_B denote the largest eigenvalue of the respective Laplacian, and if $b, c \in \mathbb{R}$ and $\tau \in \mathbb{Z}^+$, then the zero solution of (3) is asymptotically stable for $\mathcal{G}_A$ whenever it is asymptotically stable for $\mathcal{G}_B$. Figure 6 shows the shrinking of the stability region as $\lambda_{\max}$ increases.

We next show that all bipartite graphs have the same stability region, which in fact equals the minimal stability region $\mathrm{SR}_{\min}$. An important implication of this property is that the stability region for bipartite graphs is independent of the delay τ. Recall from Lemma 1 that $\lambda_{\max} = 2$ for bipartite graphs.

Theorem 5. $\mathrm{SR}_{\tau,2} = \mathrm{SR}_{\min}$ *for all* $\tau \in \mathbb{Z}^+$. *Hence, the stability regions of all bipartite*[2] *graphs are identical and independent of the delay* τ *or the graph size* n.

Proof. Let $(b, c) \in \mathrm{SR}_{\tau,2}$. Assume first that τ is odd. By Theorem 2, b and c satisfy

$$|b - c| - 1 < -c(1 - \lambda) \tag{34}$$

for $\lambda = 0$ and $\lambda = 2$. If $c \geq 0$, we use (34) with $\lambda = 0$; otherwise we use it with $\lambda = 2$, to

[2]More generally, the statement holds for all graphs which have a connected and nontrivial bipartite component.

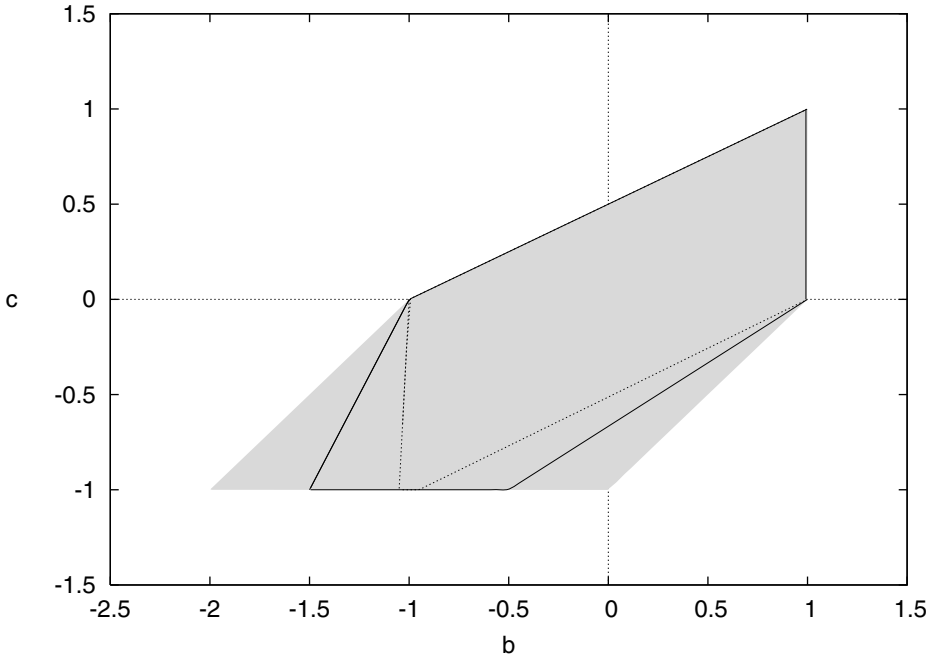


Figure 6. *The effect of the network topology: stability regions for* $\lambda_{\max} = 1$ *(gray area),* $\lambda_{\max} = 1.5$ *(inside solid lines), and* $\lambda_{\max} = 1.95$ *(inside dotted lines). As* $\lambda_{\max}$ *approaches* 2*, the stability region approaches* $\mathrm{SR}_{\min}$*. A value of* $\tau = 1$ *is used in the plots.*

obtain

$$|b - c| - 1 < -|c|. \tag{35}$$

Similarly, for τ positive and even, the condition (17) with $\lambda = 0$ and $\lambda = 2$ gives

$$|b| < 1 \quad \text{and} \quad |b - 2c| < 1. \tag{36}$$

The inequalities in (36) also hold for the case $\tau = 0$, as can be seen by setting $\lambda_{\max} = 2$ in (30). By (14) it follows that $\mathrm{SR}_{\tau,2} \subset \mathrm{SR}_{\min}$. Moreover, $\mathrm{SR}_{\min} \subset \mathrm{SR}_{\tau,2}$ by Theorem 1. Hence, $\mathrm{SR}_{\tau,2} = \mathrm{SR}_{\min}$. ■

Corollary 1. *If* $|b| > 1$*, then the zero solution of* (3) *is unstable for a bipartite graph for any choice of parameters. Hence, an unstable fixed point of the map* f *cannot be stabilized in a bipartite network configuration.*

Proof. By Theorem 5, the stability region of a bipartite graph is $\mathrm{SR}_{\min}$, which is a subset of the strip $|b| < 1$ by (14). It follows that if $|b| > 1$, then the zero solution of the linear equation (3) is unstable. ■

Recalling the spectral properties summarized in Lemma 1, the above results imply that, for a given value of delay τ, a complete graph has the best stability characteristics among all graphs of the same size, and its stability improves slightly with increasing graph size. On the other hand, bipartite graphs have the worst stability characteristics among all graphs of any size and for any value of delay. In fact, a bipartite graph configuration is incapable of

stabilizing an unstable fixed point of the individual map. All other graph configurations lie between the two extremes of bipartite and complete graphs.

5. Nonlinear and higher-dimensional maps. We summarize the implications of the foregoing results for the nonlinear system (1), which we reproduce here for convenience:

$$x_i(t+1) = f(x_i(t)) + \frac{1}{d_i}\sum_{j=1}^{n} a_{ij} g(x_i(t), x_j(t-\tau)), \quad i = 1, \ldots, n. \tag{37}$$

Furthermore, we briefly discuss some extensions to higher-dimensional maps.

In section 3 we have seen that stabilization is possible only for odd values of the delay. Moreover, Theorem 3 implies that the choice of $\tau = 1$ gives the largest stability region in the (b, c) parameter plane. It follows that the inequalities (33), i.e.,

$$\lambda_{\max} - 3 < f'(x^*) < -1 \tag{38}$$

and

$$-1 < c < \frac{f'(x^*) + 1}{(2 - \lambda_{\max})}, \tag{39}$$

are necessary conditions for stability in the case $|f'(x^*)| > 1$. From a bifurcation point of view, (38) implies that the network can stabilize x^* in case of a flip bifurcation in the isolated map ($f'(x^*) = -1$) but not in a saddle-node, transcritical, or pitchfork bifurcation ($f'(x^*) = 1$). In particular, an appropriate network configuration can suppress chaotic oscillations originating from a period doubling route to chaos. Furthermore, (38) can be satisfied only if $\lambda_{\max} < 2$, that is, if no nontrivial component of the graph $\mathcal{G}$ is bipartite (viz. Corollary 1). In short, the ingredients of a stabilization scheme involve a flip bifurcation, a nonbipartite network configuration, and an odd coupling delay.

As an example, consider the well-known quadratic family of maps $f_\rho(x) = \rho x(1-x)$ on the unit interval with parameter $\rho \in [0, 4]$. There are two fixed points, one at zero with $f'_\rho(0) = \rho$, and the other one at $x^*_\rho = (\rho - 1)/\rho$ with $f'(x^*_\rho) = 2 - \rho$. The origin undergoes a transcritical bifurcation at the parameter value $\rho = 1$ and is unstable for $1 < \rho \leq 4$ for the map f_ρ. Hence by the preceding analysis, the origin cannot be stabilized in any network configuration and by any diffusive-type coupling. On the other hand, x^*_ρ loses its stability at $\rho = 3$ through a flip bifurcation and is an unstable fixed point of f_ρ for $3 < \rho \leq 4$. By (38), x^*_ρ can be stabilized for parameter values in the range

$$3 < \rho < 5 - \lambda_{\max}$$

with a choice of c satisfying

$$-1 < c < \frac{3 - \rho}{(2 - \lambda_{\max})}.$$

The largest stability region is provided by large complete graphs, as observed in section 4, where $\lambda_{\max} \to 1$ as $n \to \infty$. In this case, one can obtain stability for up to $\rho = 4$, i.e., well

into the fully chaotic regime. (Although the analysis is inconclusive about the case $\rho = 4$, numerical solution with complete graphs having self-connections indicates stability also for the value $\rho = 4$.)

A similar analysis can be extended to m-dimensional maps, i.e., when $f : \mathbb{R}^m \to \mathbb{R}^m$ and $g : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$. In this case, linearizing and expressing in the eigenbasis of the Laplacian leads to the equations

$$\mathbf{u}(t+1) = (\mathbf{B} - \mathbf{C})\mathbf{u}(t) + (1 - \lambda_i)\mathbf{C}\mathbf{u}(t - \tau), \quad i = 1, \dots, n, \tag{40}$$

where $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{m \times m}$, and λ_i are the eigenvalues of the Laplacian as before. Hence, the spatially homogeneous fixed point $\mathbf{x}^*$ is exponentially stable if and only if the zero solution is asymptotically stable in (40) for $i = 1, \dots, n$. The depiction of detailed stability regions will not be attempted here due to the large number of parameters. We confine ourselves to some special cases that allow for some immediate conclusions.

In the classical coupled map lattice model, the interaction function is $g(x, y) = \kappa(f(y) - f(x))$, where the scalar $\kappa \in \mathbb{R}$ plays the role of a coupling strength. Using $\mathbf{C} = \kappa\mathbf{B}$ in (40), and assuming that $\mathbf{B}$ is diagonalizable and has eigenvalues b_j, one sees that $\mathbf{x}^*$ is exponentially stable if and only if all roots of the equation

$$s^{\tau+1} - b_j(1 - \kappa)s^\tau - b_j\kappa(1 - \lambda_i) = 0 \tag{41}$$

are inside the unit circle for $i = 1, \dots, n$, $j = 1, \dots, m$. Another common interaction type is linear diffusion, where $g(x, y) = \kappa(y - x)$ for some $\kappa \in \mathbb{R}$, so that $\mathbf{C} = \kappa\mathbf{I}$. With the same assumptions on $\mathbf{B}$ as above, we obtain

$$s^{\tau+1} - (b_j - \kappa)s^\tau - \kappa(1 - \lambda_i) = 0, \tag{42}$$

whose roots determine the stability. In case the eigenvalues b_j are real, both (41) and (42) become special cases of (9), and the results of the previous chapters are applicable.

A special case of (40) is the undelayed network. With $\tau = 0$, (40) reduces to

$$\mathbf{u}(t+1) = (\mathbf{B} - \lambda_i\mathbf{C})\mathbf{u}(t), \quad i = 1, \dots, n, \tag{43}$$

and the stability condition is that all eigenvalues of $\mathbf{B} - \lambda_i\mathbf{C}$, $i = 1, \dots, n$, have modulus less than one. Since the Laplacian always has a zero eigenvalue, a necessary condition for asymptotic stability is that $\mathbf{B}$ have all its eigenvalues inside the unit circle. In other words, unstable fixed points of isolated maps cannot be stabilized in a diffusive network in the absence of delays. This generalizes the corresponding results of section 3. It is interesting to compare (43) to the case obtained by rewriting the delayed system as a higher-dimensional system without delays. Hence, letting $\mathbf{u}(t) = (u(t), u(t-1), \dots, u(t-\tau)) \in \mathbb{R}^{\tau+1}$ in the system (3) of coupled scalar maps, and rewriting in the eigenbasis of the Laplacian, we obtain

$$\mathbf{u}(t+1) = (\mathbf{B} - \mathbf{C}_1 + (1 - \lambda_i)\mathbf{C}_2)\mathbf{u}(t), \quad i = 1, \dots, n, \tag{44}$$

with

$$\mathbf{B} = \begin{pmatrix} b & \mathbf{0}_{1\times\tau} \\ \mathbf{I}_\tau & \mathbf{0}_{\tau\times 1} \end{pmatrix}, \quad \mathbf{C}_1 = \begin{pmatrix} c & \mathbf{0}_{1\times\tau} \\ \mathbf{0}_{\tau\times\tau} & \mathbf{0}_{\tau\times 1} \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} \mathbf{0}_{1\times\tau} & c \\ \mathbf{0}_{\tau\times\tau} & \mathbf{0}_{\tau\times 1} \end{pmatrix},$$

where $\mathbf{0}_{n\times m}$ denotes an $n \times m$ matrix of zeros and $\mathbf{I}_\tau$ is the identity matrix of size τ. Note the difference with (43). In (43) the coupling $\mathbf{C}$ acts only through multiplication by λ_i and so has no effect along the eigenvector corresponding to the zero eigenvalue. By contrast, in (44), even when $\lambda_i = 0$ the coupling has a nonzero contribution $\mathbf{C}_2 - \mathbf{C}_1$. This is the essence of the role of delays in stabilizing unstable fixed points.

6. Conclusion. We have presented a general analysis of the stability of equilibrium solutions of diffusively coupled scalar maps, with particular focus on the effects of coupling configuration and delays. These two factors can induce a stability change upon the fixed point of the isolated map, causing a qualitative change in the phase space dynamics. While the loss of stability is not particularly surprising in the presence of delays, the possibility of stabilization by coupling of an otherwise unstable fixed point is an interesting direction in the study of network dynamics. A natural next step is the stabilization of periodic solutions, since these can be expressed as the fixed points of the iterates of the original map. However, the extension requires additional work in the presence of delays, and is deferred to a future publication.

To explore the stabilizing role of delays in more detail, note that the diffusive nature of the interaction, i.e., (2), implies that in the absence of delays the coupling term vanishes whenever the states of the nodes are identical. Consequently, the synchronized dynamics coincides with the isolated dynamics of the map f. This statement is a reflection of the fact that the Laplacian has a zero eigenvalue corresponding to the eigenvector $(1, 1, \ldots, 1)$. In other words, diffusive coupling, which helps synchronization by reducing the difference between neighboring nodes, is precisely the reason for the lack of control along the synchronization direction $(1, 1, \ldots, 1)$. The situation is markedly different when delays are present. Now the effect of the coupling does *not* vanish along the synchronized direction, which gives it a chance of stabilizing an unstable fixed point of the isolated map. The analysis presented in this paper gives the precise conditions under which such stabilization can occur. We mention in passing that exact synchronization of chaotic maps is still possible under delays, facilitated by the fact that the synchronized dynamics in this case exhibits smaller Lyapunov exponents [6].

The ability of time delays to suppress oscillatory behavior has been observed in continuous-time limit-cycle oscillators [14, 15], where the mechanism essentially involves controlling a Hopf bifurcation. The results of the present paper generalize this effect to networks of arbitrary maps, including chaotic ones, in discrete time. Here it is the flip bifurcation which induces the oscillatory behavior of the map, which is reversed by the collective action of the network, where the parity of the delay has a significant role. Similar to the continuous-time case [11], the effect of the network topology is characterized solely by the largest eigenvalue of the graph Laplacian. On the one hand, the analysis here is more general in the sense that it does not require being near a bifurcation. On the other hand, it assumes scalar units, symmetrical coupling, and a single value for the delay, in contrast to some continuous-time results [11]. The case of multiple delays, especially if they are of different parity, requires different techniques. In case of differential equations, distributed delays are known to be able to enhance stability for single equations [16] as well as networks [17]. To the extent that multiple delays in maps can be considered as the counterpart of distributed delays, one might anticipate further stabilization effects in such general networks; however, this is yet to be

quantified.

Appendix. We prove several intermediate results used in the paper.

Lemma 3. *Let b and c be arbitrary real numbers. Then,*

$$|b-c|+|c|<1 \tag{45}$$

if and only if

$$|b|<1 \quad \textit{and} \quad |b-2c|<1. \tag{46}$$

Proof. If (45) holds, then

$$|b| = |b-c+c| \le |b-c|+|c| < 1$$

and

$$|b-2c| = |b-c-c| \le |b-c|+|c| < 1,$$

yielding (46). Similarly, writing (46) as

$$|b-c+c|<1 \quad \text{and} \quad |b-c-c|<1$$

and expanding gives

$$-1-c<b-c<1-c,$$

$$-1+c<b-c<1+c.$$

Thus,

$$-1+|c|<b-c<1-|c|,$$

which implies (45). ∎

Lemma 4. *Let $p>1$ and $0<x\le 2\pi/(p+1)$. Then*

$$p\sin x > \sin(px).$$

Proof. Define $g(x)=p\sin x-\sin(px)$. Using the identity

$$\cos(z_1-z_2)-\cos(z_1+z_2)=2\sin z_1\sin z_2,$$

we calculate

$$\begin{aligned} g'(x) &= p(\cos x-\cos(px)) \\ &= 2p\sin\left(\frac{p+1}{2}x\right)\sin\left(\frac{p-1}{2}x\right) \\ &> 0, \end{aligned}$$

provided that $0 < x < 2\pi/(p+1)$. Since $g(0) = 0$, it follows that $g(x) > 0$ for $0 < x \le 2\pi/(p+1)$. ■

Lemma 5. *Let τ be a positive number and $0 \le a < (\tau+1)/\tau$. Then there exists a unique number $\Phi = \Phi(\tau, a)$ in the interval $(0, \frac{\pi}{\tau+1})$ satisfying*

$$\frac{\sin((\tau+1)\Phi)}{\sin(\tau\Phi)} = a. \tag{47}$$

Furthermore, $\partial\Phi/\partial\tau < 0$.

Proof. We let

$$h_\tau(\Phi) = \frac{\sin((\tau+1)\Phi)}{\sin(\tau\Phi)}.$$

Then $h_\tau : (0, \frac{\pi}{\tau+1}] \to \mathbb{R}$ is differentiable, and

$$\begin{aligned} h'_\tau(\Phi) &= \frac{(\tau+1)\cos((\tau+1)\Phi)\sin(\tau\Phi) - \tau\sin((\tau+1)\Phi)\cos(\tau\Phi)}{\sin^2(\tau\Phi)} \\ &= \frac{-\tau\sin\Phi + \cos((\tau+1)\Phi)\sin(\tau\Phi)}{\sin^2(\tau\Phi)}. \end{aligned}$$

The sign of $h'_\tau(\Phi)$ is determined by the numerator

$$-\tau\sin\Phi + \cos((\tau+1)\Phi)\sin(\tau\Phi) = \frac{1}{2}\left[-(2\tau+1)\sin\Phi + \sin((2\tau+1)\Phi)\right], \tag{48}$$

where we have used the identity

$$2\sin x\cos y = \sin(x+y) + \sin(x-y).$$

We apply Lemma 4 to conclude that the right-hand side of (48) is negative. Thus, h_τ is strictly decreasing over $(0, \frac{\pi}{\tau+1}]$. Furthermore, $h_\tau(\pi/(\tau+1)) = 0$ and

$$\lim_{\Phi\to 0^+} h_\tau(\Phi) = \frac{\tau+1}{\tau}.$$

Thus, for any a satisfying $0 \le a < (\tau+1)/\tau$, the equation $h_\tau(\Phi) = a$ has a unique solution, which proves the first statement of the theorem. Now fix $a \in [0, (\tau+1)/\tau)$ and consider (47) for varying τ. Implicit differentiation with respect to τ and rearranging give

$$\begin{aligned} \frac{\partial\Phi}{\partial\tau} &= \Phi\frac{\sin((\tau+1)\,\Phi)\cos\tau\Phi - \cos((\tau+1)\,\Phi)\sin\tau\Phi}{(\tau+1)\cos((\tau+1)\,\Phi)\sin\tau\Phi - \tau\sin((\tau+1)\,\Phi)\cos\tau\Phi} \\ &= \Phi\frac{\sin\Phi}{-\tau\sin\Phi + \cos((\tau+1)\,\Phi)\sin\tau\Phi}. \end{aligned}$$

The sign of the last expression is determined by the denominator, which is identical to the left-hand side of (48), which was shown above to be negative. Hence, $\partial\Phi/\partial\tau < 0$. ■

Finally, we list some results from the theory of higher order difference equations related to the study of the characteristic equation

$$s^{k+1} - a_1 s^k + a_2 = 0, \tag{49}$$

where $a_1, a_2 \in \mathbb{R}$ and k is a positive integer. A sufficient condition for stability was given by Clark [18] (see also [19]).

Lemma 6. *All roots of* (49) *lie inside the unit circle, provided* $|a_1| + |a_2| < 1$.

Necessary and sufficient conditions for the stability of (49) were given by Kuruklis [19].

Lemma 7. *All roots of* (49) *lie inside the unit circle if and only if*

$$|a_1| < \frac{k+1}{k}$$

and one of the following is satisfied:

$$k \text{ is odd} \quad \text{and} \quad |a_1| - 1 < a_2 < (a_1^2 + 1 - 2|a_1| \cos \Phi)^{\frac{1}{2}}$$

$$\text{or}$$

$$k \text{ is even}, \quad |a_1 - a_2| < 1, \quad \text{and} \quad |a_2| < (a_1^2 + 1 - 2|a_1| \cos \Phi)^{\frac{1}{2}},$$

where Φ *is the unique solution of* $\sin((k+1)\Phi)/\sin(k\Phi) = |a_1|$ *in the interval* $(0, \pi/(k+1))$.

In the statement of the result in [19], a_1 was assumed to be nonnegative; however, the proof given works also without this restriction.

Acknowledgment. Part of this work was completed in the summer of 2005, when Ö. Karabacak visited the Max Planck Institute for Mathematics in the Sciences on a summer internship. He acknowledges the generous support of the Max Planck Society and Istanbul Technical University, which made the visit possible.

REFERENCES

[1] K. Kaneko, *Period-doubling of kink-antikink patterns, quasi-periodicity in antiferro-like structures and spatial intermittency in coupled map lattices—toward a prelude to a field theory of chaos*, Progr. Theoret. Phys., 72 (1984), pp. 480–486.
[2] K. Kaneko, *Theory and Applications of Coupled Map Lattices*, Wiley, New York, 1993.
[3] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization—A Universal Concept in Nonlinear Science*, Cambridge University Press, Cambridge, UK, 2001.
[4] L. M. Pecora and T. L. Carroll, *Synchronization in chaotic systems*, Phys. Rev. Lett., 64 (1990), pp. 821–824.
[5] J. Jost and M. P. Joy, *Spectral properties and synchronization in coupled map lattices*, Phys. Rev. E, 65 (2002), paper 016201.
[6] F. M. Atay, J. Jost, and A. Wende, *Delays, connection topology, and synchronization of coupled chaotic maps*, Phys. Rev. Lett., 92 (2004), paper 144101.
[7] D. V. Ramana Reddy, A. Sen, and G. L. Johnston, *Time delay induced death in coupled limit cycle oscillators*, Phys. Rev. Lett., 80 (1998), pp. 5109–5112.
[8] F. M. Atay, *Total and partial amplitude death in networks of diffusively coupled oscillators*, Phys. D, 183 (2003), pp. 1–18.
[9] K. Bar-Eli, *On the stability of coupled chemical oscillators*, Phys. D, 14 (1985), pp. 242–252.

[10] D. G. Aronson, G. B. Ermentrout, and N. Kopell, *Amplitude response of coupled oscillators*, Phys. D, 41 (1990), pp. 403–449.
[11] F. M. Atay, *Oscillator death in coupled functional differential equations near Hopf bifurcation*, J. Differential Equations, 221 (2006), pp. 190–209.
[12] S. Elaydi, *An Introduction to Difference Equations*, 3rd ed., Springer, New York, 2005.
[13] F. R. K. Chung, *Spectral Graph Theory*, American Mathematical Society, Providence, RI, 1997.
[14] F. M. Atay, *Delayed-feedback control of oscillations in non-linear planar systems*, Internat. J. Control, 75 (2002), pp. 297–304.
[15] F. M. Atay, *Oscillation control in delayed feedback systems*, Lecture Notes in Control and Inform. Sci., 273 (2002), pp. 103–116.
[16] S. Bernard, J. Belair, and M. C. Mackey, *Sufficient conditions for stability of linear differential equations with distributed delay*, Discrete Contin. Dynam. Systems, 1 (2001), pp. 233–256.
[17] F. M. Atay, *Distributed delays facilitate amplitude death of coupled oscillators*, Phys. Rev. Lett., 91 (2003), paper 094101.
[18] C. W. Clark, *A delay-recruitment model of population dynamics, with an application to baleen whale populations*, J. Math. Biol., 3 (1976), pp. 381–391.
[19] S. A. Kuruklis, *The asymptotic stability of* $x_{n+1} - ax_n + bx_{n-k} = 0$, J. Math. Anal. Appl., 188 (1994), pp. 719–731.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 529–551

Phase Boundaries as Electrically Induced Phosphenes*

Jonathan D. Drover† and G. Bard Ermentrout‡

Abstract. A model is presented of experiments where electrical stimulation of the eye of human subjects results in the perception of evenly spaced lines, or phosphenes. The model is a two-dimensional grid of integrate-and-fire oscillators that captures the important experimental characteristics of line-creation when a sinusoidal current injection is used. The spatio-temporal behavior of the lines, once formed, is also reproduced. A reduced model consisting of an evolution/convolution equation on the real line is analyzed, and it is shown that stationary solutions with arbitrarily located discontinuities exist and are linearly stable. Traveling waves are numerically shown to exist when the coupling is both sufficiently strong and biased, which accounts for the movement of the lines in the experiments.

Key words. phase boundaries, phosphenes, traveling waves

AMS subject classifications. 37N25, 34C15, 92B05

DOI. 10.1137/050646469

1. Introduction.

A phosphene is the sensation of light produced from within the nervous system rather than from an external source. Examples of phosphenes range from the stars seen when one is hit on the head to the geometric patterns perceived during hallucinations. Phosphenes can be induced by direct electrical stimulation of the visual pathway, including the eyeball, leading to the preliminary design of prosthetic visual devices for patients with severely limited vision [14].

Claussen [6] was the first to discover that sinusoidal electrical stimulation of the retina produced a variety of complex visual sensations. In experiments two decades later, Carpenter [4] studied these electrical phosphenes in much greater detail. If the eyeball is stimulated by high frequency (100 Hz) alternating current and a dark bar is passed through visual space, it leaves in its wake a series of thin contours which slowly move over time. These disappear almost instantly when the electrical stimulation is turned off. The production and movement of these contours is the subject of this paper.

While electrical stimulation of the eye is rather unnatural, it has recently been used to study the effects of electromagnetic radiation on the human nervous system [1]. Furthermore, Wilms et al. [23] have used direct electrical stimulation to estimate the spatial resolution of the retina as mapped onto the visual cortex. Unusual and abnormal stimuli have long been used as probes of the visual system, as they can often provide information about processing that is not available with more natural stimuli. For example binocular rivalry experiments (different images presented to the left and right eyes, leading to an alternation in the perceived

*Received by the editors November 30, 2005; accepted for publication (in revised form) by T. Kaper May 22, 2006; published electronically October 4, 2006.
http://www.siam.org/journals/siads/5-4/64646.html
†Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 (drover@njit.edu).
The work of this author was supported by an NSF grant while attending the University of Pittsburgh.
‡Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260 (bard@pitt.edu).

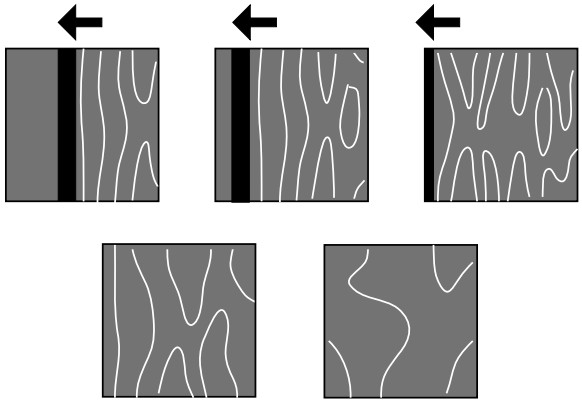


Figure 1. *The patterns reported by subjects in Carpenter's experiments (redrawn after* [4]*).*

image) led to insights into how visual forms are processed in the cortex [24]. Classic illusions such as Mach bands (illusory light or dark areas at the ends of a luminence ramp) [20] led to the theory of lateral inhibition in the retina.

In this paper, we present a new model for the formation of the so-called contour phosphenes elicited through electrical stimulation of the eyeball. Our model is based on experiments demonstrating 1:2 phase-locking of retinal cells to periodic stimuli [7], together with anatomical and physiological evidence for electrical coupling between the cells in the retina. In section 2, we introduce the experimental protocol, describe the phenomena, and introduce our model. We start with a one-dimensional (1D) system to illustrate the basic mechanism, and then turn to a two-dimensional (2D) model which reproduces both the formation of the contours and their subsequent movements. To gain better mathematical insight into the existence of contours and their stability, in section 3, we introduce a simplified 1D continuum model for which we can prove the existence of alternating domains much like those in the experiments and simulations. We prove that these are stable and also explore the onset of movement of the boundaries.

2. Electrically induced phosphenes. In this section, we model the spatio-temporal patterns reported by subjects in an experiment described in [4]. We begin by presenting a brief description of the experiment and the illusion.

A subject's eyes are submerged in a saline bath, and alternating electrical current is passed through the bath. The subject views a uniformly lit screen while receiving the stimulation. A dark object is passed through the field of vision. Subjects report line phosphenes or contours (illusions of light) in the wake of the trailing edge of the dark object. Carpenter shows that there is one line created for every complete cycle of the driving stimulus coincident with the moving edge, yielding the temporal and spatial periodicity. Thus a slowly advancing edge will produce many more lines than a rapidly advancing one. The result, upon full passage of the edge across the medium, is a set of evenly spaced line phosphenes. These lines slowly evolve in time, moving in various directions and occasionally interacting with each other to form loops (Figure 1). Later in this section, we will specify the rules of movement described in Carpenter's paper.

2.1. A model for contour formation. According to Carpenter [4] and Brindley [3], the genesis of these phosphenes is in the retina itself. This contrasts with the presumed cortical origin of more complex phosphenes seen during flicker, occular pressure, or drug ingestion [21]. Thus, we will assume that all of the dynamics are occurring within the retina.

Carpenter suggests a hypothesis as to the nature of the illusions: *The lines represent nodes, separating areas of the retina that are responding in antiphase to each other.* We take this as our main hypothesis and create a biophysically plausible implementation using recent physiological data on the retina. The basic idea is to induce bistability at each spatial location in the retina so that the phosphenes become the boundaries between the two different stable states. Given that the stimulus is periodic, one way to achieve a de facto bistability is to assume that the relevant retinal cells are able to fire only on every other cycle of the stimulus. That is, each cell is locked in a 1:2 (one spike per two stimulus cycles) manner to the 100 Hz current. Thus, the bistability is between cells firing on the "odd" cycle of the 100 Hz current and cells firing on the "even" cycle. There are many candidate cells in the retina: bipolar cells, amacrine cells, horizontal cells, and ganglion cells (whose output goes to the central nervous system). Our model does not depend on which type of cell is firing, only on the fact that the cells generating the activity are coupled. As there are electrical (gap) junctions between most of the cell types in the retina [17, 16, 15], we will not speculate as to which neurons are involved in the phosphenes as perceived in human subjects.

In [7], Crevier and Meister studied human electroretinograms (ERGs) and salamander retinal cell responses to periodic pulses of light. In the human studies, the authors found that there is a period doubling (1:2 locking) of the ERG at between 30 to 70 Hz photic stimuli. In order to explore the origin of this, these authors looked at the neurophysiology of salamander retinas since their eyes have similar structure to those of humans. They found that the photoreceptors (the first stage of vision) are able to maintain the 1:1 locking with the stimulus, but that ganglion cells (last stage in the retina) can follow only in a 1:2 manner. Thus, the inability to follow the stimulus in a 1:1 manner occurs between the photoreceptors and the ganglion cells—this includes all the cell types mentioned above: bipolar, amacrine, and horizontal.

For simplicity, we model each cell (or cluster of cells that behave identically) as an integrate-and-fire neuron with adaptation [8]:

$$\frac{dx}{dt} = -x - z + A\sin\left(\frac{2\pi t}{T}\right), \tag{1}$$

$$\frac{dz}{dt} = -\frac{z}{\tau}. \tag{2}$$

The variables x and z are real. The parameter A is the amplitude of the driving stimulus, and T is the period. We assume that both A and T are positive. The driving stimulus is the analogue to the alternating current in the experiments. The reset criterion is given by

$$x(t^-) = x_{spike} \to x(t^+) = x_{reset}; \qquad z(t^+) = z(t^-) + z_{jump}.$$

For the remainder of this section, we assume the parameter values to be $x_{spike} = \pi$, $x_{reset} = -\pi$, and $z_{jump} = 1$.

The purpose of the refractory variable, z, is to make the 1:2 phase-locking more robust. In the absence of the explicit refractory variable, the oscillations approach threshold during every cycle of the driving current, and 1:2 locking occurs only in a very narrow range of parameters. In [9] parameter regimes for 1:2 locking are explicitly found; that is, the neuron spikes exactly once for every two cycles of the current for a range of values of A. For the remainder of this paper, we assume that $A = 4.7$. Because one phase-locked solution exists, another must also exist; it is a translate of the first by one period of the stimulus. The result is a form of bistability, where the attractors are a pair of periodic orbits. A cell, after some transient behavior, will converge to one of these solutions, either the even cycle or the odd cycle. By symmetry, the basins of attraction of the two solutions have equal measure.

We now turn to the mechanism that causes the medium to break into the desired regions. We assume that, at the onset of the stimulus, the entire retinal area is simultaneously excited. Thus *all cells start in the same phase of the* 1:2 *locked solution.* All that the subject observes is a uniform background, since 100 Hz is well above the critical flicker fusion frequency of about 30 Hz. The input of a dark bar across the visual field causes the retinal receptors to depolarize. (The primary receptors in the retina act counterintuitively, since light hyperpolarizes or inhibits them.) This signal is inverted by the bipolar cells to produce spatially restricted inhibitory signal to the cells which are locked in the 1:2 rhythm with the electrical signal. This inhibition is sufficient to keep the cells from firing. Once the bar passes through, the cells are released from the inhibition and can start to fire again. Depending on the time at which they are released, they will be drawn into either the "odd" or the "even" 1:2 locked solution. All cells aligned in parallel with the bar of light will be released simultaneously, so that all cells in the vertical direction will have the same phase. Thus, to understand the initial creation of the phase-boundaries, we need consider only a 1D line of cells through which sweeps an inhibitory pulse.

We model this inhibition using a step function

$$\text{bar}(i,t) = \begin{cases} -d & \text{if } i/v + T_0 < t < i/v + T_0 + W, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where i is the index of the node, T_0 is the time when the sweep begins, W is the amount of time that a node is inhibited, and v is the velocity of the sweep. The parameter d is the strength of the inhibition.

We first consider a horizontal line of cells with the inhibitory sweep. The variable x_i at location i satisfies

$$\frac{dx_i}{dt} = -x_i - z_i + A\sin(\omega t) + \text{bar}(i,t) \equiv F_i(x_i, z_i, t). \tag{4}$$

The creation of the lines occurs in the wake of the traveling inhibition. Prior to the sweep, the visual field is uniform, and all cells are phase-locked to the driving stimulus. The traveling inhibition causes the nodes to deviate from the phase-locked solution. The inhibition is sufficiently long so that, upon the completion of a sweep, approximately half of the nodes are left in the basin of attraction corresponding to the opposite phase solution. Because the inhibition travels with a constant velocity, once the inhibition has passed, the medium has alternating regions firing in the odd and even cycles. The locations where there is a phase

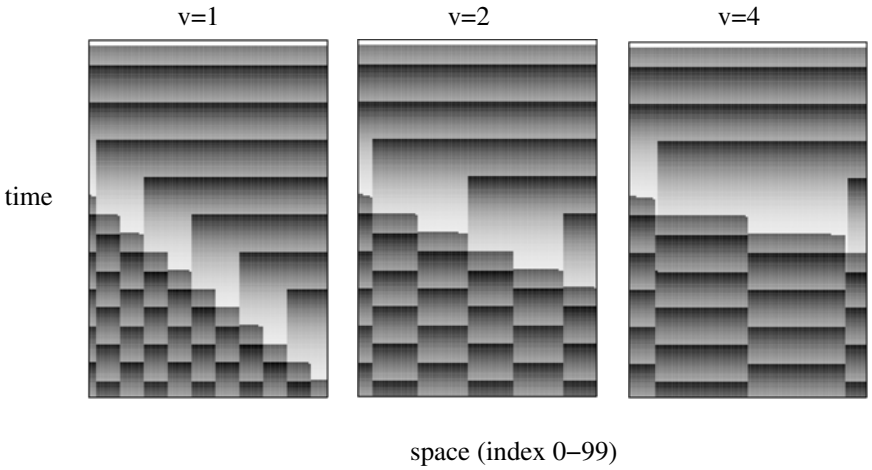


Figure 2. *Three simulations of a line of oscillators inhibited by a sweep of the bar function. In each panel the horizontal axis is the spatial index, and the vertical axis is time, increasing from top to bottom. Grayscale represents the value of the refractory variable, z_i. The velocity of the sweep is above each panel and measures the speed of the movement of the bar in index units per time. In all three panels $W = 30$ and $d = -2$.*

boundary correspond to the phosphenes in Carpenter's experiment. The phase-locked solutions have identical basins of attraction, and so the result is the appearance of evenly spaced phase boundaries. In Figure 2, the behavior of the line of oscillators is shown during and after a sweep. In Carpenter's experiments, one line was seen for every cycle of the stimulus coinciding with the trailing edge of the bar. Thus, the slower the bar moves, the more lines. In this figure, $T = 10$. For $v = 1$, there are ten cycles of the stimulus, and there should be 10 lines. Similarly, for $v = 2$ and $v = 4$, there should be 5 and 2–3 lines, respectively. These simulations confirm that our model behaves appropriately. Furthermore, the thickness of the contours is independent of the velocity of the bar. This suggests that the contours are *not* different states of the system, but rather, they are the *boundaries* between two stable states. We next model the movement of these boundaries.

2.2. Movement of the lines: Two dimensions. The previous section provided a simple but robust mechanism for partitioning a 1D "retina" into regularly spaced domains of alternating phase. Suppose, for the moment, that our situation is exactly as in section 2.1, where all the cells are uncoupled and independent, but now arranged in a 2D grid. A stimulus in the form of a long vertical bar is moved horizontally across the 2D array of cells. All cells in any column will behave exactly the same since the bar is vertical. After the bar passes, it will leave a series of vertical stripes representing the alternating domains of in- and out-of-phase oscillations much like those shown at the leading edge of the stimulus in Figure 1. However, in the experiments, the lines do not remain fixed. Rather, they move and appear to interact with each other. In this section, we suggest that the reason for the movement is that there are interactions between the neighboring cells which underlie the phosphene patterns.

Carpenter makes the following observations about the phosphenes:

1. Lines never cross through one another. Rather, they combine to form loops.
2. A line never breaks apart unless it meets another line.

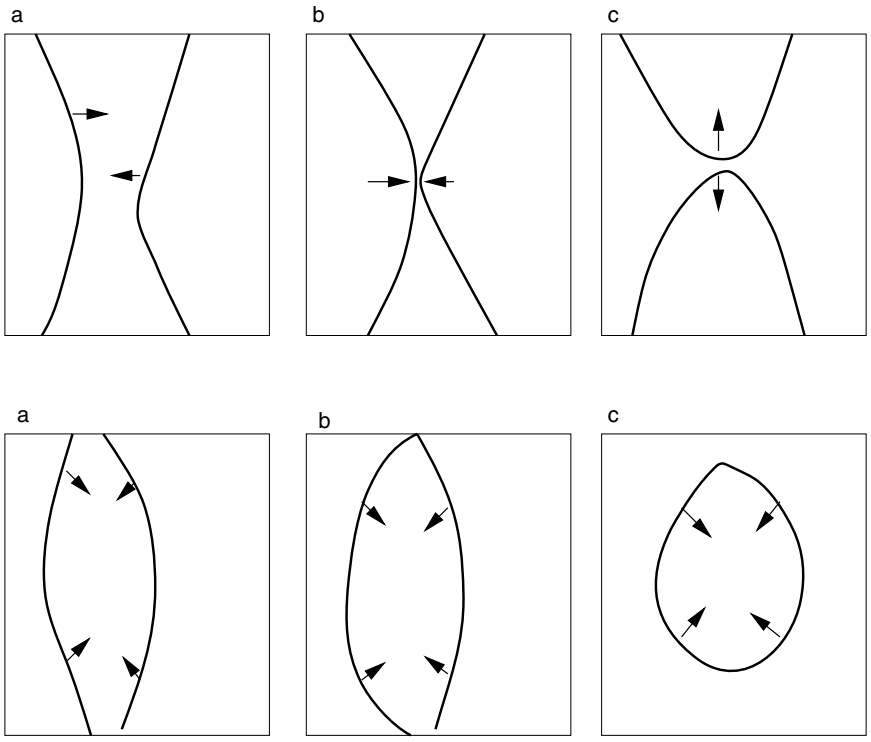


Figure 3. *Line behavior when two lines meet. The top panel depicts the case where the lines meet near their centers. The bottom panel is the case where the lines meet near the edges. In either case, the lines do not pass through one another, but instead leave patterns similar to those shown.*

3. Neighboring lines show a tendency to move in a similar manner. That is, if a given line is bulging to the right, then it is likely that a neighboring line will bulge in the same direction.

Figure 3 illustrates the first two rules.

In the model as posed so far, each cell is independent from the others, and there are no interactions. Furthermore, the model is completely homogeneous. However, there are many interactions between retinal neurons—graded chemical synapses, electrical junctions, threshold chemical synapses, etc. Thus, we can expect that, at least locally, the behavior of one neuron will influence that of another one. The detailed means of coupling is less important than its effect. We will assume that coupling is such that two coupled neurons will tend to synchronize their activities. As there is much evidence for electrical (gap) junctions in the retina [17, 15] and since such coupling can synchronize neuronal oscillations [13, 19], we will illustrate the phenomena with gap junction coupling. We note that gap junction coupling is generally modeled as a linear function of the difference between the potentials of two cells: discrete diffusion.

Coupling provides a mechanism for the first two of Carpenter's observations. (See Figure 4.) Within a domain, all cells are nearly equal and firing at the same cycle, so that in the interior the coupling has a negligible effect. However, a point near the boundary will try to synchronize with cells which are firing on opposite phases of the stimulus. With perfectly

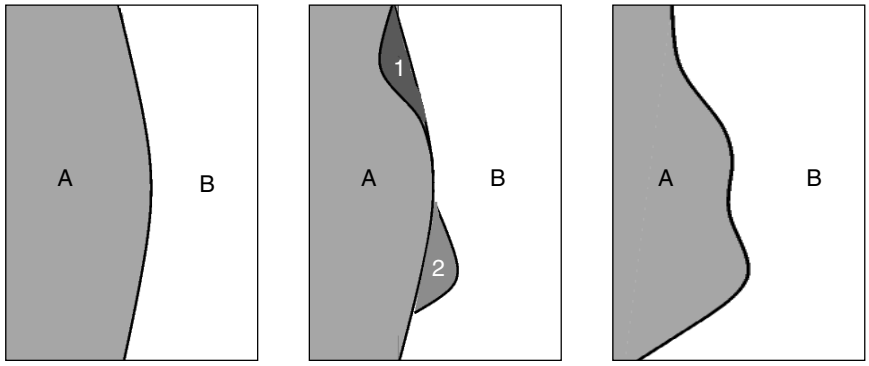


Figure 4. *Line movement mechanism. The figure depicts three moments in time, advancing from left to right. In the leftmost, the boundary is shown separating regions that are out of phase with one another. In the center of the figure, the region marked* 1 *is more strongly coupled to the cells in region B than to those in region A. The opposite is true for those cells in region* 2*. The right of the figure shows the result once the cells in region* 1 *have synchronized with those in region B and the cells in region* 2 *have synchronized with region A. This is how line movement occurs in our model.*

symmetric coupling, the battle is a "tie." However, if the coupling has any kind of asymmetry (which is generally the case), then the boundary point will be recruited into the region with the stronger effective coupling, thus moving the boundary point; one region will take over the other, and the lines will move away from the absorbing area. In the uncoupled case, the pattern is stable so that the coupling has to be sufficiently strong to begin the process of recruiting territory. In order to explain rule 3, we have to make a somewhat stronger assumption about the coupling heterogeneity. If the heterogeneity is at a "microscopic" scale, that is, essentially random from cell to cell, then we cannot expect any kind of trend in motion such as seen in Figure 4. Thus, we suppose that the coupling strength varies on a coarser scale. If the coupling has a favored direction at some point x, then nearly the same direction will be favored for a point y near x. We have no evidence of such trends in coupling, but, neither is there any compelling evidence against this. This notion could be tested by looking for any kind of asymmetry of spontaneous wave propagation in isolated retinas [12].

2.2.1. Line movement and biased coupling strength. We assume that coupling within the retina, particularly the gap junctions between horizontal cells, is not of uniform strength. We accomplish the movement of the lines by coupling a given node to its neighbors with different strengths. Since the effect of the coupling is to synchronize, a cell will approach the phase of the neighbor to which it is most strongly coupled (see Figure 6 below). It is through this mechanism that cells on a boundary synchronize with neighbors, moving the boundary itself (Figure 4).

The coupling has to have a number of characteristics, inspired by the movement "rules" stated in the previous subsection. Rules 2 and 3 make it essential that the coupling cause synchrony to be a locally stable solution. The most obvious form of coupling would be to add a term proportional to the difference between the x components of the two cells, e.g., $x_2 - x_1$. This form of coupling works very well for a pair of coupled cells and, if sufficiently strong, induces synchrony [19]. However, we have found (simulations not shown) that linear coupling

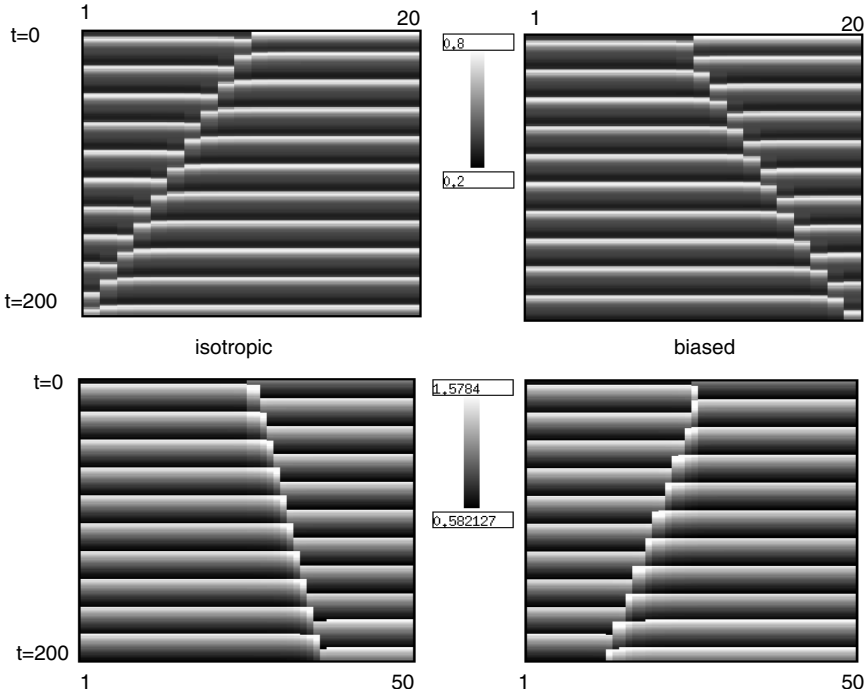


Figure 5. *Behavior of a forced array of Hodgkin–Huxley neurons (top row) with isotropic and biased linear diffusional coupling. Bottom row: Integrate-and-fire model with sinusoidal coupling, with and without bias.*

of the integrate-and-fire model in spatially organized arrays of cells does not lead to smooth patterns in which one phase takes over the other. Rather, the medium breaks up into very fine-grained spatial patterns. Thus, we do not use linear coupling between the cells, but instead coupling which depends on the sine of the difference. To justify this somewhat unusual form of coupling, we compare a 1D integrate-and-fire network with sinusoidal coupling to a 1D network of periodically forced Hodgkin–Huxley neurons. Each cell satisfies the four variable Hodgkin–Huxley equations and is forced at 100 Hz by a sinusoidal stimulus with amplitude sufficient to lead to 1:2 locking. Cells are coupled to nearest neighbors with identical coupling to the left and right cells or with a bias in one direction. Figure 5 shows a simulation when the coupling is strong enough to destroy the two-phase pattern. When the medium is isotropic, a wave is generated, and it is always in the same direction (top left). By biasing the coupling strength, we can make the wave travel in the opposite direction (top right). The same phenomenon is illustrated with the integrate-and-fire model with sinusoidal (as opposed to linear) coupling. Since this is the type of behavior we are looking for, we use sinusoidal coupling instead of linear coupling for the integrate-and-fire model.

With these considerations, we return to the full 2D model and use coupling of the form

$$cf_d \sin(x_d - x_{i,j}), \tag{5}$$

where $d = \{up, down, left, right\}$ (e.g., $f_{up} = f_{i,j-1}$). See Figure 6. The coefficients f_d are positive and discussed in detail in the next subsection. We restrict ourselves to the case of nearest neighbor coupling, although in the next section longer range coupling is allowed. The

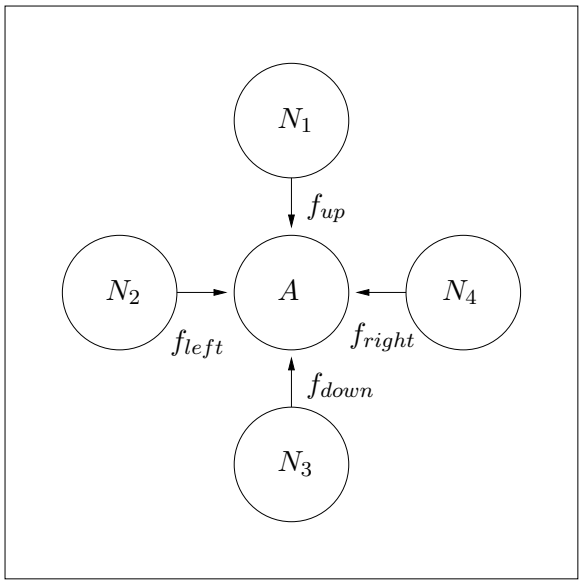


Figure 6. *Coupling diagram.*

parameter c is always positive and determines the linear strength of the coupling, and it is the same for each node. The evolution of the cells in the coupled network is governed by the equation

$$\frac{dx_{i,j}}{dt} = F_{i,j}(x_{i,j}, z_{i,j}, t) + c\sum_d f_d \sin(x_d - x_{i,j}), \tag{6}$$

where $F_{i,j}(x, z, t)$ is the local dynamics from (4) along with the dynamics for the refractory variable governed by (2) and the sum is over the four nearest neighbors. The 1D bar in (4) is replaced by the corresponding rightward moving vertical bar in the 2D system.

2.2.2. The coupling coefficients f_d. To motivate our choice of coupling, we briefly discuss the mechanism for movement of the phosphenes. Suppose that we have one oscillating cell, A, with four neighbors, N_j, $j = 1, \ldots, 4$ (Figure 6). Suppose that N_1 and N_2 are firing synchronously (with one another), and N_3 and N_4 are firing at the opposite phase. The phase to which A synchronizes depends on the relative strengths of the coupling. For example, suppose that $f_{up} + f_{left} \gg f_{right} + f_{down}$. Then, oscillating cell A will synchronize with N_1 and N_2. If the relative difference between coupling strengths is not so high, there will be an intermediate phase for A, which will in turn alter other nearby cells. Eventually, cell A will be on the interior of a region and will synchronize with the others there. The phosphenes are represented by the boundaries of regions which have different phases where they temporarily assume a parameter dependent (relative sizes of f_d), and neighbor dependent, phase and are then absorbed into the interior of a region. Once this process is complete, the line (boundary) will have moved.

Because the direction of the movement depends on the relative coupling strengths, we wish to structure the coefficients, f_d, spatially. Suppose we have a single line of oscillators. If

for every oscillator $f_{right} > f_{left}$, we expect each to synchronize with its neighbor to the right, thus moving the lines left.

From rule 3, neighboring lines tend to move in the same direction, so there should be spatial intervals in which coupling in a particular direction is favored. Thus, rather than choosing the coupling strengths independently at each spatial point, we allow them to vary in a continuous fashion. For example, if $f_{up} > f_{down}$ at location x, then the same inequality will hold for nearby oscillators. The particular choice of coupling parameters is not crucial; however, continuity is needed to satisfy rule 3. The specific coupling matrix is described in the next section.

Bias in the coupling strength plays an important role in the movement of the phosphenes. Figure 5 shows that, without bias, the phosphenes move in one direction. In particular, for the Hodgkin–Huxley model, one cycle (say the even) always takes over the other if the coupling is sufficiently strong. The reason for this (similarly for the integrate-and-fire model) is as follows. Consider a pair of uncoupled cells which are firing on alternate cycles. Turn on the coupling at a time t. If the coupling is strong enough, the cell that fires first after t will cause the other cell to fire, and thus the second cell will be pushed into firing in the same cycle as the first cell. In a network in which half the cells are set in the even and the other half in the odd cycle, the moment the coupling is turned on will determine the direction of the wave. This is an exquisite sensitivity to initial conditions as well as any small heterogeneities. Thus, by making the medium anisotropic, we allow waves that robustly travel in a preferred direction.

2.3. Simulation. Simulations were done using a fourth order Runge–Kutta integrator with constant time step, $\Delta t = 0.01$. The reset is accomplished by setting $x_{i,j}(t_k) = -\pi$ and $z_{i,j}(t_k) = z_{i,j}(t_k) + 1$ whenever $x_{i,j}(t_k) > \pi$. No interpolation is done, resulting in resets that always occur at a multiple of the time step.

All simulations are done on a 100×50 grid. The coupling strength array, $f_{i,j}$, was determined as follows:

1. Randomly choose three indices in the horizontal domain (h_1, h_2, h_3), and three more in the vertical domain (v_1, v_2, v_3). Assume that $h_1 < h_2 < h_3$ and $v_1 < v_2 < v_3$. The choice of three set indices is motivated by figures in [4]. This implementation can be extended to incorporate any number of set nodes, up to the number of nodes present in the grid.
2. Define two arrays, H and V (for horizontal and vertical), with the appropriate number of elements. For our purposes H has 100 elements (the width of the grid) and V has 50 elements (the height of the grid).
3. Assign to the array elements $H(1)$, $H(h_1)$, $H(h_2)$, $H(h_3)$, and $H(100)$ random values between 0 and 1.
4. The grid is now divided into rectangles. Divide each of these rectangles into two right triangles. In the simulations presented here, the diagonal goes from the top left to the lower right.
5. Using the three corners of the triangles, compute the value at the indices inside each triangle according to the plane the corner values define (see Figure 7).

We assume (as noted above) that there are heterogeneities in the retina, but at a coarse scale that covers many cells so that coupling strengths have some spatial correlations. While the

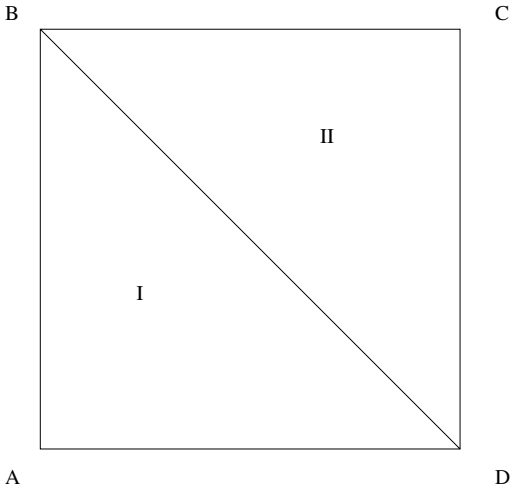


Figure 7. *A diagram that shows how the biased coupling is determined for the simulations. The value of the function is chosen randomly at the vertices A, B, C, D. The rectangle is divided into two triangles. The value of the coupling function for points in these triangles is determined using a linear interpolation of the three vertices that form the triangle. For example, if a node lies in the region marked II, the value of the relative coupling strength at that node is determined by the plane that crosses the determined points at B, C, and D.*

evidence for such large-scale heterogeneities is unclear (Marla Feller, personal communication, and [12]), by carefully looking at spontaneous propagating waves in isolated retinas, it may be possible to test this hypothesis.

Simulations were carried out using FORTRAN code, with calls to LAPACK and BLAS to do the vector operations in the Runge–Kutta integrator. The graphical output was produced using the PGPlot package. Because we use an integrate-and-fire model, we plot the recovery variable, $z_{i,j}$, as this is continuous.

Figure 8 shows output from a sample run of the simulation. In panel A, the bar (outlined in a thin dashed line) is about to complete the pass to the right. Lines (or regions) have formed on the left edge and have begun to move. In panels B and C, the lines are clearly defined and moving. In panel D, synchronous regions "poke holes" through a region in antiphase, forming two loops. Panels E and F show two of the loops annihilating themselves as the region that defines them collapses. Compare this with Figure 1.

3. A bistable evolution/convolution network. The analysis of the formation of domains, their transient stability, and the onset and direction of movement is a difficult task, given that each "cell" is governed by a nonlinear periodically driven 2D differential equation. Hence, it is useful to introduce a heuristic model which has similar qualitative features. In this section, we consider such a simplified model for the spatial network for which we can prove the existence of local phase-domains. The simple model also provides insight into the movement of the boundaries between domains and how the transition between stationary and moving patterns is effected. We will concentrate on a 1D network since this is conceptually easier to understand. However, the theorem in [2] is independent of the spatial dimension and applies, in particular, to 2D domains. The result proved below is an extension of [2], so that, with little effort, a 2D version of it is likely to hold.

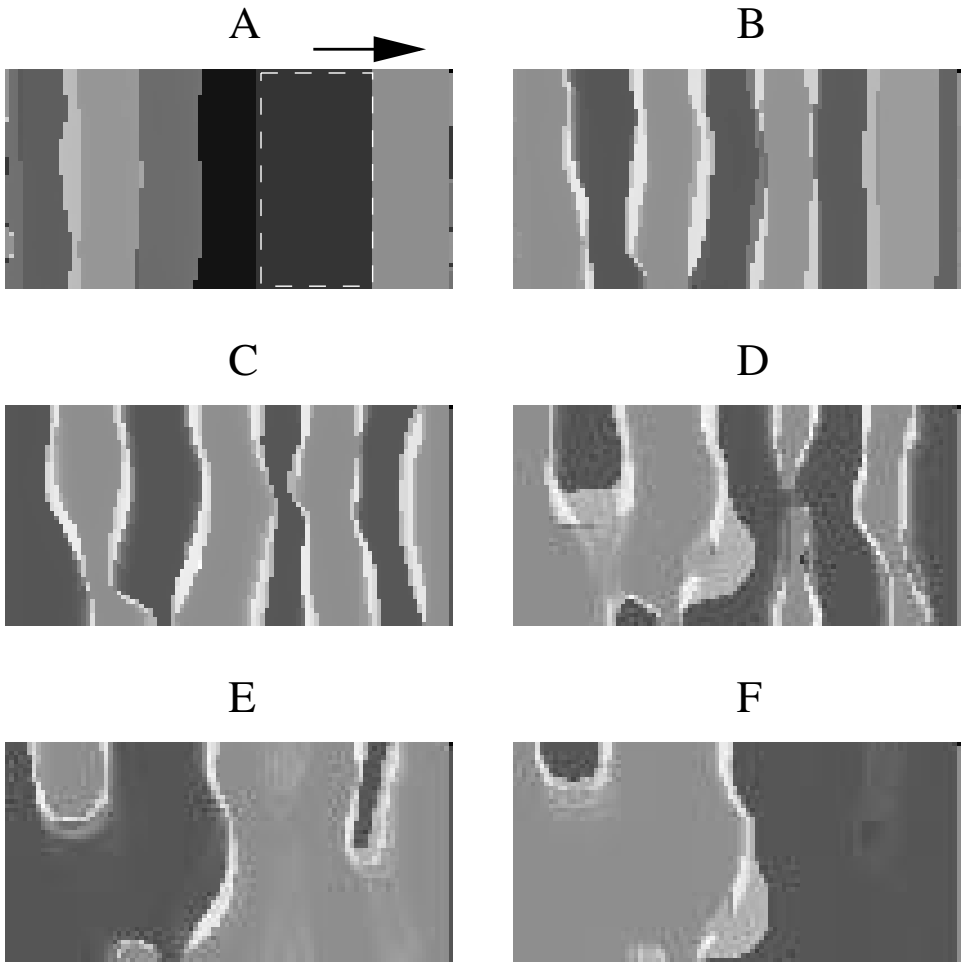


Figure 8. *Simulation of the integrate-and-fire model on a* 100×50 *oscillator grid. Each panel shows the field after a set increment of time. The variable plotted is* z*, the refractory variable. Parameter values (for* (6)*) are* $v = 1.3$*,* $A = 4.7$*,* $T = 10$*,* $\tau = 20$*,* $d = -2$*,* $W = 30$*,* $c = 7$*. Clicking on the above image displays the accompanying animation of simulated phosphenes* (64646_01.mpg [2.93MB]).

3.1. Derivation. The integrate-and-fire model with adaptation is an excitable system. With sufficiently strong periodic drive, each cell can be driven to fire at every other cycle. Conceptually, in the absence of coupling we can represent the dynamics of each cell by a scalar variable depicting its phase relative to the firing time of the stimulus. The period of the stimulus is T, but since each cell is firing only at every other cycle, the period of an individual cell is $2T$. Thus, the individual dynamics of each cell can be written in terms of the scalar variable as

$$u_t = H(u). \tag{7}$$

Zeros of $H(u)$ correspond to times with respect to the stimulus at which the cell fires. The function $H(u)$ should be T-periodic, so that if $u = \bar{u}$ is a fixed point, $H(\bar{u}) = 0$, then there

will be a second fixed point, $u = \bar{u} + T$, corresponding to u firing on the opposite cycles of the stimulus. Thus, on the interval $[0, 2T)$, there is bistability corresponding to firing on the even and odd cycles of the periodic stimulus. Furthermore, we want $\bar{u}$ to be the only stable fixed point in the interval $[0, T)$. An exact scalar equation of the form (7) can be derived from the integrate-and-fire with adaptation model if we assume that the unforced system is an intrinsic oscillator and the forcing has roughly twice the same frequency [10]. However, in the present situation, the forcing is strong and the unforced model is excitable. Thus (7) should be viewed as a simplified version of the forced excitatory cell.

Now, consider two neighboring cells which are coupled by gap junctions, that is, coupling depending only on the difference between, say, the voltages of the cells, and which vanishes when the cells are in identical states. Since $u \in [0, 2T)$, the coupling function must be $2T$-periodic. Furthermore, if both cells are in the same state, then the coupling should not contribute anything to the dynamics since the coupling is diffusion-like. This leads us to the following simplified equation for a pair of coupled cells:

$$\begin{aligned} u_1' &= H(u_1) + cD(u_2 - u_1), \\ u_2' &= H(u_2) + cD(u_1 - u_2), \end{aligned} \tag{8}$$

where H is T-periodic and D is $2T$-periodic. Before turning to the analysis of the continuum model, let us consider a simple example illustrating the general phenomena. Suppose that $T = 2$ and choose $H(u) = \sin \pi u + a \cos \pi u$ and $D(u) = \sin(\pi u/2) + q(1 - \cos(\pi u/2))$. Note that we have included cosine terms in the equations for H, D since we cannot assume any kind of intrinsic symmetry in the functions, as this can destroy the genericity of the results [18]. For small a, $u_1 \approx 1$ and $u_2 \approx -1 \equiv 3$ is a fixed point when $c = 0$. Thus u_1 fires on even cycles of the stimulus and u_2 fires on odd cycles. For c small enough this stable fixed point persists. However, for $c > c^*$, this state is lost via a saddle-node bifurcation (not shown, but see Figure 10 for a higher-dimensional analogue), and the system synchronizes at $u_1 = u_2 \approx 1$ or $u_1 = u_2 \approx -1$. That is, one domain takes over the other.

If we now imagine an array of cells arranged on a line with local coupling which can extend beyond nearest neighbors, then we obtain the following generalization of (8):

$$u_j' = H(u_j) + c \sum_k J_{j-k} D(u_k - u_j), \tag{9}$$

where J_k is a nonnegative weight for the strength of coupling. Since the strength of electrical junctions falls off with distance [13], we assume the same about J_k. Proceeding to the continuum limit, we obtain

$$\frac{\partial u}{\partial t}(x,t) = H(u(x,t)) + c \int_{\mathbb{R}} J(x-y) D(u(y,t) - u(x,t)) dy, \tag{10}$$

which we analyze in the next section.

3.2. Behavior of the continuum model. In this and the succeeding sections, we analyze (10), where $u : \mathbb{R} \times (0, \infty) \to \mathbb{R}$. The parameter c is real and positive. The function H is continuously differentiable and periodic with period T. The function D is continuously

differentiable and periodic with period $2T$. We assume, without loss of generality, that $T = 2$. Additionally, the functions satisfy

$$H(-1) = H(0) = 0,$$

$$H'(-1) < 0,$$

$$H'(0) > 0,$$

$$D(0) = 0,$$

$$D'(0) > 0.$$

The conditions on H and D guarantee that each is bounded and well defined on the entire real line. Thus, solutions to the initial value problem with $u(x, 0)$ evolving according to (10) will not blow up in finite time. Also, since the right-hand side of (10) is always defined, the first time derivative of u exists for all $t > 0$, and so the solutions will be continuous in t. In other words

$$\lim_{\Delta t \to 0} u(x, t + \Delta t) - u(x, t) = 0$$

for every $x \in \mathbb{R}$.

In [2], the authors prove the existence and stability of stationary solutions to

$$u_t = -u - \lambda f(u) + J * u, \tag{11}$$

where $\lambda > 0$, $J * u$ is the spatial convolution of J with u, and the function f is bistable. Specifically, they prove that, under conditions on the parameters, there exist stable steady state solutions that are discontinuous. In this section we prove the existence of similar solutions to (10). We remark that the proofs of existence and stability do not depend at all on the fact the the model is on a 1D domain. Indeed, all proofs hold in arbitrary domains and, in particular, the planar domain of the simulations. We focus on the 1D case, as the proof is easier to explain.

3.3. Existence of stationary solutions. We wish to prove the existence of stationary solutions that are discontinuous at arbitrary points on the real line. We begin by choosing a set, M. Denote the complement by M^c. Also, choose $\beta > 0$ such that $H'(u) < 0$ for all $u \in (1 - \beta, 1 + \beta)$. We wish to prove the existence of a solution, $U(x)$, satisfying

$$0 = H(U(x)) + c \int_{\mathbb{R}} J(x - y) D(U(y) - U(x)) dy \tag{12}$$

with

$$U(x) \in (1 - \beta, 1 + \beta) \quad \text{when } x \in M,$$

$$U(x) \in (-1 - \beta, -1 + \beta) \quad \text{when } x \in M^c.$$

We assume that the function H satisfies the following conditions:

- H is a periodic function with period 2,
- H is continuously differentiable,
- $H(1) = 0$ and $H'(1) < 0$,
- $H(0) = 0$ and $H'(0) > 0$,

and that the function D satisfies

- D is continuously differentiable,
- $D(0) = 0$ and $D'(0) > 0$,
- D is periodic with period 4 (twice the period of H).

These conditions are consistent with the analogy to the full integrate-and-fire model. Specifically, the 2-periodicity of H reflects the symmetry of the medium. In the full model, each of the pair of stimulus-induced basins of attraction were identical. Since the function H describes the intrinsic properties for a given point on the line, it is appropriate to assume higher periodicity than for D. The condition that H be continuously differentiable is for convenience in the proof that follows. The conditions on the function D reflect the properties of the coupling. The full model was developed under the assumption that the coupling encourages synchrony between neighboring cells. Since D depends on the difference in u, the $D'(0) > 0$ condition follows easily (as this is equivalent to *positive* diffusion). Also, cells that are synchronized have no effect on one another, which leads to the condition that $D(0) = 0$. Finally, the choice of fixed points of H is arbitrary and for convenience.

Let $\beta > 0$ be a number such that $H'(u) < 0$ for $u \in (1-\beta,\ 1+\beta)$, and set $\delta = -\max_{u\in(1-\beta,1+\beta)} H'(u) > 0$. Set $K = \max_{u\in\mathbb{R}} |D'(u)|$. Choose a measurable set M and denote the complement as M^c. Assume that c is small enough so that the following conditions hold:

$$
\begin{aligned}
H(1+\beta) + c\overline{D}_{MM^c} \sup_{x\in M} \int_{M^c} J(x-y)dy &\le 0,\\
H(1-\beta) + c\underline{D}_{MM^c} \sup_{x\in M} \int_{M^c} J(x-y)dy &\ge 0,\\
H(-1+\beta) + c\overline{D}_{MM^c} \sup_{x\in M^c} \int_{M} J(x-y)dy &\le 0,\\
H(-1-\beta) + c\underline{D}_{MM^c} \sup_{x\in M^c} \int_{M} J(x-y)dy &\ge 0,
\end{aligned}
\tag{13}
$$

where

$$
\overline{D}_{MM^c} = \max_{s\in(2-2\beta,2+2\beta)} D(s),
$$

$$
\underline{D}_{MM^c} = \min_{s\in(2-2\beta,2+2\beta)} D(s).
$$

These conditions quantify the competition between the intrinsic properties at a location and the coupling influence on that location. They are derived by choosing a band of width β around the stable equilibria of $u' = H(u)$ and comparing the strength of attraction with the maximum possible strength of the coupling toward the opposite equalibrium. These conditions

guarantee that the coupling strength, c, is not strong enough to force a uniform profile across the entire line for any set M, where

$$u(x,0) \in \begin{cases} (-1-\beta,-1+\beta) & \text{for} \quad x \in M, \\ (1-\beta,1+\beta) & \text{for} \quad x \in M^c. \end{cases}$$

We can now state the following theorem.

Theorem 1 (existence). *Define δ and K as above, and assume that H and D satisfy the conditions described above. Also assume that (13) holds. If $-\delta + 2cK < 0$, then there exists a solution, $U(x)$, satisfying*

$$0 = H(U(x)) + c\int_{\mathbb{R}} J(x-y)D(U(y)-U(x))dy \tag{14}$$

such that

$$\begin{aligned} U(x) &\in (1-\beta,1+\beta) \quad \text{when } x \in M, \\ U(x) &\in (-1-\beta,-1+\beta) \quad \text{when } x \in M^c. \end{aligned} \tag{15}$$

Proof. Let

$$B = \left\{ U(x) \;\middle|\; \begin{array}{rcl} U(x) \in (1-\beta,1+\beta) & \text{when} & x \in M \\ U(x) \in (-1-\beta,-1+\beta) & \text{when} & x \in M^c \end{array} \right\}$$

and define the map

$$TU(x) = U(x) + \epsilon\left[H(U(x)) + c\int_{\mathbb{R}} J(x-y)D(U(y)-U(x))dy\right]. \tag{16}$$

For ϵ sufficiently small, the conditions (13) guarantee that $T : B \to B$. Our method of proof is to show that T is a contraction mapping. This allows us to conclude that there is a solution of the type (15) that satisfies (14).

To simplify expressions, define the function

$$A(U,x) = \int_{\mathbb{R}} J(x-y)D(U(y)-U(x))dy.$$

Let $U, V \in B$. We write

$$\|TU - TV\|_\infty = \|U - V + \epsilon(H(U) - H(V) + A(U,x) - A(V,x))\|_\infty.$$

The quantity $A(U,x) - A(V,x)$ can be written

$$\int_{\mathbb{R}} J(x-y)[D(U(y)-U(x)) - D(V(y)-V(x))]dy.$$

Because D is continuously differentiable and K is a finite number, we have that

$$|D(g-h)| < K|g-h|$$

for $g, h \in \mathbb{R}$. We have the inequality

$$\begin{aligned}\|A(U,x) - A(V,x)\|_\infty &\leq \left\| cK \int_{\mathbb{R}} J(x-y)(U(y) - U(x) - V(y) + V(x))dy \right\|_\infty \\ &= \left\| -cK(U(x) - V(x)) \int_{\mathbb{R}} J(x-y) + cK \int_{\mathbb{R}} J(x-y)(U(y) - V(y))dy \right\|_\infty \\ &\leq cK \int_{\mathbb{R}} J(x-y)dy \|U(x) - V(x)\|_\infty + cK \int_{\mathbb{R}} J(x-y)dy \|U(x) - V(x)\|_\infty \\ &= \left(2cK \int_{\mathbb{R}} J(x-y)dy \right) \|U(x) - V(x)\|_\infty.\end{aligned}$$

We may write

$$\|TU - TV\| \leq \|U(x) - V(x) + \epsilon(H(U(x)) - H(V(x)))\| + 2\epsilon cK\|U(x) - V(x)\|. \tag{17}$$

Since $H'(U) > -\delta$ for $u \in (1-\beta, 1+\beta)$ we have that

$$H(U(x)) - H(V(x)) \leq -\delta(U(x) - V(x)) \leq 0$$

for x such that $U(x) > V(x)$ and

$$0 \leq H(U(x)) - H(V(x)) \leq -\delta(U(x) - V(x))$$

for x such that $U(x) < V(x)$. Since the value of $H(U(x)) - H(V(x))$ has the opposite sign as $U(x) - V(x)$ it follows that

$$\|U(x) - V(x) + \epsilon(H(U(x)) - H(V(x)))\|_\infty \leq \|U(x) - V(x) - \epsilon\delta(U(x) - V(x))\|_\infty. \tag{18}$$

We may choose ϵ small enough so that $\epsilon\delta < 1$. Substituting (18) into the right-hand side of (17) gives

$$\begin{aligned}\|TU - TV\|_\infty &\leq (1 - \epsilon\delta)\|U(x) - V(x)\|_\infty + 2\epsilon cK\|U(x) - V(x)\|_\infty, \\ \|TU - TV\|_\infty &\leq (1 - \epsilon(\delta - 2cK))\|U(x) - V(x)\|_\infty.\end{aligned}$$

Thus, if $\delta > 2cK$, then T is a contraction mapping, and hence there is a steady state solution to (10).

The quantity $-\delta + 2cK$ is a comparison of the strength of the attraction to the stable fixed points of the function H to the strength of the coupling. If δ is sufficiently large, then the attraction to the stable roots of H is strong enough to counter the coupling. On the other hand, for strong enough coupling, the network will assume a more uniform profile and the solutions will not remain in bands of width β around the fixed points for the uncoupled system.

It is important to note that Theorem 1 provides sufficient conditions for the existence of these solutions; however, they are not, in general, necessary. The Lipschitz condition used to bound the effect of the coupling can be somewhat generous, given that the solutions lie in the β bands.

We remark that the existence proof is virtually identical to that in [2]; our diffusion function is somewhat different, but once we impose the Lipschitz condition, the proof proceeds identically.

3.4. Stability. We now prove stability of the above patterns with the same hypotheses as needed for their existence.

Theorem 2 (linear stability). *Assume that for a set M, a number β, and a parameter c the conditions in Theorem 1 are satisfied. The resulting stationary solution is linearly stable.*

Proof. Let $U(x)$ be the steady state solution to (10). Linearizing around the steady state, we obtain

$$\partial_t w(x,t) = H'(U(x))w(x,t) + c\int_{\mathbb{R}} J(x-y)D'(U(y)-U(x))[w(y,t)-w(x,t)]dy. \tag{19}$$

We rewrite (19) as

$$\partial_t w(x,t) = f(x)w(x,t) + c\int_{\mathbb{R}} J(x-y)D'(U(y)-U(x))w(y,t)\ dy,$$

where

$$f(x) = H'(U(x)) + c\int_{\mathbb{R}} J(x-y)D'(U(x)-U(y))\ dy.$$

We know that $H'(U(x)) < -\delta$ for all x, and D is Lipschitz with constant K. Thus, $f(x) < -\delta + cK \equiv -b$ for all x, and if c is small enough, then $-b < 0$. Taking absolute values, we see that $|w(x,t)|$ is less than $v(x,t)$, where $v(x,t)$ is nonnegative and satisfies

$$v_t = -f(x)v(x,t) + cK\int_{\mathbb{R}} J(x-y)v(y,t)\ dy.$$

Here we have used again the fact that D is Lipschitz and also that $J(x-y)$ is nonnegative. The right-hand linear operator clearly preserves nonnegativity. Consider the following equation:

$$z_t = -bz(x,t) + cK\int_{\mathbb{R}} J(x-y)z(y,t)\ dy. \tag{20}$$

Solutions to this problem can be found with Fourier transforms. Since $J(x) \geq 0$ and is symmetric and integrable, the Fourier transform, $\hat{J}$, exists, is real, and is less than or equal to $\hat{J}(0) = 1$. Solutions to (20) have the form $z(x,t) = \exp(\lambda t + i\ell x)$, where ℓ is real and

$$\lambda = -b + cK\hat{J}(\ell).$$

Thus, if $b > cK$, then all solutions to (20) decay to zero. We claim that if $z(x,0) = v(x,0) > 0$, then $z(x,t) > v(x,t)$ for all positive t. (We note that if $z(x_0,0) = v(x_0,0) = 0$ for some value of x_0, then both z_t and v_t will be positive due to the positive convolution term, so that both $z(x_0,t)$ and $v(x_0,t)$ will be positive for any positive t. Thus, we assume that the initial data are strictly positive.) If we can prove that $z > v$, then, since z decays to zero, so does $v(x,t)$ and thus so do all solutions to (19). Since $f(x) < -b$ for all x, it is clear that at $t = 0$, $v_t(x,0) < z_t(x,0)$ so that up to some time τ, $v(x,t) < z(x,t)$. Suppose at $t = \tau$ there is an x_0 with $v(x_0,\tau) = z(x_0,\tau)$. At that point of intersection, we must have $z_t(x_0,\tau) < v_t(x_0,\tau)$. But, since $z(x,\tau) \geq v(x,\tau)$ for all x,

$$\int_{\mathbb{R}} J(x-y)z(y,\tau)\ dy \geq \int_{\mathbb{R}} J(x-y)v(y,\tau)\ dy$$

and $-bz(x_0,\tau) > f(x_0)v(x_0,\tau)$, so from the evolution equations, $z_t(x_0,\tau) > v_t(x_0,\tau)$, a contradiction. Recalling that $b = \delta - cK$, we see that a sufficient condition for stability is that $-\delta + 2cK < 0$.

As we noted at the end of the existence proof, the quantity $-\delta + 2cK$ measures the competition between the attraction to the two different stable states (odd/even cycles of the stimulus) and the coupling which attempts to make the network uniform. We can use this result to motivate a more general result that is much more akin to the simulations in section 2. The parameter δ is related to the attractivity of the uncoupled states. Suppose that instead of the scalar function H we return to the general uncoupled periodically driven system:

$$\frac{du}{dt} = F(u,t),$$

where F is a general vector function which is T-periodic in its second argument. Suppose that $u_0(t)$ is a stable $2T$-periodic function representing a 1:2 locked solution. Since it is stable, all the Floquet exponents have negative real parts. Let $-\delta$ denote the real part of the exponent closest to zero real part. Then, we conjecture that there are regions locked to the even and odd cycles which persist as long as $\delta > cK$, where c is the coupling strength and K is a coupling function dependent value.

3.5. Traveling waves. In the previous sections, we proved the existence and linear stability of stationary solutions with discontinuities, as long as the coupling between cells is sufficiently weak. The experiments and our model show movement of the lines which indicates that the coupling strength must be fairly large. We can view the theory laid out in the last two subsections as being the mechanism for setting up the domains and assuring that they are stable. In this section we explore what happens in the simple model when the coupling increases.

Once again, reconsider (10), which we repeat here:

$$u_t(x,t) = H(u(x,t) + c\int_{\mathbb{R}} J(x-y)D(u(y,t) - u(x,t))dy.$$

We will relax our assumption that $J(x)$ is symmetric. We first prove the following proposition.

Proposition 1. *Suppose that $H(\bar{u}) = 0$, $D(0) = 0$, $D'(0) > 0$, $J(x) \geq 0$, $c > 0$, and $\int_{\mathbb{R}} J(y)\, dy = 1$. Then $u(x,t) = \bar{u}$ is a linearly stable solution to (10).*

Proof. Clearly, $\bar{u}$ is a solution to this equation. Linearizing about $\bar{u}$, we see that the linearized equation satisfies

$$v_t = -\delta v - cD'(0) + cD'(0)\int_{\mathbb{R}} J(x-y)v(y,t)\, dy.$$

Here $-\delta = H'(\bar{u})$. Solutions to this equation have the form

$$v(x,t) = e^{\lambda t}e^{ikx}.$$

The eigenvalue λ satisfies

$$\lambda = -\delta + cD'(0)[\hat{J}(k) - 1],$$

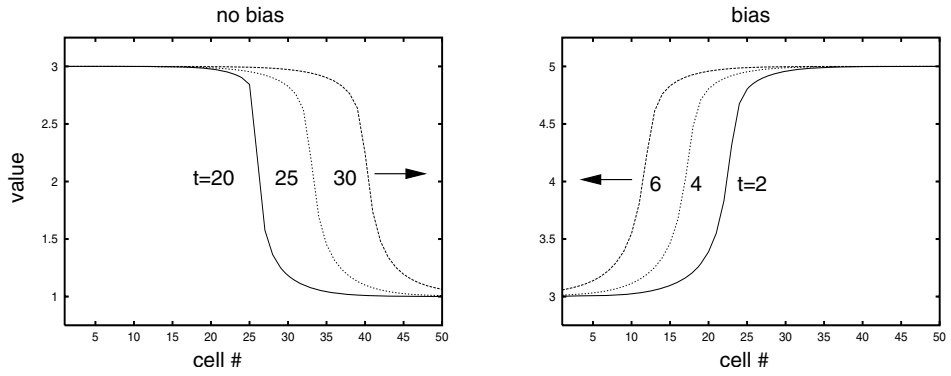


Figure 9. *Traveling wave simulation for* (10) *with* $H(u) = \sin \pi u$, $D(u) = \sin \pi u/2 + 0.25(1 - \cos \pi u/2)$, *and* $J(x) = \exp(-|x + p|)$ *on a grid of* $N = 50$ *cells. Coupling strength* $c = 6$. *Left:* $p = 0$, *no bias, shows the profiles in which the state* 3 *(equivalent to* -1 *on the periodic domain* $[0, 4)$*) takes over. Right: bias* $p = 0.25$ *leads to a wave traveling in the opposite direction.*

where $\hat{J}(k)$ is the Fourier transform of $J(x)$. Since $J(x) \geq 0$, $\Re\hat{J}(k) \leq \hat{J}(0) = 1$, so that

$$\Re\lambda < 0$$

for all k as long as $cD'(0) > 0$ and $\delta > 0$. Hence the constant state is stable.

Our assumptions on H presume that there are two stable constant solutions; thus, we expect that if the coupling is sufficiently strong, there will be a wave front switching from one state to the other. Indeed, if $D(u) = u$ was linear, then the existence, uniqueness, and stability of wave fronts would follow from a theorem of Chen [5]. Figure 9 shows an example simulation in which the -1 state takes over the $+1$ state when the coupling is large enough. This movement is due to a mechanism akin to that seen in Figure 5 for the nonbiased case. The interaction $J(x)$ is symmetric, but the function $D(u)$ is not a purely odd function. We can reverse the tendency for the $+1$ state to take over by allowing for asymmetric coupling. The right-hand side of Figure 9 shows a wave in which -1 takes over by making the coupling stronger from the right.

To gain some insight into the transition from stationary states to traveling waves, we use AUTO to explore the existence and stability of the stationary state as a function of the coupling strength. We consider $N = 20$ cells with nearest neighbor coupling:

$$u_j' = H(u) + c[(1 + s)D(u_{j-1} - u_j) + (1 - s)D(u_{j+1} - u_j)], \quad j = 1, \ldots, 20. \tag{21}$$

We set $u_0 = u_1$ and $u_{21} = u_{20}$ as the boundary conditions and choose functions $H(u) = \sin \pi u - 0.1 \cos \pi u$ and $D(u) = \sin \pi u/2 + 0.25(1 - \cos \pi u/2)$. We have added the cosine terms to make sure that the results do not depend on the oddness of the functions H, D. The parameter c is the overall coupling strength, and s is the measure of asymmetry. We choose $s = 0.1$ so that the left cell has a greater effect than the right cell. When $c = 0$, there are 2^N stable fixed points corresponding to each cell taking a value of either of the two stable roots to $H(u) = 0$ on the interval $[0, 4)$. (Since the cosine term for H is small, these roots are close to 1 and $3 = -1$.) Figure 10A shows the how the steady states evolve for

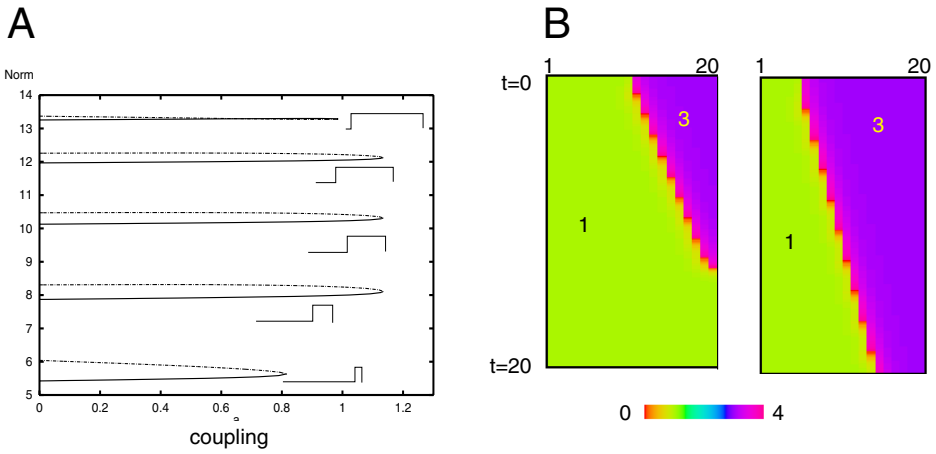


Figure 10. (A) *Bifurcation diagram for* (21) *for five different equilibrium states as the coupling, c, increases. Equilibria when* $c = 0$ *are shown to the right corresponding to cells in the* "1" *or* "3" *states.* (B) *Evolution of traveling waves for coupling larger than the critical value. Time increases downward, and cell index is horizontal. Coupling is stronger from the left so that the left state takes over the right. Left figure is for* $c = 1.3$ *such that the initial state has half the cells in the* "1" *state and the other half in the* "3" *state. Right panel is the same, but only the first five cells are in the* "1" *state.*

five different configurations. In each of the cases the first m cells are in the state near 1, and the remaining are in the state near 3. In all cases, as the coupling increases there is little quantitative change in the equilibria, but at a critical value of the coupling, the fixed point is lost at a saddle-node. If the distribution of cells is very unbalanced (say, only one cell in state 1 and the rest in state 3), then the amount of coupling required to destroy the steady state is small compared to the more balanced case. An analogue of this is seen in Figure 8C,D, where the thin region between two similar domains is quickly absorbed. Figure 10B shows the evolution of the stationary state when the coupling is larger than the saddle-node value. The left figure shows that in the balanced case, as expected from the coupling bias, the "1" state takes over the "3" state. Even when there is a 3:1 advantage of "3" state, the bias is able to overcome this, and a traveling wave results. When only the leftmost cell is in the "1" state and the other cells are in the "3" state, then the wave travels to the left—the bias is unable to overcome the huge unbalance.

4. Discussion. We have derived a simple model for complex visual effects due to the direct stimulation of the eyeball. The model is based on the simple idea that if there is phase-locking between a stimulus which fires twice for every response of the cell, then there is a natural bistability in the medium: firing on the odd or even cycles of the stimuli. We have suggested that visual stimuli are able to push the phase-locked retinal cells into different basins of attraction and that the boundaries between the bistable domains account for the thin line illusions. This notion allows us to make several predictions. First, the thickness of the boundaries should be independent of any stimulus properties—the "boundary layer" is, rather, a function of the intrinsic coupling between retinal cells involved in the illusion. Stronger coupling should lead to thicker boundaries. Second, more complex patterns should be easy to create. For example, a visual stimulus consisting of an expanding annulus should

result in a series of concentric circular patterns which will gradually disperse. Third, the illusion should disappear instantly when the electrical stimulus is cut off, since the cells will no longer fire and there will thus be no phase boundary. We have not completely explored how differently shaped electrical stimuli will affect the patterns. However, one important prediction that we can make is that if the electrical stimulation is at a sufficiently low frequency, there will be no illusion, since the retinal cells will be able to fire in a 1:1 manner and so no phase-boundaries can develop. This would allow us to distinguish our model from the model of Willis, which we discuss briefly below.

In addition to a biologically "realistic" model, we abstracted the mechanism to a scalar bistable medium with nonlocal but "weak" coupling. We proved a theorem about the existence and stability of patterned states, and then explored how asymmetric coupling which is sufficiently strong can lead to traveling waves that connect the stable states of the bistable medium, similar to those in [11].

There has been one other attempt to model these curious phosphenes [22]. The Willis model captured the idea of bistability but did not model the slow movement of the phosphenes over time. Willis' model is based on a piecewise linear firing rate model, which proposes that the interactions between two classes of cells (receptors and horizontal cells) switch sign (excitation becomes inhibition and vice versa) during the two phases of electrical excitation. Unless the nonlinearities of the Willis model are exactly balanced, one of the two states should have a larger basin of attraction than the other. Thus, with any coupling at all, in the Willis model, we would expect the phosphenes to have a strong tendency to move in a preferred direction. This is not reported in the experiments. Our model is considerably simpler and makes fewer assumptions about the cells and their connectivity. The two models could be experimentally distinguished by recording from ganglion or horizonal cells during the phosphenes. While this is unlikely, given that the subjects are humans, our model is more compatible with both the physiological data in salamanders and the psychophysical data in humans [7]. Furthermore, the Willis model would predict that the phosphenes persist at lower frequencies, say, 50 Hz, since the bistability is based on a different mechanism from ours. This is easy to test in humans.

While the illusions we have modeled are unusual and a consequence of unnatural stimuli, such experiments can often be used to shed light on basic biological structure and function. For example, by probing the retina with periodic stimuli, we can determine the frequency response of the neurons. The width of the phosphenes tells us something about the degree of intrinsic connectivity between retinal cells. Finally, there is a good deal of recent interest in direct stimulation of the nervous system as a means of providing visual prosthesis [14]. Thus, models such as the present one can provide a simple framework for exploring how direct electrical stimuli are visually interpreted.

REFERENCES

[1] D. Attwell, *Interaction of low frequency electric fields with the nervous system: The retina as a model system*, Radiat. Prot. Dosimetry, 106 (2003), pp. 341–348.

[2] P. Bates and A. Chmaj, *An integrodifferential model for phase transitions: Stationary solutions in higher space dimensions*, J. Statist. Phys., 95 (1999), pp. 1119–1139.

[3] G. S. Brindley, *The site of electrical excitation of the human eye*, J. Physiol., 127 (1955), pp. 189–200.
[4] R. H. S. Carpenter, *Contour-like phosphenes from electrical stimulation of the human eye: Some new observations*, J. Physiol., 229 (1973), pp. 767–785.
[5] X. Chen, *Existence, uniqueness, and asymptotic stability of traveling waves in nonlocal evolution equations*, Adv. Differential Equations, 2 (1997), pp. 125–160.
[6] J. Claussen, *Visual sensations (phosphenes) produced by AC sine wave stimulation*, Acta. Psychiatr. Neurol. Scand. Suppl., 94, pp. 1–10.
[7] D. Crevier and M. Meister, *Synchronous period-doubling in flicker vision of salamander and man*, J. Neurophysiol., 79 (1998), pp. 1869–1878.
[8] P. Dayan and L. Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, MA, 2001.
[9] J. Drover, *The Interplay of Intrinsic Dynamics and Coupling in Spatially Distributed Neuronal Networks*, Ph.D. thesis, Mathematics Department, University of Pittsburgh, Pittsburgh, PA, 2005; available online at http://etd.library.pitt.edu/ETD/available/etd-08162005-135431.
[10] G. B. Ermentrout, *n:m phase-locking of weakly coupled oscillators*, J. Math. Biol., 12 (1981), pp. 327–342.
[11] B. Ermentrout, X. Chen, and Z. Chen, *Transition fronts and localized structures in bistable reaction-diffusion equations*, Phys. D, 108 (1997), pp. 147–167.
[12] S. I. Firth, C.-T. Wang, and M. B. Feller, *Retinal waves: Mechanisms and function in visual system development*, Cell Calcium, 37 (2005), pp. 425–432.
[13] J. R. Gibson, M. Beierlein, and B. W. Connors, *Functional properties of electrical synapses between inhibitory interneurons of neocortical layer* 4, J. Neurophysiol., 93 (2005), pp. 467–480.
[14] L. E. Hallum, G. J. Suaning, and N. H. Lovell, *Contribution to the theory of prosthetic vision*, ASAIO J., 50 (2004), pp. 392–396.
[15] S. Hidaka, T. Kato, and Y. Hashimoto, *Structural and functional properties of homologous electrical synapses between retinal amacrine cells*, J. Integr. Neurosci., 4 (2005), pp. 313–340.
[16] E. H. Hu and S. A. Bloomfield, *Gap junctional coupling underlies the short-latency spike synchrony of retinal alpha ganglion cells*, J. Neurosci., 23 (2003), pp. 6768–6777.
[17] G. T. Kenyon and D. W. Marshak, *Gap junctions with amacrine cells provide a feedback pathway for ganglion cells within the retina*, Proc. Biol. Sci., 265 (1998), pp. 919–925.
[18] N. Kopell and G. B. Ermentrout, *Symmetry and phaselocking in chains of weakly coupled oscillators*, Comm. Pure Appl. Math., 39 (1986), pp. 623–660.
[19] T. J. Lewis and J. Rinzel, *Self-organized synchronous oscillations in a network of excitable cells coupled by gap junctions*, Network, 11 (2000), pp. 299–320.
[20] F. Ratliff, *Mach Bands: Quantitative Studies on Neural Networks of the Retina*, Holden-Day, San Francisco, 1965.
[21] C. W. Tyler, *Some new entoptic phenomena*, Vision Res., 18 (1978), pp. 1633–1639.
[22] J. B. Willis, *The modeling of neural circuitry*, Progr. Neurobiol., 26 (1986), pp. 93–118.
[23] M. Wilms, M. Eger, T. Schanze, and R. Eckhorn, *Visual resolution with epi-retinal electrical stimulation estimated from activation profiles in cat visual cortex*, Vis. Neurosci., 20 (2003), pp. 543–555.
[24] H. R. Wilson, *Computational evidence for a rivalry hierarchy in vision*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 14499–14503.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 552–574

Existence and Wandering of Bumps in a Spiking Neural Network Model*

Carson C. Chow† and S. Coombes‡

Abstract. We study spatially localized states of a spiking neuronal network populated by a pulse-coupled phase oscillator known as the lighthouse model. We show that in the limit of slow synaptic interactions in the continuum limit the dynamics reduce to those of the standard Amari model. For nonslow synaptic connections we are able to go beyond the standard firing rate analysis of localized solutions, allowing us to explicitly construct a family of coexisting one-bump solutions and then track bump width and firing pattern as a function of system parameters. We also present an analysis of the model on a discrete lattice. We show that multiple width bump states can coexist, and uncover a mechanism for bump *wandering* linked to the speed of synaptic processing. Moreover, beyond a wandering transition point we show that the bump undergoes an effective random walk with a diffusion coefficient that scales exponentially with the rate of synaptic processing and linearly with the lattice spacing.

Key words. spiking neural network, bump solutions, working memory, lighthouse model

AMS subject classification. 92C20

DOI. 10.1137/060654347

1. Introduction.

A goal of theoretical neuroscience is to develop a tractable model of a spiking neuronal network. This must necessarily involve a single cell model, capable of generating spikes of activity (so-called action-potentials), that when connected into a synaptic network can generate the rich repertoire of behavior seen in a real nervous system. For all of the popular conductance-based single neuron models, and also the simpler integrate-and-fire (IF) variety, a full understanding of network dynamics has proved elusive. In essence this is because we have not yet developed an appropriate mathematical framework for understanding the neurodynamics of spiking networks. To date, progress in this area has been restricted to firing rate neural models [1, 2, 3, 4], which cannot adequately capture known spike-train correlations, or the analysis of phase-locked states and instabilities of homogeneous steady states in spiking IF networks [5, 6, 7, 8, 9, 10].

The lighthouse model of Haken is a candidate single neuron model that falls between spiking neuron models and firing rate descriptions (see [11] and references therein). We show

*Received by the editors March 15, 2006; accepted for publication (in revised form) by D. Terman May 23, 2006; published electronically November 3, 2006. This work was supported in part by the Intramural Research Program of the NIH, NIDDK (CCC). This work was performed by an employee of the U. S. Government or under U. S. Government contract. The U. S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U. S. Government purposes.

http://www.siam.org/journals/siads/5-4/65434.html

†Laboratory of Biological Modeling, NIDDK, National Institutes of Health, Bethesda, MD 20892 (carsonc@mail.nih.gov).

‡Department of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK (stephen.coombes@nottingham.ac.uk). This author receives continuing support from the EPSRC via the award of an Advanced Research Fellowship (GR/S60914/01).

that in the limit of slow synaptic interactions it reduces to the classic Wilson–Cowan and Amari firing rate models. For fast synaptic interactions, it displays some of the complex properties that have been observed in simulations of spiking neuronal networks. Importantly the lighthouse model is sufficiently simple that it may also be analyzed at the network level even for fast synaptic responses. Hence, a detailed investigation into the network dynamics of the lighthouse model may pave the way to the development of a specific soluble spiking neurodynamics.

With this in mind we turn our attention to spatially localized bumps of persistent activity, which have been linked to working memory (the temporary storage of information within the brain) [12, 13, 14]. In many models of working memory, transient stimuli are encoded by feature-selective persistent neural activity. Such stimuli are imagined to induce the formation of a spatially localized bump of persistent activity which coexists with a stable uniform state (with low firing rate). The most popular mathematical formulations of such models assume long-range inhibition with local recurrent excitation and invoke a population level description in terms of a rate model (see, for example, [15, 16]). Although interesting in their own right and studied by a number of authors (surveyed in [17]), such models are useful only for describing systems with slow synaptic interactions.

The study of localized states in fully spiking network models with fast synaptic interactions has typically been possible only with the aid of numerical simulation. For example, the work of Laing and Chow [18] on IF networks shows a number of interesting behaviors that would be absent in a firing rate study. These include the observation that (i) a bump solution can exist in a spiking network (coexisting with the quiescent state), provided that the neurons fire asynchronously within the bump, and (ii) a bump can lose stability with increasing rate of synaptic processing to a wandering bump or a traveling wave. Wandering was also observed in a model of working memory that had populations of excitatory and inhibitory spiking neurons [19].

It is precisely these observations that we revisit in this paper within the context of the lighthouse model of a spiking neural network. Not only does the lighthouse model support behavior similar to that seen in the IF network, it is sufficiently simple to admit further mathematical analysis. In this regard it is a minimal model of a spiking neural network. Although to date only patterns of synchronized activity have been analyzed [11] it is clearly a model that may be studied further in more interesting scenarios, such as the one we now present. Importantly this allows us to uncover the mechanism that governs the transition from a stable to a wandering bump.

In the next section we introduce the lighthouse model and survey some of its known properties. For simplicity we first study the model in the continuum limit, which allows us to more easily show its correspondence with the Amari model in the limit of slow synaptic processing. Next we introduce the definitions of a bump, a stationary bump, and an equiperiod bump so that we may carefully formulate appropriate existence conditions for bumps and wandering bumps. We then show how one may analyze spatially localized states of the continuum lighthouse model for nonslow synaptic processing and explicitly construct a family of spiking bump solutions. We follow the continuum description with a treatment of dynamics on a lattice, including bump wandering. Finally, we discuss natural extensions of the work in this paper.

2. The lighthouse model. The lighthouse model was introduced by Haken as a pulse-coupled model of a spiking neural network whose behavior could be explicitly analyzed at the network level [20, 21]. Much is now known about phase-locked solutions and the effect of noise and delays on stability, summarized in the book by Haken [11]. As we will show, the lighthouse model exhibits a transition to wandering like the IF model but is easier to analyze. Here we briefly outline the lighthouse model before presenting our new results on bump solutions.

The lighthouse model is a hybrid between a phase-oscillator and a firing rate model. Indexing a lighthouse neuron with $i \in \mathbb{Z}$, the dynamical description of a discrete network is cast as a system of ordinary differential equations for the phase variable $\theta_i = \theta_i(t) \in [0, 1)$, $t \in \mathbb{R}^+$, in the form

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = H(u_i - h), \tag{2.1}$$

with

$$u_i(t) = \sum_{j=1}^{N} w_{ij} E_j(t), \qquad E_i(t) = \sum_m \eta(t - T_i^m), \tag{2.2}$$

and $H(\cdot)$ is interpreted as a firing rate function, with threshold h. Haken uses the Naka–Rushton sigmoidal function (though many of the analytical results for the lighthouse model are obtained for a linear firing rate function). Throughout this paper we work with the Heaviside step function such that $H(x) = 1$ for $x \geq 0$ and is zero otherwise. The quantity u_i is recognized as the input to neuron i, mediated by the synaptic processing of a set of incoming spikes. The spike time T_j^m marks the mth firing of neuron j, given by the times for which $\theta_j(t) = 1$. Each time neuron i fires, a synaptic pulse $\eta(t)$ (with $\eta(t) = 0$ for $t < 0$) is transmitted to other neurons $j = 1, \ldots, N$ with coupling weight w_{ij}. We shall take the synaptic response to be normalized, such that $\int_0^\infty \eta(t)\mathrm{d}t = 1$. When the input to a neuron exceeds the threshold, h, the phase will advance towards threshold at a constant rate, which we have arbitrarily set to 1. We note that the activity of a given neuron need not be periodic. Depending on the nature of the inputs, it could also be quiescent or fire irregularly.

We consider two versions of the lighthouse model. In the first version, which is the model originally proposed by Haken, whenever the input is below threshold or drops below threshold, the phase immediately resets to zero. In this case the neuron can fire only if it receives input that remains above threshold for an entire period. In the second version the phase is not reset, except when the neuron fires. We will see that the two versions have different possible bump solutions and dynamics. In particular, the instant reset version does not have wandering bumps.

We consider lateral inhibition-type synaptic coupling on a one-dimensional lattice. The system has two length scales, the lattice spacing and the coupling footprint. We take the array length to be much longer than either of these two length scales. We can index the N neurons in the array by a distance via $x = i\delta x$, where δx is the lattice spacing. The continuum limit is given by $N \to \infty$, $\delta x \to 0$ with $\delta x N$ a constant. To make the connection to the continuum limit transparent we rewrite the coupling function as $w_{ij} = w([i - j]\delta x)\delta x$. A continuum

version of the lighthouse model (along the whole real line) is obtained under the replacement $(\theta_i, u_i, T_i^m, w((i-j)\delta x)) \to (\theta(x,t), u(x,t), T^m(x), w(x-y))$, $x \in \mathbb{R}$, and is governed by the dynamics

$$\frac{\partial \theta(x,t)}{\partial t} = H(u(x,t) - h), \tag{2.3}$$

with

$$u(x,t) = \sum_m \int_0^\infty \mathrm{d}s\eta(s) \int_{-\infty}^\infty \mathrm{d}y w(x-y)\delta(s - t + T^m(y)). \tag{2.4}$$

Here the function $w(x)$ plays the role of the connectivity kernel, which we shall assume to be homogeneous so that $w(x) = w(|x|)$. For concreteness we will use the explicit functions $\eta(t) = \alpha \exp(-\alpha t)H(t)$ and the "wizard hat" $w(x) = A\exp(-a|x|) - \exp(-|x|)$.

2.1. Connection to the Amari model. We first consider the continuum limit for slow synapses and show that it is identical to the classical Amari model. Taking the time average of (2.4) gives

$$\begin{aligned}\langle u(x,t)\rangle &= \frac{1}{T}\int_t^{t+T} \mathrm{d}s \int_{-\infty}^\infty \mathrm{d}y w(x-y) \int_0^\infty \mathrm{d}s'\eta(s') \sum_m \delta(s' - s + T^m(y)) \\ &= \int_{-\infty}^\infty \mathrm{d}y w(x-y) \int_0^\infty \mathrm{d}s\eta(s)\Omega(y, t-s), \end{aligned} \tag{2.5}$$

where $\Omega(x,t) = M(x,t)/T$ is the firing rate and $M(x,t) = \int_t^{T+t} \sum_m \delta(s + T^m(x))\mathrm{d}s$ is the number of firing events at position x within a window T.

We next recognize that the firing rate is given by

$$\Omega(x,t) = \frac{\partial \theta(x,t)}{\partial t} = H(u(x,t) - h). \tag{2.6}$$

Substituting (2.6) into (2.5) gives

$$\langle u(x,t)\rangle = \int_{-\infty}^t \mathrm{d}s\eta(t-s) \int_{-\infty}^\infty \mathrm{d}y w(x-y)H(u(y,s) - h). \tag{2.7}$$

For slow synaptic interactions (such that $\eta(t)$ is approximately constant over the timescale T), then we have the condition $u(x,t) = \langle u(x,t)\rangle$. For the specific choice $\eta(t) = \alpha \mathrm{e}^{-\alpha t}H(t)$, (2.7) may be written in the form

$$\frac{1}{\alpha}\frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_{-\infty}^\infty \mathrm{d}y w(x-y)H(u(y,t) - h), \tag{2.8}$$

which is recognized as the well-known Amari model [22]. The equivalence between the lighthouse model and the Amari equation is only strictly true in the limit of $\alpha \to 0$, where $u(x)$ is constant in time. For nonzero α we will see that the dynamics between the two models deviate substantially.

3. Bump solutions. For a threshold $h > 0$ neurons will not fire spontaneously. However, for a lateral-inhibition-type coupling function, a group of neurons could mutually excite each other to sustain activity. Indeed for this choice of interaction function the Amari model given by (2.8) is known to support spatially localized solutions, commonly referred to as *bumps*. Moreover, their stability is insensitive to variation in the synaptic rate parameter α. It is therefore of interest to ascertain whether the analogue of such solutions can be found in the full spiking lighthouse network, and whether they are also stable to variations in α. As we shall see, the answers to these questions are "yes" and "no," respectively.

Definitions. We are interested in analyzing the existence and dynamics of localized pulses of activity or bumps. A bump is defined as a group of contiguous neurons that fire together. However, given the discontinuous nature of spiking events and that the bump location may change in time, it is necessary to make this idea more precise with the following definitions.

Definition 1. *A* bump *at time t is the set of all contiguous neurons that fired in the time interval $[t-1, t]$.*

Definition 2. *A* stationary bump *is a bump which persists in a single location for all time; i.e., the same set of neurons continue to fire indefinitely.*

Definition 3. *An* equiperiod bump *is a stationary bump for which the neurons in the bump all fire with the same rate.*

On a lattice any of the above states are invariant to shifts by arbitrary lattice spacings. As we will demonstrate, the position or the width of the bump need not be constant in time.

We first consider conditions for the existence of a continuum equiperiod bump in the limit of infinitely slow decaying synapses ($\alpha \to 0$). In this case each neuron in the bump receives constant above threshold input, and thus all neurons within the bump will fire with rate 1. All neurons outside of the bump receive below threshold input and never fire. We suppose that a bump exists on the domain $L = [x_1, x_2]$. This implies that $u(x,t)$ is above threshold on L, below threshold outside of L, and equal to the threshold at $x = x_1$ and x_2. From (2.4)

$$u(x,t) = \int_{x_1}^{x_2} w(x-y)P(t-\psi(y))\mathrm{d}y, \tag{3.1}$$

where $P(t) = \sum_m \eta(t-m)$ is a periodic function (with period 1) and $\psi(y)$ is the location-dependent phase. If $w(x)$ is symmetric, we need to consider only one edge. For infinitely slow synapses, $P(t) = 1$. This leads to the Amari existence condition

$$h = \int_0^{\Delta} w(y)\mathrm{d}y, \qquad \Delta = x_2 - x_1. \tag{3.2}$$

We note that in the Amari limit the bump solution is insensitive to the phases of the neurons. For a standard lateral-inhibition-type coupling there are two bump solutions that arise in a saddle-node bifurcation [3]. Importantly, it is possible to construct a class of bump solutions for the full continuum lighthouse model, and thus take a step beyond the standard Amari-style analysis of localized states.

3.1. Continuum bump solutions in the lighthouse model. For the original lighthouse model with instant phase reset whenever activity drops below threshold, an equiperiod bump

is expected since neurons within the bump will fire with rate one and those outside of the bump will not fire. For the second version without the phase reset condition, phases may ride on long subthreshold plateaus. In this case the generic solution is a spatially localized solution with an "interior" and "exterior" region. Within the interior, neurons fire with rate 1. In the exterior, the firing rate falls off to zero as the bump edge is approached. However, the contribution from the exterior region only weakly perturbs the location of the inner edge, so that the solution in the absence of a phase reset may be regarded as a perturbation of that with a phase reset.

Here we focus on the class of (symmetric) maximally firing one-bump solutions (with firing rate unity) for the instant reset model parameterized by

$$T^m(x) = m + [\beta|x| \mod 1], \tag{3.3}$$

such that $u(x,t) \geq h$ for $x \in [x_1, x_2]$ for all t, and $u(x,t) < h$ otherwise. Here we take $x_1 = -\Delta/2 = -x_2$. We are free to choose such a coordinate system by translational invariance of the continuum model. In this case $u(x,t)$ is given by (3.1) with $\psi(y) = \beta|y|$. Introducing the Fourier transform $\widetilde{P}(k) = \int_{-\infty}^{\infty} \mathrm{d}t P(t)\mathrm{e}^{-ikt}$ means that we may write (3.1) as

$$u(x,t) = \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi} \widetilde{P}(k) F(k,x) \mathrm{e}^{ikt}, \tag{3.4}$$

where

$$F(k,x) = \int_{x_1}^{x_2} w(x-y)\mathrm{e}^{-ik\beta|y|}\mathrm{d}y. \tag{3.5}$$

Note that $u(x,t)$ has period 1 (inherited from $P(t)$). It is also continuous if β is nonzero. For slow synapses $P(t)$ may be regarded as a constant so that $\widetilde{P}(k) = 2\pi\delta(k)$. In this case (3.4) becomes time-independent, with $u(x,t) = F(0,x)$. Moreover, the width of the bump is independent of β and is given implicitly by $h = F(0, \Delta/2)$. This solution is identical to the one-bump solution describing the time-independent localized state of the standard Amari model [3]. In general, however, we proceed with the evaluation of (3.4) using the result that

$$\widetilde{P}(k) = 2\pi\widetilde{\eta}(k) \sum_n \delta(k - 2\pi n). \tag{3.6}$$

Substitution into (3.4) gives us the following Fourier series representation for $u(x,t)$:

$$u(x,t) = \sum_n \widetilde{\eta}(2\pi n) F(2\pi n, x) \mathrm{e}^{2\pi i n t}. \tag{3.7}$$

To determine the width of a bump in a self-consistent fashion we must enforce the condition that at the boundaries $x = x_j$ the infimum of u (over t) must equal the threshold h. Hence, the existence condition for a bump in the continuum lighthouse model is simply

$$\inf_t u(\Delta/2, t) = h. \tag{3.8}$$

The bump width Δ is then determined by

$$h = \int_0^{\Delta} w(y)\mathrm{d}y + 2\mathrm{Re}\left(\sum_{n>0} \widetilde{\eta}(2\pi n) g(2\pi n) \mathrm{e}^{2\pi i n t^*}\right), \tag{3.9}$$

where $g(k) = F(k, \Delta/2)$. Here t^* denotes the value of t for which $u(\Delta/2, t)$ has a minimum. The value of t^* is calculated by differentiating (3.7) and then solving

$$0 = \mathrm{Re}\left(i \sum_{n>0} \widetilde{\eta}(2\pi n) g(2\pi n) n \mathrm{e}^{2\pi i n t^*}\right), \tag{3.10}$$

subject to $\partial_{tt} u(\Delta/2, t)|_{t=t^*} > 0$. In some sense we may regard the second term on the right-hand side of (3.9) as a correction to the standard Amari firing rate model description that takes into account the effects of nonslow synaptic processing.

For our choice of temporal and spatial kernels a short calculation gives $\widetilde{\eta}(k) = \alpha/(\alpha + ik)$ and

$$\begin{aligned} g(k) = A &\left\{ \frac{\mathrm{e}^{-ik\beta\Delta/2} - \mathrm{e}^{-a\Delta/2}}{a - ik\beta} + \frac{\mathrm{e}^{-a\Delta/2} - \mathrm{e}^{-a\Delta}\mathrm{e}^{-ik\beta\Delta/2}}{a + ik\beta} \right\} \\ &- \frac{\mathrm{e}^{-ik\beta\Delta/2} - \mathrm{e}^{-\Delta/2}}{1 - ik\beta} + \frac{\mathrm{e}^{-\Delta/2} - \mathrm{e}^{-\Delta}\mathrm{e}^{-ik\beta\Delta/2}}{1 + ik\beta}. \end{aligned} \tag{3.11}$$

In the limit $\alpha \to 0$ the β-independent shape of a bump is given explicitly by

$$u(x) = \begin{cases} W(x_1 - x, x_2 - x), & x < x_1, \\ W(0, x - x_1) + W(0, x_2 - x), & x_1 \le x \le x_2, \\ W(x - x_2, x - x_1), & x > x_2, \end{cases} \tag{3.12}$$

where

$$W(x_a, x_b) = \int_{x_a}^{x_b} w(y)\mathrm{d}y = \frac{A}{a}\left[\mathrm{e}^{-ax_a} - \mathrm{e}^{-ax_b}\right] - \left[\mathrm{e}^{-x_a} - \mathrm{e}^{-x_b}\right], \qquad x_b > x_a > 0, \tag{3.13}$$

and Δ satisfies $h = \int_0^{\Delta} w(y)\mathrm{d}y$. Figure 3.1 shows an example of the shape of $u(x)$ given by (3.12). This is the classic Amari bump. When $u(x) > h$, the neuron fires with rate 1. Of the two branches of solutions to $\Delta = \Delta(h)$, it is known that the branch with largest Δ is stable for all α in the Amari model [3, 23, 24, 25]. To see the effects of a nonzero choice of α in the lighthouse model we need only to calculate remaining terms from the right-hand side of (3.9).

We evaluate (3.9) and (3.10) numerically to determine the pair (Δ, t^*) as a function of system parameters. An example of this is shown in Figure 3.2, where for a fixed (nonzero) value of α we plot $\Delta = \Delta(h)$ for the synchronous solution ($\beta = 0$). As for the standard Amari solution we see two branches of solutions with a saddle-node bifurcation with increasing h. Moreover, taking α as the bifurcation parameter shows that such solutions cannot exist if α is chosen too large, as once again there is a saddle-node bifurcation, seen in the middle panel of Figure 3.2. In the right most panel of Figure 3.2 we effectively combine the information

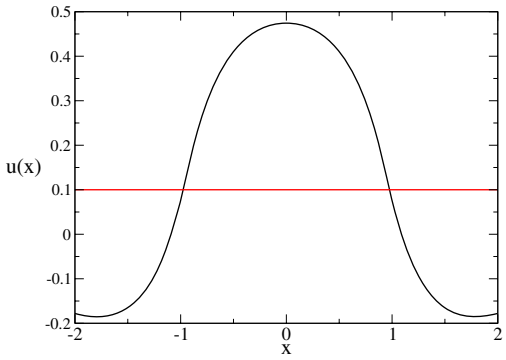


Figure 3.1. *A plot of $u(x)$ of a bump with $\alpha = 0$, $A = a = 2$, $h = 0.1$ given by (3.12). For $u(x) \geq 0.1$, the neuron will fire with rate 1.*

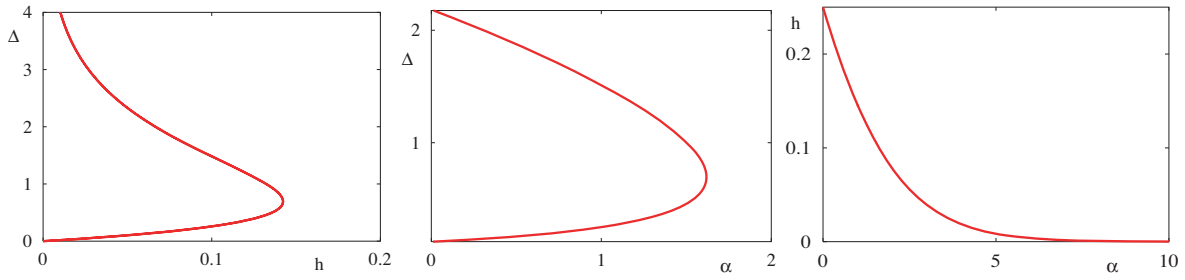


Figure 3.2. *Left: Bump width as a function of h for the synchronous solution ($\beta = 0$) with $\alpha = 1$. Middle: Bump width as a function of α for the synchronous solution with $h = 0.1$. Right: Existence window (below the curve) for the synchronous solution in the (α, h) plane obtained by the numerical continuation of the limit point in the left-hand plot. In all plots $A = a = 2$.*

from the other two panels into an *existence* plot, such that synchronous solutions are found below the curve shown in the (α, h) plane. In Figure 3.3 we show a plot of how the solution with the largest width varies as a function of β. This theoretical curve is compared with direct numerical simulations of the instant rest lighthouse model and found to be in excellent agreement.

We see in Figure 3.3 that as β is increased there is a general trend towards increasing width with a weak oscillation imposed. We now explain this behavior. The function $g(2\pi n)$ has a periodic modulation in β of fundamental period $\beta = 2/\Delta$. Hence, from (3.10), t^* will oscillate as β is varied. From (3.9) we expect a similar modulation of $\Delta = \Delta(\beta)$. Note that when $\beta = 2/\Delta$ the firing times are asynchronous, and the system may be said to be in a so-called splay state. The synchronous state ($\beta = 0$) has an infimum at $t^* = \Delta$, and hence all other nonsynchronous states are such that $t^*_{\beta=0} > t^*_{\beta\neq 0}$, at least avoiding resonances where $2\beta/\Delta = n$. Hence, from (3.7), $\inf_t u(x, t^*_{\beta=0}) < \inf_t u(x, t^*_{\beta\neq 0})$. Since $u(x,t)$ is a decreasing function of x (in the neighborhood of the bump edges where $x = \pm\Delta/2$), the synchronous solution has the smallest width as a consequence of condition (3.8). As β increases, the phases

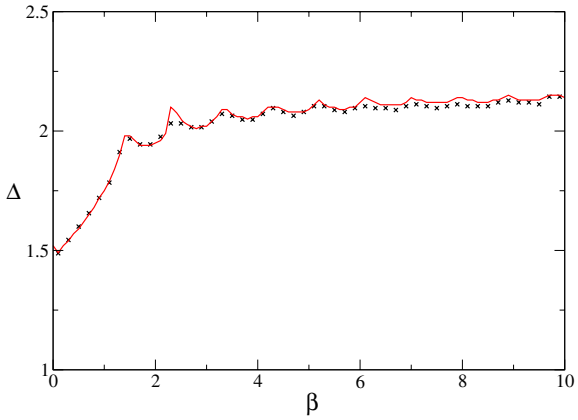


Figure 3.3. *A plot of the theoretical curve (red) for $\Delta = \Delta(\beta)$, with $h = 0.1$ and other parameters as in the leftmost panel of Figure 3.2. Black crosses denote results from direct numerical simulations of the lighthouse model with instant reset.*

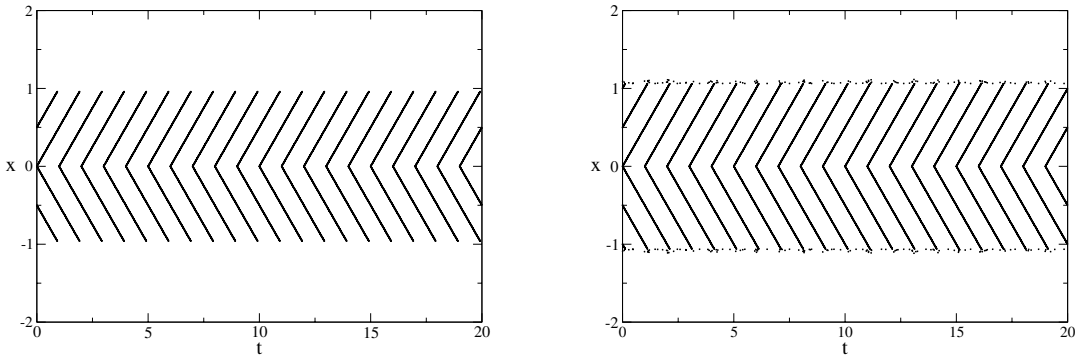


Figure 3.4. *A direct numerical simulation of the continuum model approximated by a finite system with lattice spacing $\delta x = 0.004$ on $N = 1000$ grid points. The dots in the raster plot indicate the times of firing events. Data is shown for the parameters of Figure 3.3 with $\beta = 2$. The left and right panels show results with and without phase reset, respectively.*

become more *splayed* and pass through resonances with the width Δ, ultimately resulting in the local peaks seen in Figure 3.3.

An example of a spiking bump state is shown in Figure 3.4, with the left and right panels showing simulation results with and without phase reset. The interior and exterior regions are clearly visible in the no phase reset model. Moreover, simulations show that with increasing α, bumps can lose stability and begin to wander in exactly the same fashion as seen in earlier work of Laing and Chow on IF networks [18]. Examples of this will be presented in section 3.4. Direct simulation results are obtained by approximating the continuum model as a finite system sampled at a discrete set of grid points, i.e., a lattice model. We now proceed with a

treatment of the lighthouse model dynamics on a lattice.

3.2. Bumps on a lattice: Slow synapses. We first derive the existence conditions of a bump on a discrete lattice. Suppose neuron j fires periodically with period 1 and has a phase $\psi_j \in [0,1)$, i.e., $T_j^m = m + \psi_j$. The periodic synaptic output of neuron j is given by $E_j(t) = P(t - \psi_j)$. In the limit of infinitely slow synapses $P(t - \psi_j) = 1$. Without loss of generality we fix one end of the bump at neuron index 0. The bump is then specified by the set $i \in \{0, \dots, m\}$, where

$$u_i = \sum_{j=0}^{m} w_{ij} \geq h, \qquad i \in \{0, \dots, m\}, \tag{3.14}$$

and $u_i < h$ elsewhere. The phase does not appear in the existence condition, so any configuration of phases is allowed. Thus a family of solutions satisfying condition (3.14) defines an equiperiod bump at a given location. To make this condition more explicit we introduce the following two additional definitions:

Definition 4. *The* edge *of a contiguous set of firing neurons is a neuron in the set that has one neighbor in the set and one neighbor not in the set.*

Definition 5. *The* next edge *of a set of contiguous firing neurons is a neuron that is not in the set but has one neighbor that is an edge.*

Thus the edge neurons $i = 0, m$ must satisfy $u_0 \geq h$, $u_m \geq h$, and the next edge neurons $i = -1, m+1$ must satisfy $u_{-1} < h, u_{m+1} < h$. Unlike the continuum version, the discrete bump must satisfy two independent conditions; i.e., the edge neuron is above threshold and the next edge neuron is below threshold. If w_{ij} is symmetric, the bump will be symmetric, and thus we need to consider only one edge. Following the Amari strategy, we construct two existence functions:

$$\phi_m^{ne} = \sum_{j=0}^{m} w_{-1,j} = \sum_{j=1}^{m+1} w_{0,j}, \tag{3.15}$$

$$\phi_m^{e} = \sum_{j=0}^{m} w_{0,j}. \tag{3.16}$$

A bump of width m will exist for any m for which $\phi_m^{ne} < h \leq \phi_m^e$. In the continuum limit, the existence condition is $\phi_m^{ne} = \phi_m^e = h$.

We now consider conditions on the coupling weight and threshold for which this existence condition can be satisfied. We note that if $\phi_m^{ne} < \phi_m^e$ for some m, then we can choose h to lie between the existence curves. This can be satisfied in a number of ways. One is if $w_{i-1,j} - w_{i,j} < 0$ for $i < j$, i.e., the weight function is monotonically decreasing in $|i - j|$. This can be proved by comparing the terms individually in the sums (3.15) and (3.16). We note that the weight function needs only to decrease monotonically for $|i - j| \leq m + 2$. If w_{ij} is positive at $i = j$ and decreases in $|i - j|$, then ϕ_m^{ne} and ϕ_m^e will increase until w_{ij} becomes negative, whereupon ϕ_m^{ne} and ϕ_m^e will begin to decrease. The existence curves will then be concave functions. Hence, if h is initially greater than the maximum of the existence curves, no bump solutions exist. Bump solutions will appear if h is lowered so that it intersects with

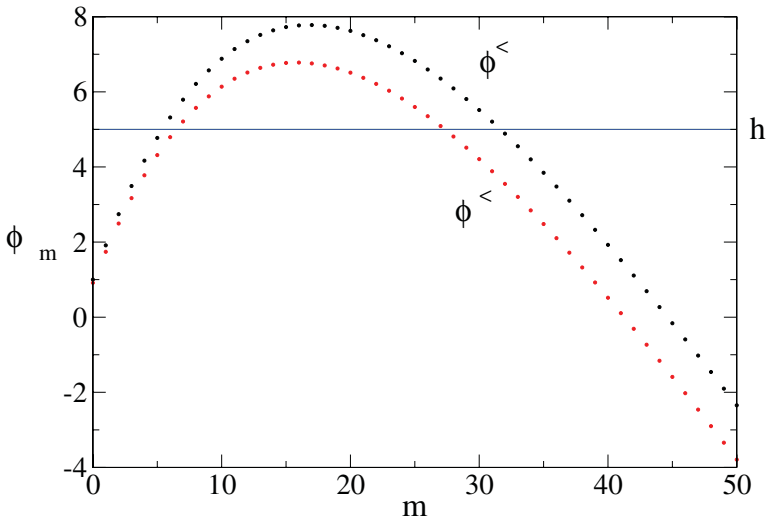


Figure 3.5. *Existence functions for* $w_{ij} = 2e^{-0.05|i-j|} - e^{-0.01|i-j|}$. *In this example, multiple bump states are possible for the large bump but not the small bump.*

the existence curves. In the continuum limit, if h is initially above the existence curve and then reduced, two bump states will arise from a saddle-node bifurcation. However, in the discrete case, a set of states can arise simultaneously or in succession. We summarize the conditions for the existence of a bump in Proposition 1.

Proposition 1. *A symmetric bump or set of bumps can exist for some threshold* $h > 0$ *if* w_{ij} *is* (a) *symmetric,* (b) *maximal and strictly positive at* $i = j$, *and* (c) *decreasing in* $|i - j|$ *(for some large enough finite set around* $i = j$*). These are sufficient but not necessary conditions.*

Figure 3.5 shows an example of the existence functions. We immediately can see that unlike in the continuum version, there is a possibility that sets of "small" and "large" bump solutions can exist. Multiple bump solutions are possible if the slope of the existence function is not too steep. More precisely, for any $m > 0$ and $k > 1$ satisfying $\phi^e_m > \phi^e_{m-k}$ and $\phi^e_{m-k} > \phi^{ne}_m$, there are k bump solutions with widths ranging from $m - k$ to m for the small set of bumps. Using (3.15) and (3.16), these conditions become $\sum_{j=m-k+1}^{m} w_{0,j} > 0$ and $w_{0,0} - \sum_{j=m-k+1}^{m+1} w_{0,j} > 0$. For $\phi^e_{m-k} > \phi^e_m$ and $\phi^e_m > \phi^{ne}_{m-k}$, there are k solutions with widths ranging from $m-k$ to m for the large set of bumps. In terms of the coupling function, the conditions are $\sum_{j=m-k+1}^{m} w_{0,j} < 0$ and $w_{0,0} + \sum_{j=m-k+2}^{m} w_{0,j} > 0$. Thus k solutions are possible for the small bump set if the self-excitation to a neuron exceeds the excitation arriving from k neurons a bump width away. For the large bump set, multiple solutions are possible if self-excitation dominates the inhibition arriving from k neurons a bump width away.

3.2.1. Stability. A given equiperiod bump with infinitely slow synaptic processing ($\alpha = 0$) is degenerate to arbitrary phase arrangements. Hence, any perturbation of the phase will not affect a bump. Additionally, infinitesimally small perturbations to the input to any neuron will have no effect, since generically no neuron in the network has an input that is arbitrarily close to threshold. Thus all firing states in a lattice network are stable to small enough perturbations. Hence, the more relevant question is whether or not the bump states are

attractors with a nontrivial basin of attraction. Consider a lateral-inhibition-like coupling function that yields a single large and small bump. In the Amari model, the large bump is stable and the small bump is not. As we have discussed, in the discrete model, both are stable to infinitesimal perturbations. In order to address the nonlinear stability question we need to address the full dynamics.

We examine the ensuing dynamics of threshold crossing perturbations to the edge neuron and next edge neuron. Consider first the small bump. Here, the recurrent excitation between the neurons in the bump is just sufficient to sustain the bump. If the edge neuron is suppressed, the internal neurons will receive subthreshold input and cease firing. The bump will then collapse to the zero state. Conversely, if the next edge is induced to fire, it will induce its neighbor and the opposite next edge to fire. Thus the set of firing neurons will spread. These dynamics can be directly deduced from Figure 3.5. If neuron $m+1$ begins to fire, the input to neuron -1 will cross threshold and begin to fire. On the other hand, if neuron m stops firing, the edge neuron 0 will stop firing as well.

Now consider the large bump where the opposite edges inhibit each other. If the next edge neuron fires, it suppresses the opposite edge neuron. If the edge neuron stops firing, inhibition is lessened on the opposite next edge, and it is induced to begin firing. Thus, a large bump of width m is stable, although it could shift location if given a strong enough perturbation. Dynamically, if too many neurons fire, they will suppress each other and the bump will shrink. If too few neurons fire, the inhibition will be released and the silent neurons will be induced to fire. The mechanism for this stability is a discretized version of the continuum case first established by Amari [3]. These observations will be relevant later when we consider wandering. It is important to note that stability and existence are inherently intertwined. Henceforth, when we refer to a stable neuron we imply nonlinear stability.

3.3. Bumps on a lattice: Fast synapses. We now consider the case where the synaptic decay rate α is nonzero. Simulations on a discrete lattice show that a stable bump is possible for small enough α. However, if the synaptic decay rate α is increased sufficiently, a stationary bump will not exist. For the no phase reset version of the model, for large enough α, the location of the bump can shift in a seemingly random way which we call wandering. However, as we will discuss in section 3.4, the instant reset version does not support wandering for large α. In the instant reset model, a neuron must receive continuous above-threshold input to fire, so a finite set of neurons cannot maintain self-sustained firing if the synaptic decay rate is so fast that the sum total of the input to any neuron that exceeds threshold has a time duration less than 1. As such, the instant reset model either has a stationary equiperiod bump solution or the zero solution where no neuron fires.

Hence in this section we examine possible stationary solutions for the no phase reset version of the lighthouse model for nonzero synaptic decay rate. In this version, each time a neuron receives above threshold input its phase advances towards threshold. As shown in the proposition below, if α is extremely large, then no self-sustaining bump can exist. However, there can be a range of α for which stationary and wandering bumps can exist.

Proposition 2. *For a fixed coupling function and lattice spacing, there exists an α large enough so no self-sustained activity can exist in a finite set of neurons. Hence no bump can exist for large enough α.*

We can prove this by considering a set of m neurons that form a self-sustaining set. Whenever a neuron receives suprathreshold input, its phase is advanced at rate 1 for the duration of time that the input exceeds threshold. Hence, suprathreshold input must total a time of 1 before the neuron will fire. The maximum rate at which a neuron can fire in the lighthouse model is 1. To sustain activity in a set of neurons, the net output of the neurons in the set must provide sufficient input to maintain firing in all the neurons. The input to a neuron is the weighted sum of synaptic pulses from all neurons in the bump. For any fixed threshold, the synaptic decay rate α can be made large enough so that the suprathreshold duration for the synaptic input due to a given firing event can be made arbitrarily short. When the duration is so short that a neuron requires more than m firing event inputs to fire, then the network cannot sustain itself.

We now investigate the dynamics for α nonzero but not large enough so that no bump can exist. For α nonzero and all neurons firing periodically with rate 1, the synaptic input to a neuron is no longer constant in time but is given by

$$u_i(t) = \sum_{j=0}^{m} w_{ij} P(t - \psi_j), \tag{3.17}$$

where $P(t) = P(t+1)$. For our specific example $P(t) = \alpha \mathrm{e}^{-\alpha t}/(1 - \mathrm{e}^{-\alpha})$, $0 \le t < 1$. The input function $u_i(t)$ now depends on the phases of the firing times of the neurons. As before, in order for a stationary equiperiod bump to exist, the input to the edge neuron must be entirely above threshold, and the input to the next edge must be entirely below threshold. We can thus re-express the existence conditions in terms of the following generalized existence functions:

$$\phi_m^{ne}(\alpha) = \sup_t \sum_{j=0}^{m} w_{-1,j} P(t - \psi_j), \tag{3.18}$$

$$\phi_m^{e}(\alpha) = \inf_t \sum_{j=0}^{m} w_{0,j} P(t - \psi_j). \tag{3.19}$$

Again we require

$$\phi_m^{ne}(\alpha) < h \le \phi_m^{e}(\alpha) \tag{3.20}$$

for an equiperiod bump.

We see for our specific choice of $\eta(t)$ that as α increases, the maximum of $P(t)$ increases and the minimum decreases. Thus for large enough α the existence conditions will be violated and no equiperiod bump can exist. We can show this more explicitly. When α increases from zero, $u_i(t)$ on the edge may not remain above h for an entire period. As an example, consider a synchronous bump where all neurons fire together. First we define the sum

$$W_i^m = \sum_{j=0}^{m} w_{ij}. \tag{3.21}$$

If all the neurons in the bump are synchronized, we can shift the phase so that $u_i(t) = W_0^m P(t)$. On $[0,1)$, the minimum of $P(t)$ is $\alpha e^{-\alpha}/(1-e^{-\alpha})$, and the maximum is $\alpha/(1-e^{-\alpha})$. The existence condition of an equiperiod bump will not be satisfied when

$$\phi_m^{ne} = W_{-1}^m \frac{\alpha}{1-e^{-\alpha}} \geq h \tag{3.22}$$

or

$$\phi_m^0 = W_0^m \frac{\alpha e^{-\alpha}}{1-e^{-\alpha}} < h. \tag{3.23}$$

Thus, for each fixed α, there is a critical lattice spacing, below which there is no equiperiod bump. Consider the difference $d_m = W_0^m - W_{-1}^m$. By translational invariance of w_{ij}, $d_m = w_{00} - w_{0,m+1}$. Recall that $w_{ij} = w(|i-j|\delta x)\delta x$ for a fixed function $w(x)$. Consider keeping the bump width Δ fixed in space while reducing δx. This implies that the number of neurons in the bump m must scale as $m = \Delta/\delta x$. We can then write $d_m = (w(0) - w(\Delta + \delta x))\delta x$. As expected, d_m scales as δx. Hence, as the lattice spacing decreases, W_0^m approaches W_{-1}^m, which implies that $\phi_m^{ne}(\alpha)$ will approach $\phi_m^e(\alpha)$. When the two existence curves meet, the equiperiod bump will no longer exist.

The crucial fact for the nonexistence of the equiperiod bump is that if the input function $u_i(t)$ is not constant, then it will have a maximum and a minimum. The minimum of the edge must remain above threshold, while the maximum of the next edge must remain below threshold for an equiperiod bump to exist. Figure 3.8 (below) shows an example of $u_i(t)$ of the edge and next edge neurons of a bump situated above and below threshold, respectively. As the lattice spacing decreases, the maximum of the next edge will approach the minimum of the edge and eventually violate the condition for the existence of an equiperiod bump. Hence, even if the neurons do not fire synchronously, the existence condition for α nonzero will still be violated for α large enough or δx small enough. For any phase arrangement ψ_j, $u_i(t)$ will always have a time-dependence on a finite lattice. The input function $u_i(t)$ will be constant only if $P(t)$ is constant, and $P(t)$ is constant only if $\alpha = 0$.

In fact, the input function $u(t)$ cannot be constant even in the continuum. We can see this by considering (3.1). Clearly if $\psi(y)$ is differentiable, then $u(x,t)$ is not a constant unless either $P(t)$ or $w(x)$ are constant. If $\psi(y)$ is discontinuous on a countable set in the interval, then the integral can be written as a sum of integrals over separate intervals between the discontinuous points. Each piece is not constant in time, so the same argument that $u(x,t)$ cannot be constant applies. We summarize our results (for the no phase reset lighthouse model) in the following propositions.

Proposition 3. *For a fixed lattice spacing δx, there exists a large enough α such that there is no equiperiod bump. Conversely, for a fixed $\alpha > 0$, there exists a small enough δx so that there is no equiperiod bump.*

Proposition 4. *A continuum equiperiod bump cannot exist for $\alpha > 0$ if neither $w(x)$ nor $P(t)$ are constants.*

Propositions 3 and 4 do not negate the possibility of a nonequiperiod bump. During the time that the input is suprathreshold, the phase of the neuron will advance at rate 1. When the accumulated time of the intervals of suprathreshold input is equal to 1, the neuron

will reach threshold and fire. Thus we could have an α such that a set of interior neurons receive above threshold input while a set of adjacent neurons receive above threshold input intermittently. These neurons will advance towards the firing threshold at a rate slower than 1. This could result in a bump solution with an interior region of neurons firing with rate 1 and an exterior region where the rate decreases to zero as the edge is approached. An example of a nonequiperiod bump is shown in Figure 3.4 (right panel).

Determining the width of a nonequiperiod bump is difficult because the firing rate in the exterior region must be calculated explicitly from the neuronal dynamics. However, we can present an argument that nonequiperiod bumps can exist for $\alpha > 0$ for a suitably chosen coupling kernel. We first introduce the following definition.

Definition 6. *The* interior edge *of a bump is the neuron nearest to the edge that is firing with period* 1. *It divides the bump into an* interior *region where the firing rate is* 1 *and an* exterior *region where the firing rate is nonzero but not* 1.

Suppose we choose w_{ij} so that a symmetric equiperiod bump with all neurons firing in phase on the set $\{0, \dots, m\}$ exists for $\alpha = 0$ and so $\phi_m^{ne} < h \leq \phi_m^e$. We want to show that there exists an $\alpha > 0$ such that the original next edge neurons of this bump will fire and become the new edge of a nonequiperiod bump. The original edge becomes an interior edge. Hence we need to show that there exists an α such that (1) $\inf \sum_{j=0}^{m+2} w_{1,j} P_j(t - \psi_j) \geq h$, (2) $\sup \sum_{j=0}^{m+2} w_{0,j} P_j(t - \psi_j) > h$, and (3) $\phi_{m+2}^{ne} < h$.

Consider a stable large bump so that w_{ij} is positive for $|i-j| > n$ where $n < m$ and negative elsewhere, so that neurons in the bump give a positive stimulus to their near neighbors. We also suppose that they give a weaker negative stimulus to the neurons on the opposite side of the bump. We can then choose h such that it is larger than ϕ_{m+2}^e to satisfy condition (3). We note that for a neuron firing with period T, $1/T - \alpha/2 \leq P(t) \leq 1/T + \alpha/2$. Thus, we can choose an α such that the original next edge begins to fire at an arbitrarily low rate which satisfies condition (2). Each time this neuron fires it gives a stimulus to each neuron in the network. For a symmetric bump, the two next edges fire synchronously, so the net effect is a positive impulse on the original edge. Since the input is already above threshold, this added input will not violate condition (1). The new next edge will also receive a positive stimulus. However, for a small enough α it can remain below threshold, so that condition (3) is still not violated. Thus we have a nonequiperiod bump with m neurons in the interior and two neurons in the exterior.

3.4. Bump wandering states. It was found in simulations of the IF model that as the synaptic decay rate α is increased, the stationary bump state can destabilize and begin to wander [18]. For even larger α, traveling waves can arise, and finally for fast enough α no sustained activity exists. In simulations of the no phase reset lighthouse model, we find that for increasing α there is a transition to wandering, and for high enough α a bump cannot be self-sustained. However, traveling waves are not observed in the lighthouse model. Figure 3.6 shows an example of wandering in a numerical simulation of the lighthouse model.

We can see how wandering may occur by considering the conditions for firing near the edge of the bump. For increased α, the firing rate of the edge neuron could decrease, while the next edge could begin to receive suprathreshold input (for some of the time) and thus begin to fire. The next edge firing would then give inhibitory input to the opposite edge and might

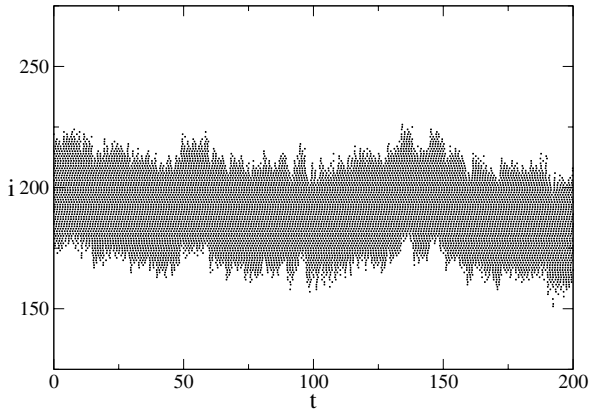


Figure 3.6. *Raster of a wandering bump. Parameters are* $N = 400$, $h = 0.1$, $w_{ij} = 2.1\mathrm{e}^{-|i-j|/60} - 2\mathrm{e}^{-|i-j|/75}$ *and* $\alpha = 3.5$.

cause it to cease firing. In this case the bump could experience a state of frustration where there is no satisfactory equilibrium state to which it can settle. The consequence is that a set of contiguous firing neurons could exist, but the set might not be of a fixed width or at a fixed location. The result could be a wandering set of contiguous neurons.

We have shown previously that, for large enough α, the edge neurons will decrease their firing rate and next edge neurons will receive above threshold input and begin to fire. Recall that a bump is defined as the set of contiguous neurons that fired within the last window of period one. Given that the edge and next edge neurons fire at a slower rate than the interior neurons, the bump width for time intervals over this window will not be stationary. However, the bulk of the bump could still remain in place. If α is not too large, the neurons in the interior of the bump will receive above threshold input continuously and fire with period 1. With this in mind, we make the concept of wandering more precise with the following definition.

Definition 7. *A bump is* wandering *if the position of the interior edge is neither time stationary nor periodic.*

We argue that a wandering state can exist by first demonstrating that there is a non-stationary attractor state comprised of a contiguous set of firing neurons. As we discussed in section 3.2.1, there exists an attractor for a set of bump widths. Consider a stable large bump of width m for $\alpha = 0$ with existence functions of the generic form in Figure 3.5. For such a bump, contiguous sets of firing neurons broader than m will result in strong inhibition to the edge neurons, and the set will diminish. Conversely, for sets smaller than m, but not smaller than the small bump, inhibition will be reduced to the next edge neurons, inducing them to fire and the set will expand. For α nonzero, the inputs to each neuron will no longer be constant in time but will remain bounded. However, for nonzero α there will be a range of α for which this attractor property will still hold.

To show that wandering can exist we must first establish the following proposition.

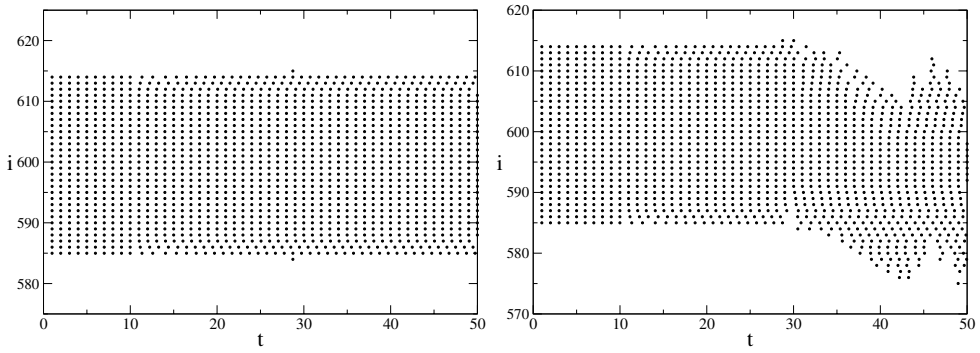


Figure 3.7. *Rasters for* $\alpha = 2.789$ *with* 0.001 *difference in* θ_i *initial conditions. Parameters are* $N - 1200$, $w_{ij} = 2.1\mathrm{e}^{-|i-j|/60} - 2\mathrm{e}^{-|i-j|/75}$, *and* $h = 0.1$

Proposition 5. *The firing of the interior edge can be perturbed by the firing of a single exterior neuron and depends sensitively on the firing phase.*

We first consider a nonequiperiod bump where the interior has more neurons than the exterior and the input from exterior neurons is positive to the adjacent interior edge but negative to the opposite interior edge. For small enough α, the input to the interior edge is bounded away from the threshold h. As α increases, two things can occur. One is that the infimum of the interior edge input will decrease. The second is that the maximum amplitude of inputs from exterior neurons will increase. When α is large enough such that the input from an exterior neuron is large enough to drop the interior edge input below threshold for some temporal duration, then the interior neuron's firing will be delayed and it will no longer fire with rate one.

For a bump with a finite number of neurons, the local infima of the interior edge comprise a finite set. Input from the exterior region will arrive as individual positive or negative exponentially decaying pulses. Hence, there should exist a range of α such that an exterior neuron can slow the firing rate of the interior edge only if it fires in a time window that overlaps with an interior edge infimum. This window can be made arbitrarily small by adjusting α. When the interior edge slows, it no longer has rate 1 and hence no longer belongs to the interior set. Conversely, a next edge neuron could be receiving input that is above threshold except at a few isolated locations around the infima. When these neurons receive a positive input that overlaps with an infimum, they can fire with rate 1 and thus become an interior edge. Hence, the location of the interior depends sensitively on the firing phase of the exterior neurons.

This also explains why there is no wandering for the instant reset lighthouse version. Here, when the input to an interior neuron drops below threshold for no matter how brief a time, the phase will be instantly reset to zero, and thus the neuron will no longer fire. Additionally, while the next edge neurons receive intermittent above threshold input, their phases do not advance because they are reset each time the input drops below threshold. Hence, they can never fire. Thus infinitesimal perturbations can only shrink a bump so no wandering can occur.

Figure 3.7 shows the sensitivity to initial conditions for a bump that is initially near

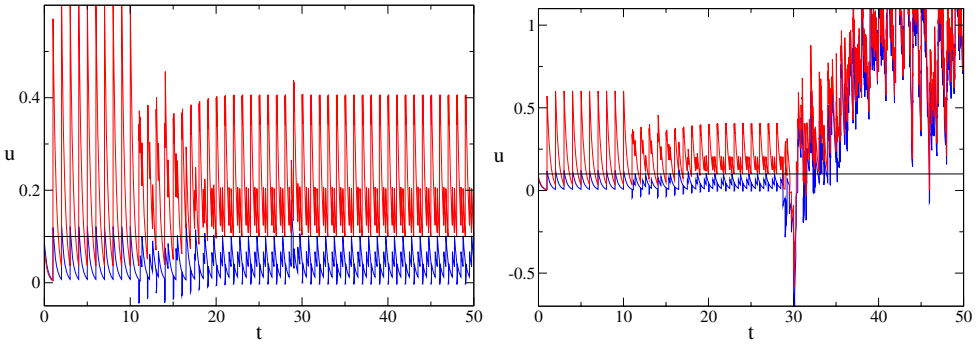


Figure 3.8. *Inputs to one edge (red) and next edge (blue) for* $w_{ij} = 2.1\mathrm{e}^{-|i-j|/60} - 2\mathrm{e}^{-|i-j|/75}$, $h = 0.1$, *and* $\alpha = 2.789$ *for the same initial conditions as in Figure* 3.7.

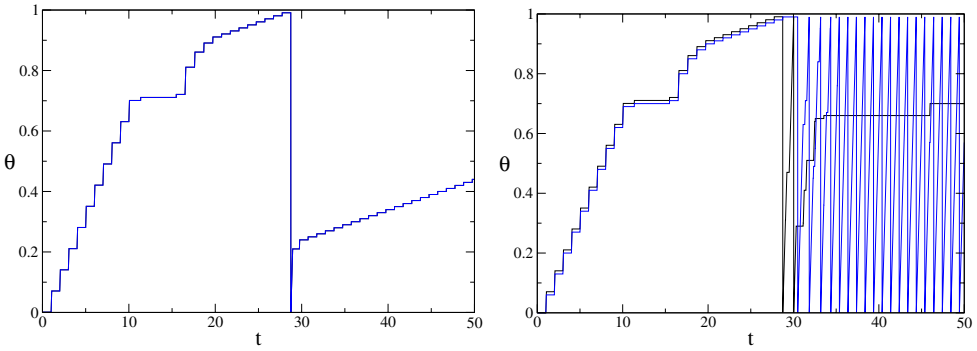


Figure 3.9. *Voltage of opposite next edges for* $w_{ij} = 2.1\mathrm{e}^{-|i-j|/60} - 2\mathrm{e}^{-|i-j|/75}$, $h = 0.1$, *and* $\alpha = 2.789$ *for the same initial conditions as in Figure* 3.7.

synchronous in a simulation. The bump was driven above threshold for the first 20 time units. The initial phases were randomized within a small neighborhood of zero phase ($\theta_i(t = 0) < 0.01$). In the left-hand panel of Figure 3.7, a stable bump is present. Note that the edge neurons fire at a slightly slower rate than the interior neurons and the next edge fires with a very long period. The initial conditions in the left-hand panel differ from those on the right by randomizing the phases in a small neighborhood of the zero phase that is 0.1% larger, but a wandering bump arises. Thus there is a neighborhood of initial phases around zero that tend to lead to a stationary bump. Initial conditions outside of this neighborhood lead to wandering. For voltages initialized to random values between 0 and 1, wandering would be generic. The crucial event that induced wandering occurs at time 30 when one of the next edges fires but the other does not. If both fire together, then their effects on the interior edges essentially cancel out and the stable bump will persist. Figure 3.8 shows the input to the lower edge and next edge. In the left-hand graph, we can see that the inputs to the neurons just barely cross threshold, but the bump can tolerate such excursions. However, in the right-hand graph, one of the crossings disrupts the bump, and wandering ensues. Figure 3.9 shows the voltages of the two opposite next edges. On the left we see the voltages reach threshold together and thus fire together. However, on the right, one of the next edges reaches threshold first and gives an inhibitory pulse to the opposite next edge, delaying its firing. This

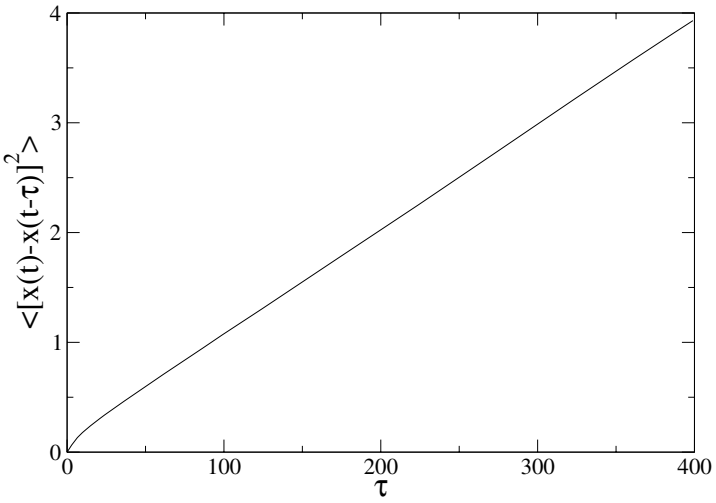


Figure 3.10. *An example of* $\langle [x(t)-x(t-\tau)]^2 \rangle$ *versus* τ *for a numerical simulation with* $N = 400$, $\delta x = 1$, $h = 0.1$, $w_{ij} = 2.1\mathrm{e}^{-|i-j|/60} - 2\mathrm{e}^{-|i-j|/75}$, *and* $\alpha = 1.4$. *The graph was obtained from an average over* 20 *trial runs of length* $20,000$ *time units. Linear scaling with* τ *indicates that wandering is a diffusion process.*

breaking of symmetry allows wandering to take place.

The implication of Proposition 5 is that time stationary bumps do not exist for sufficiently large α, since arbitrarily small rearrangements of the phases could change the input to the neurons and shift the interior edge. However, this does not show that the bumps will wander. They could breathe in and out at a given location or move back and forth periodically. What we need to show is that the bump can shift either left or right, and that the shift does not depend on history. Given that an interior edge can become an exterior neuron and vice versa, we see that shifts could occur in either direction. For a stable large bump, the edges on opposite sides of the bump inhibit each other. Thus, if say the left inner edge moves left, inhibition will increase on the right edge and force it to move left as well. If the inner edge moves right, inhibition will be lessened on the right edge and it will move right as well. For the exponential synaptic function, if $\alpha > 1$, the influence of a neuronal firing decays quickly beyond one period. Thus, past shifts should have little or no effect on future shifts. Hence, after each period, the bump is as likely to move left as right. Thus, wandering is equivalent to a random walk on a lattice. This implies that the position of the inner edge of a bump x will obey $\langle (x(t) - x(t-\tau))^2 \rangle = D\tau$, with $D \propto \delta x^2/T_s$, where δx is the lattice spacing and T_s is the average time for a perturbation of a firing neuron to cause a right- or leftward shift. The graph of $\langle (x(t) - x(t-\tau))^2 \rangle$ versus τ in Figure 3.10 for a direct numerical simulation of the lighthouse model verifies that that wandering is a random walk (i.e., diffusion process).

Figure 3.11 shows a plot of the diffusion constant for different δx. We see that it increases linearly in δx. We can see why this should be true by examining the four key quantities that govern wandering: (1) the amplitude of the input due to the firing of one neuron, (2) the distance to threshold for the interior edge and next interior edge neurons, (3) the average shift size of the bump, and (4) the average time step per shift. The amplitude of the input to location x from a single neuron at y is given by $\eta(t)w(x-y)\delta x$. The distances to threshold are

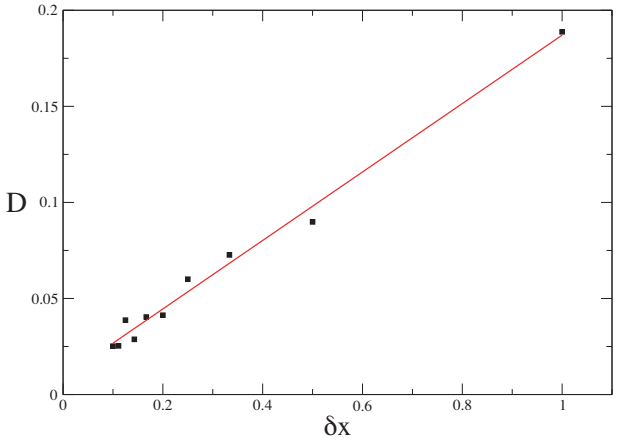


Figure 3.11. *Diffusion coefficient as a function of δx, with $\alpha = 2.5$ and the other parameters as in Figure 3.10.*

given by $d_T^{ne} = \phi_m^{ne}(\alpha) - h$ and $d_T^e = \phi_m^e(\alpha) - h$. For small δx, given the smoothness of $u(x,t)$ in x, d_T^{ne} and d_T^e both should scale as δx. As δx decreases, the amplitude of an individual input from a neuron will decrease, but so too will the distance to threshold. Hence, as the lattice spacing decreases, the threshold α for initiating wandering will not change appreciably. However, the minimum shift the interior edge can take also scales as δx. Thus, at the lattice length scale, wandering of the interior edge will take place regardless of the spacing. However, on the global length scale of the bump, the wandering step size will decrease with δx. Finally, each step is potentially induced by the firing of a neuron within the bump. Therefore, it is reasonable to expect that the step time T_s will scale with δx since the number of neurons in the bump scales as δx^{-1}. Hence, the result is that wandering decreases to zero as δx approaches zero. Figure 3.12 shows the dependence of the diffusion coefficient on α. Interestingly, the dependence seems to be $D \propto \mathrm{e}^{k\alpha}$, for some constant k. Thus $D(\alpha, \delta x) \propto e^{k\alpha}\delta x$.

4. Discussion. In this paper we have revisited the lighthouse model within the context of localized reverberatory activity that can underlie the formation of attractor bump states, both stationary and nonstationary (wandering). In particular, we have shown how this model can be viewed as a natural extension of the Amari model to include the effects of nonslow synaptic processing. Although the model is clearly a highly simplified and abstract model of a neuron, it does possess network behavior that is reminiscent of more biological neuron models.

One prediction of our work, supported by direct numerical simulations, is the possibility that bump states with differing widths can coexist. This simple prediction would not be forthcoming in a purely firing rate model (that has no knowledge of underlying spiking patterns), and would seem to be a hard quantitative prediction to make from the analysis of IF networks.

Moreover, the lighthouse model's simplicity lends itself naturally to both continuum and

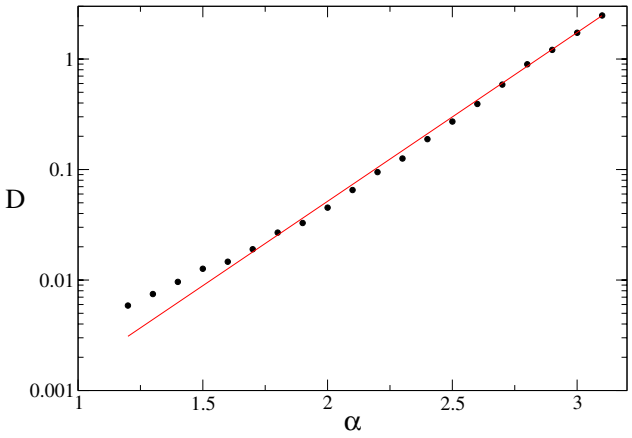


Figure 3.12. *Diffusion coefficient as a function of* α *for* $w_{ij} = 2.1\mathrm{e}^{-|i-j|/60} - 2\mathrm{e}^{-|i-j|/75}$, $\delta x = 1$, $N = 400$, *and* $h = 0.1$.

lattice studies. Indeed by analyzing both we have been able to clarify the role of discreteness in contributing to emergent bump dynamics. In particular, we have shown that multiple width bumps can coexist on a discrete lattice and that wandering bumps can coexist with stationary bump states. Moreover, by tracking the dynamics of bump edges, we have seen a form of competition giving rise to an equal probability that a bump will move left or right, resulting in an effective random walk of the bump. Our studies suggest a diffusion coefficient that scales with the exponential of the rate of synaptic processing and linearly with the lattice spacing. Interestingly, our numerical simulations of the IF model (data not shown) also seem to exhibit similar scaling for the diffusion coefficient on α and δx. Although this result warrants further investigation, it does give further evidence that the lighthouse model can generate dynamics consistent with that of other popular spiking neural networks.

A number of natural extensions of the work presented here suggest themselves. One would be to study the dynamics of moving bumps. In contrast to an IF network, the lighthouse model does not generate waves in response to a brief localized initial stimulus. To obtain a traveling bump in the first instance one would need to augment the lighthouse model to include a form of spike frequency adaptation, as in the work of Hansel and Sompolinsky [26] and Pinto and Ermentrout [23], or perhaps an asymmetry in the synaptic footprint, as in the work of Zhang [27], or even an adaptive threshold, as in the work of Coombes and Owen [28]. The persistence of solutions in the presence of noise is also of interest. We say so mainly because the addition of noise is known to reduce wave speed to turn a traveling bump into a stationary bump, so that in this sense noise can lead to stabilization [29]. Another important study would be to treat heterogeneous networks, say with a distribution of natural single neuron frequencies and thresholds, i.e., by considering $\dot{\theta}_i = \omega_i H(u_i - h_i)$. Even for a globally coupled network this might shed light on how heterogeneity can often lead to the emergence of long time-scale oscillations, as in the network studies of Smolen, Rinzel, and Sherman [30]. Moreover,

deviation from the lattice is completely equivalent to manipulating the coupling function w. The influence of this type of heterogeneity could certainly destroy bump solutions. Including them in the lighthouse model would open up a whole new avenue of research, which might initially be tackled using techniques developed in [31, 32, 33].

REFERENCES

[1] H. R. Wilson and J. D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophys. J., 12 (1972), pp. 1–24.

[2] H. R. Wilson and J. D. Cowan, *A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue*, Kybernetik, 13 (1973), pp. 55–80.

[3] S. Amari, *Characteristics of random nets of analog neuron-like elements*, IEEE Trans. Systems Man Cybernet., 2 (1972), pp. 643–657.

[4] S. Amari, *Homogeneous nets of neuron-like elements*, Biolog. Cybernet., 17 (1975), pp. 211–220.

[5] C. van Vreeswijk, *Partial synchronization in populations of pulse-coupled oscillators*, Phys. Rev. E, 54 (1996), pp. 5522–5537.

[6] D. Hansel, G. Mato, and C. Meunier, *Synchrony in excitatory neural networks*, Neural Comput., 7 (1995), pp. 2307–2337.

[7] W. Gerstner, *Time structure of the activity in neural network models*, Phys. Rev. E, 51 (1995), pp. 738–758.

[8] P. C. Bressloff and S. Coombes, *Dynamics of strongly coupled spiking neurons*, Neural Comput., 12 (2000), pp. 91–129.

[9] C. C. Chow, *Phase-locking in weakly heterogeneous neuronal networks*, Phys. D, 118 (1998), pp. 343–370.

[10] N. Brunel, *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons*, J. Comput. Neurosci., 8 (2000), pp. 183–208.

[11] H. Haken, *Brain Dynamics: Synchronization and Activity Patterns in Pulse-coupled Neural Nets with Delays and Noise*, Springer-Verlag, New York, Berlin, 2002.

[12] C. L. Colby, J. R. Duhamel, and M. E. Goldberg, *Oculocentric spatial representation in parietal cortex*, Cerebral Cortex, 5 (1995), pp. 470–481.

[13] P. S. Goldman-Rakic, *Cellular basis of working memory*, Neuron, 14 (1995), pp. 477–485.

[14] J. Tegner, A. Compte, and X.-J. Wang, *The dynamical stability of reverberatory neural circuits*, Biolog. Cybernet., 87 (2002), pp. 471–481.

[15] C. R. Laing, W. C. Troy, B. Gutkin, and G. B. Ermentrout, *Multiple bumps in a neuronal model of working memory*, SIAM J. Appl. Math., 63 (2002), pp. 62–97.

[16] B. S. Gutkin, C. R. Laing, C. C. Chow, G. B. Ermentrout, and C. L. Colby, *Turning on and off with excitation: The role of spike-timing asynchrony and synchrony in sustained neural activity*, J. Comput. Neurosci., 11 (2001), pp. 121–134.

[17] S. Coombes, *Waves, bumps and patterns in neural field theories*, Biolog. Cybernet., 93 (2005), pp. 91–108.

[18] C. Laing and C. C. Chow, *Stationary bumps in networks of spiking neurons*, Neural Comput., 13 (2001), pp. 1473–1494.

[19] A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang, *Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model*, Cerebral Cortex, 10 (2000), pp. 910–923.

[20] H. Haken, *Phase locking in the lighthouse model of a neural net with several delay times*, Progr. Theoret. Phys., 139 (2000), pp. 96–111.

[21] H. Haken, *Quasi-discrete dynamics of a neural net: The lighthouse model*, Discrete Dynamics in Nature and Society, 4 (2000), pp. 187–200.

[22] S. Amari, *Dynamics of pattern formation in lateral-inhibition type neural fields*, Biolog. Cybernet., 27 (1977), pp. 77–87.

[23] D. J. Pinto and G. B. Ermentrout, *Spatially structured activity in synaptically coupled neuronal networks:* I. *Travelling fronts and pulses*, SIAM J. Appl. Math., 62 (2001), pp. 206–225.

[24] Y. Guo and C. C. Chow, *Existence and stability of standing pulses in neural networks:* I. *Existence*, SIAM J. Appl. Dynam. Syst., 4 (2005), pp. 217–248.

[25] Y. Guo and C. C. Chow, *Existence and stability of standing pulses in neural networks:* II. *Stability*, SIAM J. Appl. Dynam. Syst., 4 (2005), pp. 249–281.
[26] D. Hansel and H. Sompolinsky, *Modeling feature selectivity in local cortical circuits*, in Methods in Neuronal Modeling, From Ions to Networks, 2nd ed., MIT Press, Cambridge, MA, 1998, pp. 499–567.
[27] K. Zhang, *Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory*, J. Neurosci., 16 (1996), pp. 2112–2126.
[28] S. Coombes and M. R. Owen, *Bumps, breathers, and waves in a neural network with spike frequency adaptation*, Phys. Rev. Lett., 94 (2005), paper 148102.
[29] C. R. Laing and A. Longtin, *Noise-induced stabilization of bumps in systems with long-range spatial coupling*, Phys. D, 160 (2001), pp. 149–172.
[30] P. Smolen, J. Rinzel, and A. Sherman, *Why pancreatic islets burst but single beta cells do not. The heterogeneity hypothesis*, Biophys. J., 64 (1993), pp. 1668–1680.
[31] V. K. Jirsa and J. A. S. Kelso, *Spatiotemporal pattern formation in neural systems with heterogeneous connection topologies*, Phys. Rev. E, 62 (2000), pp. 8462–8465.
[32] P. C. Bressloff, *Traveling fronts and wave propagation failure in an inhomogeneous neural network*, Phys. D, 155 (2001), pp. 83–100.
[33] P. C. Bressloff, *Spatially periodic modulation of cortical patterns by long-range horizontal connections*, Phys. D, 185 (2003), pp. 131–157.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 575–597

Pulses in Nonlinearly Coupled Schrödinger Equations I. A Homoclinic Flip Bifurcation*

Russell K. Jackson†

Abstract. In this work, we describe a new mechanism for the generation of multipulse solutions in a class of nonlinearly coupled Schrödinger equations. Many novel pulses have been observed in such systems both numerically and experimentally, but, until now, an understanding of their origins has been lacking. The particular bifurcation studied here is spurred by the passage through degeneracy of a one-component pulse in orbit-flip configuration. We provide a straightforward geometric analysis, demonstrating the production not only of a multicomponent 1-pulse nearby the original one-component pulse, but also of an entire family of alternating N-pulses, for all positive integers N.

Key words. homoclinic bifurcation, standing waves, N-pulses, linearization, degeneracy, cusp horseshoe

AMS subject classifications. 37K50, 70K44, 34C37, 35Q51

DOI. 10.1137/050637686

1. Introduction and statement of results.

Equations of nonlinear Schrödinger (NLS) type provide models in various optical contexts for the propagation of a single electro-magnetic field. In order to describe phenomena in which multiple fields are present, such models are augmented to include two or more NLS equations, coupled together through the nonlinearity. For example, such coupled equations describe the effects of birefringence in an optical fiber [1] and incoherent beam interaction within a photorefractive slab [2].

The solitary wave solutions (solitons) often found in such NLS equations have garnered much recent attention; these solitons are ideally suited to serve as information carriers in modern optical systems. The coupling required to model multiple fields opens the possibility of finding novel solitary wave solutions not seen when only a single field is present. In fact, a variety of multicomponent and multipulse solitary waves have been observed for such coupled systems in a number of works, both numerically [1, 2] and even experimentally [3]. However, an analytic explanation for the presence of such multipulse solutions has been desired [4] and until now has been absent. In the current paper, we identify the properties that make these coupled systems ripe for interesting behavior and we provide a straightforward geometric analysis detailing the unfolding of a family of multipulse solutions.

*Received by the editors August 5, 2005; accepted for publication (in revised form) by T. Kaper May 15, 2006; published electronically November 14, 2006. This work was supported by a grant from the National Science Foundation under award number DMS-0202542. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.
http://www.siam.org/journals/siads/5-4/63768.html
†Mathematics Department, U.S. Naval Academy, Annapolis, MD 21402 (rkjackso@usna.edu).

We consider a general coupled system of the form

$$\begin{array}{rcrcrcrcl} \mathrm{i}u_z & + & u_{xx} & + & h_1(|u|^2,|w|^2)u & - & u & = & 0, \\ \mathrm{i}w_z & + & w_{xx} & + & h_2(|u|^2,|w|^2)w & - & r^2w & = & 0, \end{array} \tag{1}$$

where $0 < r < 1$ is a normalized propagation constant and $h_1(\cdot)$ and $h_2(\cdot)$ model the material properties. We further assume that (1) has a Hamiltonian structure. Both Haelterman and Sheppard's model [1] for birefringence in optical fibers, where

$$\begin{aligned} h_1(|u|^2,|w|^2) &= (1-B)|u|^2 + (1+B)|w|^2, \\ h_2(|u|^2,|w|^2) &= (1+B)|u|^2 + (1-B)|w|^2, \end{aligned} \tag{2}$$

and Ostrovskaya and Kivshar's model [2] for incoherent beam interaction in a biased photorefractive crystal, where

$$h_1(|u|^2,|w|^2) = h_2(|u|^2,|w|^2) = \frac{|u|^2+|w|^2}{1+s(|u|^2+|w|^2)}, \tag{3}$$

can be written in this way.

Stationary waves are real-valued, pulse-like solutions that do not vary with z. In system (1), a stationary wave can then be interpreted as an orbit homoclinic to the origin in the related system of ordinary differential equations

$$\begin{array}{rcrcrcl} u'' & = & -h_1(|u|^2,|w|^2)u & + & u, \\ w'' & = & -h_2(|u|^2,|w|^2)w & + & r^2w, \end{array} \tag{4}$$

where $'$ indicates differentiation with respect to x. For reasonable incarnations of the nonlinearity $h_1(\cdot)$—including $h_1(\cdot)$ nonnegative (a focusing nonlinearity), s, and B (above) not too large—there is a symmetric pair of stationary waves of the form $(U_0(x),0)$ and $(-U_0(x),0)$ which are independent of the parameter r. Our bifurcation values will be values of the parameter r strictly between 0 and 1 for which these one-component stationary waves are *degenerate*, where degeneracy is defined precisely in Property 5 of section 2. For such parameter values, we prove the following result.

Theorem 1. *Suppose that the ordinary differential equation* (4) *supports a single-component pulse* $(U_0(x),0)$*, and that this pulse is degenerate for a particular parameter* $r = r^*$ *with* $0 < r^* < 1$*. Then, under the condition defined in* (19)*, there exists a multicomponent pulse near* $(U_0(x),0)$ *for all* r *in a small parameter neighborhood with* $r > r^*$*. Additionally, for every* r *in this parameter neighborhood, there also exists an* N*-pulse solution, alternating in the sign of first component, for every positive integer* N*.*

Examples of the multicomponent 1-pulse and the 2-pulse are shown in Figure 1.

In section 2, we discuss our geometric method of proof and outline the properties of (4) and of the *fast* one-component pulse $(U_0(x),0)$ that are pertinent to the bifurcation analysis. Additionally, we provide a sketch of the bifurcation values for both of our motivating examples: the fiber birefringence model (2) and the incoherent beam interaction model (3). Also of special note is the simplification of the flow in a neighborhood of the origin achieved by an application of Theorem 1.2 of Banyaga, de la Llave, and Wayne [5].

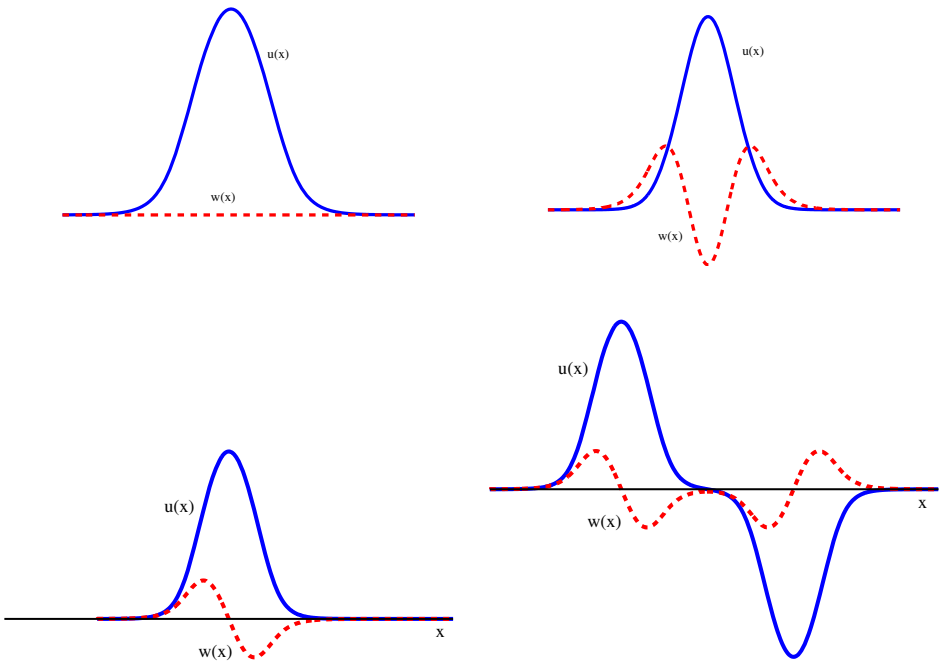


Figure 1. *Theorem 1 guarantees the existence of a variety of stationary waves near a value of the parameter r where the single-component pulse $(U_0(x), 0)$ is degenerate. The top left panel shows the single-component pulse in (3) with $s = 0.8$. The top right panel shows a bifurcating one-pulse with $r^2 = 0.45$. The bottom left panel shows another bifurcating one-pulse when r^2 is increased to 0.65. The bottom right panel shows an alternating 2-pulse for this same parameter value $r^2 = 0.65$. Such alternating concatenated N-pulses exist nearby in the phase space for all N.*

In section 3, we construct Poincaré cross-sections Σ_{in} and Σ_{out} transverse to the pulse $U_0(x)$ in phase space; this allows us to consider the flow near the fixed point at the origin and the excursion far from the origin separately. The behaviors in these regions are reflected in the transition maps Φ_{near} and Φ_{far} between these sections. A careful study of the composition of these maps yields the proof of Theorem 1. We note especially that the flow near the origin, as reflected in the map Φ_{near}, differs substantially depending upon whether the bifurcation parameter satisfies $r^* < 1/2$ or $r^* \geq 1/2$; however, the conclusion of Theorem 1 holds in either case.

2. The procedure and pertinent properties. We are interested in stationary waves of (4) bifurcating from the solution $U_0(x)$, for functions $h_1(\cdot)$ where such a single component wave exists. There are several different parameter regimes (in r) where we might expect nontrivial behavior near such a wave, resulting in a bifurcation of additional pulses. For instance, near $r = 1$, the problem displays a resonance in which all of the eigenvalues at the origin have the same magnitude. This is one possible bifurcation value, and has been studied by Yew, Sandstede, and Jones [6]. Their analysis near this resonance uses the method developed sequentially by Hale [7], Lin [8], and Sandstede [9], sometimes called the homoclinic Lyapunov–Schmidt method (and, in any case, abbreviated as the HLS method). Another

possible bifurcation may be expected at the other extreme, near $r = 0$. Here, the origin has a pair of zero eigenvalues and ceases to be hyperbolic. This limit is of interest for further study but will not be considered in this paper. We remark that the work of Champneys and Yang [10] provides some insight into the dynamics in this regime.

In this paper, we consider intermediate bifurcation values: parameter values r between 0 and 1 for which the fast stationary wave solution $U_0(x)$ is *degenerate*, as defined in Property 5 below.

We prove Theorem 1 using a geometric method, in the spirit of Homburg, Kokubu, and Krupa [11]. The usual geometric method proceeds by defining sections that divide the flow near a homoclinic orbit into two regions: a *near* region around the critical point, where the dynamics are dominated by the linearized flow; and a *far* region, away from the critical point, where the behavior is governed by rectilinear flow. The full behavior can then be determined by considering the interplay between these two regimes. This method has been used successfully to analyze other *homoclinic flip* bifurcations: for instance, the *orbit flip* and the *inclination flip*; see Homburg, Kokubu, and Krupa [11], Nii [12], or Homburg and Krauskopf [13], for example. Several works have also analyzed degenerate homoclinic solutions in other contexts, including Vanderbauwhede [14] and Fiedler and Turaev [15].

In the more analytic HLS method, piecewise continuous solutions are defined between sections with jumps allowed only in a specified direction within these sections. Homoclinic orbits correspond to solutions where the jumps can all be set to 0. Despite the increased generality of the HLS method, the geometric mechanisms for multipulse production are often hidden behind detailed remainder estimates. Analyses of this type can be seen for various bifurcations in Sandstede [9], Yew [16], Sandstede, Jones, and Alexander [17], Knobloch [18], or van Gils, Krupa, and Tchistiakov [19]. The work of Knobloch [18] is especially pertinent, as it also examines degenerate pulse solutions in the presence of additional structure.

In the first part this section, we outline the properties of the ordinary differential equation (4) and of its one-component pulse solution $(U_0(x), 0)$ that impact the bifurcation. Before proceeding, we implement the standard convention of reducing a second order system (4) to an equivalent system of first order equations:

$$\begin{aligned} u' &= p, \\ w' &= q, \\ p' &= u - h_1(|u|^2, |w|^2)\, u, \\ q' &= r^2 w - h_2(|u|^2, |w|^2)\, w. \end{aligned} \tag{5}$$

We sometimes abbreviate this system as

$$v' = f(v)$$

for v in $\mathbb{R}^4$. For $r > 0$, this system has a hyperbolic fixed point at the origin with eigenvalues given by -1, $-r$, r, and 1. The stable and unstable manifolds of this fixed point are both two dimensional.

Property 1. *The system* (5) *is Hamiltonian; i.e., there is a function* $H(u, w, p, q)$ *such that*

$$
\begin{aligned}
u' &= -\tfrac{\partial H}{\partial p},\\
w' &= -\tfrac{\partial H}{\partial q},\\
p' &= \tfrac{\partial H}{\partial u},\\
q' &= \tfrac{\partial H}{\partial w}.
\end{aligned}
\tag{6}
$$

Property 2. *The system* (5) *is time-reversible; i.e., there exists an involution* $R : \mathbb{R}^4 \to \mathbb{R}^4$ *with* $R^2 = I$ *such that*

$$f(R(u, w, p, q)) = -R\ f(u, w, p, q).$$

In particular, R *acts as* $R(u, w, p, q) = (u, w, -p, -q)$.

Property 3. *There is a pair of additional symmetries* S_1 *and* S_2 *such that, for* $i = 1, 2$,

$$f(S_i(u, w, p, q)) = S_i\ f(u, w, p, q).$$

These are $S_1(u, w, p, q) = (-u, w, -p, q)$ *and* $S_2(u, w, p, q) = (u, -w, p, -q)$.

Regarding Property 1, in the model given in (2) for pulse propagation in a birefringent fiber, the Hamiltonian has the form

$$
\begin{aligned}
H_{\mathrm{HS}}(u, w, p, q) &= -\tfrac{1}{2}(p^2 + q^2) + \tfrac{1}{2}(u^2 + (rw)^2)\\
&\quad - \tfrac{1}{4}\left((1 - B)u^4 + 2(1 + B)u^2w^2 + (1 - B)w^4\right).
\end{aligned}
\tag{7}
$$

In the case of the model in (3) for incoherently interacting beams in a photorefractive crystal, we have

$$
\begin{aligned}
H_{\mathrm{OK}}(u, w, p, q) &= -\tfrac{1}{2}(p^2 + q^2) + \tfrac{1}{2}(u^2 + (rw)^2)\\
&\quad - \tfrac{1}{2s^2}\left(s(u^2 + w^2) - \ln(1 + s(u^2 + w^2))\right).
\end{aligned}
\tag{8}
$$

It follows from Property 3 that the two dimensional subspaces

$$U = \{(u, w, p, q) \mid w = q = 0\}$$

and

$$W = \{(u, w, p, q) \mid u = p = 0\}$$

are invariant. These subspaces correspond to one of the two fields being identically zero; in either of these subspaces, the system behaves as a simple Newtonian vector field in a plane. In the subspace U, for many variations of the focusing nonlinearity $h_1(\cdot)$, there is a symmetric pair of homoclinic orbits: we denote the solution with positive u-part $U_0(x)$ and call this pulse the *fast pulse*. With a slight abuse of notation, we often refer to the solutions $(U_0(x), 0)$ of (4) and $(U_0(x), 0, U_0'(x), 0)$ of (5) both simply as $U_0(x)$. Note that both the fast pulse $U_0(x)$ and its partner $S_1U_0(x)$ are independent of the parameter r. Similarly, for a reasonable variety of focusing nonlinearities $h_2(\cdot)$, when $r > 0$, there is a symmetric pair of homoclinic orbits in

W. The pulse with positive w-part is called the *slow pulse*, and both the slow pulse and its reflection via S_2 vary with the parameter r.

Here we also take note of several important properties of the fast pulses $U_0(x)$ and $S_1U_0(x)$ within the four dimensional phase space of (5).

Property 4. *For $r < 1$, the elementary pulses $U_0(x)$ and $S_1U_0(x)$ are in* orbit flip configuration. *That is, as $x \to \pm\infty$, $U_0(x)$ and $S_1U_0(x)$ approach the origin in phase space via a direction other than the one associated with the eigenvalues of smallest real part.*

Property 5. *Depending upon the functions $h_1(\cdot)$ and $h_2(\cdot)$, for particular values of the parameter r between 0 and 1, the pulses $U_0(x)$ and $S_1U_0(x)$ may be* degenerate, *i.e.,* $\dim(T_{U_0(x)}W^u(0) \cap T_{U_0(x)}W^s(0)) > 1$.

Property 4 holds because the fast pulse $U_0(x)$ lies in the invariant subspace U and so must approach the origin as $x \to \pm\infty$ tangent to the eigenvectors corresponding to the eigenvalues 1 and -1. However, a generic solution in the stable or unstable manifold in the $\mathbb{R}^4$ approaches the origin as $x \to \pm\infty$ tangent to an eigenvector associated with the eigenvalues $\pm r$.

Property 5 says that, for particular values of the parameter r, the global stable and unstable manifolds are tangent along the pulse $U_0(x)$. In general, the presence of any homoclinic orbit indicates an intersection of at least one dimension of the spaces $T_{U_0(x)}W^u(0)$ and $T_{U_0(x)}W^s(0)$; in particular, these spaces will always have the vector $U_0'(x)$ in common. The degeneracy specified in Property 5 indicates that, for particular values of the parameter r, there may be yet another vector common to both tangent spaces. Because this is a property of the tangent spaces about a pulse, such parameter values can be isolated by studying the linearization about the fast pulse $U_0(x)$ (or its twin $S_1U_0(x)$). This will be pursued in detail in the subsection below.

The degeneracy described in Property 5 is sometimes called a *homoclinic flip*, as the orientation of the global unstable manifold of the origin, followed along the solution $U_0(x)$, can be expected to change as r passes through such a value.

2.1. Degeneracy and Sturm–Liouville. In this subsection, we consider the problem of identifying values of the parameter r for which the pulse $U_0(x)$ is degenerate, as defined in Property 5, above. This turns out to be a *Sturm–Liouville* problem, the analysis of which is well known. For completeness, we include a review of this theory that is in keeping with our geometric outlook.

We write the solution $U_0(x)$ of (5) in component form as $(U_0(x), 0, U_0'(x), 0)$. Linearizing this equation about this pulse yields the nonautonomous system

$$
\begin{aligned}
U' &= P,\\
W' &= Q,\\
P' &= \left(1 - h_1(U_0^2(x), 0) - 2\,U_0^2(x)\; D_1h_1(U_0^2(x), 0)\right)\; U,\\
Q' &= \left(r^2 - h_2(U_0^2(x), 0)\right)\; W.
\end{aligned}
\tag{9}
$$

This decouples into the two smaller systems

$$\begin{cases} U' &= P, \\ P' &= \left(1 - h_1(U_0^2(x),0) - 2\ U_0^2(x)\ D_1 h_1(U_0^2(x),0)\right)\ U, \end{cases} \tag{10}$$

$$\begin{cases} W' &= Q, \\ Q' &= \left(r^2 - h_2(U_0^2(x),0)\right)\ Q. \end{cases} \tag{11}$$

The first system (10) has a bounded solution given by $(U_0'(x), U_0''(x))$; this solution captures the trivial intersection of the tangent spaces induced by the mere existence of a homoclinic orbit. The degeneracy described in Property 5 can then be interpreted as a simultaneous bounded solution of the second system (11). We wish to understand the conditions under which we may expect such a degeneracy to occur for some $r > 0$.

We simplify the system (11) for consideration in two ways. First, since these equations are linear, we projectivize the flow. To this end, we introduce the radial and angular variables, ρ and θ, satisfying

$$\begin{aligned} \rho^2 &= W^2 + Q^2, \\ \tan(\theta) &= \tfrac{Q}{W}. \end{aligned}$$

It is a simple exercise to show that this substitution yields the equivalent system

$$\begin{aligned} \rho' &= \left(1 + (r^2 - h_2(U_0^2(x),0))\right)\ \sin(2\theta)\ \rho, \\ \theta' &= \left(r^2 - h_2(U_0^2(x),0)\right)\ \cos^2(\theta) - \sin^2(\theta). \end{aligned} \tag{12}$$

Note especially that the equation for θ is independent of ρ. We can therefore consider only the second equation of (12)—a nonautonomous flow on the real projective space $\mathbb{R}P^1$.

Second, since we are interested in the behavior of orbits as $x \to \pm\infty$, we compactify the evolution variable. In order to clarify the asymptotic behavior, we use Terman's convention and set

$$\tau = \tanh(\kappa x),$$

where κ is chosen to be appropriately small, e.g., $0 < \kappa < \frac{1}{2}$, so that

$$U_c^2(\tau) = U_0^2\left(\frac{1}{2\kappa}\ln\left(\frac{1+\tau}{1-\tau}\right)\right)$$

can be extended in a C^1 fashion from the open interval $(-1,1)$ to its closure. Hence we have

$$\begin{aligned} \theta' &= \left(r^2 - h_2(U_c^2(\tau),0)\right)\cos^2(\theta) - \sin^2(\theta), \\ \tau' &= \kappa(1-\tau^2), \end{aligned} \tag{13}$$

which is an autonomous flow on $\mathbb{R}P^1 \times [-1,1]$. The spaces $\tau = -1$ and $\tau = +1$ are now invariant and carry the asymptotic flows. In these spaces we simply have

$$\begin{aligned} \theta' &= r^2\cos^2(\theta) - \sin^2(\theta), \\ \tau' &= 0. \end{aligned}$$

For $r > 0$, the asymptotic system has two fixed points between $-\pi/2$ and $\pi/2$; these are $\theta_+ = +\tan^{-1}(r)$ and $\theta_- = -\tan^{-1}(r)$. Solutions of the linearized equation (11) that decay as $x \to -\infty$ correspond to solutions of (13) that tend to $(\theta_+, -1) \in \mathbb{R}P^1 \times (-1, 1]$ as $x \to -\infty$. Likewise, solutions of the linearized equation (11) that decay as $x \to +\infty$ correspond to solutions of (13) that tend to $(\theta_-, +1) \in \mathbb{R}P^1 \times [-1, 1)$ as $x \to +\infty$. And hence a bounded solution of (11) corresponds to an orbit of (13) connecting $(\theta_+, -1)$ and $(\theta_-, +1)$.

We then have the following shooting problem. As $r > 0$ is allowed to vary, we follow the nontrivial solution θ_r that approaches $(\theta_+, -1)$ as $x \to -\infty$. As x increases, this solution will wind around the cylinder for $\tau \in (-1, 1)$, and for particular values of r it may meet up exactly with $(\theta_-, +1)$ as $x \to +\infty$. We can count the number of these connecting values exactly by invoking a monotonicity argument. In particular, consider two solutions $\theta_{r_1}(x)$ and $\theta_{r_2}(x)$ that approach $(\theta_+, -1)$ as $x \to -\infty$ for parameter values $r = r_1$ and $r = r_2$, respectively. The following result can be deduced easily.

Lemma 2. *For each x, if $r_1 < r_2$, then $\theta_{r_1}(x) < \theta_{r_2}(x)$.*

Also, for r large, there is no winding. In particular, we have the following lemma.

Lemma 3. *For r sufficiently large, $0 < \theta_r(x) < \pi/2$ for all x.*

Since there is no winding for r large, and $\theta(x)$ changes monotonically in r, the number of times the trajectory leaving $(\theta_+(r), -1)$ winds around the cylinder for $r \to 0$ gives an exact count of the number of degenerate parameter values satisfying $r > 0$.

For Haelterman and Sheppard's model for birefringence in optical fibers, as given in (2) and by the Hamiltonian (7), when $B = 0$ there are no parameter values r between 0 and 1 for which $U_0(x)$ is degenerate. In the limit $B = 0$, the equations (2) become the integrable Manakov equations, and in this case exact solutions can be calculated; see, for instance, Li and Promislow [20]. As B is increased from zero, a degenerate parameter r (in the sense of Property 5) bifurcates immediately from $r = 0$. As B is increased further, the continuation of this parameter r for which $U_0(x)$ is degenerate increases until it is no longer in the range $0 < r < 1$. However, the pulse $U_0(x)$ becomes taller and wider with increasing B, and more and more degenerate parameter values bifurcate from $r = 0$. As $B \to 1$, the height and width of the stationary pulse $U_0(x)$ approach infinity and the family of degenerate parameter values r grows; see Figure 2, panel 1.

The behavior is similar for Ostrovskaya and Kivshar's model for incoherent beam propagation in a photorefractive material, as given in (3) and by the Hamiltonian (8). The case $s = 0$ here is again the Manakov system, but the structure differs from the previous model as the saturability s is allowed to vary. In this case, $r = 1$ is always degenerate; when $r = 1$, the equations for u and w in (4) are identical and $(U_0(x), U_0'(x))$ always provides a bounded solution of (11). Since $U_0(x)$ is nowhere zero, it follows that there cannot be any degenerate parameter values for parameter values $r > 1$. Here, as the pulse $U_0(x)$ becomes taller and wider with increasing s, degenerate parameter values again bifurcate from $r = 0$. In this case, though, these degenerate parameter values accumulate at $r = 1$. In fact, the number of degenerate parameter values between zero and one approaches infinity as the saturability s of the medium approaches one; see Figure 2, panel 2.

2.2. A change of coordinates near the origin. One difficulty common to the geometric method is the need to assume that the vector field f can be linearized about the origin, that

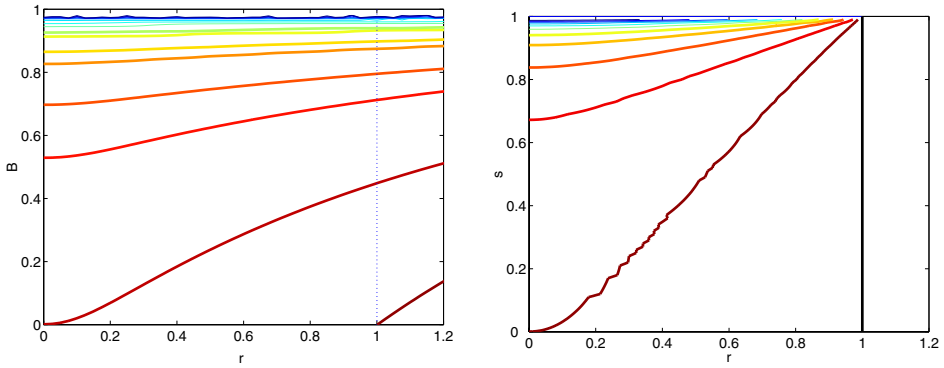


Figure 2. *Degeneracy of the pulse $U_0(x)$ can be reduced to a Sturm–Liouville problem. The curves in panel 1 indicate parameter values for which the pulse $U_0(x)$ is degenerate for the nonlinearity given in the Haelterman–Sheppard model of (2). Panel 2 shows the parameter values for which $U_0(x)$ is degenerate for the Ostrovskaya–Kivshar model of (3).*

is, that there is a "smooth-enough" function φ so that

$$\varphi^{-1} \circ f \circ \varphi(x) = Df(0)\ x$$

for x in some neighborhood of the origin. The classical result that would yield this property is the Sternberg linearization theorem. Unfortunately, this theorem requires that the eigenvalues of the fixed point be nonresonant; and in system (5) such an assumption will necessarily be violated. In fact, such a linearization is in general not possible for the vector field given by (5). However, through the use of the Birkhoff normal form and Theorem 1.2 of Banyaga, de la Llave, and Wayne [5], we are nonetheless able to reduce (5) to a much simpler system.

To begin, we first rewrite system (5) with the aid of the canonical transformation:

$$\begin{aligned} \tilde{v}_{ss} &= \tfrac{1}{\sqrt{2}}(p-u), & \tilde{v}_s &= \tfrac{1}{\sqrt{2r}}(q-rw), \\ \tilde{v}_{uu} &= \tfrac{1}{\sqrt{2}}(p+u), & \tilde{v}_u &= \tfrac{1}{\sqrt{2r}}(q+rw). \end{aligned} \tag{14}$$

In these new coordinates, the eigenvectors associated with -1, $-r$, r, and 1 are tangent to the coordinate axes $\tilde{v}_{ss}$, $\tilde{v}_s$, $\tilde{v}_u$, and $\tilde{v}_{uu}$, respectively. In this coordinate system, the Hamiltonian takes the form

$$H = \tilde{v}_{uu}\tilde{v}_{ss} + r\tilde{v}_u\tilde{v}_s + O(\|(\tilde{v}_{uu}, \tilde{v}_u, \tilde{v}_{ss}, \tilde{v}_s)\|^4).$$

According to Theorem 1.2 of Banyaga, de la Llave, and Wayne [5], we then have the following.

Lemma 4. *Suppose that the C^∞ ordinary differential equation (5) preserves a Hamiltonian H, as described in Property 1. For $0 < r < 1$ and $l \geq 1$, let k be chosen so that*

$$k \ \geq\ \frac{(1+r)^2}{r^2}\ (l+1)\ +\ 2.$$

Then, in a neighborhood of the origin, there is a C^l canonical change of coordinates such that,

in the new coordinates, the Hamiltonian H takes the truncated form

$$H_k = v_{uu}v_{ss} + rv_uv_s + \sum_{j=2}^{k}\sum_{m=0}^{j} a_{jm}\,(v_{uu}v_{ss})^m\,(v_uv_s)^{j-m} \tag{15}$$

for some prescribed constants a_{jm}.

In these new coordinates, the equations of motion become

$$\begin{array}{rcrcl} v'_{uu} &=& v_{uu} &+& \sum_{j=2}^{k}\sum_{m=1}^{j} m\ a_{jm}\ (v_{uu}v_{ss})^{m-1}\,(v_uv_s)^{j-m}\,v_{uu}, \\ v'_{ss} &=& -v_{ss} &-& \sum_{j=2}^{k}\sum_{m=1}^{j} m\ a_{jm}\ (v_{uu}v_{ss})^{m-1}\,(v_uv_s)^{j-m}\,v_{ss}, \\ v'_{u} &=& rv_u &+& \sum_{j=2}^{k}\sum_{m=0}^{j-1}(j-m)\ a_{jm}\ (v_{uu}v_{ss})^{m}\,(v_uv_s)^{j-m-1}\,v_u, \\ v'_{s} &=& -rv_s &-& \sum_{j=2}^{k}\sum_{m=0}^{j-1}(j-m)\ a_{jm}\ (v_{uu}v_{ss})^{m}\,(v_uv_s)^{j-m-1}\,v_s. \end{array} \tag{16}$$

We note especially that the equations in (16) are integrable. That is, in addition to H_k being conserved along any trajectory, the quantities $(v_{uu}v_{ss})$ and (v_uv_s) are also conserved. This may be quickly verified by noting that

$$\frac{\mathrm{d}}{\mathrm{d}x}(v_uv_s) \;=\; v'_uv_s + v_uv'_s \;=\; 0$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}(v_{uu}v_{ss}) \;=\; v'_{uu}v_{ss} + v_{uu}v'_{ss} \;=\; 0.$$

It then follows that the summations

$$\sum_{j=2}^{k}\sum_{m=1}^{j} m\ a_{jm}\ (v_{uu}v_{ss})^{m-1}\,(v_uv_s)^{j-m}$$

and

$$\sum_{j=2}^{k}\sum_{m=0}^{j-1}(j-m)\ a_{jm}\ (v_{uu}v_{ss})^{m}\,(v_uv_s)^{j-m-1}$$

in (16) are also constant along trajectories. Thus, for any initial condition

$$(v_{uu}(0), v_{ss}(0), v_u(0), v_s(0))$$

there are constants given by

$$\gamma_1 = \sum_{j=2}^{k}\sum_{m=1}^{j} m\ a_{jm}\ (v_{uu}(0)v_{ss}(0))^{m-1}\,(v_u(0)v_s(0))^{j-m}$$

and

$$\gamma_2 = \sum_{j=2}^{k}\sum_{m=0}^{j-1}(j-m)\ a_{jm}\ (v_{uu}(0)v_{ss}(0))^{m}\,(v_u(0)v_s(0))^{j-m-1}$$

so that the solutions of (16) can be written as

$$
\begin{aligned}
v_{uu}(x) &= \mathrm{e}^{(1+\gamma_1)x}\, v_{uu}(0),\\
v_{ss}(x) &= \mathrm{e}^{-(1+\gamma_1)x}\, v_{ss}(0),\\
v_{u}(x) &= \mathrm{e}^{(r+\gamma_2)x}\, v_{u}(0),\\
v_{s}(x) &= \mathrm{e}^{-(r+\gamma_2)x}\, v_{s}(0).
\end{aligned}
\tag{17}
$$

In short, even though the system (5) is not smoothly linearizable near the origin, any set of initial conditions in a neighborhood about 0 evolves in a similarly predictable manner. Each trajectory behaves as though it were controlled by a linear system, but the growth and decay rates of this system vary slightly based on the initial conditions. The variations γ_1 and γ_2 approach zero as $\|(v_{uu}(0), v_{ss}(0), v_u(0), v_s(0))\| \to 0$.

3. A geometric proof of Theorem 1.

In this section, we use the properties laid out in the previous section to analyze the flow in $\mathbb{R}^4$ near $U_0(x)$ and to provide a proof of the bifurcation described in Theorem 1. We break this procedure down into the following four steps:

1. We define Poincaré cross-sections Σ_{in} and Σ_{out} transverse to the pulse $U_0(x)$ inside the region where the change of coordinates guaranteed by Lemma 4 is valid. This allows us to separate the behavior near the origin in phase space, where the behavior is *almost* linear, and the behavior away from the fixed point, where initial conditions evolve simply via rectilinear flow. Additionally, we use the Hamiltonian to reduce the effective dimensions of Σ_{in} and Σ_{out} for our search.

2. Let $\Phi_{\text{near}} : \Sigma_{\text{in}} \to \Sigma_{\text{out}} \cup S_1\Sigma_{\text{out}}$ be the local transition map induced by the evolution equation (5). In this regime, near the origin, the behavior is dominated by linear effects, and the orbit flip configuration described in Property 4 plays a key role.

3. Let Φ_{far} from Σ_{out} to Σ_{in} represent the global transition map. Away from any fixed points, the behavior is described by simple rectilinear flow. The properties of the transition map Φ_{far} are heavily influenced by the nearby parameter r^* where the pulse $U_0(x)$ is degenerate, as described in Property 5.

4. Using the reversibility described in Property 2, we identify a pair of windows, Ω_{near} in Σ_{out} and Ω_{far} in Σ_{in}, that are pulled back to reflections of themselves by the maps Φ_{near}^{-1} and Φ_{far}^{-1}, respectively. These important regions help us to identify critical features of the Poincaré map $\Phi_{\text{far}} \circ \Phi_{\text{near}}$ from Σ_{in} to $S_1\Sigma_{\text{in}}$ and therefore to isolate a mechanism for the creation of N-pulses.

3.1. Definition of the Poincaré sections Σ_{in} and Σ_{out}.

For any fixed parameter $0 < r < 1$, there is an integer k so that a C^3 canonical transformation takes system (5) into the truncated form given in (16) on some neighborhood of the origin. For simplicity, we normalize our variables so that the unit box is in the interior of this neighborhood. We then define the sections Σ_{in} and Σ_{out} as

$$\Sigma_{\text{in}} = \{(v_{uu}, v_{ss}, v_u, v_s) \mid v_{ss} = 1\}$$

and

$$\Sigma_{\text{out}} = \{(v_{uu}, v_{ss}, v_u, v_s) \mid v_{uu} = 1\}.$$

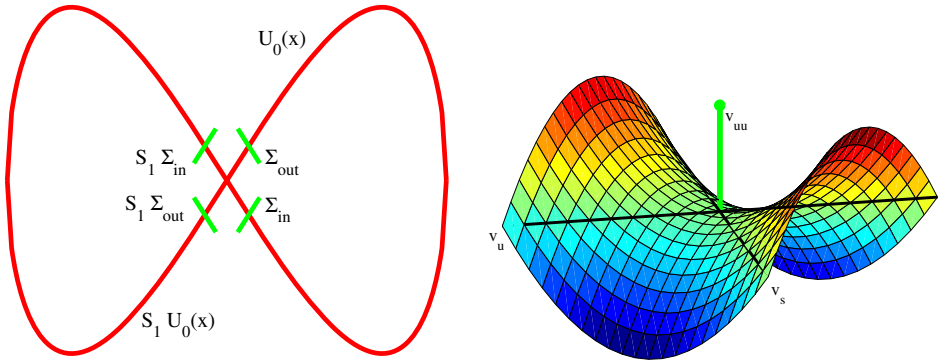


Figure 3. *The cross-sections Σ_{in}, $S_1\Sigma_{\text{in}}$, Σ_{out}, and $S_1\Sigma_{\text{out}}$ are set up transverse to the homoclinic orbits $U_0(x)$ and $S_1U_0(x)$. The first panel shows the two dimensional invariant subspace U containing the pulses $U_0(x)$ and $S_1U_0(x)$; the intersection of the Poincaré sections with this space are shown. The second panel shows the two dimensional submanifold within the three dimensional section Σ_{in} for which $\{H = 0\}$; the axis v_{uu} corresponds to the set $\Sigma_{\text{in}} \cap U$ shown in the first panel.*

Because the pulse $U_0(x)$ is in orbit-flip configuration (Property 4), these three dimensional sections Σ_{in} and Σ_{out} are transverse to the homoclinic orbit $U_0(x)$ and intersect $U_0(x)$ at the points $(0, 1, 0, 0)$ and $(1, 0, 0, 0)$, respectively. The homoclinic orbit $S_1U_0(x)$ meets the sections $S_1\Sigma_{\text{in}}$ and $S_1\Sigma_{\text{out}}$ similarly. See Figure 3.

Note further that any solution to the system (16) that approaches the origin as $x \to \pm\infty$ must lie on the same level set of the Hamiltonian H_k as the origin itself, i.e., $H_k = 0$ on such sets. Considering the intersections of the sets Σ_{in} and Σ_{out} with the level set of the Hamiltonian $\{(v_{uu}, v_{ss}, v_u, v_s) \mid H_k(v_{uu}, v_{ss}, v_u, v_s) = 0\}$, we can restrict our attention to two dimensional manifolds within either Σ_{in} or Σ_{out}. In particular, for v_{uu}, v_u, and v_s sufficiently small, the implicit function theorem guarantees that any point $(v_{uu}, 1, v_u, v_s)$ in the set Σ_{in} satisfying $H_k = 0$ can be written as $(v_{uu}(v_s, v_u), 1, v_u, v_s)$. Similarly, any point in Σ_{out}, again with v_{ss}, v_u, and v_s sufficiently small, that satisfies $H_k = 0$ can be written as $(1, v_{ss}(v_s, v_u), v_u, v_s)$. In this way, the maps $\Phi_{\text{far}} : \Sigma_{\text{out}} \to \Sigma_{\text{in}}$ and $\Phi_{\text{near}} : \Sigma_{\text{in}} \to \Sigma_{\text{out}} \cup S_1\Sigma_{\text{out}}$ can be specified simply as maps from a subset of $\mathbb{R}^2$ into $\mathbb{R}^2$. In what follows, we will restrict our analysis to the level set $H_k = 0$ inside Σ_{in} and Σ_{out}. With a slight abuse of notation, we will continue to refer to these restricted sets as Σ_{in} and Σ_{out} and treat the maps Φ_{near} and Φ_{far} as transformations of the two variables v_s and v_u.

3.2. Properties of the near map Φ_{near}.

In this subsection, we describe the transition map Φ_{near} that details the passage of trajectories near the origin. The flow in this region is dominated by the presence of the fixed point, and we have already seen that in this regime the flow is *almost* linear, as given in (17).

The map Φ_{near} takes the section Σ_{in} to $\Sigma_{\text{out}} \cup S_1\Sigma_{\text{out}}$. (More precisely, we consider only those initial conditions in Σ_{in} that leave the unit box in $\mathbb{R}^4$ through Σ_{out} or $S_1\Sigma_{\text{out}}$.) First, observe that whether Φ_{near} takes an initial condition in this domain to Σ_{out} or $S_1\Sigma_{\text{out}}$ is determined solely by the value of v_{uu} associated with (v_s, v_u) in Σ_{in}. If $v_{uu} > 0$, then

$\Phi_{\text{near}}(v_s, v_u) \in \Sigma_{\text{out}}$. Likewise, if $v_{uu} < 0$, then $\Phi_{\text{near}}(v_s, v_u) \in S_1\Sigma_{\text{out}}$. In particular, since

$$v_{uu} = -rv_sv_u + O(v_s^2 v_u^2),$$

initial conditions (v_s, v_u) in the first and third quadrants are taken to $S_1\Sigma_{\text{out}}$, while initial conditions in the second and fourth quadrants are taken to Σ_{out}; see the second panel of Figure 3.

Evolving via (17), one can quickly determine the behavior of the map Φ_{near}. In particular, since $|v_{uu}| = 1$ in either exit slice Σ_{out} or $S_1\Sigma_{\text{out}}$, the elapsed time T between the slices is determined to be

$$\begin{aligned} 1 &= |v_{uu}|\mathrm{e}^{(1+\gamma_1)T}, \\ |v_{uu}|^{-1} &= \mathrm{e}^{(1+\gamma_1)T}, \\ -\ln|v_{uu}| &= (1+\gamma_1)T, \\ T &= \tfrac{-1}{1+\gamma_1}\ln|v_{uu}|. \end{aligned}$$

Since γ_1 and γ_2 are both $O(v_sv_u)$, the map Φ_{near} takes the form

$$\begin{aligned} &\Phi_{\text{near}}(v_s, v_u) \\ &= \left(v_s\mathrm{e}^{-(r+\gamma_2)T}, v_u\mathrm{e}^{(r+\gamma_2)T}\right) \\ &= \left(v_s\mathrm{e}^{\frac{r+\gamma_2}{1+\gamma_1}\ln|v_{uu}|}, v_u\mathrm{e}^{-\frac{r+\gamma_2}{1+\gamma_1}\ln|v_{uu}|}\right) \\ &= \left(v_s|v_{uu}|^{\frac{r+\gamma_2}{1+\gamma_1}}, v_u|v_{uu}|^{-\frac{r+\gamma_2}{1+\gamma_1}}\right) \\ &= \left(v_s|rv_sv_u + O(v_s^2v_u^2)|^{r+O(v_sv_u)}, v_u|rv_sv_u + O(v_s^2v_u^2)|^{-r+O(v_sv_u)}\right). \end{aligned}$$

Although it may not be readily apparent from this equation, the qualitative behavior of Φ_{near} varies dramatically depending upon whether or not the parameter r satisfies $r < 1/2$. In particular, we will note the differing behavior of Φ_{near} in two separate examples:

1. for short curves in Σ_{in} terminating on the stable manifold at $(v_s, v_u) = (0, 0)$, and
2. for short curves in Σ_{in} terminating elsewhere along the stable manifold, i.e., where $v_s \neq 0$.

3.2.1. Φ_{near} acting on curves approaching $(0, 0)$. For the first case, we consider a curve Γ in Σ_{in} parameterized as $(v_s, v_u) = (t, at+O(t^2))$ for some $a \neq 0$ and t small. We now consider the manner in which the map Φ_{near} acts upon Γ:

$$\begin{aligned} &\Phi_{\text{near}}(t, at + O(t^2)) \\ &= \left(t|art^2 + O(t^3)|^{r+O(t^2)}, (at + O(t^2))|art^2 + O(t^3)|^{-r+O(t^2)}\right) \\ &= \Big(t|t|^{2(r+O(t^2))}|ar + O(t)|^{r+O(t^2)}, \\ &\qquad (at|t|^{-2(r+O(t^2))} + O(t^{2-2r}))|ar + O(t)|^{-r+O(t^2)}\Big) \\ &\to \left(|ar|^r\ (\operatorname{sign} t)\ |t|^{1+2r}, a|ar|^{-r}\ (\operatorname{sign} t)\ |t|^{1-2r}\right), \end{aligned}$$

as $t \to 0$. For $r < 1/2$, $v_s = O((v_u)^{\frac{1+2r}{1-2r}})$, and so $\Phi_{\text{near}}(\Gamma)$ approaches the origin tangent to the v_u-axis. On the other hand, when $r > 1/2$, the first component of $\Phi_{\text{near}}(\Gamma)$ again goes to zero, but the second component is unbounded. In this case the portion of Γ nearest the origin is taken by Φ_{near} out of any small neighborhood of $(0,0)$ in Σ_{out}.

3.2.2. Φ_{near} acting on curves approaching $(b,0)$, $b \neq 0$. In the second case, we consider a curve Γ parameterized by $(b - ct + O(t^2), t)$ for $b, c \neq 0$, under the action of the map Φ_{near}:

$$\begin{aligned}
&\Phi_{\text{near}}(b - ct + O(t^2), t) \\
&= \left((b - ct + O(t^2))|brt + O(t^2)|^{r+O(t)}, t|brt + O(t^2)|^{-r+O(t)}\right) \\
&= \Big((b - ct + O(t^2))|t|^{r+O(t)}|br + O(t)|^{r+O(t)}, \\
&\qquad\qquad t|t|^{-r+O(t)}|br + O(t)|^{-r+O(t)}\Big) \\
&\to \left(b|br|^r\ |t|^r, |br|^{-r}\ (\operatorname{sign} t)\ |t|^{1-r}\right),
\end{aligned}$$

as $t \to 0$. For any $0 < r < 1$, $\Phi_{\text{near}}(\Gamma)$ approaches the origin. However, the direction of approach again depends upon whether or not $r > 1/2$. If $r < 1/2$, then since $v_u = O((v_s)^{\frac{1-r}{r}})$, $\Phi_{\text{near}}(\Gamma)$ approaches zero tangent to the v_s-axis. On the other hand, for $r > 1/2$, we have $v_s = O((v_u)^{\frac{r}{1-r}})$ and $\Phi_{\text{near}}(\Gamma)$ approaches zero tangent to the v_u-axis.

3.3. Properties of the far map Φ_{far}. In this subsection, we describe the transition map Φ_{far} that details the behavior near the pulse $U_0(x)$ during its excursion far from the origin. Since there are no fixed points in this region, the behavior is governed by rectilinear flow, and the map Φ_{far} is a local diffeomorphism.

For r values nearby a parameter $r = r^*$ where the twin homoclinic orbits $U_0(x)$ and $S_1U_0(x)$ are degenerate, as in Property 5, the nearby degeneracy will play a major role in defining the behavior of the far map.

Note that the coordinate system given by v_{uu}, v_{ss}, v_u, and v_s in section 2.2 is valid only in a small neighborhood of the origin. The pulse $U_0(x)$ likely leaves this region, and thus we would need to switch to a different coordinate system as we follow the flow near the pulse. However, the sections Σ_{out} and Σ_{in} both lie within the region where these coordinates are available, and so in writing the map Φ_{far} from Σ_{out} to Σ_{in} it is most convenient to use this same coordinate system.

Now, suppose that at the parameter value $r = r^*$, the homoclinic orbit $U_0(x)$ is degenerate, as specified in Property 5. At such a parameter value, Φ_{far} takes the local unstable manifold $W^u_{\text{loc}}(0)$ of the origin, which coincides with the v_u-axis in Σ_{out}, to a curve that is tangent to the local stable manifold $W^s_{\text{loc}}(0)$, i.e., the v_s-axis in Σ_{in}. Conversely, Φ^{-1}_{far} takes the v_s-axis in Σ_{in} to a curve in Σ_{out} that is tangent at the origin to the v_u-axis. In particular, as a result of the reversibility in Property 2, the following two conditions hold for the map Φ_{far} when $r = r^*$:

$$D_2\Phi^{(1)}_{\text{far}}(0,0) = \pm 1 \quad \text{and} \quad D_2\Phi^{(2)}_{\text{far}}(0,0) = 0,$$

where the notation $D_i\Phi^{(j)}_{\text{far}}$ refers to the derivative of the jth component of Φ_{far} with respect to the ith variable.

For parameters $r > r^*$, Lemma 2 purports that a tangent vector to the local unstable manifold $W^u_{\text{loc}}(0)$ undergoes less spinning between Σ_{out} and Σ_{in} than it does for $r = r^*$; so for ϵ small and positive and for $r = r^* + \epsilon$, it holds that

$$\operatorname{sign} D_2\Phi^{(1)}_{\text{far}}(0,0) = \operatorname{sign} D_2\Phi^{(2)}_{\text{far}}(0,0), \tag{18}$$

where $D_2\Phi^{(1)}_{\text{far}}(0,0) = \pm 1 + O(\epsilon)$ and $D_2\Phi^{(2)}_{\text{far}}(0,0) = O(\epsilon)$.

Additionally, regardless of the value of r, all of the derivatives of even order vanish because of the symmetry given in Property 3. For r near r^*, we make the following assumption regarding the map Φ_{far}:

$$\operatorname{sign} \mathrm{D}_{222}\Phi^{(2)}_{\text{far}}(0,0) = -\operatorname{sign} D_2\Phi^{(1)}_{\text{far}}(0,0). \tag{19}$$

We would like to discuss this assumption in a bit more depth—it seems quite natural and can be *roughly* considered in the context of the winding discussed in subsection 2.1. Recall that the winding of the invariant manifolds along the pulse $U_0(x)$ is seen to decrease with increasing r in (13). However, the inclusion of a nontrivial w component for any fixed r will, in essence, work to increase the winding by increasing the influence of the nonlinearity. In particular, had we included the nonlinear terms when deriving (13), the equation for θ would have been

$$\theta' = \left(r^2 - h_2(U_c^2(\tau), \rho^2\cos^2(\theta))\right)\cos^2(\theta) \ - \ \sin^2(\theta).$$

This equation, which was independent of ρ in the linear regime, now shows a dependence upon ρ both explicitly in the second component of the nonlinearity as $\rho^2\cos^2(\theta)$ and implicitly in the first component, as it influences the equation for $U_c^2(\tau)$, which is no longer decoupled. For a focusing nonlinearity $h_2(\cdot)$, the explicit action of the nontrivial second component is to increase the winding in comparison with the linear case. Although the action through the change in the profile of $U_c(\tau)$ is more subtle, we believe, but do not prove, that the condition (19) holds in general.

3.4. The grand finale. In this subsection, we combine the properties of the transition maps Φ_{far} and Φ_{near} in order to complete the proof of Theorem 1.

Suppose that the one-component pulse $U_0(x)$ is degenerate for a parameter $r = r^*$, with $0 < r^* < 1$ as described in Property 5. Lemma 4 guarantees that, since $r^* > 0$, there is a constant k so that the ordinary differential equation (5) can be reduced to its resonant normal form of degree k in a neighborhood of the origin, for any r between r^* and 1. This neighborhood can be chosen uniformly in the interval between r^* and 1, and since the constants a_{jm} vary smoothly with r, the sections Σ_{in} and Σ_{out} inside this neighborhood also vary smoothly with r.

Recall that $D_2\Phi^{(2)}_{\text{far}}(0,0) = 0$ when $r = r^*$. Now, suppose that the condition given in (19) is satisfied for $r = r^*$. Continuity dictates that this same relationship must then hold for all r in a small neighborhood of r^*. In particular, $D_{222}\Phi^{(2)}_{\text{far}}(0,0)$ is nonzero on this neighborhood and therefore can be bounded away from zero on some (possibly smaller) neighborhood.

$\Phi^{(2)}_{\text{far}}$ can then be seen to have the form

$$\Phi^{(2)}_{\text{far}}(0,h) = h\ D_2\Phi^{(2)}_{\text{far}}(0,0) \ + \ \frac{h^3}{6}\ D_{222}\Phi^{(2)}_{\text{far}}(0,0) + O(h^5).$$

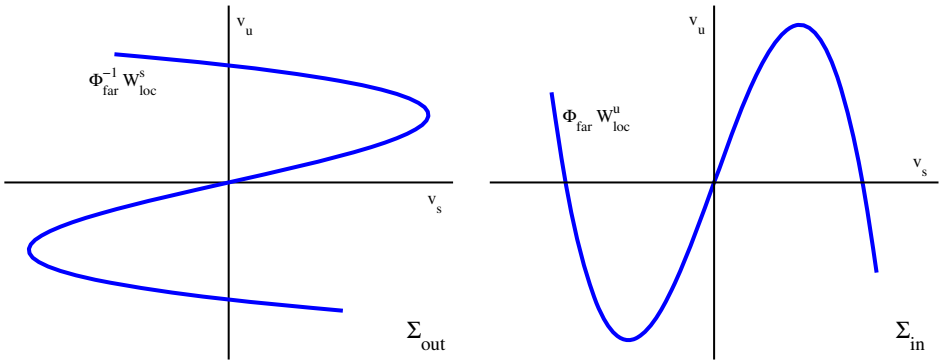


Figure 4. *The map Φ_{far}. When $r = r^*$, the map Φ_{far} takes the local unstable manifold W^u_{loc} in Σ_{out} to a curve that is tangent to the v_s-axis in Σ_{in}. For r slightly beyond r^*, these curves are no longer tangent and a symetric pair of intersections of $\Phi_{\text{far}} W^u_{\text{loc}}$ and the v_s-axis appear. (Note that the extreme scaling obscures the tangency in this figure; in the right panel, the v_u-axis is blown up roughly 100-fold versus the v_s-axis; the opposite holds on the left.) Because of the reversibility, the curves $\Phi^{-1}_{\text{far}} W^s_{\text{loc}}$ and the v_u-axis intersect similarly in Σ_{out}.*

Since $D_2\Phi^{(2)}_{\text{far}}(0,0)$ and $D_{222}\Phi^{(2)}_{\text{far}}(0,0)$ are of opposite sign on some interval $r^* < r < r^* + \epsilon_1$, with $D_2\Phi^{(2)}_{\text{far}}(0,0) = 0$ when $r = r^*$ and $D_{222}\Phi^{(2)}_{\text{far}}(0,0)$ bounded away from zero, we quickly see that on some possibly smaller interval $r^* < r < r^* + \epsilon_2$ there is a positive value for h such that $\Phi^{(2)}_{\text{far}}(0,h) = \Phi^{(2)}_{\text{far}}(0,-h) = 0$. These roots represent nontrivial intersections in Σ_{in} of the local stable manifold W^s_{loc} of the origin and the global unstable manifold W^u (in particular, the image under Φ_{far} of the local unstable manifold W^u_{loc}). Therefore, these intersections can be identified as the multicomponent 1-pulse (and its reflection via the symmetry S_2) described in the theorem. See Figure 4.

Note that we were able to use properties of the far map Φ_{far} alone to eke out the bifurcation of a multicomponent 1-pulse from the "fast" single component pulse $U_0(x)$ near a parameter value $r = r^*$ where $U_0(x)$ is degenerate. In order to find the alternating N-pulses, we will be required to study the interaction of the transition maps Φ_{far} and Φ_{near}.

Before continuing further, we describe the vital role played by the reversibility, as described in Property 2. (Both the reversibility involution R and the symmetries S_1 and S_2 can be easily translated into the new coordinate system.) Because the Poincaré sections Σ_{in} and Σ_{out} themselves are reflections of one another through the reversibility involution, reversibility forces a particular symmetry in the maps Φ_{far} and Φ_{near} in relation to their inverses. Specifically, for any pairs (ξ_s, ξ_u) in Σ_{out} and (ζ_s, ζ_u) in Σ_{in}, if Φ_{far} takes (ξ_s, ξ_u) to (ζ_s, ζ_u), then Φ^{-1}_{far} must take the reflection (ξ_u, ξ_s) in Σ_{in} back to the point (ζ_u, ζ_s) in Σ_{out}. Likewise, if Φ_{near} takes (ζ_s, ζ_u) to (ξ_s, ξ_u), then Φ^{-1}_{near} must take (ζ_u, ζ_s) back to (ξ_u, ξ_s).

For r in the interval where a multicomponent 1-pulse can be found, i.e., $r^* < r < r^* + \epsilon_2$, we can now identify two small regions, one in Σ_{in} and one in Σ_{out}, that are pulled back by the reverse flow to their reflections in Σ_{out} and $S_1\Sigma_{\text{in}}$, respectively.

First, consider a region Ω_{far} in Σ_{in}, defined as a bounded region between $\Phi_{\text{far}} W^u_{\text{loc}}$ and W^s_{loc}. In the right-hand side of Figure 4, this is the area between $\Phi_{\text{far}} W^u_{\text{loc}}$ and the v_s-axis. Φ^{-1}_{far} clearly takes this region to the one bounded between W^u_{loc} (the v_u-axis) and $\Phi^{-1}_{\text{far}} W^s_{\text{loc}}$

in Σ_{out}; see the left-hand side of Figure 4. However, each of these two curves is related to the preimage of the other via the reversibility involution, and so Ω_{far} is simply taken to its reflection $R\Omega_{\text{far}}$ by Φ_{far}^{-1}. (Be careful to note that this identification as a reflection holds as a set and not pointwise.)

Second, we consider the subset of $\Phi_{\text{far}}^{-1}(\Omega_{\text{far}})$ in Σ_{out} that lies within $S_1\Phi_{\text{near}}(\Omega_{\text{far}})$. In particular, we let

$$\Omega_{\text{near}} = \Phi_{\text{far}}^{-1}(\Omega_{\text{far}}) \cap S_1\Phi_{\text{near}}(\Omega_{\text{far}}).$$

The qualitative shape of Ω_{near} will vary depending upon whether or not $r > 1/2$; however, in either case, Ω_{near} is taken to its reflection $R\ S_1\Omega_{\text{near}}$ in $S_1\Sigma_{\text{in}}$ by the inverse of the local transition map, Φ_{near}^{-1}. See the central panel of either Figure 5 or Figure 6.

Note that both the reflected preimage $S_1\Phi_{\text{near}}^{-1}(\Omega_{\text{near}})$ and the image $\Phi_{\text{far}}(\Omega_{\text{near}})$ lie within Ω_{far} in Σ_{in}. We remark that Ω_{near} is a subset of $\Phi_{\text{far}}^{-1}(\Omega_{\text{far}})$ and that the return map $\Phi_{\text{near}} \circ \Phi_{\text{far}}$ takes Ω_{near} to a different subset of $S_1\Phi_{\text{far}}^{-1}(\Omega_{\text{far}})$. Careful consideration of curves within Ω_{near} under the action of this composite map leads to the proof of Theorem 1.

We will now complete the proof of Theorem 1, considering the two cases separately. First, we consider the case where $r^* < 1/2$. Then we will consider the case when $r^* \geq 1/2$.

3.4.1. The case $r^* < 1/2$. Suppose that the parameter r^* for which the pulse $U_0(x)$ is degenerate satisfies $r^* < 1/2$. For r in the interval $r^* < r < r^* + \epsilon_2$, we have seen that a multicomponent 1-pulse can be found near $U_0(x)$. We further choose ϵ_3 with $0 < \epsilon_3 \leq \epsilon_2$ so that ϵ_3 satisfies $r^* + \epsilon_3 < 1/2$. We now prove that there are alternating N-pulses near $U_0(x)$ for any r satisfying $r^* < r < r^* + \epsilon_3$. In particular, we will prove the following lemma.

Lemma 5. *Suppose that $r^* < r < r^* + \epsilon_3$ with $r^* + \epsilon_3 < 1/2$. Consider any simple smooth curve Γ_0 in the window Ω_{near} in Σ_{out} having the following two properties:*

i. *Γ_0 originates at $(0,0)$ tangent to W^u_{loc}, i.e., the v_u-axis.*
ii. *Γ_0 terminates at a point on $\Phi_{\text{far}}^{-1}W^s_{\text{loc}}$ distinct from the origin.*

Then the image $\Phi_{\text{near}} \circ \Phi_{\text{far}}(\Gamma_0)$ in $S_1\Sigma_{\text{out}}$ also contains a simple smooth curve Γ_1 within $S_1\Omega_{\text{near}}$ satisfying these same properties.

The proof of this lemma concludes the proof of Theorem 1 in this case. We have already described the multicomponent 1-pulse as an intersection between the curve $\Phi_{\text{far}}W^u_{\text{loc}}$ and the v_s axis within the Poincaré section Σ_{in}. For $r^* < r < r^* + \epsilon_3$, we will see that the loop $\Phi_{\text{near}} \circ \Phi_{\text{far}}W^u_{\text{loc}}$ contains a curve satisfying the two properties of the lemma. The intersection of this curve with $\Phi_{\text{far}}^{-1}W^s_{\text{loc}}$ represents the alternating 2-pulse of the theorem. The lemma then implies that every iterate $(\Phi_{\text{near}} \circ \Phi_{\text{far}})^{(N-1)}W^u_{\text{loc}}$, $N \geq 2$, also intersects $\Phi_{\text{far}}^{-1}W^s_{\text{loc}}$ nontrivially, and we immediately identify these intersections with the desired N-pulses.

Proof. We begin by examining the shape of the region Ω_{near} in Σ_{out} for this case where $r < 1/2$. We describe only the portion of Ω_{near} in the first quadrant of Σ_{out}; the portion in the third quadrant is equivalent by way of the symmetry S_2.

Now Ω_{near} is contained within the set $S_1\Phi_{\text{near}}(\Omega_{\text{far}})$. Ω_{far} was defined to be the set bounded between the v_s-axis and the curve $\Phi_{\text{far}}W^u_{\text{loc}}$ in Σ_{in} (recall the second panel of Figure 4, and now also the first and last frames of Figure 5). Via the preliminary analysis of section 3.2, the portion of the bounding curve $\Phi_{\text{far}}W^u_{\text{loc}}$ that approaches the origin is taken by Φ_{near} to a curve approaching the origin and tangent to the v_u-axis in $S_1\Sigma_{\text{out}}$. Since the portion of $\Phi_{\text{far}}W^u_{\text{loc}}$

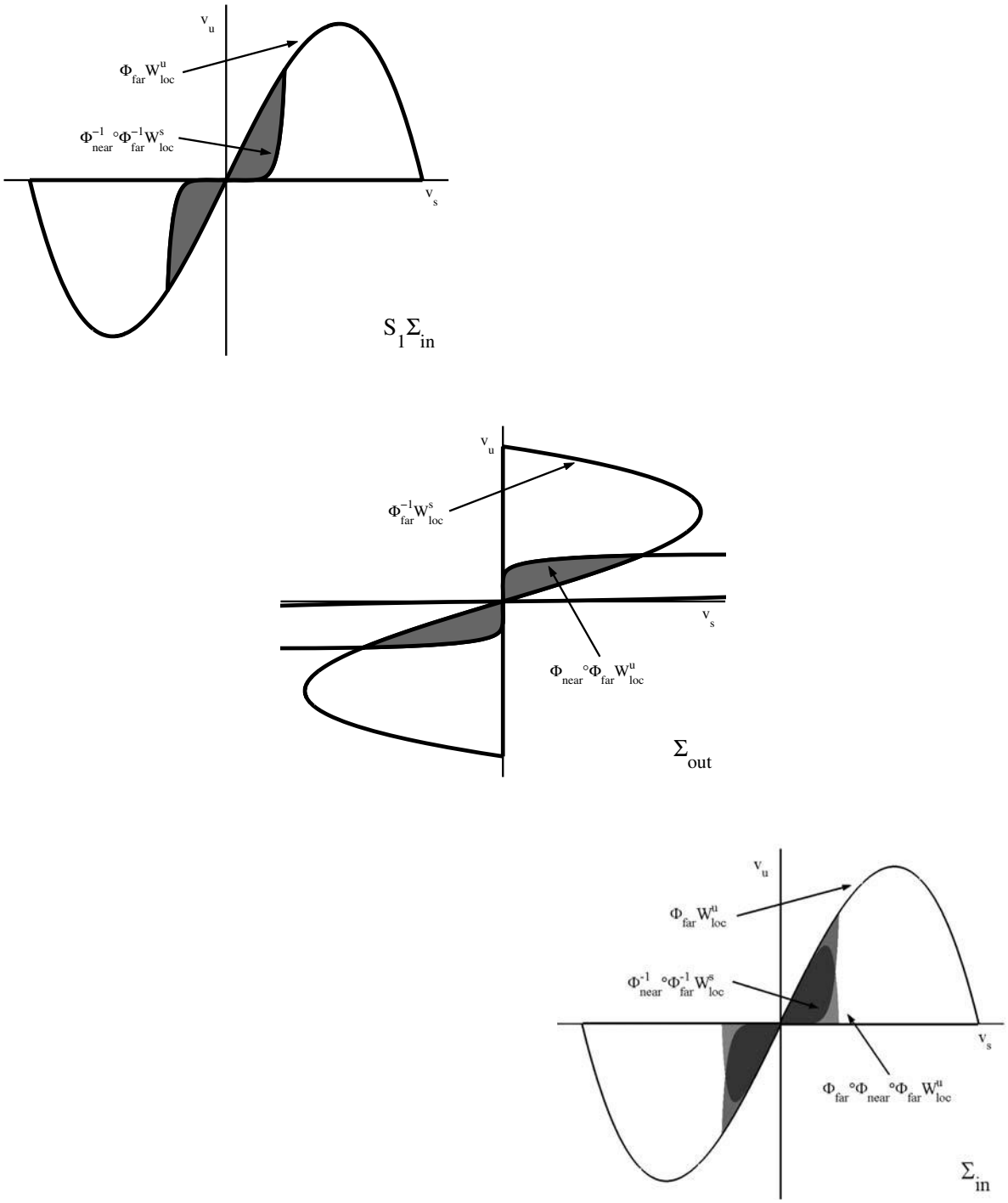


Figure 5. *The flow between successive sections in the case $r^* < 1/2$. The shaded area in the middle frame is Ω_{near}. This area is pulled back via Φ^{-1}_{near} to its reflection in $S_1\Sigma_{\text{in}}$, shown in the top frame. Ω_{near} is also mapped forward by Φ_{far} to another subset of Ω_{far}, as seen in the bottom frame. The manner of intersection of the sets $S_1\Phi^{-1}_{\text{near}}(\Omega_{\text{near}})$ and $\Phi_{\text{far}}(\Omega_{\text{near}})$ as shown in this final frame is important to the creation of N-pulses.*

near the nontrivial intersection point with W^s_{loc} is taken by Φ_{near} to a curve approaching the origin and tangent to the v_s-axis in $S_1\Sigma_{\text{out}}$, the image $\Phi_{\text{near}} \circ \Phi_{\text{far}} W^u_{\text{loc}}$ is a closed loop in the first quadrant of $S_1\Sigma_{\text{out}}$. See, for example, the central frame of Figure 5, where this closed loop extends out of the right-hand boundary of the figure. The region $\Phi_{\text{near}}(\Omega_{\text{far}})$ is bounded by this closed loop, and Ω_{near} is the intersection of the reflection of this region $S_1\Phi_{\text{near}}(\Omega_{\text{far}})$ with $\Phi^{-1}_{\text{far}}(\Omega_{\text{far}})$ in Σ_{out}, shown as the shaded region of the central panel of Figure 5.

The boundary of $\Phi^{-1}_{\text{near}}(\Omega_{\text{near}})$ in $S_1\Sigma_{\text{in}}$ contains a portion of the curve $\Phi^{-1}_{\text{near}} \circ \Phi^{-1}_{\text{far}} W^s_{\text{loc}}$ within $S_1\Omega_{\text{far}}$ that leaves the origin tangent to the v_s-axis before leaving $S_1\Omega_{\text{far}}$ through the curve $\Phi_{\text{far}} W^u_{\text{loc}}$. See the upper frame of Figure 5. Note that $\Phi^{-1}_{\text{near}} \circ \Phi^{-1}_{\text{far}} W^s_{\text{loc}}$ must leave $S_1\Omega_{\text{far}}$ since it returns to the origin tangent to the v_u-axis. In fact, the point where $\Phi^{-1}_{\text{near}} \circ \Phi^{-1}_{\text{far}} W^s_{\text{loc}}$ leaves $S_1\Omega_{\text{far}}$ through $\Phi_{\text{far}} W^u_{\text{loc}}$ can be identified with the alternating 2-pulse, as promised above.

Suppose that Γ_0 is a simple smooth curve in the window Ω_{near} in Σ_{out} originating tangent to W^u_{loc} at $(0,0)$, and terminating at another point on $\Phi^{-1}_{\text{far}} W^s_{\text{loc}}$. Γ_0 is taken by the transition map Φ_{far} to a curve $\Phi_{\text{far}}(\Gamma_0)$ within Ω_{far} in Σ_{in} that is tangent to $\Phi_{\text{far}} W^u_{\text{loc}}$ at the origin and meets the manifold W^s_{loc}, i.e., the v_s-axis, at another point.

The end of $\Phi_{\text{far}}(\Gamma_0)$ approaching $(0,0)$ is tangent to $\Phi_{\text{far}} W^u_{\text{loc}}$ and lies within $S_1\Phi^{-1}_{\text{far}}(\Omega_{\text{near}})$. Except for the point $(0,0)$, however, the v_s-axis is not in the set $S_1\Phi^{-1}_{\text{far}}(\Omega_{\text{near}})$ and therefore, since $\Phi_{\text{far}}(\Gamma_0)$ terminates on this axis, it must cross $\Phi^{-1}_{\text{near}} \circ \Phi^{-1}_{\text{far}} W^s_{\text{loc}}$ at some point.

Under the action of Φ_{near}, $\Phi_{\text{far}}(\Gamma_0)$ is then clearly taken to a curve in Ω_{near} that leaves $(0,0)$ tangent to the v_u-axis and intersects $\Phi^{-1}_{\text{far}} W^s_{\text{loc}}$ at some other point. ■

3.4.2. The case $r^* \geq 1/2$. Now, suppose that the parameter r^* for which the pulse $U_0(x)$ is degenerate satisfies $r^* \geq 1/2$. For r in the interval $r^* < r < r^* + \epsilon_2$, we have seen that a multicomponent 1-pulse can be found near $U_0(x)$. Here, we further choose ϵ_3 with $0 < \epsilon_3 \leq \epsilon_2$ so that ϵ_3 satisfies $r^* + \epsilon_3 < 1$. We now prove that there are alternating N-pulses near $U_0(x)$ for any r satisfying $r^* < r < r^* + \epsilon_3$. In particular, we will prove the following lemma.

Lemma 6. *Suppose that $r^* < r < r^* + \epsilon_3$ with $r^* \geq 1/2$ and $r^* + \epsilon_3 < 1$. Consider any simple smooth curve Γ_0 in the window Ω_{near} in Σ_{out} with the following two properties:*

i. *Γ_0 originates at $(0,0)$.*

ii. *Γ_0 terminates at a point on $\Phi^{-1}_{\text{far}} W^s_{\text{loc}}$, distinct from the origin.*

Then the image $\Phi_{\text{near}} \circ \Phi_{\text{far}}(\Gamma_0)$ in $S_1\Sigma_{\text{out}}$ also contains a simple smooth curve Γ_1 within $S_1\Omega_{\text{near}}$ satisfying these same properties.

In this case, we need not assume that Γ_0 is tangent to W^u_{loc}, i.e., the v_u-axis at $(0,0)$. For $r > 1/2$, any curve in Ω_{near} originating at $(0,0)$ automatically satisfies this property. As above, the proof of this lemma concludes the proof of Theorem 1 for this case.

Proof. The region Ω_{near} in Σ_{out} has a qualitatively different shape when $r > 1/2$ than it does otherwise. We again describe only the portion of Ω_{near} in the first quadrant of Σ_{out}.

In this case, recalling once more the preliminary analysis of section 3.2, the portion of $\Phi_{\text{far}} W^u_{\text{loc}}$ that approaches the origin in $S_1\Sigma_{\text{in}}$ is taken by Φ_{near} out of any bounded region of Σ_{out}. Since the portion of $\Phi_{\text{far}} W^u_{\text{loc}}$ near the nontrivial intersection point with W^s_{loc} is now taken by Φ_{near} to a curve approaching the origin and tangent to the v_u-axis in Σ_{out}, the image $\Phi_{\text{near}} \circ \Phi_{\text{far}} W^u_{\text{loc}}$ is an unbounded curve in the first quadrant of Σ_{out}. See, for example,

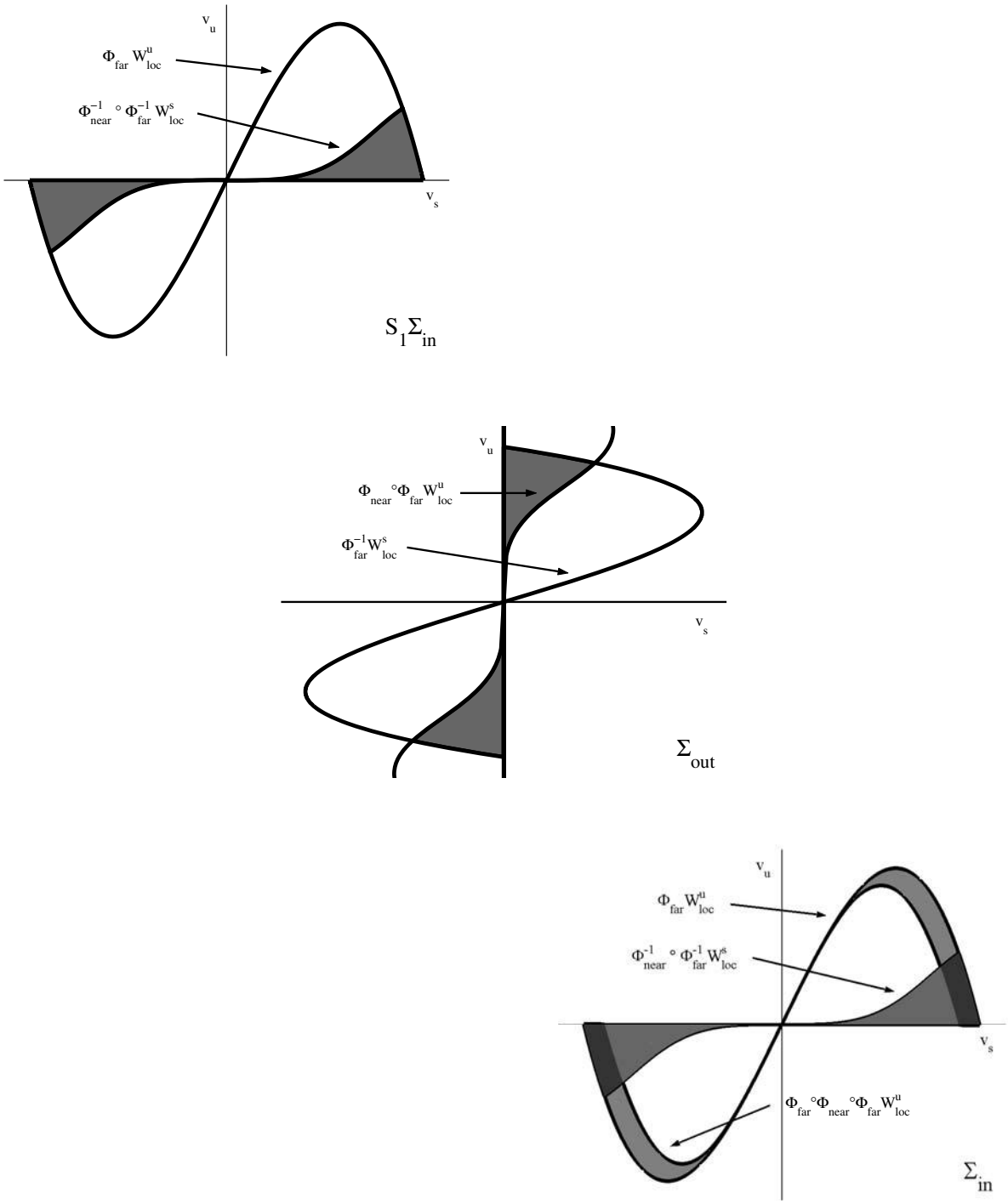


Figure 6. *The flow between the successive sections for the case $r^* \geq 1/2$. The shaded area in the middle frame is Ω_{near}. This area is again pulled back via Φ_{near}^{-1} to its reflection in $S_1\Sigma_{\text{in}}$, shown in the top frame and mapped forward by Φ_{far} to another subset of Ω_{far}, as seen in the bottom frame. The manner of intersection of the sets $S_1\Phi_{\text{near}}^{-1}(\Omega_{\text{near}})$ and $\Phi_{\text{far}}(\Omega_{\text{near}})$ as shown in this final frame is important to the creation of N-pulses.*

the central panel of Figure 6. The region $\Phi_{\text{near}}(\Omega_{\text{far}})$ is bounded between this curve and the v_u-axis, and Ω_{near} is the intersection of this region's reflection $S_1\Phi_{\text{near}}(\Omega_{\text{far}})$ with $\Phi_{\text{far}}^{-1}(\Omega_{\text{far}})$, now shown as the shaded region in the central panel of Figure 6.

The set $\Phi_{\text{near}}^{-1}(\Omega_{\text{near}})$ in $S_1\Sigma_{\text{in}}$ is again a reflection of Ω_{near} in Σ_{out} via the reversibility involution R, as shown in the upper left panel of Figure 6. The boundary of $\Phi_{\text{far}}^{-1}(\Omega_{\text{near}})$ contains a portion of the curve $\Phi_{\text{near}}^{-1} \circ \Phi_{\text{far}}^{-1} W_{\text{loc}}^s$ within $S_1\Omega_{\text{far}}$ that leaves the origin tangent to the v_s-axis before leaving $S_1\Omega_{\text{far}}$ through the curve $\Phi_{\text{far}} W_{\text{loc}}^u$. This follows since $\Phi_{\text{near}}^{-1} \circ \Phi_{\text{far}}^{-1} W_{\text{loc}}^s$ must leave the bounded region $S_1\Omega_{\text{far}}$, and the crossing again indicates the alternating 2-pulse in the theorem.

Suppose that Γ_0 is a simple smooth curve in the window Ω_{near} in Σ_{out} connecting $(0,0)$, with another point on $\Phi_{\text{far}}^{-1} W_{\text{loc}}^s$. Γ_0 is taken by the transition map Φ_{far} to a curve $\Phi_{\text{far}}(\Gamma_0)$ within Ω_{far} in Σ_{in} that is tangent to $\Phi_{\text{far}} W_{\text{loc}}^u$ at the origin and meets the manifold W_{loc}^s, i.e., the v_s-axis, at another point.

The end of $\Phi_{\text{far}}(\Gamma_0)$ near the nontrivial intersection with the v_s-axis lies within $S_1\Phi_{\text{far}}^{-1}(\Omega_{\text{near}})$. However, $\Phi_{\text{far}}(\Gamma_0)$ is tangent at the origin to the curve $\Phi_{\text{far}} W_{\text{loc}}^u$, while any curve within $\Phi_{\text{far}}^{-1}\Omega_{\text{near}}$ must be tangent to the v_u-axis. So $\Phi_{\text{far}}(\Gamma_0)$ must have left $S_1\Phi_{\text{far}}^{-1}(\Omega_{\text{near}})$, crossing $\Phi_{\text{near}}^{-1} \circ \Phi_{\text{far}}^{-1} W_{\text{loc}}^s$ at some point before its approach to the origin.

Under the action of Φ_{near}, $\Phi_{\text{far}}(\Gamma_0)$ is then clearly taken to a curve in Ω_{near} connecting $(0,0)$ with some other point on $\Phi_{\text{far}}^{-1} W_{\text{loc}}^s$. This completes the proof. ■

4. Discussion. We have shown the mechanism for the bifurcation of not only a multicomponent 1-pulse but also alternating N-pulses, for all positive N, near a parameter value $r = r^*$ where the original fast single component pulse is degenerate. The geometry of the bifurcation differed significantly depending upon whether or not the parameter r^* was smaller than $1/2$. In fact, when $r^* \geq 1/2$, the bifurcation can be described as a *degenerate cusp horseshoe*, where the degeneracy here refers to the property that $S_1\Phi_{\text{far}}^{-1}(\Omega_{\text{near}})$ intersects $\Phi_{\text{far}}^{-1}(\Omega_{\text{near}})$ at the single point $(0,0)$ for the left-hand side of the horseshoe; reference the final panel of Figure 6.

We have made no claim concerning the uniqueness of the N-pulses described in Theorem 1. In fact, especially in the regime near $r^* = 1/2$, we expect that there may be multiple alternating N-pulses present—and in this regime especially, multiple intersections of the curves $\Phi_{\text{far}} W_{\text{loc}}^u$ and $\Phi_{\text{near}}^{-1} \circ \Phi_{\text{far}}^{-1} W_{\text{loc}}^s$, as the diagram in the final panel of Figure 5 morphs into that of Figure 6 with changing system parameters.

Moreover, we do not exclude the possibility that a whole host of other N-pulses may exist. We proved the existence of N-pulses alternating in the first component but have not ruled out the possibility of different configurations. It seems likely that there may in fact be pulses near

$$\sum_{i=1}^{N} (-1)^{s_i}\, U_0(x - T_i)$$

for any binary sequence $s_1 s_2 \ldots s_N \in \Sigma_n$ and appropriately chosen pulse spacings T_i. Such pulses will, of course, intersect Σ_{in} outside of the window Ω_{near} that we defined and will require a more careful analysis.

Additionally, there may be secondary bifurcations where the multicomponent 1-pulse and the N-pulses described here become degenerate themselves as they are continued farther from the original bifurcation value $r = r^*$. However, since these multicomponent pulses no longer respect the orbit-flip configuration described in Property 4, we do not expect a cascade of N-pulses as in the case studied in this paper [21, 22]. Such a secondary bifurcation should be covered by the results of Knobloch [18].

Finally, it is often possible to glean information about the stability properties of a pulse by carefully examining the mechanism that spurred the pulse into existence in the first place. It would be interesting to understand the differing stability properties of the fast one-component pulse and the bifurcating multicomponent 1-pulse, as well as any variations in the stability properties that arise for the N-pulses from the different geometries of the cases $r^* < 1/2$ and $r^* \geq 1/2$. Numerical evidence has suggested that the continuation of the multicomponent 1-pulse may be stable for a range of parameter values beyond $r = r^*$ [23]. Any subsequent secondary bifurcations may then still be of physical importance, as these newly bifurcating pulses may be stable, and their presence may explain the eventual loss of stability of the multicomponent 1-pulse in some cases. In the sequel to this work [24] we study these very problems, examining an instability criterion for the many pulses appearing via the bifurcation studied in the current paper, as well as those spawned in secondary bifurcations.

Acknowledgments. Special thanks to my thesis advisor, Christopher K. R. T. Jones, as well as to Shunsaku Nii, for their guidance and introduction to these topics. Thanks also to Björn Sandstede for his interest and advice. Finally, thanks to C. Eugene Wayne for his careful reading and for his reference to (and explanation of) [5].

REFERENCES

[1] M. Haelterman and A. Sheppard, *Bifurcation phenomena and multiple soliton-bound states in isotropic Kerr media*, Phys. Rev. E (3), 49 (1994), pp. 3376–3381.

[2] E. A. Ostrovskaya and Y. S. Kivshar, *Multihump optical solitons in a saturable medium*, J. Opt. B Quantum Semiclass. Opt., 1 (1999), pp. 77–83.

[3] M. Mitchell, M. Segev, and D. N. Christodoulides, *Observation of multihump multimode solitons*, Phys. Rev. Lett., 80 (1998), pp. 4657–4660.

[4] K. Y. Kolossovski, A. V. Buryak, V. V. Steblina, A. R. Champneys, and R. A. Sammut, *Higher-order nonlinear modes and bifurcation phenomena due to degenerate parametric four-wave mixing*, Phys. Rev. E (3), 62 (2000), pp. 4309–4317.

[5] A. Banyaga, R. de la Llave, and C. E. Wayne, *Cohomology equations near hyperbolic points and geometric versions of Sternberg linearization theorem*, J. Geom. Anal., 6 (1996), pp. 613–649.

[6] A. C. Yew, B. Sandstede, and C. K. R. T. Jones, *Instability of multiple pulses in coupled nonlinear Schrödinger equations*, Phys. Rev. E (3), 61 (2000), pp. 5886–5892.

[7] J. K. Hale and K. Sakamoto, *Existence and stability of transition layers*, Japan J. Appl. Math., 5 (1988), pp. 367–405.

[8] X.-B. Lin, *Using Mel′nikov's method to solve Šilnikov's problems*, Proc. Roy. Soc. Edinburgh Sect. A, 116 (1990), pp. 295–325.

[9] B. Sandstede, *Verzweigungstheorie homokliner Verdopplungen*, Ph.D. thesis, University of Stuttgart, Stuttgart, Germany, 1993.

[10] A. R. Champneys and J. Yang, *A scalar nonlocal bifurcation of solitary waves for coupled nonlinear Schrödinger systems*, Nonlinearity, 15 (2002), pp. 2165–2192.

[11] A. J. Homburg, H. Kokubu, and M. Krupa, *The cusp horseshoe and its bifurcations in the unfolding of an inclination-flip homoclinic orbit*, Ergodic Theory Dynam. Systems, 14 (1994), pp. 667–693.
[12] S. Nii, *N-homoclinic bifurcations for homoclinic orbits changing their twisting*, J. Dynam. Differential Equations, 8 (1996), pp. 549–572.
[13] A. J. Homburg and B. Krauskopf, *Resonant homoclinic flip bifurcations*, J. Dynam. Differential Equations, 12 (2000), pp. 807–850.
[14] A. Vanderbauwhede, *Bifurcation of degenerate homoclinics*, Results Math., 21 (1992), pp. 211–223.
[15] B. Fiedler and D. Turaev, *Coalescence of reversible homoclinic orbits causes elliptic resonance*, in Nonlinear Dynamics, Bifurcations and Chaotic Behavior, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 6 (1996), pp. 1007–1027.
[16] A. C. Yew, *Multipulses of nonlinearly coupled Schrödinger equations*, J. Differential Equations, 173 (2001), pp. 92–137.
[17] B. Sandstede, C. K. R. T. Jones, and J. C. Alexander, *Existence and stability of N-pulses on optical fibers with phase-sensitive amplifiers*, Phys. D, 106 (1997), pp. 167–206.
[18] J. Knobloch, *Bifurcation of degenerate homoclinic orbits in reversible and conservative systems*, J. Dynam. Differential Equations, 9 (1997), pp. 427–444.
[19] S. A. van Gils, M. Krupa, and V. Tchistiakov, *Homoclinic twist bifurcation in a system of two coupled oscillators*, J. Dynam. Differential Equations, 12 (2000), pp. 733–806.
[20] Y. A. Li and K. Promislow, *Structural stability of non-ground state traveling waves of coupled nonlinear Schrödinger equations*, Phys. D, 124 (1998), pp. 137–165.
[21] B. Sandstede, *Center manifolds for homoclinic solutions*, J. Dynam. Differential Equations, 12 (2000), pp. 449–510.
[22] S.-N. Chow, W. Liu, and Y. Yi, *Center manifolds for smooth invariant manifolds*, Trans. Amer. Math. Soc., 352 (2000), pp. 5179–5211.
[23] E. A. Ostrovskaya, Y. S. Kivshar, D. V. Scryabin, and W. J. Firth, *Stability of multihump optical solitons*, Phys. Rev. Lett., 83 (1999), pp. 296–299.
[24] R. K. Jackson and C. K. R. T. Jones, *Pulses in nonlinearly coupled Schrödinger equations* II. *An instability criterion*, SIAM J. Appl. Dyn. Syst., submitted.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 598–633

Three Is a Crowd: Solitary Waves in Photorefractive Media with Three Potential Wells*

Todd Kapitula†, P. G. Kevrekidis‡, and Zhigang Chen§

Abstract. In this paper we analytically, numerically, and experimentally study the dynamics of waves in photorefractive media in the presence of a potential with three wells. The results contained herein are also immediately applicable to the study of Bose–Einstein condensates in the weak interaction limit. Motivated by the recent theoretical and experimental efforts in the case of two wells, we systematically analyze the ways in which the bifurcation analysis of steady states and the stability picture are modified in the presence of a third potential well. In particular, it is shown that the presence of a third well causes all bifurcations to be of saddle-node type. Our analytical results are based on a Lyapunov–Schmidt reduction in the modes of the underlying linear problem. We corroborate the analytical predictions with numerical results which are based on fixed point methods. Finally, we illustrate how these findings may be related to experimental observations obtained in strontium-barium-niobate crystals.

Key words. Lyapunov–Schmidt reduction, Hamiltonian systems, solitary waves, stability

AMS subject classifications. 35P05, 37K45, 37K50, 78-05

DOI. 10.1137/05064076X

1. Introduction.

In the past few years, experimental progress in soft condensed matter physics [1, 3] and nonlinear optics [6, 31] has rekindled theoretical and experimental interest in looking at "small lattices" consisting of only a few potential wells and understanding the interesting phenomena, such as symmetry breaking, that arise in such settings [5, 13]. In turn, these experimental efforts have furthered theoretical investigations to address double well potentials from an analytical point of view, using tools such as dynamical systems theory [20], Galerkin decompositions, and Lyapunov–Schmidt reductions [45, 24, 31, 44], as well as exact solutions [34]. However, most of these investigations have focused on the dimer case. On the other hand, very few of these studies have focused on few-site lattices with three or more potential wells. Some examples of this type, but restricted to discrete systems, are in the works of [12, 17, 22].

*Received by the editors September 30, 2005; accepted for publication (in revised form) by E. Wayne May 25, 2006; published electronically November 14, 2006.
http://www.siam.org/journals/siads/5-4/64076.html

†Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131 (kapitula@math.unm.edu). This author was partially supported by the NSF under grant DMS-0304982 and by the ARO under grant 45428-PH-HSI.

‡Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-4515 (kevrekid@math.umass.edu). This author was supported by NSF-DMS-0204585, NSF-DMS-0505663, and NSF-CAREER grants.

§Department of Physics and Astronomy, San Francisco State University, San Francisco, CA 94132, and TEDA Applied Physical School, Nankai University, Tianjin, China (zchen@stars.sfsu.edu). This author's work was supported by the NSF, NSFC, and AFOSR.

Arguably, it is important to understand how the behavior of the system changes between the double-well potential and the infinite lattice. It is clear that there are some drastic differences between them. For instance, in a double-well potential the recent studies of [20, 24, 31] showed that the asymmetric state localized in one of the wells bifurcates from the symmetric one. This bifurcation results in the instability of the symmetric ("uniform" in this case) state, leading to the asymmetric ("localized" in this setting) waveform becoming the ground state of the system. This picture was directly confirmed by the experiments of [1, 3, 6, 31]. On the other hand, in the case of the infinite lattice it is well known that this picture may break down. For example, when considering focusing nonlinearities (to which we will restrict ourselves herein), it is well known that the uniform solution is *always* unstable for the infinite lattice due to modulational instability [19, 47], while the localized state is always the stable ground state of the one-dimensional problem [48]. These conclusions indicate that the picture with $N = 2$ wells is somewhat different than that of $N \to \infty$ wells. This observation, along with the experimental feasibility of variable numbers of wells that we discuss below, suggests that it would be particularly interesting to examine how this transition between "few-site" and "many-site" lattices occurs. As a particularly important and enlightening example, due to its nontrivial differences from the double-well case, we focus here on the case of $N = 3$ potential wells for the model problem discussed in section 2. The results contained herein are also immediately applicable to the study of weakly interacting attractive Bose–Einstein condensates in the presence of the three-well potential (see the review articles [4, 29, 37] and the references therein). The details will be left to the interested reader.

Our qualitative and quantitative findings illustrate that in the setting of $N = 3$ potential wells, any state with multiple in-phase pulses is *always* unstable even when such a solution may exist; furthermore, each adjacent pair of in-phase pulses produces a pair of purely real eigenvalues for the linearized problem. The only configurations that can be linearly stable are either single pulse states or ones where the adjacent pulses are out-of-phase. However, if there exists a pair of out-of-phase pulses, then for the linearized problem there exists a pair of purely imaginary eigenvalues with negative Krein sign; hence, these solutions are not ground states. This result is akin to the infinite lattice case [2, 36, 42]. We also find that the bifurcations that lead to the generation of nonlinear branches of solutions are qualitatively different than those of the $N = 2$ problem; more specifically, we typically find saddle-node bifurcations instead of the pitchfork bifurcation of the $N = 2$ case. (A cartoon of the situation is depicted in Figure 3.) We derive these results on the basis of a Lyapunov–Schmidt reduction of the nonlinear problem using the bound states of the underlying linear problem. Furthermore, we complement the existence results with the analysis of the linear stability problem for the emerging solutions.

The paper is organized as follows. In section 2 we present the mathematical framework and the connection of the nonlinear partial differential equation model to the physical problem of photorefractive crystals. In section 3 solutions to the relevant stationary nonlinear problem are constructed via Lyapunov–Schmidt theory. In section 4 the spectral stability of these solutions is determined. In fact, under a suitable assumption regarding existence, we also consider the stability of solutions associated with N_{p}-wells for any finite $N_{\mathrm{p}} \in \mathbb{N}$. In section 5 we show how the analytic findings have excellent agreement with the full numerical results. In section 6 we discuss the relevance of our theoretical findings with experimental results

obtained in photorefractive strontium-barium-niobate (SBN) crystals. We present some final thoughts in section 7. In Appendix A the point spectrum of the associated linear Schrödinger operator given in (2.7) is determined. This technical section is necessary in that the results contained therein allow one to more easily analyze the bifurcation equations in section 3. In Appendix B the spectra of the linear operator of (2.7) is explicitly determined in the case of a single square-well potential. Finally, in Appendix C a technical result is proved regarding the finding of roots of a special sixth-order polynomial which arises in the discussion in section 3.

2. Model equation. Recently, the study of light dynamics in photonic structures, such as materially fabricated photonic crystals (PCs) and optically induced photonic lattices in nonlinear media, has witnessed a large number of ground-breaking advances; e.g., see [21, 32]. Among the many phenomena heavily explored are nonlinear effects associated with propagation, localization, and discretization of light in optically induced photonic lattices [10], including the formation of lattice solitons in one [14, 38] and two [9, 15, 35] dimensions, and discrete vortex solitons [16, 39].

The model that we will use herein is based on a one-dimensional PDE describing the propagation of light in a photorefractive crystal [11] (also see the recent exposition of [49]). More specifically, we examine the dynamics of a probe beam that propagates in a photorefractive crystal (e.g., SBN), being extraordinarily polarized, while a strong ordinarily polarized beam creates an effective lattice potential for the probe. The equation for the spatial evolution of a slowly varying amplitude U of the probe beam is

$$\mathrm{i}q_z + \frac{1}{2k_0 n_e}\Delta q - \frac{1}{2}k_0 n_e^2 r_{33}\frac{E_0}{1+I_0(\boldsymbol{x})+|q|^2}q = 0. \tag{2.1}$$

In (2.1), z and x are the propagation distance and transverse coordinate, respectively; k_0 is the wavenumber of the probe beam in the vacuum; n_e is the refractive index along the extraordinary axis; r_{33} is the electro-optic coefficient for the extraordinary polarization; E_0 is the bias electric field; and $I_0(\boldsymbol{x}) \geq 0$ is the intensity of the ordinarily polarized beam, subject to modulation in the transverse direction (all intensities are normalized with respect to the crystal's dark irradiance, I_d). Measuring z in units of $2k_0 n_e$ and E_0 in units of $1/(k_0^2 n_e^4 r_{33})$, (2.1) can be cast in a dimensionless form,

$$\mathrm{i}q_z + \Delta q - \frac{E_0}{1+I_0(\boldsymbol{x})+|q|^2}q = 0, \tag{2.2}$$

with $z \in \mathbb{R}^+$ and $\boldsymbol{x} \in \mathbb{R}^n$ for some $n \in \mathbb{N}$.

Let us now put (2.2) into a form which is more convenient for analysis. If one sets

$$q := (1+I_0)^{1/2}\tilde{q}, \tag{2.3}$$

then, upon dropping the tilde, (2.2) becomes

$$\mathrm{i}q_z + \Delta q - \frac{E_0}{(1+I_0(\boldsymbol{x}))(1+|q|^2)}q = 0. \tag{2.4}$$

Looking for bound states of the form

$$q(\boldsymbol{x},z) := Q(\boldsymbol{x})\mathrm{e}^{-\mathrm{i}\omega z} \tag{2.5}$$

yields the steady-state problem

$$\mathcal{L}_\omega Q - \frac{E_0}{1+I_0(\boldsymbol{x})}\frac{|Q|^2}{1+|Q|^2}Q = 0, \tag{2.6}$$

where

$$\mathcal{L}_\omega := \mathcal{L} - \omega, \qquad \mathcal{L} := -\Delta + \frac{E_0}{1+I_0(\boldsymbol{x})}. \tag{2.7}$$

Henceforth it will be assumed that $n = 1$. The specific form for the potential $I_0(x)$ will now be given. Let $G(x) : \mathbb{R} \mapsto \mathbb{R}^+$ be even, and assume that for some $C, a \in \mathbb{R}^+$ it satisfies

$$|G(x)| \leq C\mathrm{e}^{-x^2/2a^2}, \quad x \in \mathbb{R}. \tag{2.8}$$

For $L \in \mathbb{R}^+$ sufficiently large, we will consider in this paper symmetric potentials of the form

$$I_0(x) = V_0\left(G(x) + \sum_{j=1}^{N}[G(x+2jL) + G(x-2jL)]\right) \tag{2.9}$$

(odd number of potential wells) or

$$I_0(x) = V_0\sum_{j=1}^{N}[G(x+(2j-1)L) + G(x-(2j-1)L)] \tag{2.10}$$

(even number of potential wells), where $V_0 \in \mathbb{R}^+$. In other words, the multiwell potential will be modeled by several evenly spaced single-well potentials; furthermore, the interaction between adjacent wells is of $\mathcal{O}(\mathrm{e}^{-L^2/2a^2})$. If $N = 0$, then it will be implicitly assumed that for (2.9) one has that

$$I_0(x) = V_0 G(x). \tag{2.11}$$

The following assumption regarding $\sigma_{\mathrm{p}}(\mathcal{L})$ will be used throughout the rest of this article. The implications of Assumption 2.1 regarding the stability of bifurcating nonlinear waves will be discussed in detail in section 4.1.

Assumption 2.1. Assume that $\mathcal{L}$ has a finite number of point eigenvalues, and let the eigenvalues of $\mathcal{L}$ for the single-well potential given in (2.11) be labeled $\gamma_0 < \gamma_1 < \cdots < \gamma_n < E_0 \in \mathbb{R}$. It will be assumed that for a given $0 \leq k \leq n$,

(a) $\gamma_j - \gamma_k \neq \gamma_k - \gamma_\ell$ for all choices of eigenvalues $\gamma_\ell < \gamma_k < \gamma_j$,

(b) $\gamma_\ell > 2\gamma_k - E_0$ for $\ell = 0, \ldots, k-1$.

Remark 2.2. Note that, in particular, Assumption 2.1 automatically holds if $k = 0$. Hence, the assumption is truly needed only when considering the existence and stability of nonlinear excited states.

3. Existence. Consider the existence of real-valued solutions to (2.6). Since the potential is even, solutions to (2.6) are invariant under the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$-symmetry induced by the operators $\mathcal{R}_1$ and $\mathcal{R}_2$, where

$$[\mathcal{R}_1 q](x) := q(-x), \qquad [\mathcal{R}_2 q](x) := -q(x). \tag{3.1}$$

Each nonlinear solution found subsequently will generate a family of solutions which is invariant under the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$-symmetry; in other words, if Q is a solution, then so are $[\mathcal{R}_1 Q]$, $[\mathcal{R}_2 Q]$, and $[\mathcal{R}_1 \mathcal{R}_2 Q]$.

As in Appendix A, let $\Phi(x)$ represent a normalized eigenfunction for the operator $\mathcal{L}$ given in (2.7), with the single-well potential given in (2.11). The eigenfunction, which by the assumption on $G(x)$ is either even or odd, is clearly independent of ω; hence, for this eigenfunction one can let the parameter $\omega < E_0$ be chosen so that the associated eigenvalue for $\mathcal{L}_\omega$ is $\lambda = 0$. When considering the potential given in (2.9), we will henceforth be interested in the case that $N = 1$. In this case, upon applying the result of (A.13) to (A.11), one sees that the exponentially small eigenvalues of $\mathcal{O}(\mathrm{e}^{-2c_\omega L})$ and $c_\omega := \sqrt{E_0 - \omega}$, as well as their associated eigenfunctions, are given to leading order by

$$\begin{aligned} \lambda_1 &= \sqrt{2}\, a_L^{\pm}, & q_1 &:= \frac{1}{2}[\Phi(x+2L) + \sqrt{2}\,\Phi(x) + \Phi(x-2L)], \\ \lambda_2 &= 0, & q_2 &:= \frac{1}{2}[-\sqrt{2}\,\Phi(x+2L) + \sqrt{2}\,\Phi(x-2L)], \\ \lambda_3 &= -\sqrt{2}\, a_L^{\pm}, & q_3 &:= \frac{1}{2}[-\Phi(x+2L) + \sqrt{2}\,\Phi(x) - \Phi(x-2L)]. \end{aligned} \tag{3.2}$$

Here

$$a_L^{\pm} := \pm 2\Phi(L)\Phi'(L),$$

and one takes a_L^+ if $\Phi(x)$ is even, and a_L^- otherwise. In (3.2) note that if $\Phi(x)$ is even, then $\lambda_1 < \lambda_2 < \lambda_3$; otherwise, the reverse ordering is taken. All of the eigenfunctions in (3.2) have been scaled to have norm one. The expressions presented in (3.2) will be heavily used in all of the subsequent analysis.

3.1. Lyapunov–Schmidt reduction. The idea is to apply a Lyapunov–Schmidt reduction to (2.6). Since $|\lambda_j| = \mathcal{O}(\mathrm{e}^{-2c_\omega L})$, it will initially be assumed that the semisimple eigenvalue $\lambda = 0$ for $\mathcal{L}_\omega$ has geometric multiplicity three. For $0 < \epsilon \ll 1$ write

$$\begin{aligned} Q &= (xq_1 + yq_2 + zq_3)\epsilon^{1/2} + \mathcal{O}(\epsilon), \\ \omega &= \gamma_k + \Delta\omega\, \epsilon + \mathcal{O}(\epsilon^2); \end{aligned} \tag{3.3}$$

the parameter ϵ is introduced so that the nonlinear term is small. In the expansion for ω given in (3.3) we are explicitly assuming that the parameter is free (see (2.6)). Note that to leading order the power is given by

$$\begin{aligned} P(Q) &:= \int_{-\infty}^{+\infty} Q^2(t)\, \mathrm{d}t \\ &= (x^2 + y^2 + z^2)\epsilon + \mathcal{O}(\epsilon^{3/2}). \end{aligned} \tag{3.4}$$

It is also important to note, if $\Phi(x)$ is a positive bound state, the fact that to leading order

$$\langle Q, q_1\rangle = x\,\epsilon^{1/2}, \quad \langle Q, q_2\rangle = y\,\epsilon^{1/2}, \quad \langle Q, q_3\rangle = z\,\epsilon^{1/2},$$

and the expansions in (3.2) then yield that the value of x corresponds to a projection onto a bound state of three in-phase pulses, the value of y corresponds to a projection onto a bound state of two out-of-phase pulses, and the value of z corresponds to a projection onto a bound state of three out-of-phase pulses. Now define the inner product

$$\langle f, g\rangle_{I_0} := \int_{-\infty}^{+\infty} \frac{1}{1+I_0(x)} f(x)g(x)\,\mathrm{d}x,$$

and set

$$g^i_{jk\ell} := \langle q_j q_k q_\ell, q_i\rangle_{I_0},$$

where the eigenfunction q_α for $\alpha \in \{1,2,3\}$ is given in (3.2). Upon applying the standard Lyapunov–Schmidt reduction theory, one sees that the assumption that the eigenvalue is semisimple yields the set of existence equations

$$\begin{aligned}
0 &= \Delta\omega\, x + E_0(g^1_{111}x^3 + 3g^1_{122}xy^2 + 3g^1_{133}xz^2 + 3g^1_{113}x^2z + g^1_{333}z^3 + 3g^1_{223}x^2y),\\
0 &= \Delta\omega\, y + E_0(3g^1_{122}x^2 + 6g^1_{223}xz + g^2_{222}y^2 + 3g^2_{233}z^2)y,\\
0 &= \Delta\omega\, z + E_0(3g^1_{133}x^2z + g^1_{113}x^3 + 3g^1_{333}xz^2 + 3g^1_{223}xy^2 + 3g^2_{233}y^2z + g^3_{333}z^3).
\end{aligned} \tag{3.5}$$

In other words, if (x_0, y_0, z_0) is a solution to (3.5), then by (3.3) the associated nonlinear solution is given by $Q \sim (x_0q_1 + y_0q_2 + z_0q_3)\epsilon^{1/2}$. The fact that the eigenfunctions are either even or odd was implicitly used in setting up (3.5).

Now let us use the functional form of the eigenfunctions given in (3.2) to simplify (3.5). Upon using the fact that the tails of $\Phi(x)$ are of $\mathcal{O}(\mathrm{e}^{-c_\omega L})$, one quickly sees that up to an exponentially small error term,

$$g^1_{133} = g^3_{333} = 3g^1_{113} = 3g^1_{333} = \frac{3}{2}g^1_{122} = \frac{3}{2}g^2_{233} = -\frac{3}{2}g^1_{223} = \frac{3}{4}g^2_{222} = g^1_{111}. \tag{3.6}$$

Upon defining

$$\Gamma := \frac{3}{4g^1_{111}}\frac{1}{E_0}, \tag{3.7}$$

and substituting the result of (3.6) into (3.5), one gets that (3.5) can be rewritten as

$$\begin{aligned}
0 &= \frac{4}{3}\Gamma\Delta\omega\, x + x^3 + x^2z + 3xz^2 + \frac{1}{3}z^3 + 2(x-z)y^2,\\
0 &= \left(\Gamma\Delta\omega + \frac{3}{2}(x-z)^2 + y^2\right)y,\\
0 &= \frac{4}{3}\Gamma\Delta\omega\, z + \frac{1}{3}x^3 + 3x^2z + xz^2 + z^3 - 2(x-z)y^2.
\end{aligned} \tag{3.8}$$

As already stated, (3.8) is valid only in the case that $\lambda = 0$ is semisimple with multiplicity three. In order to take into account the eigenvalue unfolding given in (3.2), the equations must be appropriately modified. As stated in [18, Proposition IX.1.3], via the use of singularity theory it is known that the universal unfolding associated with (3.8) contains minimally *eight* free parameters. To the best of our knowledge this unfolding has not yet been written down. Consequently, we take the following approach. Consider the subspace which has as a basis the set $\{q_1, q_2\}$. If one were to perform a Lyapunov–Schmidt reduction with this set, then upon using [18, Theorem X.2.4], one would get the universal unfolding of the bifurcation equations to be of the form

$$\begin{aligned} 0 &= \frac{4}{3}\Gamma(\Delta\omega + \Delta\lambda)x + x^3 + 2xy^2, \\ 0 &= \left(\Gamma\Delta\omega + \frac{3}{2}x^2 + y^2\right) y. \end{aligned} \tag{3.9}$$

In (3.9) the parameter $\Delta\lambda \propto \lambda_2 - \lambda_1$ is an unfolding parameter. Since $a_L^{\pm}\Delta\lambda > 0$, one has that $\Delta\lambda > 0$ if $\Phi(x)$ is even, and is negative otherwise. One arrives at a system similar to (3.9) if the basis $\{q_2, q_3\}$ is taken. Thus, it seems reasonable to conclude that a one-parameter unfolding of (3.8) which is consistent with the results presented in Appendix A and which preserves the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$-symmetry is given by

$$\begin{aligned} 0 &= \frac{4}{3}\Gamma(\Delta\omega + \Delta\lambda)x + x^3 + x^2 z + 3xz^2 + \frac{1}{3}z^3 + 2(x - z)y^2, \\ 0 &= \left(\Gamma\Delta\omega + \frac{3}{2}(x - z)^2 + y^2\right) y, \\ 0 &= \frac{4}{3}\Gamma(\Delta\omega - \Delta\lambda)z + \frac{1}{3}x^3 + 3x^2 z + xz^2 + z^3 - 2(x - z)y^2. \end{aligned} \tag{3.10}$$

In (3.10), $\Delta\lambda$ is an eigenvalue unfolding parameter that satisfies

$$\Delta\lambda \propto \lambda_2 - \lambda_1 = \lambda_3 - \lambda_2. \tag{3.11}$$

The remainder of this section will be devoted to the study of (3.10). While the results for $\Delta\lambda \neq 0$ will not be rigorous in predicting the bifurcation points in parameter space, the numerical results presented in section 5 will justify the effort. A cartoon for the subsequently described situation is given in Figure 1. In all that follows, the expressions for the nonlinear wave and the total power will be given only to leading order. The power follows from (3.4), while the expressions for the wave follow upon combining (3.2) with (3.3).

3.2. Solution: $x = z = 0$. In this subsection we consider the existence of odd solutions, i.e., solutions which are fixed points for $\mathcal{R}_1\mathcal{R}_2$. When $x = z = 0$ the solution to (3.10) is particularly easy: $y^2 = -\Gamma\Delta\omega$, with

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(-\Phi(x + 2L) + \Phi(x - 2L)\right), \qquad P(Q) = -\Gamma\Delta\omega\,\epsilon.$$

This solution corresponds to the curve labeled $(-0+)$ in Figure 1 and is a continuation from the linear limit.

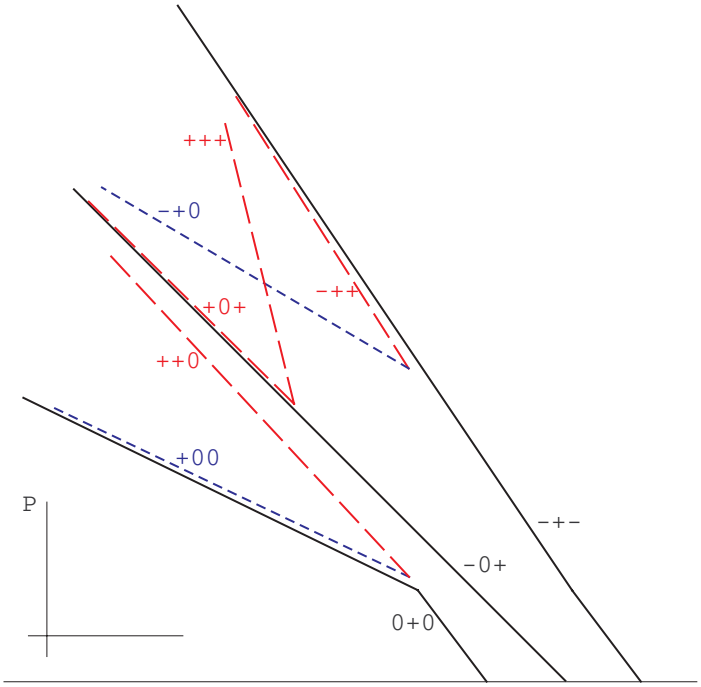


Figure 1. *The bifurcation diagram associated with* (3.10) *in the case that* $\Delta\lambda > 0$. *The* "+" *corresponds to* $\Phi(\cdot)$ *sitting in the potential well, the* "−" *corresponds to* $-\Phi(\cdot)$ *sitting in the potential well, and the* "0" *implies that the potential well is empty.*

3.3. Solution: $y = 0$. In this subsection we consider the existence of even solutions, i.e., solutions which are fixed points for $\mathcal{R}_1$. When $y = 0$, (3.10) reduces to the system

$$\begin{aligned} 0 &= \frac{4}{3}\Gamma(\Delta\omega + \Delta\lambda)x + x^3 + x^2 z + 3xz^2 + \frac{1}{3}z^3, \\ 0 &= \frac{4}{3}\Gamma(\Delta\omega - \Delta\lambda)z + \frac{1}{3}x^3 + 3x^2 z + xz^2 + z^3. \end{aligned} \tag{3.12}$$

Set

$$\Delta\eta := \frac{\Delta\lambda}{\Delta\omega}, \tag{3.13}$$

and write $z = \mu x$. Substitution of the above ansatz into (3.12) yields that

$$x^2 = -4\frac{1+\Delta\eta}{3+3\mu+9\mu^2+\mu^3}\Gamma\Delta\omega, \tag{3.14}$$

with

$$g_z(\mu, \Delta\eta) := \mu^4 + 6\mu^3 - 6\mu - 1 - \Delta\eta(\mu^4 + 12\mu^3 + 6\mu^2 + 12\mu + 1) = 0. \tag{3.15}$$

3.3.1. $\Delta\lambda = 0$. Upon setting $\Delta\lambda = 0$ in (3.15), one has that

$$g_z(\mu, 0) = (\mu^2 - 1)(\mu^2 + 6\mu + 1).$$

If $\mu = 1$, then $x^2 = -\Gamma\Delta\omega/4$ with

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\,\Phi(x), \quad P(Q) = -\frac{1}{2}\Gamma\Delta\omega\,\epsilon, \tag{3.16}$$

i.e., the solution curve labeled $(0+0)$ in Figure 1, while if $\mu = -1$, then $x^2 = -\Gamma\Delta\omega/2$ with

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(\Phi(x+2L) + \Phi(x-2L)\right), \quad P(Q) = -\Gamma\Delta\omega\,\epsilon, \tag{3.17}$$

i.e., the solution curve labeled $(+0+)$ in Figure 1. Now suppose that $\mu^2 + 6\mu + 1 = 0$. One sees that (3.14) reduces to

$$x^2 = \frac{1}{4\mu}\Gamma\Delta\omega, \tag{3.18}$$

and the subsequent profiles are given by

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(\Phi(x+2L) + \Phi(x) + \Phi(x-2L)\right), \quad P(Q) = -\frac{3}{2}\Gamma\Delta\omega\,\epsilon, \tag{3.19}$$

($\mu = -3 + \sqrt{8}$) and

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(-\Phi(x+2L) + \Phi(x) - \Phi(x-2L)\right), \quad P(Q) = -\frac{3}{2}\Gamma\Delta\omega\,\epsilon \tag{3.20}$$

($\mu = -3 - \sqrt{8}$). These solutions correspond to the curves labeled $(+++)$ and $(-+-)$, respectively, in Figure 1.

3.3.2. $\Delta\lambda \neq 0$. In order to study the solution set to (3.15), it is natural to first determine those points $(\mu_c, \Delta\eta_c)$ for which

$$g_z(\mu_c, \Delta\eta_c) = \frac{\partial}{\partial\mu}g_z(\mu, \Delta\eta) = 0.$$

One initially sees that $\partial_\mu g_z = 0$ only if

$$\Delta\eta = \frac{2\mu^3 + 9\mu^2 - 3}{2\mu^3 + 18\mu^2 + 6\mu + 6}. \tag{3.21}$$

Substituting the result of (3.21) into (3.15) and simplifying then yields

$$g_z^{\mathrm{b}}(\mu) := \mu^6 + 2\mu^5 + 15\mu^4 + 48\mu^3 + 15\mu^2 + 2\mu + 1 = 0. \tag{3.22}$$

Upon using the result of Proposition C.1, one finds that the real-valued roots of g_z^{b} are approximately given by

$$\mu \in \{-0.388, -2.574\}.$$

Substitution of these values into (3.21) yields that

$$\mu \sim -0.388 \implies \Delta\eta_{\mathrm{c}} \sim -0.281, \qquad \mu \sim -2.547 \implies \Delta\eta_{\mathrm{c}} \sim 0.297. \tag{3.23}$$

Upon substituting the result of (3.23) into (3.14), one sees that the solution is consistent; i.e., $x^2 \in \mathbb{R}^+$ only if $\Delta\omega = \Delta\eta_{\mathrm{c}}/\Delta\lambda < 0$. If $\Delta\lambda < 0$, so that $\Delta\eta_{\mathrm{c}} > 0$, then upon consulting (3.23), one sees that upon the appropriate substitution into (3.14),

$$x^2 \sim -0.137\Gamma\Delta\omega, \tag{3.24}$$

and upon using (3.3), one sees that the resulting wave is given by

$$Q(x) \sim \epsilon^{1/2}\sqrt{|\Gamma\Delta\omega|}\left(0.66\Phi(x+2L) - 0.41\Phi(x) + 0.66\Phi(x-2L)\right), \quad P(Q) \sim -1.045\,\Gamma\Delta\omega\,\epsilon. \tag{3.25}$$

If $\Delta\lambda > 0$, then by (3.14) one has that

$$x^2 \sim -0.918\Gamma\Delta\omega, \tag{3.26}$$

with

$$Q(x) \sim \epsilon^{1/2}\sqrt{|\Gamma\Delta\omega|}\left(0.67\Phi(x+2L) + 0.41\Phi(x) + 0.67\Phi(x-2L)\right), \quad P(Q) \sim -1.057\,\Gamma\Delta\omega\,\epsilon. \tag{3.27}$$

One finally has that the solutions defined by (3.17) and (3.19), i.e., the solutions (+0+) and (+++), terminate as a consequence of a saddle-node bifurcation at either the solution defined by (3.25) ($\Delta\lambda < 0$) or that defined by (3.27) ($\Delta\lambda > 0$).

3.4. Solution: $x, y, z \neq 0$. We finally consider those solutions which are neither odd nor even in the spatial variable. Solving the second equation of (3.10) yields

$$y^2 = -\Gamma\Delta\omega - \frac{3}{2}(x-z)^2. \tag{3.28}$$

Substituting the result of (3.28) into the first and third equations of (3.10) then yields the system

$$\begin{aligned} 0 &= \Gamma[(\Delta\omega - 2\Delta\lambda)x - 3\Delta\omega z] + 3x^3 - 15x^2 z + 9xz^2 - 5z^3, \\ 0 &= \Gamma[-3\Delta\omega x + (\Delta\omega + 2\Delta\lambda)z] - 5x^3 + 9x^2 z - 15xz^2 + 3z^3. \end{aligned} \tag{3.29}$$

Upon using (3.13) and writing $z = \mu x$, one sees that a solution to (3.29) must satisfy

$$x^2 = -\frac{1 - 3\mu - 2\Delta\eta}{3 - 15\mu + 9\mu^2 - 5\mu^3}\Gamma\Delta\omega, \tag{3.30}$$

with

$$g_y(\mu, \Delta\eta) := 2\mu^4 - 12\mu^3 + 12\mu - 2 - \Delta\eta(5\mu^4 - 12\mu^3 + 30\mu^2 - 12\mu + 5) = 0. \tag{3.31}$$

Note that (3.28) can now be rewritten as

$$y^2 = -\Gamma\Delta\omega - \frac{3}{2}(1-\mu)^2 x^2,$$

where x^2 is given in (3.30), i.e.,

$$y^2 = -\left(1 - \frac{3}{2}(1-\mu)^2 \frac{1 - 3\mu - 2\Delta\eta}{3 - 15\mu + 9\mu^2 - 5\mu^3}\right)\Gamma\Delta\omega. \tag{3.32}$$

3.4.1. $\Delta\lambda = 0$. Upon setting $\Delta\lambda = 0$ in (3.31), one has that

$$g_y(\mu, 0) = (\mu^2 - 1)(\mu^2 - 6\mu + 1).$$

If $\mu = 1$, then $x^2 = -\Gamma\Delta\omega/4$ and $y^2 = -\Gamma\Delta\omega$ with

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(-\Phi(x+2L) + \Phi(x) + \Phi(x-2L)\right), \quad P(Q) = -\frac{3}{2}\Gamma\Delta\omega\,\epsilon, \tag{3.33}$$

i.e., the solution curve labeled $(-++)$ in Figure 1, while if $\mu = -1$, then $x^2 = -\Gamma\Delta\omega/8$ and $y^2 = -\Gamma\Delta\omega/4$ with

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\,\Phi(x+2L), \quad P(Q) = -\frac{1}{2}\Gamma\Delta\omega\,\epsilon, \tag{3.34}$$

i.e., the solution curve labeled $(+00)$ in Figure 1. Now suppose that $\mu^2 - 6\mu + 1 = 0$. One sees that (3.30) and (3.32) reduce to

$$x^2 = -\frac{1}{8\mu}\Gamma\Delta\omega, \qquad y^2 = -\frac{1}{4}\Gamma\Delta\omega, \tag{3.35}$$

and the subsequent profiles are given by

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(-\Phi(x+2L) + \Phi(x)\right), \quad P(Q) = -\Gamma\Delta\omega\,\epsilon \tag{3.36}$$

($\mu = 3 + \sqrt{8}$), i.e., the solution curve labeled $(-+0)$ in Figure 1, and

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\Delta\omega|}{2}}\left(\Phi(x+2L) + \Phi(x)\right), \quad P(Q) = -\Gamma\Delta\omega\,\epsilon \tag{3.37}$$

($\mu = 3 - \sqrt{8}$), i.e., the solution curve labeled $(++0)$ in Figure 1.

3.4.2. $\Delta\lambda \neq 0$. In order to study the solution set to (3.31), it is natural to first determine those points $(\mu_c, \Delta\eta_c)$ for which

$$g_y(\mu_c, \Delta\eta_c) = \frac{\partial}{\partial\mu} g_y(\mu, \Delta\eta) = 0.$$

One initially sees that $\partial_\mu g_y = 0$ only if

$$\Delta\eta = \frac{2\mu^3 - 9\mu^2 + 3}{5\mu^3 - 9\mu^2 + 15\mu - 3}. \tag{3.38}$$

Substituting the result of (3.38) into (3.31) and simplifying then yields

$$g_y^{\mathrm{b}}(\mu) := \mu^6 + \frac{10}{3}\mu^5 - \frac{51}{3}\mu^4 + \frac{164}{9}\mu^3 - \frac{51}{3}\mu^2 + \frac{10}{3}\mu + 1 = 0. \tag{3.39}$$

Upon using the result of Proposition C.1, one finds that the real-valued roots of g_y^{b} are given by

$$\mu \in \left\{1 \pm \frac{2}{3}\sqrt{3}, -3 \pm 2\sqrt{3}\right\}.$$

Substitution of these values into (3.38) yields that

$$\mu = 1 \pm \frac{2}{3}\sqrt{3} \implies \Delta\eta_c = -\frac{1}{2}, \qquad \mu = -3 \pm 2\sqrt{3} \implies \Delta\eta_c = \frac{1}{2}. \tag{3.40}$$

Upon substituting the result of (3.40) into (3.30) and (3.32), one sees that the solution is consistent, i.e., $x^2, y^2 \in \mathbb{R}^+$ only if $\Delta\omega = \Delta\eta_c/\Delta\lambda < 0$. First suppose that $\Delta\lambda > 0$, so that upon consulting (3.40) one sees that one must have $\mu \in \{1 - 2\sqrt{3}/3,\ 1 + 2\sqrt{3}/3\}$. If $\mu = 1 + 2\sqrt{3}/3$, then

$$x^2 = -\frac{9}{32}\left(1 - \frac{\sqrt{3}}{3}\right)\Gamma\Delta\omega, \quad y = \pm\left(\frac{10}{3} + \frac{16}{9}\sqrt{3}\right)^{1/2} x, \quad z = \left(1 + \frac{2}{3}\sqrt{3}\right)x, \tag{3.41}$$

which, upon using (3.3), yields

$$\begin{aligned} Q(x) &\sim \epsilon^{1/2}\sqrt{|\Gamma\Delta\omega|}\left(-0.82\Phi(x + 2L) + 0.77\Phi(x) + 0.42\Phi(x - 2L)\right), \\ P(Q) &= -\left(1 + \frac{1}{4}\sqrt{3}\right)\Gamma\Delta\omega\,\epsilon. \end{aligned} \tag{3.42}$$

One has that the solutions defined by (3.33) and (3.36), i.e., the solutions $(-++)$ and $(-+0)$, terminate as a consequence of a saddle-node bifurcation at the solution defined by (3.42). If $\mu = 1 - 2\sqrt{3}/3$, then

$$x^2 = -\frac{9}{32}\left(1 + \frac{\sqrt{3}}{3}\right)\Gamma\Delta\omega, \quad y = \pm\left(\frac{10}{3} - \frac{16}{9}\sqrt{3}\right)^{1/2} x, \quad z = \left(1 - \frac{2}{3}\sqrt{3}\right)x, \tag{3.43}$$

with

$$Q(x) \sim \epsilon^{1/2}\sqrt{|\Gamma\Delta\omega|}\left(0.15\Phi(x+2L)+0.40\Phi(x)+0.62\Phi(x-2L)\right),$$
$$P(Q) = -\left(1-\frac{1}{4}\sqrt{3}\right)\Gamma\Delta\omega\,\epsilon. \tag{3.44}$$

One has that the solutions defined by (3.34) and (3.37), i.e., the solutions $(+00)$ and $(++0)$, terminate as a consequence of a saddle-node bifurcation at the solution defined by (3.44).

If $\Delta\lambda < 0$, then for $\mu = -3+2\sqrt{3}$ one has that

$$x^2 = -\frac{9}{32}\left(1+\frac{5}{9}\sqrt{3}\right)\Gamma\Delta\omega, \quad y = \pm\left(6-\frac{8}{3}\sqrt{3}\right)^{1/2}x, \quad z = \left(-3+2\sqrt{3}\right)x, \tag{3.45}$$

with

$$Q(x) \sim \epsilon^{1/2}\sqrt{|\Gamma\Delta\omega|}\left(-0.82\Phi(x+2L)-0.77\Phi(x)+0.42\Phi(x-2L)\right),$$
$$P(Q) = -\left(1+\frac{1}{4}\sqrt{3}\right)\Gamma\Delta\omega\,\epsilon. \tag{3.46}$$

If $\mu = -3-2\sqrt{3}$, then

$$x^2 = -\frac{9}{32}\left(1-\frac{5}{9}\sqrt{3}\right)\Gamma\Delta\omega, \quad y = \pm\left(6+\frac{8}{3}\sqrt{3}\right)^{1/2}x, \quad z = \left(-3-2\sqrt{3}\right)x, \tag{3.47}$$

with

$$Q(x) \sim \epsilon^{1/2}\sqrt{|\Gamma\Delta\omega|}\left(0.15\Phi(x+2L)-0.40\Phi(x)+0.62\Phi(x-2L)\right),$$
$$P(Q) = -\left(1-\frac{1}{4}\sqrt{3}\right)\Gamma\Delta\omega\,\epsilon. \tag{3.48}$$

The bifurcation diagram is exactly as described in the previous for the case $\Delta\lambda > 0$.

4. Spectral stability of nonlinear solutions.

4.1. General spectral results.

The purpose of this subsection is to provide a broad theoretical framework for the analysis of section 4.3. The general results presented in [26, 27, 41] will be used. Upon taking real and imaginary parts via $q := u+iv$, and linearizing (2.4) about a real-valued solution Q, one has the eigenvalue problem

$$L_+u = -\lambda v, \qquad L_-v = \lambda u, \tag{4.1}$$

where

$$L_+ := \mathcal{L}_\omega - 3\epsilon\frac{E_0}{1+I_0(x)}Q^2 + \mathcal{O}(\epsilon^{3/2}), \qquad L_- := \mathcal{L}_\omega - \epsilon\frac{E_0}{1+I_0(x)}Q^2 + \mathcal{O}(\epsilon^{3/2}).$$

Eigenvalue problems of this type have been studied extensively (see [26] for a bibliography). For (4.1) let k_{r} represent the number of real positive eigenvalues, k_{c} the number of eigenvalues

with both positive real and imaginary parts, and k_{i}^{-} the number of purely imaginary eigenvalues with positive imaginary part and negative Krein sign. The Krein signature of a simple eigenvalue $\lambda \in \mathrm{i}\mathbb{R}^{+}$ is given by

$$K := \operatorname{sign}(\langle u_\lambda, L_+ u_\lambda \rangle), \tag{4.2}$$

where the associated eigenfunction of (4.1) is given by (u_λ, v_λ) (see [26, section 2.2] for more details). Let $\mathrm{n}(L_\pm)$ correspond to the number of negative eigenvalues of $L_\pm$ and $\mathrm{z}(L_\pm) = \dim(\ker(L_\pm))$. The result of [26, Theorem 3.3] (also see [40]) states that

$$k_{\mathrm{r}} + 2k_{\mathrm{i}}^{-} + 2k_{\mathrm{c}} = [\mathrm{n}(L_+) - \mathrm{n}(D)] + \mathrm{n}(L_-), \tag{4.3}$$

where

$$D := \partial_\omega \langle Q, Q \rangle; \tag{4.4}$$

furthermore,

$$k_{\mathrm{r}} \geq |[\mathrm{n}(L_+) - \mathrm{n}(D)] - \mathrm{n}(L_-)|. \tag{4.5}$$

As a consequence of the existence results in section 3, i.e., the fact that $\Delta\omega < 0$, one has that for all of the waves under consideration, $\mathrm{n}(D) = 1$. Thus, (4.3) and (4.5) can be rewritten as

$$\begin{aligned} k_{\mathrm{r}} + 2k_{\mathrm{i}}^{-} + 2k_{\mathrm{c}} &= [\mathrm{n}(L_+) - 1] + \mathrm{n}(L_-), \\ k_{\mathrm{r}} &\geq |[\mathrm{n}(L_+) - 1] - \mathrm{n}(L_-)|. \end{aligned} \tag{4.6}$$

Note that if both $\mathrm{n}(L_\pm)$ are either even or odd, then as a consequence of (4.6) one has that $k_{\mathrm{r}} \geq 1$.

The perturbation calculation presented in section 4.2 may be insufficient for fully determining the spectral stability of the solutions, as it describes only those eigenvalues for $L_\pm$ and (4.1) of $\mathcal{O}(\epsilon)$. It is possible that $\mathcal{O}(1)$ eigenvalues of opposite sign collide, and hence create an oscillatory instability. However, as will now be seen, this possibility is precluded under Assumption 2.1 for $\epsilon > 0$ sufficiently small.

First consider (2.6) under the assumption of the single-well potential given in (2.11). Following the argument given in [23], it can be shown that for each $0 \leq k \leq n$, and for $0 < \epsilon \ll 1$, there is a unique value $\omega = \gamma_k + \mathcal{O}(\epsilon)$ and a corresponding real-valued nonlinear solution $Q_k = \epsilon^{1/2} q_k + \mathcal{O}(\epsilon)$ to (2.6), where q_k is the eigenfunction of $\mathcal{L}$ associated with the eigenvalue γ_k. Furthermore, for this wave $\mathrm{n}(D) = 1$. When discussing the spectral stability of this nonlinear wave one has that

$$\mathrm{n}(L_+) = k + 1, \qquad \mathrm{n}(L_-) = k,$$

which by (4.6) implies that

$$k_{\mathrm{r}} + 2k_{\mathrm{i}}^{-} + 2k_{\mathrm{c}} = 2k$$

[23, section 3.2]. Under this scenario it was further shown in [23, appendix] that the eigenvalues $\gamma_1 - \omega, \ldots, \gamma_{k-1} - \omega, \gamma_{k+1} - \omega, \ldots, \gamma_n - \omega$ of $\mathcal{L}_\omega$ map to purely imaginary eigenvalues for

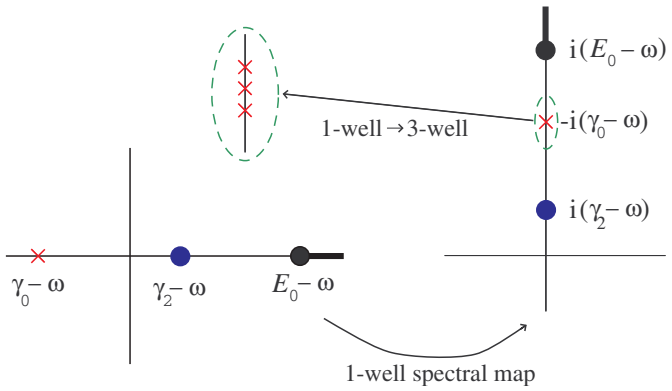


Figure 2. *The left-hand part of the figure gives the nonzero spectrum of $\mathcal{L}_\omega$ in the case of a 1-well potential with $\omega = \gamma_1$. The right-hand part demonstrates the manner in which the spectrum of $\mathcal{L}_\omega$ is mapped to that for (4.1) under Assumption 2.1. In this figure the cross represents a simple purely imaginary eigenvalue with negative Krein sign, and the circle represents a simple purely imaginary eigenvalue with positive Krein sign. The inset demonstrates the manner in which the single eigenvalue with negative sign becomes three eigenvalues with negative sign when the number of potential wells increases from one to three.*

the eigenvalue problem in (4.1); furthermore, the eigenvalues $\gamma_1 - \omega, \dots, \gamma_{k-1} - \omega$ map to $\mp \mathrm{i}(\gamma_\ell - \omega) + \mathcal{O}(\epsilon)$, $\ell = 0, \dots, k-1$, and have negative Krein sign, whereas the eigenvalues $\gamma_{k+1} - \omega, \dots, \gamma_n - \omega$ map to purely imaginary eigenvalues $\pm \mathrm{i}(\gamma_\ell - \omega) + \mathcal{O}(\epsilon)$, $\ell = k+1, \dots, n$, and have positive sign.

Assume that Assumption 2.1(a) holds, i.e., that the eigenvalues are not in resonance. This circumstance can be precluded via a proper choice of the potential, such as that described in Appendix B. As a consequence of Assumption 2.1(a) one has that all of the $\mathcal{O}(1)$ eigenvalues given above are simple. If Assumption 2.1(b) holds, then one further has that none of the eigenvalues resides in the continuous spectrum when $\epsilon = 0$, and consequently will not do so for $\epsilon > 0$ sufficiently small. Because the system is Hamiltonian, one can then conclude that all of the eigenvalues remain on the imaginary axis for ϵ sufficiently small. It is interesting to note that if Assumption 2.1(b) does not hold, then an eigenvalue with negative sign is embedded in the essential spectrum, and it is generic that upon perturbation to the nonlinear state the perturbed eigenvalue will have nonzero real part (e.g., see [33, 41]).

Now consider the case of the N-well potential. By [46] it is known that associated with each simple $\mathcal{O}(1)$ eigenvalue given above for the 1-well potential there will exist a cluster of N eigenvalues that are $\mathcal{O}(\mathrm{e}^{-2c_\omega L})$ close to the original 1-well eigenvalue. Assumption 2.1(a) guarantees that each cluster of N eigenvalues will be isolated from all other clusters, while Assumption 2.1(b) ensures that none of the eigenvalues will reside in the continuous spectrum. Because the system is Hamiltonian, upon using an energy argument along the lines presented in [26, section 3], one knows that each cluster of eigenvalues will be purely imaginary. An illustration of the above discussion is given in Figure 2.

As a consequence of the above discussion, when applying the result of (4.6) under Assumption 2.1, it is sufficient to consider only those small $\mathcal{O}(\epsilon)$ eigenvalues which are generated by the perturbation from the linear state. The task of locating these small eigenvalues for general Hamiltonian systems will be considered in the next subsection.

4.2. General perturbation results. Consider the eigenvalue problem

$$J\mathcal{L}u = \lambda u, \tag{4.7}$$

where J is a nonsingular skew-symmetric operator and $\mathcal{L}$ is a symmetric operator of the form

$$\mathcal{L} = \mathcal{L}_0 + \epsilon\mathcal{L}_1 + \mathcal{O}(\epsilon^2); \quad \mathcal{L}_0 := \operatorname{diag}(L_0, L_0),\ \mathcal{L}_1 := \begin{pmatrix} L_1^+ & L_1^{\mathrm{c}} \\ L_1^{\mathrm{c}} & L_1^- \end{pmatrix}.$$

It will be assumed that $\mathrm{z}(L_0) = n \in \mathbb{N}$ and that the orthonormal basis for $\ker(L_0)$ is given by

$$\ker(L_0) = \operatorname{Span}\{\phi_1, \dots, \phi_n\}.$$

One then has that $\mathrm{z}(\mathcal{L}_0) = 2n$, with

$$\ker(\mathcal{L}_0) = \operatorname{Span}\{\phi_1^+, \dots, \phi_n^+, \phi_1^-, \dots, \phi_n^-\}; \quad \phi_j^+ := (\phi_j, 0)^{\mathrm{T}},\ \phi_j^- := (0, \phi_j)^{\mathrm{T}}.$$

Upon using regular perturbation theory (see [28, Theorem 8.2.6]) and Lyapunov–Schmidt theory (see the proof of [26, Theorem 4.1]), one can write the perturbed eigenfunctions and eigenvalues as

$$u = u_0 + \epsilon u_1 + \mathcal{O}(\epsilon^2), \qquad \lambda = \epsilon\lambda_1 + \mathcal{O}(\epsilon^2),$$

where $u_0 \in \ker(\mathcal{L}_0)$ and the pair (u_1, λ_1) solves the equation

$$\mathcal{L}_0 u_1 = (\lambda_1 J^{-1} - \mathcal{L}_1)u_0. \tag{4.8}$$

Upon writing

$$u_0 = \sum_{j=1}^{n} c_j^+ \phi_j^+ + \sum_{j=1}^{n} c_j^- \phi_j^- \tag{4.9}$$

and applying the standard solvability condition, one then sees that (4.8) can be solved if and only if

$$(\lambda_1 \boldsymbol{J} - \boldsymbol{S})\boldsymbol{c} = \boldsymbol{0}, \quad \boldsymbol{c} := (\boldsymbol{c}^+, \boldsymbol{c}^-)^{\mathrm{T}}, \tag{4.10}$$

where

$$\boldsymbol{J} := \begin{pmatrix} \boldsymbol{J}^+ & \boldsymbol{J}^{\mathrm{c}} \\ -(\boldsymbol{J}^{\mathrm{c}})^{\mathrm{T}} & \boldsymbol{J}^- \end{pmatrix}; \quad (\boldsymbol{J}^{\pm})_{ij} := \langle J^{-1}\phi_i^{\pm}, \phi_j^{\pm}\rangle,\ (\boldsymbol{J}^{\mathrm{c}})_{ij} := \langle J^{-1}\phi_i^-, \phi_j^+\rangle,$$

and

$$\boldsymbol{S} := \begin{pmatrix} \boldsymbol{S}^+ & \boldsymbol{S}^{\mathrm{c}} \\ \boldsymbol{S}^{\mathrm{c}} & \boldsymbol{S}^- \end{pmatrix}; \quad \boldsymbol{S}_{ij}^{\pm} := \langle L_1^{\pm}\phi_i, \phi_j\rangle,\ \boldsymbol{S}_{ij}^{\mathrm{c}} := \langle L_1^{\mathrm{c}}\phi_i, \phi_j\rangle.$$

Note that $\boldsymbol{S}$ is symmetric and that $\boldsymbol{J}$ is skew-symmetric; furthermore, if J is in canonical form, i.e.,

$$J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix},$$

then

$$\boldsymbol{J} = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}.$$

Remark 4.1. If one considers the eigenvalue problem

$$\mathcal{L}u = \lambda u,$$

then up to $\mathcal{O}(\epsilon)$ one sees that

$$(\lambda_1 \mathbb{1} - \boldsymbol{S})\boldsymbol{c} = \boldsymbol{0}. \tag{4.11}$$

As an observation, which will be applicable to the real-valued solutions discussed in section 3, suppose that $L_1^{\mathrm{c}} = 0$, which implies that $\boldsymbol{S}^{\mathrm{c}} = \boldsymbol{0}$ and that $L_1^{\pm}$ are such that

$$\boldsymbol{S}^{\pm} = a_{\pm}\mathbb{1} + b_{\pm}\boldsymbol{S}_0, \tag{4.12}$$

where $\boldsymbol{S}_0 \in \mathbb{R}^{n\times n}$ and $a_{\pm}, b_{\pm} \in \mathbb{R}$. The linearization given in (4.1) satisfies this assumption with $a_{\pm} = -\Delta\omega$, $b_- = 1$, and $b_+ = 3$. If $\{s_1^0, \dots, s_n^0\} = \sigma(\boldsymbol{S}_0)$, then one can easily show that for (4.11), $\{a_{\pm} + b_{\pm}s_1^0, \dots, a_{\pm} + b_{\pm}s_n^0\} = \sigma(\boldsymbol{S}^{\pm})$. Now suppose that J is in canonical form. One can then easily show that for (4.10) the eigenvalues are given by

$$\lambda_j^{\pm} = \pm \mathrm{i}\sqrt{(a_- + b_- s_j^0)(a_+ + b_+ s_j^0)}. \tag{4.13}$$

Thus, in this case the eigenvalues will be either in $\mathbb{R}$ or in $\mathrm{i}\mathbb{R}$. Since the underlying system is Hamiltonian, if the eigenvalues are simple, this situation will hold true to all orders in the expansion. Finally, if $\lambda \in \mathrm{i}\mathbb{R}$ is simple, then the Krein signature is given by

$$K = \operatorname{sign}(a_- + b_- s_j^0). \tag{4.14}$$

This follows immediately upon using the diagonal structure of $\boldsymbol{S}$ and the result of [26, Theorem 3.3].

4.3. Application: N_{p} potential wells. As we will see herein, the perturbation results presented in section 4.2 can be used to determine the spectral stability of waves in the case of N_{p} potential wells, where, following the notation of Appendix A, one has that $N_{\mathrm{p}} \in \{2N, 2N+1\}$ with $N \geq 1$. All that is required is the following assumption on the form of the wave, which has been shown in section 3 to be true in the event that $N_{\mathrm{p}} = 3$. An analysis along the lines presented in section 3 also demonstrates its validity in the case that $N_{\mathrm{p}} = 2$.

Assumption 4.2. In the case of an odd number of potential wells one has that to leading order the nonlinear wave is of the form

$$Q(x) = \epsilon^{1/2}\sqrt{\frac{|\Gamma\,\Delta\omega|}{2}}\sum_{\ell=1}^{N_{\mathrm{p}}} \delta_\ell \Phi(x + 2t_{\mathrm{r}}(\ell)L) + \mathcal{O}(\epsilon),$$

where $\delta_\ell \in \{-1, 0, 1\}$, $\Delta\omega < 0$, and $t_{\rm r}(\cdot) : \mathbb{Z} \mapsto \mathbb{Z}$ is monotone decreasing with $t_{\rm r}(1) = N$ and $t_{\rm r}(N_{\rm p}) = -N$. If the number of potential wells is even, then the wave is of the above form with the mapping

$$2t_{\rm r}(\ell) \mapsto 2t_{\rm r}(\ell) - 1,$$

and $t_{\rm r}(N_{\rm p}) = -N + 1$. Any bifurcations occur for

$$|\Delta\omega| = \mathcal{O}(\mathrm{e}^{-2c_\omega L}).$$

In other words, as proved to be true in section 3 for the case $N_{\rm p} = 3$, we are assuming in Assumption 4.2 that each potential well is filled with the same 1-well eigenfunction, and that all sequences $\{\delta_1, \ldots, \delta_{N_{\rm p}}\}$ are achievable. Furthermore, it is being assumed that any saddle-node or pitchfork bifurcations are being controlled by the exponentially small interactions between adjacent potential wells.

For a given sequence $\{\delta_1, \ldots, \delta_{N_{\rm p}}\}$ extract all of the nonzero elements to form the subsequence $\{\delta_1^{\rm s}, \ldots, \delta_{N_{\rm f}}^{\rm s}\}$, where $1 \le N_{\rm f} \le N_{\rm p}$ is the number of filled potential wells. For example, if the sequence is given by $\{-1, 0, +1, 0, 0, +1\}$ ($N_{\rm p} = 6$), then the associated subsequence is given by $\{-1, +1, +1\}$ ($N_{\rm f} = 3$). Now set

$$\begin{aligned} p_{\rm i} &:= \{\#\text{ times } \delta_\ell^{\rm s}\delta_{\ell+1}^{\rm s} = +1,\ \ell = 1, \ldots, N_{\rm f} - 1\}, \\ p_{\rm o} &:= \{\#\text{ times } \delta_\ell^{\rm s}\delta_{\ell+1}^{\rm s} = -1,\ \ell = 1, \ldots, N_{\rm f} - 1\}. \end{aligned} \tag{4.15}$$

In (4.15) the quantity $p_{\rm i}$ counts the number of adjacent pulses which are in-phase, while $p_{\rm o}$ counts the number of adjacent pulses which are out of phase. Note that

$$p_{\rm i} + p_{\rm o} = N_{\rm f} - 1.$$

Upon letting $k_{\rm i}^+$ represent the number of purely imaginary eigenvalues in the upper half of the complex plane with positive Krein signature, the following theorem can now be proved. The implications for the case of $N_{\rm p} = 3$ are depicted in Figure 3.

Theorem 4.3. *Under Assumption 4.2, if $\epsilon > 0$ is sufficiently small, then for the $\mathcal{O}(\epsilon)$ eigenvalues one has that*

$$k_{\rm r} = p_{\rm i}, \quad k_{\rm i}^- = p_{\rm o}, \quad k_{\rm i}^+ = N_{\rm p} - N_{\rm f}, \quad k_{\rm c} = 0.$$

Furthermore, the eigenvalues in the upper half-plane with positive signature are given by $\mathrm{i}|\Delta\omega|\epsilon + \mathcal{O}(\epsilon^{3/2})$, whereas all of the other nonzero eigenvalues are $\mathcal{O}(\epsilon^{3/2})$.

Proof. For notational ease, and without loss of generality, it will be assumed that there is an odd number of potential wells. Recall that the location of the $\mathcal{O}(\epsilon)$ spectrum is given in section 4.2.

Following (4.10), one has that

$$(\boldsymbol{S}_0)_{jk} = E_0\langle Q^2, q_j q_k\rangle_{I_0}, \quad j, k = 1, \ldots, N_{\rm p}.$$

Since L is sufficiently large so that the interaction between adjacent pulses is exponentially small, upon ignoring these exponentially small terms one has that to leading order

$$Q^2(x) = -\epsilon\frac{\Gamma\,\Delta\omega}{2}\sum_{\ell=1}^{N_{\rm p}} \delta_\ell^2\Phi^2(x + 2t_{\rm r}(\ell)L).$$

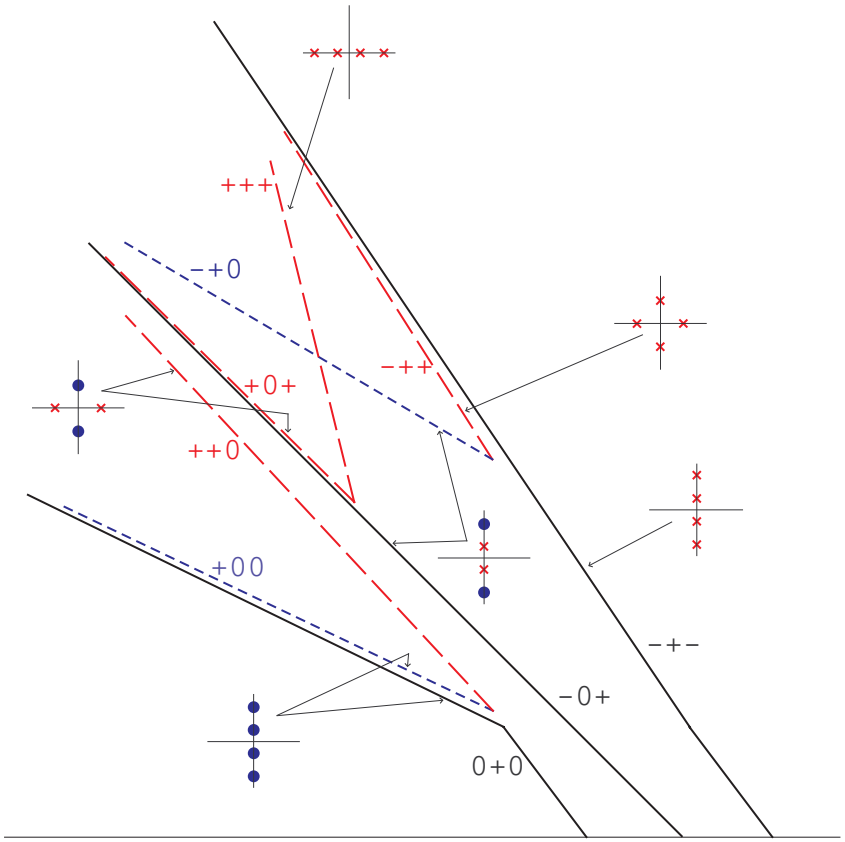


Figure 3. *The spectral stability diagram associated with the solutions found in section 3. The crosses represent either real eigenvalues, or purely imaginary eigenvalues with negative Krein sign. The circle represents a purely imaginary eigenvalue with positive Krein sign. See the caption associated with Figure 1 for further details.*

For the eigenvectors given in (A.13) associated with the $N_{\rm p}$-well eigenvalue problem, denote the matrix whose columns are these vectors by $\boldsymbol{E} \in \mathbb{R}^{N_{\rm p}\times N_{\rm p}}$. For the wave given in Assumption 4.2, for each ℓ such that $\delta_\ell = 0$ replace row ℓ in $\boldsymbol{E}$ with the zero row, and denote this new matrix by $\boldsymbol{E}_{\rm r}$. Using the expression for the eigenfunctions given in (A.10) or (A.11) then yields that to leading order

$$\begin{aligned}(\boldsymbol{S}_0)_{jk} &= E_0\langle Q^2, q_j q_k\rangle_{I_0}\\ &= -\frac{1}{2}E_0\Gamma\,\Delta\omega\langle \boldsymbol{E}_{\rm r}\boldsymbol{e}_j, \boldsymbol{E}_{\rm r}\boldsymbol{e}_k\rangle\,\langle \Phi^2, \Phi^2\rangle_{I_0},\end{aligned}$$

where $\boldsymbol{e}_i \in \mathbb{R}^{N_{\rm p}}$ denotes the ith unit vector. In other words, after ignoring the exponentially small terms, one has to leading order

$$\boldsymbol{S}_0 = -\frac{1}{2}E_0\Gamma\,\Delta\omega\langle \Phi^2, \Phi^2\rangle_{I_0}\boldsymbol{E}_{\rm r}^{\rm T}\boldsymbol{E}_{\rm r}. \tag{4.16}$$

Write $\sigma(\boldsymbol{S}_0) = \{s_1\Delta\omega, \ldots, s_m\Delta\omega\}$. Upon following the discussion leading to (4.13), one sees that to leading order the $\mathcal{O}(\epsilon)$ eigenvalues in $\sigma(L_\pm)$ for $j = 1, \ldots, N_\mathrm{p}$ are given by

$$\sigma(L_-) \,:\, -\Delta\omega(1+s_j)\epsilon, \qquad \sigma(L_+) \,:\, -\Delta\omega(1+3s_j)\epsilon. \tag{4.17}$$

Since $\ker(\boldsymbol{E}_\mathrm{r}^T\boldsymbol{E}_\mathrm{r}) = \ker(\boldsymbol{E}_\mathrm{r})$, the formulation of $\boldsymbol{S}_0$ in (4.16) implies that

$$\mathrm{z}(\boldsymbol{S}_0) = \mathrm{z}(\boldsymbol{E}_\mathrm{r}^\mathrm{T}\boldsymbol{E}_\mathrm{r}) = N_\mathrm{p} - N_\mathrm{f}. \tag{4.18}$$

To leading order, the eigenfunctions of $L_\pm$ are in $\mathrm{Span}\{\Phi(x + 2t_\mathrm{r}(\ell)L)\}$. Since $L_-Q = 0$, by applying (4.17) one then has that the eigenvalue associated with $s_j = -1$ has minimal geometric multiplicity N_f. Combining this with (4.18) then yields that $\sigma(\boldsymbol{S}_0) = \{0, -\Delta\omega\}$, where $-\Delta\omega$ is an eigenvalue with geometric multiplicity N_f, and 0 is an eigenvalue with geometric multiplicity $N_\mathrm{p} - N_\mathrm{f}$. As a consequence of this spectral decomposition of $\boldsymbol{S}_0$, as well as (4.17) and Sturm–Liouville theory, one can finally conclude that

$$\mathrm{n}(L_+) = N_\mathrm{f}, \quad \mathrm{n}(L_-) = p_\mathrm{o}. \tag{4.19}$$

Upon using the result of (4.19) in (4.6), one can conclude that for (4.1),

$$k_\mathrm{r} \geq [N_\mathrm{f} - 1] - p_\mathrm{o} = p_\mathrm{i} \tag{4.20}$$

and

$$k_\mathrm{r} + 2k_\mathrm{i}^- + 2k_\mathrm{c} = [N_\mathrm{f} - 1] + p_\mathrm{o} = p_\mathrm{i} + 2p_\mathrm{o}. \tag{4.21}$$

Now, there exist a total of $2N_\mathrm{p} - 2$ eigenvalues of $\mathcal{O}(\epsilon)$, which implies that at $\mathcal{O}(\epsilon)$,

$$2k_\mathrm{r} + 2(k_\mathrm{i}^- + k_\mathrm{i}^+) + 4k_\mathrm{c} = 2N_\mathrm{p} - 2. \tag{4.22}$$

Now recall (4.13), which shows that to leading order the $\mathcal{O}(\epsilon)$ eigenvalues for (4.1) are given by

$$\lambda_j^\pm = \pm\mathrm{i}\,\Delta\omega\sqrt{(1+s_j)(1+3s_j)}\,\epsilon; \tag{4.23}$$

furthermore, the Krein signature of purely imaginary eigenvalues is given by $\mathrm{sign}(1 + s_j)$. If $s_j = 0$, then the associated eigenvalue is purely imaginary with positive sign. There are $2(N_\mathrm{p} - N_\mathrm{f})$ eigenvalues of this type, i.e.,

$$k_\mathrm{i}^+ = N_\mathrm{p} - N_\mathrm{f}. \tag{4.24}$$

If $s_j = -1$, then to leading order (4.23) yields no information, so that a more detailed analysis is necessary. In conclusion, at $\mathcal{O}(\epsilon)$ one has that

$$k_\mathrm{i}^+ = N_\mathrm{p} - N_\mathrm{f}, \qquad k_\mathrm{i}^- = k_\mathrm{r} = k_\mathrm{c} = 0. \tag{4.25}$$

As a consequence of the above discussion one has that the eigenvalues satisfying (4.21) are necessarily $\mathcal{O}(\epsilon^{3/2})$. Furthermore, by (4.25) one has that the total remaining number of nonzero small eigenvalues of $\mathcal{O}(\epsilon^{3/2})$ satisfies

$$2k_{\mathrm{r}} + 2(k_{\mathrm{i}}^{-} + k_{\mathrm{i}}^{+}) + 4k_{\mathrm{c}} = (2N_{\mathrm{p}} - 2) - 2(N_{\mathrm{p}} - N_{\mathrm{f}}) \\ = 2(p_{\mathrm{i}} + p_{\mathrm{o}}). \tag{4.26}$$

Combining (4.21) with (4.26) yields the relationship

$$k_{\mathrm{r}} - p_{\mathrm{i}} + 2(k_{\mathrm{i}}^{+} + k_{\mathrm{c}}) = 0;$$

thus, as a consequence of (4.20), one necessarily has that

$$k_{\mathrm{r}} = p_{\mathrm{i}}, \qquad k_{\mathrm{i}}^{+} = k_{\mathrm{c}} = 0.$$

Substitution of the above into (4.21), as well as (4.24), yields the desired final counts for $k_{\mathrm{i}}^{\pm}$. ∎

Remark 4.4. We expect that a more definitive determination of the location of the $\mathcal{O}(\epsilon^{3/2})$ point spectra will require the type of analysis presented in [46].

Remark 4.5. As a consequence of Theorem 4.3, (4.20), and the discussion presented in [26, section 2.2], one has that for a fixed value of ϵ sufficiently small there is the upper bound

$$k_{\mathrm{c}} \leq \min\{N_{\mathrm{p}} - N_{\mathrm{f}}, p_{\mathrm{o}}\}$$

of $\mathcal{O}(\epsilon)$ eigenvalues. In other words, for ϵ sufficiently small, complex eigenvalues with nonzero real part can be created only by the collision of eigenvalues of opposite sign on the imaginary axis.

Remark 4.6. It is interesting to note that the result of Theorem 4.3 exactly matches that of [42, Theorem 3.6]. In particular, this seems to imply that there may be a strong correlation between studies of soliton solutions to a discrete nonlinear Schrödinger equation (NLS) near the anticontinuum limit and those studies initiated herein. In one sense this is not surprising, as it was shown in [29, section 3] (also see the references therein) that in the case of a periodic potential, (2.4) can formally be reduced to a discrete NLS. However, given that the potentials discussed here are not periodic, the connection is still somewhat tenuous. It would be very interesting to see whether the correlation still holds when discussing (2.4) in the case of two space dimensions. (See [43] for a discussion of discrete vortices in two-dimensional lattices.)

5. Numerical results. We now numerically examine the solutions of the full (2.2) and compare the results to the analysis presented in section 3. Nonlinear bound states of the equation are localized solutions of the form $q(x,z) = u(x)\exp(i\mu z)$, where u obeys

$$u_{xx} - \frac{E_0}{1 + I_0(x) + |u|^2}u = \mu u \tag{5.1}$$

and μ is the "chemical potential" in BEC applications or the propagation constant in optics applications. Note that u in (5.1) is Q in (2.6). Further note that $\mu = -\omega$, where ω is the

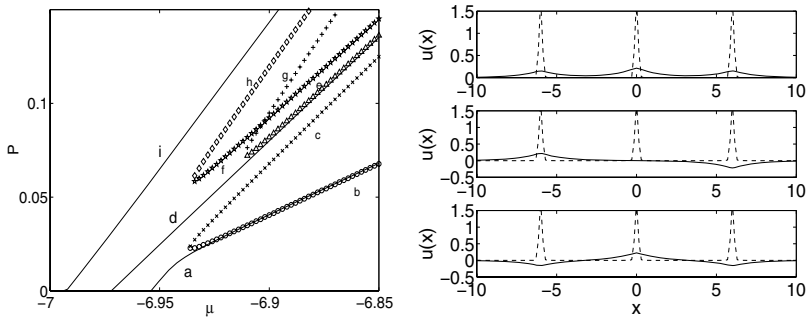


Figure 4. *The bifurcation diagram of various branches is shown on the left for the optical power P as a function of μ. Figures 5–7 show the profiles and linear stability of each of the branches (*(a) *through* (i)*, labeled accordingly below) shown in this bifurcation diagram. The right panel shows the potential (dashed line) and its three linear eigenstates (solid lines).*

frequency parameter used in the earlier sections; consequently, for the sake of comparison, the results associated with Figure 3 should be reflected as if held to a mirror.

The 3-well potential that we will examine will be of the form given in (2.9) with

$$G(x) = \exp\left(-\frac{x^2}{2\epsilon^2}\right); \tag{5.2}$$

in other words, the potential corresponds to a superposition of three Gaussian beams. With respect to (2.8) one is assuming that $C = 1$ and $\epsilon = a$. We examine the case of $E_0 = 7.5$ [49] with $L = 3$, $\epsilon = 0.1$, and $V_0 = 1/\sqrt{\pi\epsilon}$ [31]. Solutions to (5.1) with $I_0(x)$ given by (5.2) are obtained via a fixed point iteration on a finite-difference grid. The linear stability of the stationary states is subsequently determined by identifying the eigenvalues and eigenvectors $\{\lambda, (a, b)\}$ of the linearization equation. The latter are obtained using the ansatz

$$q(x, z) = \exp(i\mu z)\left\{u(x) + \delta\left[\exp(-\lambda z)a(x) + \exp(-\lambda^\star z)b^\star(x)\right]\right\}$$

in (2.2) and identifying the terms of $\mathcal{O}(\delta)$.

The bifurcation picture of the problem is presented in terms of the power of the nonlinear bound state, $P = \int_{-\infty}^{+\infty} |u(x)|^2 \mathrm{d}x$, as a function of μ. Figure 4 encapsulates the main findings of the study. It shows the full bifurcation diagram (left panel) of the various branches present in the 3-well problem and also illustrates the linear states (right panel) from which the corresponding nonlinear ones stem. It is immediately worth noting here the excellent qualitative agreement that this figure bears with Figure 1 of the existence analysis given in section 3. The ground state of the system is the linear state at the top of the right panel with $\mu = -6.959$. The first excited state (middle of the right panel) is an odd state with $\mu = -6.978$, while the second excited state (bottom of the right panel) is again an even state with $\mu = -6.998$. This means in connection with the theoretical results that $\Delta\lambda = 0.0197 \pm 0.0011$.

Now we can clearly observe on the left panel that there are three corresponding nonlinear branches that bifurcate from these linear states, namely (a) $(0 + 0)$, (d) $[\mathcal{R}_2(-0+)]$, and (i) $[\mathcal{R}_2(-+-)]$ in the relevant diagram (also see Figure 1). For the chosen parameter values there are no other localized states for the linear problem. Branch (a) persists for all values

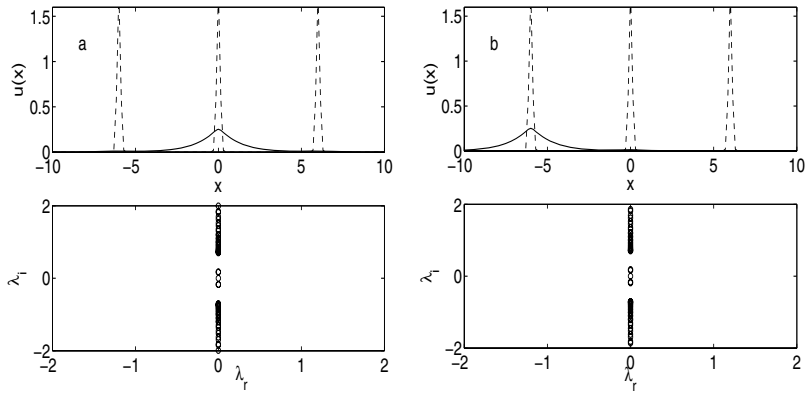


Figure 5. *The top panel shows the* 1*-pulse configurations, and the bottom shows a typical example of the spectral plane* (λ_r, λ_i) *of their eigenvalues. The labeling of the branches follows the bifurcation diagram of Figure* 4*.*

of μ, while the similar-looking one-pulse state of branch (b) (+00) (and similarly its mirror image $[\mathcal{R}_1(+00)]$) disappears through a saddle-node bifurcation with the unstable branch (c) $(++0)$. One of the most stringent tests to which we subject our theory is the comparison of the full numerical results with the critical points for the relevant saddle-node bifurcations from our analysis in section 3. In the case of branches (b) and (c), the relevant theoretical prediction for the bifurcation is $\mu = -(\lambda_2 + 2\Delta\lambda) = -6.9385 \pm 0.0023$ (based on the $\Delta\lambda$ given above). The relevant numerical finding is -6.937 ± 0.001, in excellent agreement with the analysis presented above. The single pulse branches (a) and (b) (shown in Figure 5) are stable throughout the domain of their existence, while branch (c) possesses two adjacent pulses with the same parity and hence should be unstable (also see [25] and the references therein).

On the other hand, the antisymmetric branch (d) $[\mathcal{R}_2(-0+)]$ (a twisted-mode similar to the ones of [30]; see also references therein) persists to the linear limit. Such a solution is potentially unstable due to the existence of an eigenvalue of negative Krein signature, which can lead to a quartet of complex eigenvalues upon the occurrence of a Hamiltonian Hopf bifurcation; e.g., see [26] and references therein. Finally, as regards 2-pulse states (all of which are shown in Figure 6), branch (f) $[\mathcal{R}_2(-+0)]$ contains two opposite parity pulses, which are nearest neighbors, while branch (e) (+0+) contains same parity, non–nearest-neighbor pulses. The former branch also possesses an eigenvalue pair of negative Krein sign, while the latter has a real eigenvalue pair leading to instability. Notice that branch (e) disappears in a saddle-node bifurcation with the 3-same-parity-pulse branch (g) $(+++)$ for $\mu \approx -6.911 \pm 0.001$; in this case, the theoretical prediction for the termination of the branch is $\mu = -(\lambda_2 + 3.367\Delta\lambda) = -6.9116 \pm 0.0039$, which again is in excellent agreement with the above numerical result. On the other hand, branch (f) disappears in a saddle-node bifurcation with the 3-pulse asymmetric branch (h) $[\mathcal{R}_2(--+)]$ for $\mu \approx -6.933 \pm 0.001$. In this case, the theoretical critical point is predicted to be at $\mu = -(\lambda_2 + 2\Delta\lambda) = -6.9385 \pm 0.0023$, which is again close to the obtained numerical result.

From the 3-pulse configurations, all of which are shown in Figure 7, only the symmetric

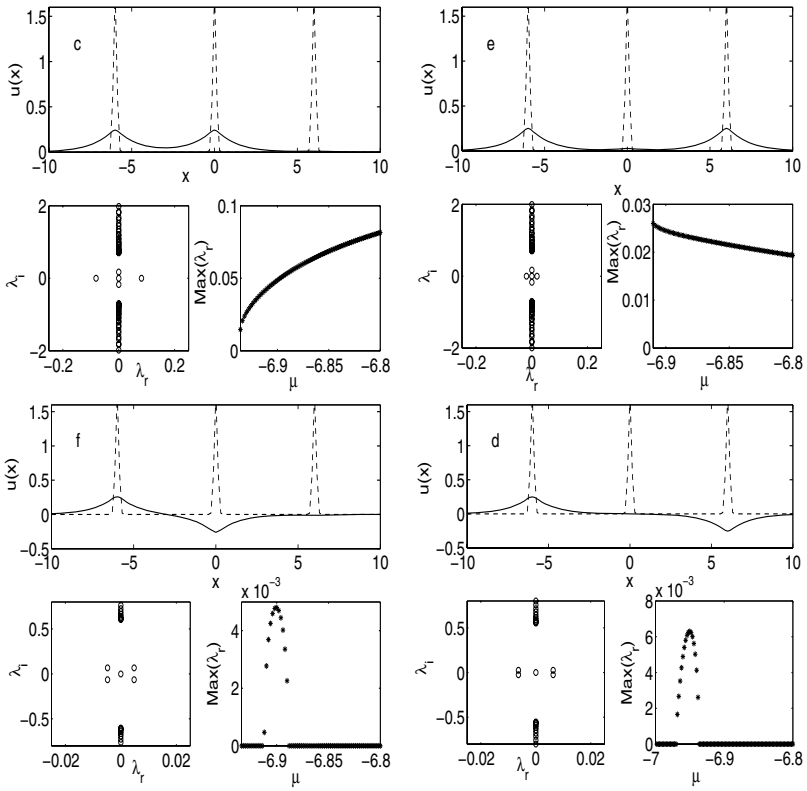


Figure 6. *Same as the previous figure, but for 2-pulse states. Once again the labeling of the branches is consonant with the bifurcation diagram of Figure 4. In the bottom-right subplot, the behavior of (the real part of) the most unstable eigenvalue is shown.*

but alternating parity branch can be stable and persist to the linear limit. Any 3-pulse pair that contains neighboring pulses of the same parity, i.e., (g) and (h), has one real eigenvalue pair per same-parity-pulse pair and is thus always unstable. Notice that all of the stability results reported above and illustrated in Figures 5, 6, and 7 fully corroborate the analytical picture given in section 4 and illustrated in Figure 3.

6. Experimental results. Admittedly, the theoretical setup presented in this study cannot be immediately connected with experiments in latices induced in photorefractive crystals; e.g., see [7, 8]. In the latter, typically a fully two-dimensional waveguide lattice is formed, as is explained in detail below. A natural question that is then posed is how this setting can be related to our results about 3-well configurations. We argue that this is a matter of scales, i.e., of propagation distances. In particular, if two beams are launched between three lattice sites (and along one dimension of the waveguide lattice), in the early stages of propagation and due to the geometric proximity, the dynamics of light will be confined within these three sites and will be effectively amenable to the description of the PDE with the three potential wells. Obviously, for later stages of the evolution, the dynamics will be far more complex, but as our experimental results will show, the three adjacent wells are still dominant. It is in that light that our experimental results presented below should be considered.

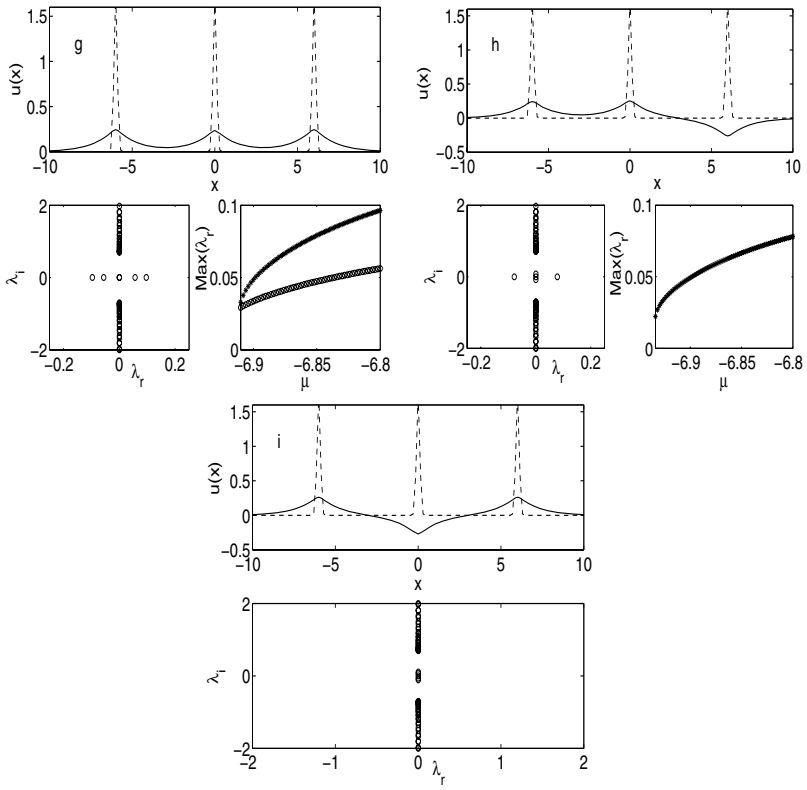


Figure 7. *Same as Figure 6, but for 3-pulse configurations.*

The experimental setup for our study is similar to those used in [7, 8]. A partially spatially incoherent light beam of wavelength $\lambda = 488$ nm is generated by converting an argon ion laser beam into a quasi-monochromatic light source with a rotating diffuser. Such a diffused laser source has the advantage of providing an incoherent beam with a controllable degree of spatial coherence and intensity suitable for experiments with partially coherent light. A biased photorefractive crystal (strontium barium niobate, SBN:60) is employed to provide a saturable self-focusing nonlinearity. To generate a two-dimensional waveguide lattice, we use an amplitude mask to spatially modulate the otherwise uniform incoherent beam after the diffuser. The mask is then properly imaged onto the input face of the crystal, thus creating an intensity pattern of a square lattice with its principal axes oriented in horizontal/vertical directions. This lattice beam is ordinarily polarized; thus it induces a nearly linear waveguide array. An extraordinarily polarized Gaussian beam split from the same laser but without passing through the diffuser is used as the probe beam propagating along with the lattice, as in most of our experiments with discrete solitons [9, 35, 39]. The probe can be a single component Gaussian beam, or it can be sent into a Mach–Zehnder interferometer to create two extraordinarily polarized probe beams propagating collinearly through the crystal. With a piezo-transducer mirror, the two beams can be made mutually incoherent or mutually coherent with either in-phase or out-of-phase relation. The probe beams are monitored with a CCD camera. In addition, a uniform incoherent beam is used as background illumination for fine-

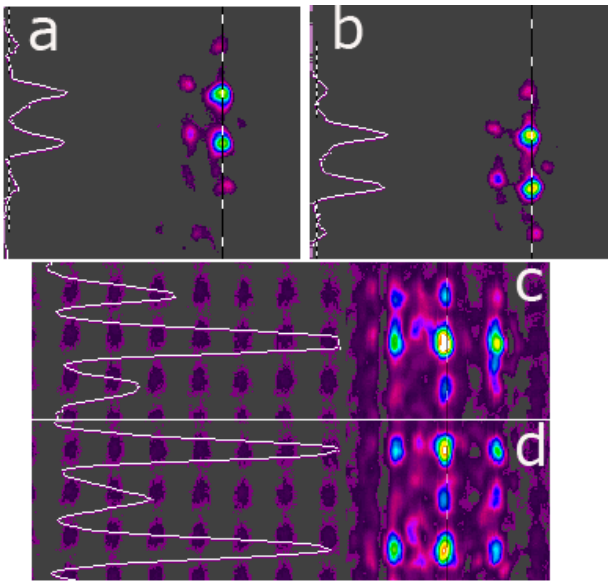


Figure 8. *Experimental results for two probe beams launched between three potential wells embedded in an optically induced two-dimensional photonic lattice. All photographs were taken at the output of a photorefractive crystal.* (a) *and* (b): *only one probe beam is present, which tunnels into two nearby lattice sites.* (c) *and* (d) *two in-phase and out-of-phase, respectively, probe beams are present, which tunnel into the central site* (c) *or two lateral sides* (d), *depending on the relative phase.*

tuning the nonlinearity.

First, a single Gaussian beam ($\lambda = 488$ nm) was used as a coherent probe beam propagating between two lattice sites in the vertical direction. When the probe beam was extraordinarily polarized, we observed spontaneous symmetry breaking (SSB) towards an asymmetric state (even though the input had a nearly symmetric profile) as the intensity of the probe beam was increased gradually while keeping all other experimental conditions unchanged, which constitutes a typical behavior expected in a double-well system [31]. To demonstrate that such SSB was induced by nonlinearity, the experiment was repeated under the same conditions, except that the 488 nm probe beam was replaced by a 632.8 nm laser beam, which is at a much less photosensitive wavelength. The SSB did not occur at this wavelength, regardless of the increase in the probe intensity.

Next, and more directly related to the theoretical work presented in this paper, two Gaussian probe beams ($\lambda = 488$ nm) exiting the Mach–Zehnder interferometer were launched between three lattice sites to test the dynamics in a 3-well potential. The two beams were launched collinearly with the lattice beam, but aimed at two intensity minima in the y-direction (rather than maxima, as in the experiments with discrete solitons) so as to have inter-site excitations. Typical experimental results are shown in Figure 8. The intensity of each probe is chosen so that each beam alone would give rise to a symmetric pattern (i.e., below the threshold for SSB in a double-well potential), as shown in the top panel of Figure 8. When two beams are launched simultaneously, the outcome of the intensity pattern is quite dependent on the relative phase between the probe beams. If the two beams are in phase,

most of the energy goes to the central site due to constructive interference, leading to a stable state of a nearly single pulse discussed before (branch (a) in Figure 4). Radiation to other lattice sites in both x- and y-directions is clearly visible due to strong intersite coupling. On the other hand, when the two beams are made out-of-phase, most of the energy goes to the two lateral sites (in the y-direction) due to destructive interference, leading to a stable state of two pulses with antiphase relation. This out-of-phase excitation corresponds to the branch (d) of the bifurcation diagram discussed in section 5. Details of this experimental work will be reported elsewhere.

As can be seen in Figure 8, the experimental context where our theoretical predictions can be tested is in settings where two light pulses are injected *between* three wells of a quasi–one-dimensional configuration. The relevant pulses can be in-phase (panel (c) of Figure 8) or out-of-phase (panel (d) of Figure 8) between them. (For comparison, the case of a single probe beam tunneling between two lattice wells is also shown in panels (a) and (b) of the figure.) In the in-phase, 3-well case, the dynamics of the evolution results in a pulse centered predominantly at the central lattice site. This is a natural consequence of our stability analysis, since states with same parity pulses (with either two or three constituent pulses) are unstable, and hence the only dynamically stable solution that such an in-phase configuration can explore is the single pulse configuration (centered at the central site, given the constructive interference of the two pulses injected between the three wells). On the other hand, the out-of-phase pulses will result in destructive interference at the central well but will lead to a stable $[\mathcal{R}_2(-0+)]$ configuration of branch (d). Notice that these types of dynamical evolution are also supported by the direct numerical results from the integration of (2.1). Such results are shown in Figure 9, emulating the experimental setting. (The in-phase case is shown in the top panels, while the out-of-phase case is given in the bottom panels.)

In concluding this section, we note that clearly the 3-mode approximation used for the analytical results obtained in this paper is not directly applicable to the experimental results of this section. However, the relevance can be seen from the predicted stability and experimentally observed robustness of two of the configurations that were shown in the 3-mode approximation and in the full numerical results of section 5, namely the single pulse $(0+0)$ as well as the anti-symmetric branch $(-0+)$. In light of these connections, we feel that these experimental results usefully corroborate the analytical and numerical predictions presented herein.

7. Final remarks. From the results presented in sections 3 and 4 we can conclude that the theoretical picture obtained by the Lyapunov–Schmidt reductions and the stability analysis agrees very well with the numerically obtained results of section 5. In addition, some of the predicted stable branches can also be observed to be of experimental relevance, as discussed in section 6. It is interesting to compare/contrast the situation in the present setting with that of the much simpler 2-well configuration, as well as with that of the infinite lattice of wells, to illustrate how the passage from $1 \to 2 \to 3 \to \cdots \to \infty$ wells occurs. The case of a single well is very simple, as a single stable mode exists and no bifurcations may occur. The addition of a second well makes the problem considerably more interesting, but due to the presence of conservation laws the cases are still pretty special [20, 45, 24, 31, 44]. For example, in this case a symmetric same parity pulse pair may be stable if one is $\mathcal{O}(\mathrm{e}^{-2c_\omega L})$ close to the

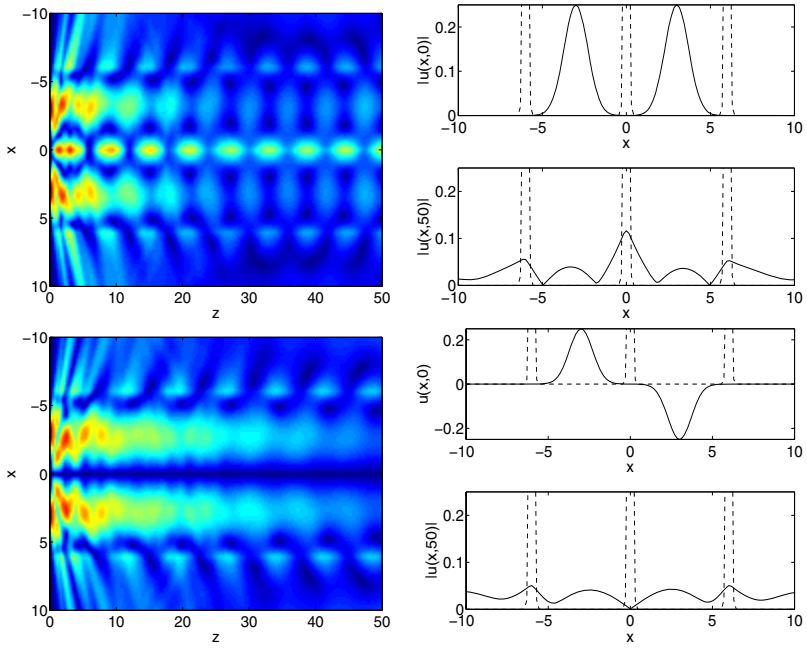


Figure 9. *The top panels show the evolution of* (2.1) *for* $q(x,z)$ *in the case of an initial condition of two in-phase pulses initialized between three wells of the lattice. The left panel shows the contour plot of the z-propagation of the field's modulus. The right panel shows the initial condition and the profile of the solution's modulus at* $z = 50$*. The bottom panels illustrate similar features, but for an initial condition with two out-of-phase pulses.*

linear limit. Furthermore, the only bifurcation that can occur is that of an asymmetric mode from the combination of the symmetric and antisymmetric states. In the 3-well case, things are starting to look considerably more like the infinite lattice. The stability of all the states examined herein coincides with the stability and potential instabilities that one would find for these states in the infinite lattice limit [42]. While some bifurcations still occur, which are spurious in the infinite well limit, these are now saddle-node, rather than the supercritical pitchfork bifurcation of the 2-well case. Indeed the saddle-node bifurcations, such as those annihilating branches like (b) and (c), will disappear in the infinite-lattice limit because of the integer-shift invariance that will "collapse" branch (b) into branch (a), while it will distinctly separate it from the two-site branch (c) [2]. We believe that the above picture illustrates how in our setting three wells is a crowd, or, in fact, how three wells is starting to better resemble an infinite lattice.

Appendix A. Spectral analysis for multiple wells. Herein we will compute $\sigma_{\mathrm{p}}(\mathcal{L}_\omega)$ in the case of a multiple-well potential, where $\mathcal{L}_\omega$ is the Schrödinger operator given in (2.7):

$$\mathcal{L}_\omega := -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{E_0}{1 + I_0(x)} - \omega, \quad x \in \mathbb{R}.$$

It will henceforth be assumed that the spectrum is known in the case of a single-well potential given in (2.11). As seen in Appendix B, this calculation is straightforward in the case of a

square-well potential. Since $\mathcal{L}_\omega$ is a Schrödinger operator, the eigenvalues for any potential must all be simple and real-valued; furthermore, since the potential is assumed to be even, it is known that the eigenfunctions will be either even or odd. For the single-well potential given in (2.11), let $\omega < E_0$ be chosen so that $\{0\} \subset \sigma_{\rm p}(\mathcal{L}_\omega)$, and let $\Phi(x)$ represent the eigenfunction associated with the simple null eigenvalue which satisfies $\langle \Phi, \Phi \rangle = 1$. Recall that Φ is either even or odd in x. It is expected that for $L \gg 1$ one can apply an appropriate modification of the results of [46] to determine the location of the $2N+1$ exponentially small eigenvalues for the potential of (2.9) with $(2N+1)$ wells, or the $2N$ exponentially small eigenvalues for the potential of (2.10) with $2N$ wells.

For the potential given in (2.11) and satisfying (2.8), rewrite the eigenvalue problem $\mathcal{L}_\omega \phi = \lambda \phi$ as

$$\boldsymbol{u}' = (\boldsymbol{A}_0(x) + \lambda \boldsymbol{B})\boldsymbol{u}, \qquad \boldsymbol{u} := \left(\phi, \phi'\right)^{\rm T} \in \mathbb{C}^2, \tag{A.1}$$

and note that as a consequence of (2.8) one has a $\boldsymbol{A}_0^\infty \in \mathbb{R}^{2\times 2}$ such that

$$\|\boldsymbol{A}_0(x) - \boldsymbol{A}_0^\infty\| \leq C{\rm e}^{-x^2/2a^2}. \tag{A.2}$$

We are assuming that (A.1) has the exponentially decaying solution $\boldsymbol{v}_0(x) := (\Phi(x), \Phi'(x))^{\rm T}$ for $\lambda = 0$, and that the matrix $\boldsymbol{A}_0^\infty$ is hyperbolic (i.e., $\lambda = 0$ is not in the essential spectrum). Let $\psi(x) = (-\Phi'(x), \Phi(x))^{\rm T}$ represent the exponentially decaying solution to the adjoint problem associated with (A.1) at $\lambda = 0$.

In order to understand the following argument, it is necessary for the reader to become acquainted with the ideas in [46, section 3]. Consider the system

$$\boldsymbol{u}' = (\boldsymbol{A}_0(x) + \boldsymbol{A}_1(x) + \lambda \boldsymbol{B})\boldsymbol{u}, \quad x \in [0, L_j], \tag{A.3}$$

together with the appropriate matching and jump conditions at $x = L_j$ and $x = 0$, respectively, where $\boldsymbol{A}_1(x)$ satisfies

$$\|\boldsymbol{A}_1(x)\| \leq C{\rm e}^{-L^2/2a^2}, \quad x \in [0, L], \quad L := \min_j\{L_j\}. \tag{A.4}$$

One thinks of $\boldsymbol{A}_1(x)$ as representing the tails of neighboring widely spaced copies of $\boldsymbol{A}_0(x)$, e.g., the matrix generated by considering the potential of (2.9). If $\boldsymbol{A}_0(x)$ satisfies (A.2), then $\boldsymbol{A}_0(x - 2L_j)$ evaluated on $[0, L_j]$ will satisfy (A.4). Note that the estimate given in (A.4) is stronger than is assumed in [46]; however, this is simply a consequence of (A.2).

In order to solve (A.3) for $|\lambda| \ll 1$, consider the ansatz

$$\boldsymbol{u}(x) = \boldsymbol{u}_j(x) := d_j \boldsymbol{v}_0(x) + \boldsymbol{w}_j(x), \tag{A.5}$$

which upon substitution into (A.3) yields the system

$$\boldsymbol{w}_j' = (\boldsymbol{A}_0(x) + \boldsymbol{A}_1(x) + \lambda \boldsymbol{B})\boldsymbol{w}_j + d_j(\boldsymbol{A}_1(x) + \lambda \boldsymbol{B})\boldsymbol{v}_0(x), \quad x \in [0, L_j]. \tag{A.6}$$

The matching condition at $x = L_j$ reads

$$\boldsymbol{w}_j(L_j) + d_j \boldsymbol{v}_0(L_j) = \boldsymbol{w}_{j+1}(-L_j) + d_{j+1}\boldsymbol{v}_0(-L_j),$$

from which one finds the leading-order solution to be

$$\boldsymbol{w}_j(L_j) = d_{j+1}\boldsymbol{v}_0(-L_j)$$

[46, equation (3.25)]. Proceeding as in [46], one gets the contribution

$$\langle \psi(L_j), \boldsymbol{v}_0(-L_j)\rangle d_{j+1} - \left(\int_0^{+\infty} \langle \psi(x), \boldsymbol{B}\boldsymbol{v}_0(x)\rangle \, \mathrm{d}x\right) \lambda d_j$$

to the jump $\langle \psi(0), \boldsymbol{u}_j^+(0) - \boldsymbol{u}_j^-(0)\rangle$ of the jumps evaluated in the ψ-direction. Indeed, the additional term $\boldsymbol{A}_1(x)\boldsymbol{v}_0(x)d_j$ that appears in (A.6) compared with [46, equation (3.9)] is of $\mathcal{O}(\mathrm{e}^{-L^2/2a^2})$, which is small relative to the leading-order $\mathcal{O}(\mathrm{e}^{-2c_\omega L_j})$, $c_\omega := \sqrt{E_0 - \omega}$. It is at this point that the proof of the stability result in [46] would fail to give the leading-order behavior if the ansatz in (A.5) had been used instead of the more accurate ansatz [46, equation (3.5)], for in that context the correction $\boldsymbol{A}_1(x)\boldsymbol{v}_0(x)$ is of $\mathcal{O}(\mathrm{e}^{-2c_\omega L_j})$.

Upon applying the above theory to the symmetric potential given in (2.9) (consequently, $L_j \equiv L$), and using the fact that the eigenfunction has been normalized, one sees that for $N \in \mathbb{N}$ one has the reduced eigenvalue problem

$$(\boldsymbol{A} - \lambda \mathbb{1})\boldsymbol{d} = 0. \tag{A.7}$$

In the case of an even number of potential wells one has that $\boldsymbol{A} \in \mathbb{R}^{2N\times 2N}$ is a symmetric tridiagonal matrix that satisfies

$$\begin{aligned} \boldsymbol{A}_{j,j} &= 0, \quad j = 1, \ldots, 2N, \\ \boldsymbol{A}_{j-1,j} &= a_L^{\pm}, \quad j = 2, \ldots, 2N, \end{aligned} \tag{A.8}$$

whereas in the event of an odd number of potential wells the matrix $\boldsymbol{A}$ is that given in (A.8), except that now $\boldsymbol{A} \in \mathbb{R}^{(2N+1)\times(2N+1)}$. Here

$$a_L^{\pm} := \pm 2\Phi(L)\Phi'(L), \qquad |a_L^{\pm}| = \mathcal{O}(\mathrm{e}^{-2c_\omega L}). \tag{A.9}$$

The constant a_L^+ in (A.8) is taken if Φ is even; otherwise, a_L^- is taken. Note that $a_L^+ < 0$ and $a_L^- > 0$. The vector $\boldsymbol{d}$ is such that for the given eigenvalue λ, the associated eigenfunction is to leading order given by

$$q = \sum_{j=0}^{N-1} [d_{j+1}\Phi(x + (2(N-j)-1)L) + d_{N+j+1}\Phi(x - (2(j+1)-1)L)] \tag{A.10}$$

(potential of (2.10)), or

$$q = d_N\Phi(x) + \sum_{j=0}^{N-1} [d_{j+1}\Phi(x + 2(N-j)L) + d_{N+j+2}\Phi(x - 2(j+1)L)] \tag{A.11}$$

(potential of (2.9)). Upon setting $m = 2N$ in the case of an even number of potential wells, or $m = 2N+1$ in the case of an odd number of potential wells, (A.7) can then be rewritten as the autonomous difference equation for $j = 1, \ldots, m$,

$$a_L(d_{j-1} + d_{j+1}) = \lambda d_j, \quad d_0 = d_{m+1} = 0. \tag{A.12}$$

If for $\ell = 1, \dots, m$ one sets

$$\theta_\ell := \frac{\pi}{m+1}\ell,$$

then the solution pair (λ_ℓ, d_ℓ) to (A.12) is given by

$$\lambda_\ell = 2a_L^{\pm}\cos\theta_\ell; \quad (d_\ell)_j = \sin(j\theta_\ell),\ j = 1, \dots, m. \tag{A.13}$$

Note that for $k = 1, \dots, N$,

$$\lambda_k = -\lambda_{m+1-k}; \tag{A.14}$$

furthermore, in the case of an odd number of potential wells one has that $\lambda_N = 0$.

Appendix B. Spectral analysis for a square-well potential. Consider the operator $\mathcal{L}$ given in (2.7) with the potential

$$I_0(x) = \begin{cases} V_0, & |x| < L, \\ 0, & |x| \geq L. \end{cases}$$

In this case the eigenvalue problem $(\mathcal{L} - \lambda)q = 0$ can be solved explicitly, and it is a standard exercise to do so. First, the operator $\mathcal{L}$ is self-adjoint, so that the spectrum is real. Second, the essential spectrum is given by

$$\sigma_{\mathrm{e}}(\mathcal{L}) = \{\lambda \in \mathbb{R} : \lambda \geq E_0\};$$

consequently, when searching for isolated point spectra, one must focus on $\lambda \in \mathbb{R}\backslash[E_0, +\infty)$. Since $I_0(x)$ is even in x, one knows that the eigenfunctions will be either odd or even; in particular, the eigenfunctions will satisfy the following initial conditions:

$$\begin{aligned} \text{even}: &\quad (q(0), q'(0)) = (1, 0), \\ \text{odd}: &\quad (q(0), q'(0)) = (0, 1). \end{aligned} \tag{B.1}$$

Set

$$\gamma := \sqrt{E_0 - \lambda}, \qquad \mu := \sqrt{\frac{E_0}{1+V_0} - \lambda}, \tag{B.2}$$

and assume that $\gamma(\lambda) \in \mathbb{R}^+$ for $\lambda < E_0$. Note that $\mu(\lambda)$ has a branch point at $\lambda = \lambda_{\mathrm{b}} := E_0/(1+V_0)$. Define the branch cut so that $\mu(\lambda) \in \mathrm{i}\mathbb{R}^+$ for $\lambda > \lambda_{\mathrm{b}}$. Upon solving the eigenvalue problem with the initial conditions given in (B.1), and defining the two Evans functions

$$\begin{aligned} E_{\mathrm{e}}(\lambda) &:= \mu(\lambda)\sinh(\mu(\lambda)L) + \gamma(\lambda)\cosh(\mu(\lambda)L), \\ E_{\mathrm{o}}(\lambda) &:= \gamma(\lambda)\frac{\sinh(\mu(\lambda)L)}{\mu(\lambda)} + \cosh(\mu(\lambda)L), \end{aligned} \tag{B.3}$$

one sees that an even eigenfunction exists if and only if $E_{\mathrm{e}}(\lambda) = 0$, and an odd eigenfunction exists if and only if $E_{\mathrm{o}}(\lambda) = 0$.

Since $\mu(\lambda), \gamma(\lambda) \in \mathbb{R}^+$ for $\lambda < \lambda_b$, one clearly has that $E_{e,o}(\lambda) \neq 0$ for $\lambda < \lambda_b$. Now assume that $\lambda_b < \lambda < E_0$. One has that $E_e(\lambda) = 0$ can be rewritten as

$$\sin(|\mu(\lambda)|\, L + \phi_e(\lambda)) = 0, \tag{B.4}$$

where

$$\tan\phi_e := -\frac{\gamma}{|\mu|}, \qquad \frac{\pi}{2} < \phi_e < \pi, \tag{B.5}$$

and $E_o(\lambda) = 0$ can be rewritten as

$$\frac{\sin(|\mu(\lambda)|\, L + \phi_o(\lambda))}{|\mu(\lambda)|} = 0, \tag{B.6}$$

where

$$\tan\phi_e := \frac{|\mu|}{\gamma}, \qquad 0 < \phi_o < \frac{\pi}{2}. \tag{B.7}$$

Solving (B.4) and (B.6) then yields that the eigenvalues are solutions of the transcendental equation

$$\lambda = \frac{E_0}{1+V_0} + \left(\frac{k\pi - \phi_{e,o}(\lambda)}{L}\right)^2, \quad k \in \mathbb{N}. \tag{B.8}$$

Regarding (B.8), one has the following observations. First, since eigenvalues must satisfy $\lambda < E_0$, one is interested only in those solutions to (B.8) which satisfy the constraint

$$\frac{1+V_0}{V_0}\left(k\pi - \phi_{e,o}(\lambda)\right)^2 < E_0 L^2; \tag{B.9}$$

if the inequality is reversed, then an isolated eigenvalue cannot exist for that value of k. As a consequence of the constraints given in (B.5) and (B.7), for a particular value of $k \in \mathbb{N}$ one is guaranteed a solution to (B.4) if

$$\frac{1+V_0}{V_0}\left(k - \frac{1}{2}\right)^2 \pi^2 < E_0 L^2, \tag{B.10}$$

and is guaranteed the nonexistence of a solution if

$$E_0 L^2 < \frac{1+V_0}{V_0}\left(k-1\right)^2 \pi^2. \tag{B.11}$$

Similarly, one is guaranteed a solution to (B.6) if

$$\frac{1+V_0}{V_0} k^2 \pi^2 < E_0 L^2, \tag{B.12}$$

and is guaranteed the nonexistence of a solution if

$$E_0 L^2 < \frac{1+V_0}{V_0}\left(k - \frac{1}{2}\right)^2 \pi^2. \tag{B.13}$$

As consequences, one then has that

(a) constraint (B.12) being satisfied automatically implies that (B.10) is satisfied,
(b) the number of eigenvalues increases as E_0 increases,
(c) for a given value of E_0, an application of (B.11) and (B.13) yields an upper bound on the number of possible eigenvalues.

Finally, direct examination of (B.8) yields that

$$\lim_{E_0 \to +\infty} \frac{\lambda}{E_0} = \frac{1}{1+V_0};$$

hence, the eigenvalues accumulate at the branch point of $E_{\mathrm{e,o}}(\lambda)$ as $E_0 \to +\infty$.

Appendix C. Exact roots of a polynomial of degree 6. Consider the sixth-order polynomial given by

$$p(x) := x^6 + ax^5 + bx^4 + cx^3 + bx^2 + ax + 1. \tag{C.1}$$

Upon noting that $p(x) = 0$ if and only if $p(x^{-1}) = 0$, one can rewrite (C.1) as

$$p(x) = \prod_{j=1}^{3} (x^2 - \beta_j x + 1), \tag{C.2}$$

where

$$\begin{aligned} \beta_1 + \beta_2 + \beta_3 &= -a, \\ \beta_1\beta_2 + \beta_2\beta_3 + \beta_3\beta_1 &= b - 3, \\ \beta_1\beta_2\beta_3 &= 2a - c. \end{aligned} \tag{C.3}$$

Note that once a solution to (C.3) has been found, then the corresponding solution to (C.1) is given by

$$x_\beta^{\pm} := \frac{1}{2}(\beta \pm \sqrt{\beta^2 - 4}). \tag{C.4}$$

In particular, the result of (C.4) implies that in order to have a real-valued solution to (C.1), one must require that a real-valued solution to (C.3) satisfy $|\beta| \geq 2$.

The system (C.3) can be solved in the following manner. Solve the first equation for β_3, i.e.,

$$\beta_3 = -(a + \beta_1 + \beta_2). \tag{C.5}$$

Substitution into the second equation yields the reduced system

$$\begin{aligned} \beta_1\beta_2 - \beta_3(\beta_3 + a) &= b - 3, \\ \beta_1\beta_2\beta_3 &= 2a - c, \end{aligned}$$

i.e.,

$$\beta_1\beta_2 = \frac{2a - c}{\beta_3}, \tag{C.6}$$

with β_3 being a root of

$$z^3 + az^2 + (b-3)z + c - 2a = 0. \tag{C.7}$$

If $\beta_3 = z \in \mathbb{R}$ is a solution to (C.6), then upon solving the system created by combining (C.5) and (C.6) and defining

$$\beta^{\pm} = -\frac{1}{2}\left(\beta_3 + a \pm \sqrt{(\beta_3 + a)^2 + 4\frac{c-2a}{\beta_3}}\right), \tag{C.8}$$

one has the solution pair

$$(\beta_1, \beta_2) = (\beta^+, \beta^-). \tag{C.9}$$

Proposition C.1. *Let $\beta_3 \in \mathbb{R}$ be a solution to (C.7). The corresponding solutions to (C.3) are given in (C.9). If $\beta_j \in \mathbb{R}$ for some $j \in \{1,2,3\}$ is such that $|\beta_j| \geq 2$, then the corresponding pair of real-valued solutions to (C.1) is given by $x^{\pm}_{\beta_j}$, where $x^{\pm}_{\beta}$ is defined in (C.4).*

Acknowledgments. P. G. K. gratefully acknowledges numerous discussions with D. J. Frantzeskakis in the initial stages of this work. We thank C. Lou and J. Xu for the experimental work at Nankai University; ZC is also with Nankai University. Finally, the authors are grateful to B. Sandstede for his invaluable help in deriving the calculation presented in Appendix A.

REFERENCES

[1] M. Albiez, R. Gati, J. Fölling, S. Hunsmann, T. Cristiani, and M. K. Oberthaler, *Direct observation of tunneling and nonlinear self-trapping in a single bosonic Josephson junction*, Phys. Rev. Lett., 95 (2005), paper 010402.

[2] G. L. Alfimov, V. A. Brazhnyi, and V. V. Konotop, *On classification of intrinsic localized modes for the discrete nonlinear Schrödinger equation*, Phys. D, 194 (2004), pp. 127–150.

[3] Th. Anker, M. Albiez, R. Gati, S. Hunsmann, B. Eiermann, A. Trombettoni, and M. K. Oberthaler, *Nonlinear self-trapping of matter waves in periodic potentials*, Phys. Rev. Lett., 94 (2005), paper 020403.

[4] V. A. Brazhnyi and V. V. Konotop, *Theory of nonlinear matter waves in optical lattices*, Mod. Phys. Lett. B, 18 (2004), pp. 627–651.

[5] R. Brout, *A Brief Course in Spontaneous Symmetry Breaking* I: *The Palelolitic Age*, hep-th/0203096, Los Alamos National Laboratory archives, Los Alamos, NM, 2002.

[6] C. Cambournac, T. Sylvestre, H. Maillotte, B. Vanderlinden, P. Kockaert, Ph. Emplit, and M. Haelterman, *Symmetry-breaking instability of multimode vector solitons*, Phys. Rev. Lett., 89 (2002), paper 083901.

[7] Z. Chen and K. McCarthy, *Spatial soliton pixels from partially coherent light*, Opt. Lett., 27 (2002), pp. 2019–2021.

[8] Z. Chen, A. Bezryadina, I. Makasyuk, and J. Yang, *Observation of two-dimensional lattice vector solitons*, Opt. Lett., 29 (2004), pp. 1656–1658.

[9] Z. Chen, H. Martin, E. D. Eugenieva, J. Xu, and A. Bezryadina, *Anisotropic enhancement of discrete diffraction and formation of two-dimensional discrete-soliton trains*, Phys. Rev. Lett., 92 (2004), paper 143902.

[10] D. N. Christodoulides, F. Lederer, and Y. Silberberg, *Discretizing light behavior in linear and nonlinear waveguide lattices*, Nature, 424 (2003), pp. 817–823.

[11] N. K. Efremidis, S. Sears, D. N. Christodoulides, J. W. Fleischer, and M. Segev, *Discrete solitons in photorefractive optically induced photonic lattices*, Phys. Rev. E, 66 (2002), paper 046602.

[12] J. C. Eilbeck, P. S. Lomdahl, and A. C. Scott, *The discrete self-trapping equation*, Phys. D, 16 (1985), pp. 318–338.
[13] F. Englert, *A Brief Course in Spontaneous Symmetry Breaking* II: *Modern Times: The BEH Mechanism*, hep-th/0203097, Los Alamos National Laboratory archives, Los Alamos, NM, 2002.
[14] J. W. Fleischer, T. Carmon, M. Segev, N. K. Efremidis, and D. N. Christodoulides, *Observation of discrete solitons in optically induced real time waveguide arrays*, Phys. Rev. Lett., 90 (2003), paper 023902.
[15] J. W. Fleischer, M. Segev, N. K. Efremidis, and D. N. Christodoulides, *Observation of two-dimensional discrete solitons in optically induced nonlinear photonic lattices*, Nature, 422 (2003), pp. 147–150.
[16] J. W. Fleischer, G. Bartal, O. Cohen, O. Manela, M. Segev, J. Hudock, and D. N. Christodoulides, *Observation of vortex-ring "discrete" solitons in 2D photonic lattices*, Phys. Rev. Lett., 92 (2004), paper 123904.
[17] R. Franzosi and V. Penna, *Chaotic behavior, collective modes and self-trapping in the dynamics of three coupled Bose-Einstein condensates*, Phys. Rev. E, 67 (2003), paper 046227.
[18] M. Golubitsky and D. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. 1, Springer-Verlag, New York, Berlin, 1985.
[19] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Clarendon Press, Oxford, UK, 1995.
[20] R. K. Jackson and M. I. Weinstein, *Geometric analysis of bifurcation and symmetry breaking in a Gross-Pitaevskii equation*, J. Statist. Phys., 116 (2004), pp. 881–905.
[21] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, eds., *Photonic Crystals: Modeling the Flow of Light*, Princeton University Press, Princeton, NJ, 1995.
[22] M. Johansson, *Hamiltonian Hopf bifurcations in the discrete nonlinear Schrödinger trimer: Oscillatory instabilities, quasi-periodic solutions and a "new" type of self-trapping transition*, J. Phys. A, 37 (2004), pp. 2201–2222.
[23] T. Kapitula and P. G. Kevrekidis, *Bose-Einstein condensates in the presence of a magnetic trap and an optical lattice*, Chaos, 15 (2005), paper 037114.
[24] T. Kapitula and P. G. Kevrekidis, *Bose-Einstein condensates in the presence of a magnetic trap and optical lattice: Two-mode approximation*, Nonlinearity, 18 (2005), pp. 2491–2512.
[25] T. Kapitula, P. G. Kevrekidis, and B. A. Malomed, *Stability of multiple pulses in discrete systems*, Phys. Rev. E, 63 (2001), paper 036602.
[26] T. Kapitula, P. Kevrekidis, and B. Sandstede, *Counting eigenvalues via the Krein signature in infinite-dimensional Hamiltonian systems*, Phys. D, 195 (2004), pp. 263–282.
[27] T. Kapitula, P. Kevrekidis, and B. Sandstede, *Addendum: Counting eigenvalues via the Krein signature in infinite-dimensional Hamiltonian systems*, Phys. D, 201 (2005), pp. 199–201.
[28] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1980.
[29] P. G. Kevrekidis and D. J. Frantzeskakis, *Pattern forming dynamical instabilities of Bose-Einstein condensates*, Mod. Phys. Lett. B, 18 (2004), pp. 173–202.
[30] P. G. Kevrekidis, A. R. Bishop, and K. Ø. Rasmussen, *Twisted localized modes*, Phys. Rev. E, 18 (2001), paper 036603.
[31] P. G. Kevrekidis, Z. Chen, B. A. Malomed, D. J. Frantzeskakis, and M. I. Weinstein, *Symmetry-breaking instability of multimode vector solitons*, Phys. Lett. A, 340 (2005), pp. 275–280.
[32] Yu. S. Kivshar and G. P. Agrawal, eds., *Optical Solitons: From Fibers to Photonic Crystals*, Academic Press, New York, 2003.
[33] Y. Li and K. Promislow, *Structural stability of non-ground state traveling waves of coupled nonlinear Schrödinger equations*, Phys. D, 124 (1998), pp. 137–165.
[34] K. W. Mahmud, J. N. Kutz, and W. P. Reinhardt, *Bose-Einstein condensates in a one-dimensional double square well: Analytical solutions of the nonlinear Schrödinger equation*, Phys. Rev. A, 66 (2002), paper 063607.
[35] H. Martin, E. D. Eugenieva, Z. Chen, and D. N. Christodoulides, *Discrete solitons and soliton-induced dislocations in partially coherent photonic lattices*, Phys. Rev. Lett., 92 (2004), paper 123902.
[36] A. Morgante, M. Johansson, G. Kopidakis, and S. Aubry, *Standing wave instabilities in a chain of nonlinear coupled oscillators*, Phys. D, 162 (2002), pp. 53–94.
[37] O. Morsch and M. Oberthaler, *Bose-Einstein condensates in optical lattices*, Rev. Mod. Phys., 78

(2006), pp. 179–215.
[38] D. Neshev, E. Ostrovskaya, Yu. Kivshar, and W. Krolikowski, *Spatial solitons in optically induced gratings*, Opt. Lett., 28 (2003), pp. 710–712.
[39] D. N. Neshev, T. J. Alexander, E. A. Ostrovskaya, Yu. S. Kivshar, H. Martin, I. Makasyuk, and Z. Chen, *Observation of discrete vortex solitons in optically induced photonic lattices*, Phys. Rev. Lett., 92 (2004), paper 123903.
[40] D. Pelinovsky, *Inertia law for spectral stability of solitary waves in coupled nonlinear Schrödinger equations*, Proc. Royal Soc. London A, 461 (2005), pp. 783–812.
[41] D. Pelinovsky and J. Yang, *Instabilities of multihump vector solitons in coupled nonlinear Schrödinger equations*, Stud. Appl. Math., 115 (2005), pp. 109–137.
[42] D. Pelinovsky, P. Kevrekidis, and D. Frantzeskakis, *Stability of discrete solitons in nonlinear Schrödinger lattices*, Phys. D, 212 (2005), pp. 1–19.
[43] D. Pelinovsky, P. Kevrekidis, and D. Frantzeskakis, *Persistence and stability of discrete vortices in nonlinear Schrödinger lattices*, Phys. D, 212 (2005), pp. 20–53.
[44] A. Sacchetti, *Nonlinear time-dependent Schrödinger equations: The Gross-Pitaevskii equation with double-well potential*, J. Evol. Equ., 4 (2004), pp. 345–369.
[45] A. Sacchetti, *Nonlinear time-dependent one-dimensional Schrödinger equation with double-well potential*, SIAM J. Math. Anal., 35 (2004), pp. 1160–1176.
[46] B. Sandstede, *Stability of multiple-pulse solutions*, Trans. Amer. Math. Soc., 350 (1998), pp. 429–472.
[47] C. Sulem and P.-L. Sulem, *The Nonlinear Schrödinger Equation*, Springer-Verlag, New York, Berlin, 1999.
[48] M. Weinstein, *Excitation thresholds for nonlinear localized modes on lattices*, Nonlinearity, 12 (1999), pp. 673–691.
[49] J. Yang, *Stability of vortex solitons in a photorefractive optical lattice*, New J. Phys., 6 (2004), p. 47.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 634–649

Geometric Relations of Absolute and Essential Spectra of Wave Trains*

Jens D. M. Rademacher†

Abstract. We analyze geometric relations of absolute and essential spectra for certain linear operators on the real line with periodic coefficients. These spectra correspond to accumulation sets of eigenvalues for increasing domain length under separated and periodic boundary conditions, respectively. The main result shows that critical isolated sets of essential spectra contain absolute spectra and yields an algorithm for its numerical computation. Linearizations of reaction diffusion systems in wave trains are used as an illustration, and we present a detailed numerical study of absolute and essential spectra for a wave train in the Schnakenberg model.

Key words. absolute spectrum, essential spectrum, wave trains, isolated spectral sets, reaction diffusion systems

AMS subject classifications. 37L15, 34L16, 76E15, 47A10

DOI. 10.1137/050635390

1. Introduction.

We consider spectral properties of certain linear operators on the real line with periodic coefficients. Posed on an interval $(-L, L)$ with periodic boundary conditions, eigenvalues of these operators accumulate, as $L \to \infty$, on the *essential spectrum* of the operator posed on the real line. However, for generic separated boundary conditions eigenvalues accumulate on the so-called *absolute spectrum* introduced in [13]. In the presence of convection, absolute and essential spectra typically differ, and the absolute spectrum may be contained in the open left half plane, while the essential spectrum intersects the open right half plane. The absolute spectrum has been characterized in [13] via a complex dispersion relation of temporal and spatial modes. It was formulated for asymptotically constant coefficients, but all main results remain valid for the periodic case; cf. [13, p. 239].

In this article we analyze the relation of absolute and essential spectra for the case of periodic coefficients. Our main result, Theorem 4.5, shows that critical isolated sets of essential spectra contain absolute spectra. The proof provides an algorithm for the detection of this absolute spectrum, which we use for numerical computations in a specific example. A weaker version of this theorem is contained in [10, Theorem 3.3].

The main application for the abstract framework are linearizations of parabolic partial differential equations (PDEs), e.g., reaction diffusion systems, in traveling waves that are asymptotically periodic or that have an extended periodic region and are asymptotically constant in space. We refer to spatially periodic traveling waves as *wave trains*. In fact, absolute and essential spectra are determined by the spatially asymptotic states and we can restrict our attention to wave trains for their study. We summarize the relevance of the absolute spectrum

*Received by the editors July 7, 2005; accepted for publication (in revised form) by M. Silber July 26, 2006; published electronically November 22, 2006. This work was supported in part by NSF grant DMS-0203301, PIMS, and SPP 1095 of the German Research Foundation.
http://www.siam.org/journals/siads/5-4/63539.html
†Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany (rademach@wias-berlin.de).

of traveling waves on large bounded domains as follows and refer the reader to [12, 13] for details.

1. *Accumulation of eigenvalues*. On a bounded domain of length L only point spectrum occurs, but most of it accumulates as $L \to \infty$: For periodic boundary conditions at the essential spectrum, but for generic separated boundary conditions at the absolute spectrum.
2. *Absolute, remnant versus convective instabilities*. If the absolute spectrum is unstable, then perturbations grow pointwise or are convected in both directions. Pointwise growth occurs when a branch point of the dispersion relation (Definition 3.1) in the absolute spectrum is unstable.
3. *Absolute instabilities are inherited*. If the profile of a traveling wave on the real line with constant asymptotics is close to another uniform steady state or wave train for a segment of length L, then, as $L \to \infty$, point spectrum accumulates at certain parts of the absolute spectrum of that state. An absolutely unstable state implies instability for all sufficiently large L.
4. *Linear spreading speeds*. The largest and smallest speeds of comoving frames in which the absolute spectrum is marginally stable yield linear predictions for spreading speeds of instabilities.

More formally, we consider linear operators on the real line as in [13], but explicitly assume periodicity of the coefficient matrix. These are a family $\mathcal{T}(\lambda)$, $\lambda \in \mathbb{C}$, where

$$\mathcal{T}(\lambda) : H^1(\mathbb{R}, \mathbb{C}^n) \to L^2(\mathbb{R}, \mathbb{C}^n), \quad u \mapsto \frac{\mathrm{d}u}{\mathrm{d}\xi} - A(\cdot; \lambda)u$$

for the usual Sobolev space H^1 of L^2 functions with weak derivative in L^2. Here $A(\xi; \lambda) \in \mathbb{R}^{n \times n}$ is a matrix-valued function that satisfies the following.

Hypothesis 1. *The matrices $A(\xi; \lambda)$ are smooth in $\xi \in \mathbb{R}$ and analytic in $\lambda \in \mathbb{C}$.*

1. Periodic coefficients. *There is $L > 0$ such that $A(\xi + L; \lambda) = A(\xi; \lambda)$ for all $\xi \in \mathbb{R}$ and $\lambda \in \mathbb{C}$.*
2. Well-posedness. *There exists $\rho \in \mathbb{R}$ such that for all $\lambda \in \mathbb{C}$ with real part $\Re(\lambda) \geq \rho$ the period map of the evolution of $\mathcal{T}(\lambda)v = 0$ has no Floquet exponent in* $\mathrm{i}\mathbb{R}$.

We use Hypothesis 1 as the abstract basis for studying absolute spectra and illustrate the results using reaction diffusion systems. In Lemma 3.3 we will show that Hypothesis 1 holds for the matrices arising in eigenvalue problems of wave trains in reaction diffusion systems in one space dimension.

Reaction diffusion systems consist of N "species" $U = (U_1, \dots U_N) \in \mathbb{R}^N$ that are spatially coupled by diffusion, $D := \mathrm{diag}(d_1, \dots, d_N)$ with $d_j > 0$, and driven by pointwise reaction kinetics $F : \mathbb{R}^N \to \mathbb{R}^N$ in the form

$$U_t = DU_{xx} + F(U). \tag{1.1}$$

We assume this equation is posed on a function space X so that $D\partial_{xx}$ can be cast as a closed and densely defined operator, and the Nemitskij operator F_{N} derived from F satisfies $F_{\mathrm{N}} \in C^1(X, X)$. For example $X = BC^0_{\mathrm{unif}}(\mathbb{R}, \mathbb{R}^N)$, domain of definition $\mathrm{dom}(D\partial_{xx}) = BC^2_{\mathrm{unif}}(\mathbb{R}, \mathbb{R}^N)$, and $F \in C^1(\mathbb{R}^N, \mathbb{R}^N)$; cf. Chapter 2 in [17].

In a comoving frame with the variable $\xi = x - ct$ and speed c system (1.1) becomes

$$U_t = DU_{\xi\xi} + cU_\xi + F(U), \tag{1.2}$$

and we call t-independent solutions *traveling waves*. These solve the spatial ordinary differential equation (ODE)

$$DU_{\xi\xi} + cU_\xi + F(U) = 0 \tag{1.3}$$

and are called *wave trains* if $U(\xi + L) = U(\xi)$ for all ξ and some $L > 0$.

In case $d_j = 0$ for one or more j, spectra of linearizations of (1.3) about traveling waves with $c \neq 0$ are typically continuous as $d_j \to 0$ [10, Theorem 3.5].

This article is organized as follows. In section 2 we define and mention the relevance of spectra for stability. Absolute and essential spectra are characterized in section 3. In section 4, we investigate the relative location of absolute and essential spectra and prove the main result. We present numerical computations of (generalized) absolute and essential spectra for a specific example in section 5.

2. Spectra and stability of wave trains. Following [13], we define the so-called essential spectrum of $\mathcal{T}$.

Definition 2.1. *We say that $\lambda \in \mathbb{C}$ lies in the essential spectrum Σ_{ess} of $\mathcal{T}$ if $\mathcal{T}(\lambda)$ is not boundedly invertible. The essential spectrum of $\mathcal{T}$ is called (strictly) stable if it lies in the (open) left half plane $\{\Re(\lambda) \leq 0\}$ and unstable if it intersects the open right half plane.*

As an illustration, consider a solution U^* to (1.3) as a steady state of (1.2). Its spectral stability is determined by the spectrum of the linearization of (1.2) about U^*, which is the linear operator

$$\mathcal{L} := D\partial_{\xi\xi} + c\partial_\xi + \partial_U F(U^*). \tag{2.1}$$

Generally, λ lies in the spectrum of $\mathcal{L}$ if the eigenvalue problem

$$\lambda V = \mathcal{L}V \tag{2.2}$$

has a bounded solution V. More precisely, we define the spectrum of U^* as follows.

Definition 2.2. *The spectrum of $\mathcal{L}$, $\Sigma(\mathcal{L})$, is the set of $\lambda \in \mathbb{C}$ for which the operator $\mathcal{L} - \lambda$ is not boundedly invertible in X. The point spectrum of U^*, $\Sigma_{\text{pt}}(\mathcal{L})$, is the set of all $\lambda \in \Sigma(\mathcal{L})$ for which $\mathcal{L} - \lambda$ is a Fredholm operator with index zero. The essential spectrum of $\mathcal{L}$ is $\Sigma_{\text{ess}}(\mathcal{L}) := \Sigma(\mathcal{L}) \setminus \Sigma_{\text{pt}}(\mathcal{L})$. We call these spectra (strictly) stable, respectively, if they lie in the (open) left half plane, and unstable if parts lie in the open right half plane.*

Definitions 2.1 and 2.2 are consistent in the sense that for a wave train U^* the spectrum of $\mathcal{L}$ equals that of $\mathcal{T}$ for the period L of U^*; see, e.g., [6, 14]. In this case, the connection of $\mathcal{L}$ and $\mathcal{T}$ is as follows. We cast (2.2) as a linear nonautonomous first order ODE in $\mathbb{R}^{2N}$

$$\dot{v} = A_{\mathcal{L}}(\xi; \lambda)v, \tag{2.3}$$

corresponding to $\mathcal{T}(\lambda)v = 0$ and, according to (1.2), it has the matrix

$$A_{\mathcal{L}}(\xi;\lambda) = \begin{pmatrix} 0 & \text{Id} \\ D^{-1}\partial_U F(U^*(\xi)) & D^{-1}c \end{pmatrix} - \lambda \begin{pmatrix} 0 & 0 \\ D^{-1} & 0 \end{pmatrix}.$$

For any traveling wave U^* of (1.1), the derivative $\partial_\xi U^*$ solves the eigenvalue problem (2.2), and zero lies in the spectrum of U^* ("Goldstone mode"). Therefore, we cannot expect asymptotic stability on unbounded or periodic domains. Nevertheless, two types of weakened nonlinear stability can be concluded from spectral properties: nonlinear stability with asymptotic phase and diffusive stability. The former follows from a simple zero eigenvalue and a spectral gap to the strictly stable rest of the spectrum: a perturbed wave converges exponentially to a selected translate of the original wave; see, e.g., [8, Chapter 5.1]. However, the required spectral gap does not occur for wave trains on the real line, where the essential spectrum comes in curves. Instead, wave trains can be stable in the sense that perturbations "diffuse" self-similarly over the wave train [16].

On the other hand, an unstable spectrum causes nonlinear instability for any traveling wave [8].

3. Essential and absolute spectra. Let $\Phi_\lambda(\xi,\zeta)$ denote the evolution of (2.3). It has a Floquet representation with L-periodic matrix $S_\lambda(\xi)$, $S_\lambda(0) = \mathrm{Id}$, of the form

$$\Phi_\lambda(\xi,0) = S_\lambda(\xi)\mathrm{e}^{R(\lambda)\xi}.$$

Definition 3.1.

- *The complex dispersion relation* $d : \mathbb{C}^2 \to \mathbb{C}$ *of* $\mathcal{T}$ *is defined by*

$$d(\lambda,\nu) = \det(R(\lambda) - \nu) = 0,$$

- *a branch point* $\lambda \in \mathbb{C}$ *at* $\nu \in \mathbb{C}$ *is such that* $d(\lambda,\nu) = \partial_\nu d(\lambda,\nu) = 0$,
- *the spatial Morse index* $i(\lambda)$ *counts the center unstable dimensions of* $R(\lambda)$,
- *eigenvalues* $\nu(\lambda)$ *of* $R(\lambda)$ *are called spatial Floquet exponents.*

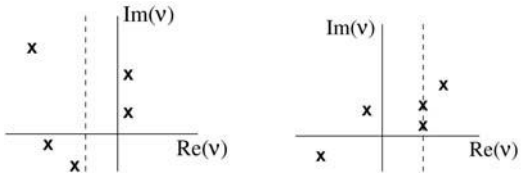
Under Hypothesis 1, the matrix $R(\lambda)$ is analytic in λ; cf., e.g., [3]. Note that spatial Floquet exponents are unique modulo $2\pi\mathrm{i}$.

It is well known that the essential spectrum of $\mathcal{T}$ consists of those λ for which a spatial Floquet exponent is purely imaginary [6], that is,

$$\lambda \in \Sigma_{\mathrm{ess}} \Leftrightarrow \exists k \in \mathbb{R} \,:\, d(\lambda, \mathrm{i}k) = 0. \tag{3.1}$$

Hence the spatial Morse index $i(\lambda)$ is constant in connected components of $\mathbb{C} \setminus \Sigma_{\mathrm{ess}}$ and whenever $\partial_\lambda d(\lambda,\nu)\partial_\nu d(\lambda,\nu) \neq 0$ in the essential spectrum, the implicit function theorem locally yields a unique smooth curve of Σ_{ess}. We expect either that singularities occur on a discrete set or that $d(\lambda,\nu)$ has a multiple factor; note that $d(\lambda,\nu)$ contains the term $R(\lambda)^n$ so that the Weierstrass preparation theorem applies (cf. [9]), which reduces $d(\lambda,\nu) = 0$ locally to roots of a polynomial with analytic coefficients in ν.

Concerning the global topology of Σ_{ess}, for constant matrix $A(\xi;\lambda) = A(\lambda)$ the essential spectrum is a connected set of curves in the closed complex plane; cf., e.g., [10]. However, for the present case with periodic matrix $A(\xi;\lambda)$, it can consist of several connected components. For instance, the essential spectrum of wave trains in reaction diffusion systems that are sufficiently close to a pulse (constant spatial asymptotics) decomposes in a certain way [7, 14]; see also Corollary 4.4 and section 5 below.



(a) (b)

Figure 1. *We plot sample configurations of* 5 *Floquet exponents and* $i_\infty = 2$ *for some* λ*:* (a) λ *not in the absolute spectrum,* (b) λ *in the absolute spectrum.*

Hypothesis 1 implies that $i(\lambda)$ is constant for all $\Re(\lambda) \geq \rho$. The absolute spectrum, introduced first in [13], is defined in terms of this constant as follows.

Definition 3.2. *The generalized absolute spectrum of* $\mathcal{T}$ *with Morse index* j *is*

$$\Sigma^j_{\text{abs}} := \{\lambda \in \mathbb{C} \mid \Re(\nu_j) = \Re(\nu_{j+1})\},$$

where $\Re(\nu_1) \geq \Re(\nu_2) \geq \cdots \geq \Re(\nu_n)$ *is counted with multiplicity. The generalized absolute spectrum,* Σ^*_{abs}*, is the union of* Σ^j_{abs} *for* $j = 1, \dots n-1$.

Finally, the absolute spectrum for $\mathcal{T}$ *satisfying Hypothesis* 1 *is defined as*

$$\Sigma_{\text{abs}} := \Sigma^{i_\infty}_{\text{abs}},$$

where i_∞ *is the constant number of Floquet exponents with positive real part given by the well-posedness in Hypothesis* 1 *for* $\Re(\lambda) \geq \rho$.

To illustrate this definition we plot sample configurations of spatial Floquet exponents ν in Figure 1. By the implicit function theorem Σ^*_{abs} consists of smooth curves away from singularities. The additional requirement of a certain Morse index for Σ^j_{abs} typically causes corners in Σ^j_{abs} at points where smooth curves in Σ^*_{abs} cross transversely. Therefore, computing Σ^*_{abs} is a natural first step for the computation of the absolute spectrum; cf. [11]. Similar to the essential spectrum, the absolute spectrum is connected for constant coefficients (cf. [11]) but is, e.g., not connected for wave trains close to a pulse; see Corollary 4.4.

We next show that the abstract framework indeed applies to wave trains in reaction diffusion systems with $i_\infty = N$. In section 2, we identified the spectrum of the linearization of a reaction diffusion system about a wave train as the essential spectrum of $\mathcal{T}$ for a certain matrix $A = A_{\mathcal{L}}$. The arising operator $\mathcal{T}$ satisfies Hypothesis 1 as follows.

Lemma 3.3. *The family of operators* $\mathcal{T}_{\mathcal{L}}(\lambda)$ *derived from* (1.1) *for a wave train* U^* *as in* (2.3) *satisfies Hypothesis* 1*. More precisely, there is* $R \in \mathbb{R}$ *such that for* $\Re(\lambda) > R$ *the spatial Morse index is* $i(\lambda) = N$*, and for all spatial Floquet exponents* ν *corresponding to* $\mathcal{T}_{\mathcal{L}}(\lambda) = 0$ *it holds that* $|\Re(\nu)| \to \infty$ *as* $\Re(\lambda) \to \infty$.

Proof. Consider $D\ddot{V} + c\dot{V} + \partial_U F(U_*(\xi))V - \lambda V = 0$. Rescaling to fast "spatial time" $\xi = \epsilon\zeta$, $' = \frac{\mathrm{d}}{\mathrm{d}\zeta}$, and $\lambda = \tilde{\lambda}/\epsilon^2$, we obtain

$$\begin{aligned} DV''/\epsilon^2 + cV'/\epsilon + \partial_U F(U_*(\epsilon\zeta))V - \tilde{\lambda}/\epsilon^2 V = 0, \\ (\epsilon \neq 0) \Leftrightarrow DV'' + \epsilon cV' + \epsilon^2 \partial_U F(U_*(\epsilon\zeta))V - \tilde{\lambda}V = 0. \end{aligned}$$

For $\epsilon = 0$ the latter "fast system" is $DV'' = \tilde{\lambda}V$ and has dispersion relation $\Pi_{j=1}^{N}(d_j\nu^2 - \tilde{\lambda}) = 0$, $d_j > 0$. For $\Re(\tilde{\lambda}) > 0$ its spatial Morse index is N and there are no spatial Floquet exponents on the imaginary axis. Therefore, for all $\Re(\tilde{\lambda}) > 0$, the fast system has exponential dichotomies on $\mathbb{R}^{\pm}$ with Morse index N and exponential rate $\min\{\Re(\sqrt{\tilde{\lambda}/d_j})\}$. By roughness of exponential dichotomies (e.g., [4]), the spatial Morse index and the exponential rate (with $o(1)_{\epsilon\to 0}$ adjustment) persist for small bounded perturbation, i.e., for $\epsilon > 0$ sufficiently small.

In particular, the exponential rates of the dichotomies in the normal "time" ξ for $0 < \epsilon \ll 1$ are bounded from below by

$$\frac{1}{2}\min\left\{\Re\left(\epsilon^{-1}\sqrt{\frac{\tilde{\lambda}}{d_j}}\right)\right\} = \epsilon^{-1}\frac{1}{2}\min\left\{\Re\left(\sqrt{\frac{\tilde{\lambda}}{d_j}}\right)\right\}.$$

Since the precise exponential rate is given by the spatial Floquet exponent ν that is closest to the imaginary axis, the lower bound implies that all spatial Floquet exponents have unbounded real part as $\Re(\lambda) \to \infty$, i.e., $\epsilon \to 0$ for fixed $\tilde{\lambda}$.

Periodicity and smoothness in ξ are evident, which completes Hypothesis 1. ∎

4. Relative location of absolute and essential spectra. A straightforward constraint for the location of the absolute spectrum is that it must lie on or to the left of the essential spectrum in the complex plane. To be definite we formulate the following lemma.

Lemma 4.1. *Under Hypothesis 1, the connected component $\Omega_\infty \subset \mathbb{C} \setminus \Sigma_{\text{ess}}$ that contains an unbounded interval of $\mathbb{R}_+$ is well defined. Any curve that connects Σ_{abs} and Ω_∞ intersects Σ_{ess}.*

Proof. By (3.1) and Hypothesis 1 we have $\{\lambda \mid \Re(\lambda) \geq \rho\} \cap \Sigma_{\text{ess}} = \emptyset$, and hence $\Omega_\infty \supset \{\lambda \mid \Re(\lambda) \geq \rho\}$. Therefore, Ω_∞ is well defined and $i(\lambda) = i_\infty$ for any $\lambda \in \Omega_\infty$. This is not possible for $\lambda \in \Sigma_{\text{abs}}$, because there $i(\lambda) \geq i_\infty + 1$ or $i(\lambda) \leq i_\infty - 1$. Since $\partial\Omega_\infty \subset \Sigma_{\text{ess}}$, the intersection claim follows. ∎

If the spatial Morse index changes at some $\lambda_0 \in \Sigma_{\text{ess}}$ while moving *within* the essential spectrum, then $\lambda_0 \in \Sigma_{\text{abs}}^*$, because two spatial Floquet exponents have the same real part, namely, zero. Typically, λ_0 is an intersection of two curves in the essential spectrum, which gives some more information about the location of absolute spectrum; see also Figure 2 and section 5.

Lemma 4.2. *Suppose that two curves in Σ_{ess} intersect at λ_0, but $\partial_\lambda d(\lambda_0, \nu) \neq 0$ for all spatial Floquet exponents $\nu \in \mathrm{i}\mathbb{R}$. Then λ_0 lies in the generalized absolute spectra with Morse indices $j_- + 1, \dots, j_+ - 1$, where $j_- = \liminf_{\epsilon\to 0}\{i(\lambda) \mid \epsilon = |\lambda - \lambda_0|\}$ and $j_+ = \limsup_{\epsilon\to 0}\{i(\lambda) \mid \epsilon = |\lambda - \lambda_0|\}$.*

Assume further that these curves intersect transversely and are $\lambda_\iota(\mathrm{i}k)$ with $\lambda_0 = \lambda_\iota(\mathrm{i}k_\iota)$, $\frac{\mathrm{d}}{\mathrm{d}k}|_{k=k_\iota}\lambda_\iota(\mathrm{i}k) \neq 0$ for $\iota = 1, 2$, and that $\Re(\nu) \neq 0$ for any other solution to $d(\lambda_0, \nu) = 0$. Then there exists a locally unique curve $C \subset \Sigma_{\text{abs}}^j$, $j = j_+ - 1 = j_- + 1$, crossing at λ_0 from the region where $i(\lambda) = j_-$ into the region where $i(\lambda) = j_+$.

Proof. Such an intersection point lies in the generalized absolute spectrum, because either two different purely imaginary spatial Floquet exponents have the same real part, or it is a branch point. Since the Morse index counts the number of unstable and purely imaginary

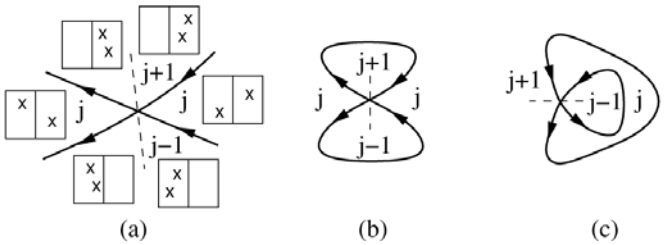


Figure 2. *We sketch intersection points in* Σ^j_{abs} *of two curves of* Σ_{ess} *(solid lines) oriented by increasing* k*, and show the Morse indices in the neighboring regions. Dashed lines indicate the curve of* Σ^j_{abs} *that typically crosses the intersection point and insets in* (a) *the distribution of* ν_j *and* ν_{j+1} *in* $\mathbb{C}$ *(vertical line is* $\mathrm{i}\mathbb{R}$*). Plots* (b) *and* (c) *illustrate the case of isola.*

Floquet exponents, it follows from the ordering of spatial Floquet exponents at λ_0 by decreasing real parts that the largest and smallest indices ℓ so that $\Re(\nu_\ell) = \Re(\nu_{\ell+1})$ are $j_+ - 1$ and $j_- + 1$.

The regularity assumptions imply that precisely two curves of essential spectrum intersect transversely so that $\mathbb{C} \setminus \Sigma_{\mathrm{ess}}$ is divided into four sectors near λ_0 and $i(\lambda) = j$ in two opposing sectors, and $i(\lambda) = j \pm 1$ in the other two, respectively. By the implicit function theorem there exists a locally unique curve $C \subset \Sigma^j_{\mathrm{abs}}$, which crosses λ_0 transversely. Finally, note that C cannot intersect the sector with index j or Σ_{ess} near λ_0. ■

Remark 1. *In a weighted space* X_η *with norm* $\sup_{\xi\in\mathbb{R}} \mathrm{e}^{\eta\xi}|u(\xi)|$ *the essential spectrum is* $\Sigma^\eta_{\mathrm{ess}} = \{\lambda \mid d(\lambda, \mathrm{i}k - \eta) = 0\}$*. For reaction diffusion systems* (1.1)*, the asymptotics of* $\Re(\nu)$ *stated in Lemma* 3.3 *imply that for any weight* η *there is a connected component* $\Omega^\eta_\infty \subset \mathbb{C}$ *with the same constant Morse index* i_∞ *and* $\Omega^0_\infty = \Omega_\infty$*. Therefore, constraints on the relative location of essential and absolute spectra hold true for the absolute spectrum and* $\Sigma^\eta_{\mathrm{ess}}$ *as well. We expect that a singularity in* Σ_{ess} *can be removed by changing* η*; hence regularity assumptions, as in Lemma* 4.2*, should not be very restrictive. In fact, it can be more efficient to numerically continue the essential spectrum in an exponential weight and thereby infer the location of the absolute spectrum than to compute the absolute spectrum itself, e.g.,* [11].

Our main result implies absolute spectrum within *isola* in Σ_{ess}, i.e., closed bounded curves. As a motivation, consider the aforementioned wave trains near a pulse on a large bounded domain. Unstable essential spectrum might "only" cause a convective instability that is not seen under separated boundary conditions. However, an instability of the point spectrum of a pulse should heuristically destabilize nearby wave trains. Indeed, point spectrum in Ω_∞ generates isola of essential spectrum for nearby wave trains [7, 14] that contain absolute spectrum on account of Corollary 4.4 below. Therefore, instabilities of point spectrum of pulses are inherited to nearby wave trains, at least on sufficiently large domains.

A Jordan curve in the complex plane is a closed, bounded curve $\gamma \subset \mathbb{C}$ without self-intersections. It is well known that such a curve divides the complex plane into an "exterior" set, which is an unbounded connected component of $\mathbb{C}$, and its complementary "interior" set, $\mathrm{int}(\gamma)$, which is a bounded connected component of $\mathbb{C}$. More generally, for an isola $\gamma \subset \mathbb{C}$,

i.e., γ is a closed bounded curve, we define

$$M(\gamma) := \bigcap\{\mathrm{int}(\gamma') \mid \gamma' \text{ is a Jordan curve and } \gamma \subset \mathrm{int}(\gamma')\},$$
$$\mathrm{int}(\gamma) := M(\gamma) \setminus \partial M(\gamma).$$

Note that $\mathrm{int}(\gamma)$ may consist of several connected components and may also be empty, e.g., if γ is an interval.

We say $\lambda_0 \in \Sigma_{\mathrm{ess}}$ is a *regular point* if $\partial_\lambda d(\lambda_0,\nu)\partial_\nu d(\lambda_0,\nu) \neq 0$ for all $\nu \in \mathrm{i}\mathbb{R}$ with $d(\lambda_0,\nu) = 0$ and a *simple regular point* if this is unique.

Proposition 4.3. *Assume Hypothesis 1 and that an isola in Σ_{ess} contains a nonempty connected component $K \subset \mathbb{C} \setminus \Sigma_{\mathrm{ess}}$ with a regular point on ∂K and $i(\lambda)$ is constant for $\lambda \in \partial K$. Then $\bar{K} \cap \Sigma^j_{\mathrm{abs}} \neq \emptyset$, where $j = i(\partial K)$ if $i(K) < i(\partial K)$, and $j = i(\partial K) - 1$ otherwise.*

The proof of this proposition follows below. Note that the spatial Morse indices $i(\partial K)$ and $i(\partial K) - 1$ distinguish an increasing or decreasing Morse index when entering K. For an isola at the origin this is (typically) determined by the group velocity $-\mathrm{d}\lambda/\mathrm{d}\nu|_{\nu=0}$.

Using the Cauchy–Riemann equations, Proposition 4.3 yields Theorem 3.3 in [10]. Together with Theorem 2.1 in [14] it directly implies the following corollary about wave trains near a pulse, which we formulate for reaction diffusion systems (1.1). Here the pulse is required to be a generic homoclinic orbit of (1.3) in the sense that the underlying parameter provides a transverse unfolding and the variational equation has a unique bounded solution (up to constant multiples); see [14].

Corollary 4.4. *Suppose U^*_L is a family of wave trains of (1.1) with the following properties. As solutions to (1.3) we have $U^*_L \to U^*_\infty$, as $L \to \infty$, which is a generic homoclinic orbit to an equilibrium (i.e., it satisfies the assumptions of Theorem 5.1 in [14]). For any λ_* in the point spectrum of U^*_∞ there are positive constants L_*, C, and κ, such that for any $L \geq L_*$ there is an isola $\gamma_L \subset \Sigma_{\mathrm{ess}}(U^*_L)$ with $\lambda_* \in \overline{\mathrm{int}(\gamma_L)}$ and $\mathrm{diam}(\gamma_L) \leq C\mathrm{e}^{-\kappa L}$. Moreover, $\mathrm{int}(\gamma_L) \cap \Sigma^*_{\mathrm{abs}}(U^*_L) \neq \emptyset$ and $\gamma_L \subset \partial\Omega_\infty(U^*_L)$ implies $\mathrm{int}(\gamma_L) \cap \Sigma_{\mathrm{abs}}(U^*_L) \neq \emptyset$.*

Remark 2. *It appears natural that isola in Ω_∞ contain branch points (Definition 3.1) in the absolute spectrum. This would be meaningful, because such points distinguish remnant and absolute instabilities. However, the problem is the Morse index: Having located a branch point by a homotopy, it seems difficult to exclude the possibility of an index change during this homotopy. Indeed absolute spectra without branch points can occur; see [11, section 5.1].*

We combine the most relevant parts of Lemma 4.2 for isola and Proposition 4.3 in the following theorem; recall the definition of $j_\pm$ in Lemma 4.2.

Theorem 4.5. *Assume Hypothesis 1 and that Σ_{ess} contains an isola γ so that an open neighborhood V of its interior satisfies $i(V \setminus \overline{\mathrm{int}(\gamma)}) \equiv j$. Suppose one of the following.*

(i) *$i(\mathrm{int}(\gamma)) \equiv j + 1$ or $i(\mathrm{int}(\gamma)) \equiv j - 1$, and $i(\lambda)$ is constant for $\lambda \in \gamma$.*
(ii) *γ self intersects at a regular point where $j_- + 1 \leq j \leq j_+ - 1$,*
(iii) *$\overline{\mathrm{int}(\gamma)} = \emptyset$ and γ contains a regular point.*

Then $\overline{\mathrm{int}(\gamma)} \cap \Sigma^j_{\mathrm{abs}} \neq \emptyset$, in particular, $\gamma \subset \partial\Omega_\infty$, implies γ contains absolute spectrum.

Note that assumption (ii) is satisfied for an exterior loop that is connected to the rest of the curve as in a figure eight shape; see Figure 2(b) and Lemma 4.2.

Proof. Case (i) follows directly from Proposition 4.3 and case (ii) immediately from Lemma 4.2.

At a regular point $\lambda_0 \in \gamma$ in case (iii), the Cauchy–Riemann equations imply that a Floquet exponent crosses the imaginary axis as λ crosses λ_0. By assumption, the Morse index cannot change when λ has crossed λ_0; hence a second Floquet exponent crosses the imaginary axis in the opposite direction. Therefore, two spatial Floquet exponents are purely imaginary at λ_0, and their indices in the ordering of decreasing real parts are necessarily j and $j+1$, so $\gamma \cap \Sigma^j_{\text{abs}} \neq \emptyset$.

Last, note that $\gamma \subset \partial\Omega_\infty$ yields $j = i_\infty$. ∎

Proof of Proposition 4.3.

Definition 4.6. *We say that a bounded curve $\lambda(s) \in \mathbb{C}$, $s \in [0, s_0]$, $s_0 > 0$, is Σ-connecting if there exists a curve $\nu(s) \in \mathbb{C}$, $s \in [0, s_0]$, such that*

- *$\lambda(0) \in \Sigma_{\text{ess}}$ is a simple regular point,*
- *$d(\lambda(s), \nu(s)) = 0$ for all $s \in [0, s_0]$,*
- *either $\partial_\nu d(\lambda(s_0), \nu(s_0)) = 0$ or $\lambda(s_0) \in \Sigma_{\text{ess}}$, $\nu(s_0) \notin \mathrm{i}\mathbb{R}$,*
- *$i(\lambda(0)) = i(\lambda(s_0))$ in case $\nu(s_0) \in \mathrm{i}\mathbb{R}$,*
- *for $s \in (0, s_0)$ any solution ν to $d(\lambda(s), \nu) = 0$ satisfies $\Re(\nu) \neq 0$.*

We prove Proposition 4.3 in two steps. First, we show that a Σ-connecting curve emanates from any regular point on ∂K. More precisely, we show that $\nu(s) = \mathrm{i}k_0 + s$ or $\nu(s) = \mathrm{i}k_0 - s$ can be used for certain k_0, which corresponds to continuing the essential spectrum in an exponentially weighted space as in Remark 1. Second, we prove that a Σ-connecting curve generally implies the presence of absolute spectrum with a certain Morse index.

To illustrate the approach, consider the scalar constant coefficient equation $u_{xx} = cu_x$. It has the dispersion relation $d(\lambda, \nu) = \nu^2 - c\nu - \lambda = 0$; hence $\Sigma_{\text{ess}} = \{\mathrm{i}ck - k^2 | k \in \mathbb{R}\}$ and $\Sigma_{\text{abs}} = (-\infty, -c^2/4]$. In this case, for fixed $k \in \mathbb{R}$, the roots of $d(\lambda, s + \mathrm{i}k) = 0$ form the Σ-connecting curve $\lambda(s) = s^2 - k^2 - cs + \mathrm{i}k(2s - c)$ with $\nu(s) = s$ and appropriately chosen s_0. This curve indeed intersects the absolute spectrum at $s = \frac{c}{2}$, where $\lambda(\frac{c}{2}) = -c^2/4 - k^2 \in \Sigma_{\text{abs}}$. In fact, $d(\lambda, \mathrm{i}k + \frac{c}{2}) = 0$ gives the entire absolute spectrum.

Lemma 4.7. *Assume the hypotheses of Proposition 4.3. For any simple regular point $\lambda_0 \in \partial K$ there exists a Σ-connecting curve with $\lambda(s) \in \overline{\text{int}(K)}$ for $s \in [0, s_0]$ and $\lambda(0) = \lambda_0$.*

Proof. The implicit function theorem provides $s_0 > 0$ and a locally unique curve $\lambda(s)$ for $|s| \in [0, s_0)$ with $\lambda(0) = \lambda_0$ and $d(\lambda(s), s + \mathrm{i}k_0) = 0$. Note that the Morse index $i(\lambda)$ changes by one at λ_0, but remains constant in K, because $K \cap \Sigma_{\text{ess}} = \emptyset$.

Assume first that $\lambda(s) \in K$ for $s > 0$. In case $\partial_\lambda d(\lambda_0, \mathrm{i}k_0 + s_0) = 0$, we can use Rouché's theorem (cf., e.g., [1]) to continue $\lambda(s)$ to larger values of s. Indeed, if roots could not be chosen continuously for increasing s, then the number of roots of $d(\cdot, \mathrm{i}k_0 + s_0)$ and $d(\cdot, \mathrm{i}k_0 + s_0 + \epsilon)$ would differ in a small neighborhood of λ_0 for any sufficiently small $\epsilon > 0$, which contradicts Rouché's theorem since $d(\cdot, \cdot)$ is smooth.

Using $K \cap \Sigma_{\text{ess}} = \emptyset$ we can choose $s_0 > 0$ minimal, so that either $\partial_\nu d(\lambda(s_0), s_0 + \mathrm{i}k_0) = 0$ or $\lambda(s_0) \in \partial K$. The assumption of constant Morse index on ∂K implies $i(\lambda(0)) = i(\lambda(s_0))$.

In case or $\lambda(s) \in K$ for $s < 0$ the same holds for smaller, maximally chosen s_0. ∎

Lemma 4.8. *Assume Hypothesis 1 and that there exists a Σ-connecting curve. Then there exists $0 < s_* \leq s_0$ such that $\lambda(s_*) \in \Sigma^j_{\text{abs}}$, where $j = i(\lambda(0))$ if $i(\lambda(s_0/2)) < i(\lambda(0))$, and $j = i(\lambda(0)) - 1$ otherwise.*

Proof. By definition of the spatial Morse index and Σ-connecting curves, either $i(\lambda(s))$ is constant for $[0, s_0]$ or $i(\lambda(0)) > i(\lambda(s))$ for $s \in (0, s_0)$. By regularity at λ_0, the definition of j

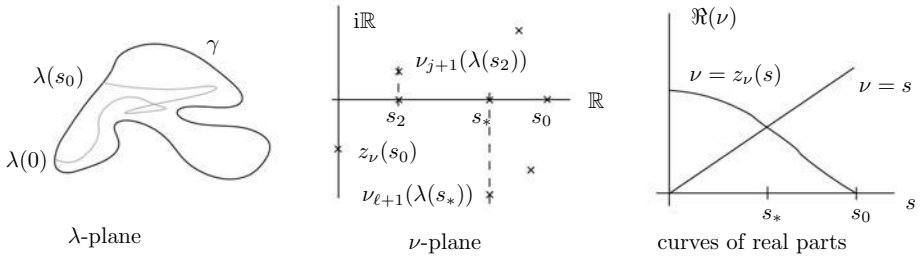


Figure 3. *Configuration of spatial Floquet exponents in the proof of Proposition 4.3.*

covers all cases and we focus on $i(\lambda(0)) = j+1$, i.e., $\Re(\nu(s)) > 0$ for $s \in (0, s_0)$, and comment on the case $i(\lambda(0)) = j$. The idea of the proof with $\nu(s) = \mathrm{i}k_0 + s$ is sketched in Figure 3.

Step 1. We show that whenever the Σ-connecting curve contains generalized absolute spectrum involving $\nu(s_2)$ for some $s_2 \in [0, s_0]$, then it also contains absolute spectrum. We thus assume that $\lambda(s_2) \in \Sigma^{\ell}_{\mathrm{abs}}$ and $\nu_\ell = \nu(s_2)$ in the ordering of decreasing real parts. Since $i(\lambda(s)) \equiv j+1$ for $s \in (0, s_0)$, no Floquet exponent crosses the imaginary axis and $\ell \in \{1, \dots, j+1\}$. The unique spatial Floquet exponents on the imaginary axis at $\lambda(0)$ are $\nu(0)$, and hence $\nu(0) = \nu_{j+1}$. Since Floquet exponents are continuous along $\lambda(s)$, there is $s_* \in [0, s_2]$ such that $\lambda(s_*) \in \Sigma^{j}_{\mathrm{abs}}$; see Figure 3(b) for an illustration.

(*Case* $i(\gamma) = j$. The real parts of two stable Floquet exponents coincide, and $\nu(0) = \nu_j$ in the ordering, so $\ell \in \{j, \dots, n\}$. Again we infer $\ell = j$ at some point.)

Step 2. We establish generalized absolute spectrum in the Σ-connecting curve and apply the first step. In particular, the first step applies if $\lambda(s_0)$ is a branch point at $\nu(s_0)$ so that we may assume this is not the case. Therefore, $\lambda(s_0) \in \Sigma_{\mathrm{ess}}$ and there is a solution $z_\nu(s_0) \in \mathrm{i}\mathbb{R}$ to $d(\lambda(s_0), \cdot) = 0$. By assumption, the Morse index is constant on ∂K and thus, if $\lambda(s_0)$ is a branch point at $z_\nu(s_0)$, then it lies in $\Sigma^{j}_{\mathrm{abs}}$. Otherwise, Rouché's theorem (and typically the implicit function theorem) yields a curve $z_\nu(s) \in \mathbb{C}$ for $s \le s_0$ near s_0 solving $d(\lambda(s), z_\nu(s)) = 0$. As in the proof of Lemma 4.7, we can choose maximal $0 \le s_1 \le s_0$ so that either $\lambda(s_1)$ is a branch point at $z_\nu(s_1)$, or $s_1 = 0$. Since the above first step applies analogously to $\lambda(s) \in \Sigma^{*}_{\mathrm{abs}}$ involving $z_\nu(s)$, and a branch point lies in this set, we next assume $s_1 = 0$.

(*Case* $i(\lambda(0)) = j$. We can choose $s_1 \ge s_0$ and $z_\nu(s_0)$ so that $\Re(z_\nu(s)) < 0$.)

The regularity of $\lambda(0)$ implies that $\mathrm{i}k_0$ is the locally unique solution to $d(\lambda(0), \nu) = 0$ for $\nu \in \mathrm{i}\mathbb{R}$, and $\lambda(0)$ that of $d(\lambda, \mathrm{i}k_0) = 0$ for $\lambda \in \mathbb{C}$. Therefore, $\Re(z_\nu(0)) > 0$ and by construction $\Re(z_\nu(s_0)) = 0$, $\Re(\nu(s_0)) > 0$, $\Re(\nu(0)) = 0$. Hence, by continuity, $\Re(z_\nu(s_2)) = \Re(\nu(s_2))$ at some $s_2 \in (0, s_0)$, so that $\lambda(s_2)$ lies in the generalized absolute spectrum at $\nu(s_2)$. Applying the first step concludes the proof. ∎

Lemmas 4.7 and 4.8 together prove Proposition 4.3. Note that Lemma 4.8 applies to any Σ-connecting curve, including ones that emanate from unbounded curves of essential spectrum.

Remark 3. *The constructive nature of the proof of Proposition 4.3 yields an algorithm for the location generalized absolute spectrum in an isola by numerical continuation (see also* [11]*): Given an isola γ, pick $\lambda \in \gamma$ and continue the solution $\lambda(s)$ to $d(\lambda, s + \mathrm{i}k_0) = 0$ in s such that*

it moves into $\mathrm{int}(\gamma)$. *We expect that* $\partial_\lambda d(\lambda,\nu)\neq 0$ *along this curve, and either a branch point occurs, or* $\lambda(s)$ *intersects* γ *for some* $|s_0|>0$. *In the latter case, compute* $d(\lambda(s_0),\mathrm{i}k_1)=0$ *with* $\mathrm{i}k_0\neq\mathrm{i}k_1$ *and continue the two Floquet exponents back along the curve* $\lambda(s)$ *until either a branch point occurs or the difference in real parts is zero.*

Thus, a point $\Sigma^\ell_{\mathrm{abs}}$ *has been located (unless a double root with respect to* λ *occurred), and in practice often* $\ell=j$; *see section 5 for examples. To handle the case* $\ell\neq j$, *we can extend the algorithm as follows. Pick* $\lambda\in\gamma$ *and find all* n *solutions to* $d(\lambda,\nu)=0$; *in particular, there is* $\nu=\mathrm{i}k$. *Then continue all solutions simultaneously as above in* s. *The theorem guarantees (unless a double root with respect to* λ *occurred) that there is* $0\leq|s|\leq|s_0|$ *such that* $\Re(\nu_j)=\Re(\nu_{j+1})$. *We refer to* [11] *for algorithms to find all initial* ν, *and for notes on the implementation in the software* AUTO [5].

5. An example. We present computations of essential and generalized absolute spectra for a wave train that occurs in the Schnakenberg model [15] using the numerical methods described in Remark 3 and [11] with the software AUTO [5]. Both absolute and essential spectrum of this wave train are unstable. We aim to illustrate the possible structures of (generalized) absolute and essential spectra of wave trains, and the use of Theorem 4.5 in calculating them.

We use the Schnakenberg model in the form

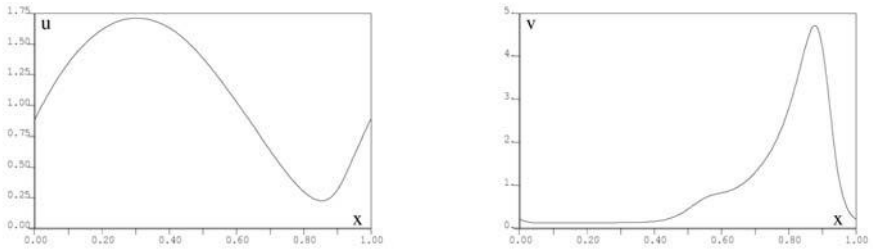
$$\begin{aligned} u_t &= D_u u_{xx} + 0.9 - uv^2,\\ v_t &= D_v v_{xx} + 0.1 + uv^2 - v, \end{aligned}$$

and consider a wave train U^* of period $L=3$ for the parameters $D_u\approx 0.45$, $D_v\approx 0.0045$ and velocity $c\approx 0.029$. The profiles of U^* are plotted in Figure 4.

This wave train can be located as follows. For $D_u\equiv 1$ the unique equilibrium $u=v=0$ undergoes a Turing bifurcation at $D_v\approx 0.12$. We continue the bifurcating Turing pattern with period $L=3$ to $D_u=1$, $D_v=0.01$ and then fix the ratio $D_v/D_u\equiv 0.01$. A bifurcation to a traveling wave train occurs as D_u decreases and we consider the solution U^* on this branch at $D_u\approx 0.45$ in the comoving frame with its speed $c\approx 0.029$.

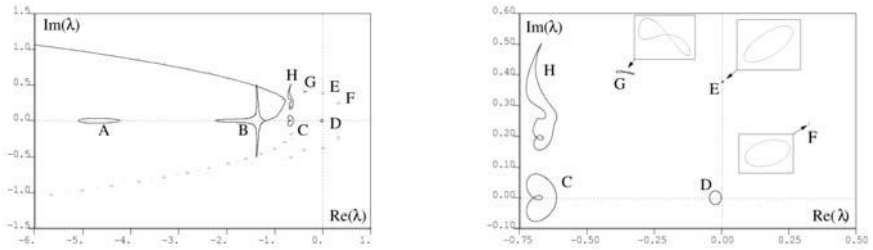
Recall that essential and generalized absolute spectra are determined by solutions of the (complex) dispersion relation (3.1). All $(\lambda,\nu)\in\mathbb{C}^2$ in the following are such solutions, viewed as local functions $\lambda(\nu)$ or $\nu(\lambda)$. The essential spectrum consists of all branches of $\lambda(\mathrm{i}k)$ and the generalized absolute spectrum consists of all λ for which two branches $\nu(\lambda)$ have the same real part. For this model there are four Floquet exponents ν at each λ, i.e., four branches of $\nu(\lambda)$, and these connect at branch points (Definition 3.1). To locate λ in the absolute spectrum (here $i_\infty=N=2$) the spatial Morse index $i(\lambda)$ is needed, and therefore all four Floquet exponents at λ.

The search for essential spectrum was guided by predictions from eigenvalues of first order finite difference approximations. Figure 5 plots the critical parts of the essential spectrum computed with AUTO [5] as well as eigenvalues computed via finite difference discretization with LAPACK [2]. The latter was performed under periodic boundary conditions on domains with length 3 (the period of U^*) for 800 grid points, and length 6 for 1600 grid points. Since any curve of the essential spectrum connects eigenvalues from lengths 3 and 6, computing these spectra helps to locate isola of the essential spectrum.



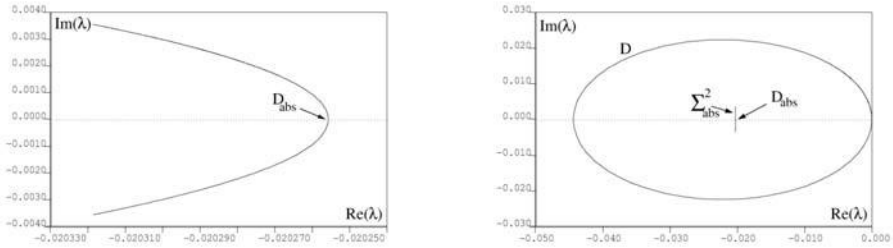
(a) (b)

Figure 4. (a) *u-component of the wave train* U^*. (b) *v-component.*



(a) (b)

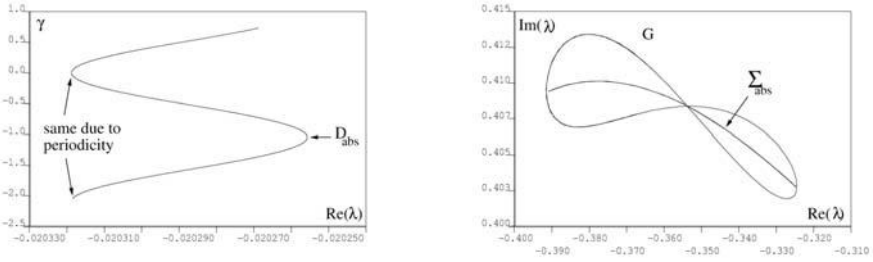
Figure 5. (a) *Parts of essential spectrum of the wave train computed with* AUTO *(black curves) and eigenvalues computed with finite differences on* $[0, 3]$ *(stars) and* $[0, 6]$ *(crosses). Letters denote isola referred to in the text.* (b) *magnifies part of* (a) *and insets magnify the indicated isola.*



(a) (b)

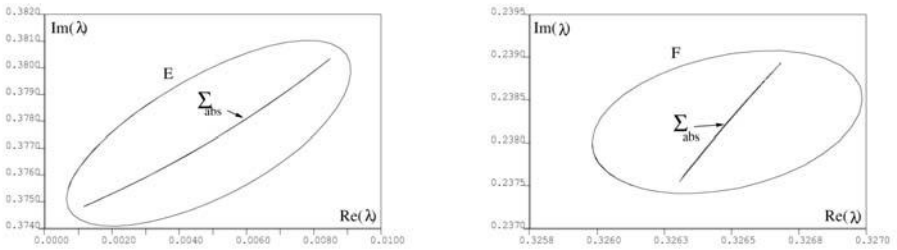
Figure 6. (a) *Curve of absolute spectrum inside isola D.* (b) *This curve together with isola D. For the point* D_{abs} *see also Figure* 9.

Here the isolas D, E, F, G lie in the boundary of the region Ω_∞, satisfy the assumptions of Theorem 4.5, and therefore contain absolute spectrum as plotted in Figures 6, 7, and 8, respectively. The isolas within the curves C and H do not have constant Morse index on their boundaries. While Theorem 4.5 cannot be applied directly, Lemma 4.2 implies that the self-intersection points in C and H lie in curves of Σ^1_{abs}, and these cross from Ω_∞ into the interior isolas; see also Figure 2. Since isola E and F are contained in the right half plane, the wave train is absolutely unstable.



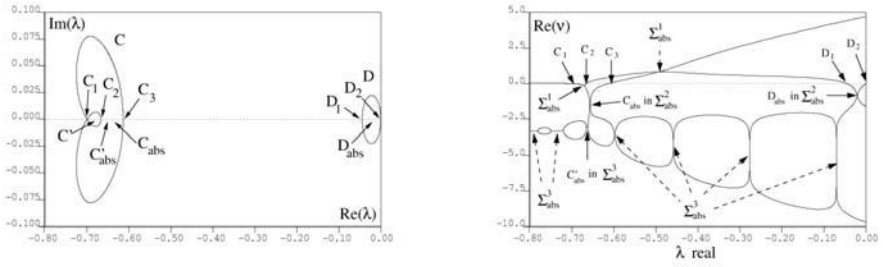
(a) (b)

Figure 7. (a) *Parameterization of the curve of absolute spectrum from Figure 6 by* $\gamma = \Im(\nu_2 - \nu_3)$ *with period* $2\pi/3 \sim 2.1$. *For* D_{abs} *see also Figures 9 and 6.* (b) *Isola G and the imbedded curve of absolute spectrum.*



(a) (b)

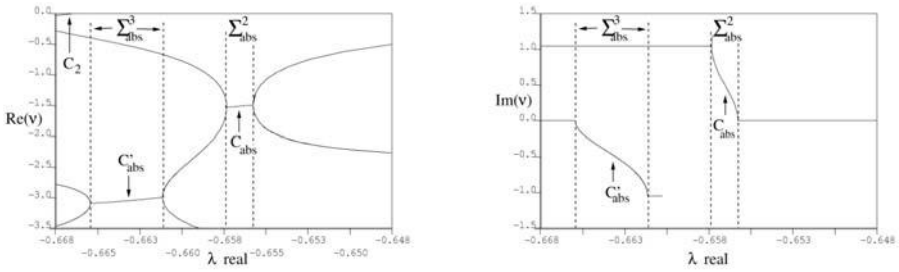
Figure 8. *Isolas and imbedded curves of absolute spectrum.* (a) *isola E* (b) *isola F.*



(a) (b)

Figure 9. (a) *Isola C and D of essential spectrum; see also Figure 5.* (b) *Branches of Floquet exponents for* $\lambda \in \mathbb{R}$ *in the range of* (a). *Intersections with essential spectrum and generalized absolute spectrum can be read off as indicated; compare also the location of points with* (a).

Details of the essential spectrum and curves of Floquet exponents for the isola C and D are plotted in Figure 9. Each crossing point of two branches $\Re(\nu(\lambda))$ in Figure 9(b) lies in the generalized absolute spectrum. For $\Re(\lambda) > -0.05$ the real parts of the Floquet exponents are separated into two unstable and two stable exponents, as predicted for the region Ω_∞. Computations on the real line fail to indicate the instability of the small isola E and F in $\mathbb{C} \setminus \mathbb{R}$; see Figure 5.



(a) (b)

Figure 10. *Branches of two spatial Floquet exponents within isola C for $\lambda \in \mathbb{R}$. Only one branch of a complex conjugate pair is plotted.* (a) *Real part, magnified from Figure 9*(b). (b) *Imaginary part with period $2\pi/3 \sim 2.1$; horizontal lines correspond to real spatial Floquet exponents.*

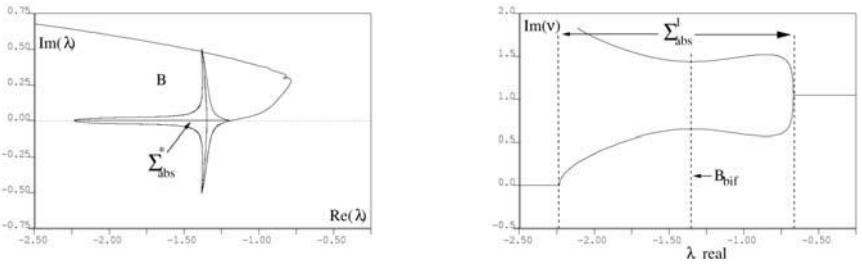
As to generalized absolute spectrum, recall $\lambda \in \Sigma^j_{\mathrm{abs}}$, if $\Re(\nu_j(\lambda)) = \Re(\nu_{j+1}(\lambda))$ for $\Re(\nu_1) \geq \cdots \geq \Re(\nu_n)$. The spatial Morse index j is one plus the number of Floquet exponents with larger real part; e.g., in Figure 9(b) it is one plus the number of curves above a point in the generalized absolute spectrum.

Due to symmetry, $\Sigma^*_{\mathrm{abs}} \cap \mathbb{R}$ consists of λ, where two $\nu(\lambda)$ are complex conjugate or coincide (branch points), or a positive and negative Floquet *multiplier* have the same modulus. In the latter case two Floquet exponents differ by the factor $\mathrm{e}^{\mathrm{i}\pi}$, which cannot occur for spatial eigenvalues of a uniform steady state. By symmetry, intervals in $\Sigma_{\mathrm{abs}} \cap \mathbb{R}$ with a complex conjugate pair terminate at branch points, while a curve of Σ_{abs} typically crosses $\mathbb{R}$ with vertical tangent at a point where multipliers in $\Sigma_{\mathrm{abs}} \cap \mathbb{R}$ have opposite sign; cf. [11, section 4.3].

Figure 6 shows a curve of absolute spectrum (index $i_\infty = N = 2$) crossing the real line at the point $\mathrm{D}_{\mathrm{abs}}$; see also Figure 9(b). This curve was predicted above using Theorem 4.5 and was computed using the algorithm of its proof as described in Remark 3. Figure 7(a) shows the parameterization of this curve of absolute spectrum by the difference of imaginary parts $\kappa = \Im(\nu_2 - \nu_3)$. Since the period of U^* is 3, the period in κ is $2\pi/3 \sim 2.1$. There are branch points at $\kappa = j2\pi/3$, $j \in \mathbb{Z}$, the endpoints of the curve of absolute spectrum. At $\kappa = -\pi/3$, $j \in \mathbb{Z}$, the Floquet exponents are real, which corresponds to the point $\lambda = \mathrm{D}_{\mathrm{abs}} \in \mathbb{R}$. Since the Floquet exponents' real parts plotted in Figure 9(b) differ for $\lambda \in \mathrm{int}(\mathrm{D}) \cap \mathbb{R} \setminus \{\mathrm{D}_{\mathrm{abs}}\}$, there is no further Σ^*_{abs} in D that intersects the real line.

We next investigate some intervals of generalized absolute spectrum on the real axis. Since we strive for illustration, we disregard the points of Σ^*_{abs} that are marked with dotted arrows in Figure 9(b).

As mentioned above, Theorem 4.5 does not apply to isola C directly. However, for suitable exponential weights (see Remark 1, e.g. $\eta = 1.1$), part of isola C continues in η to a Jordan curve C' $\in \mathrm{C} \cup \Omega^\eta_\infty$. In this weighted space, Theorem 4.5 applies and predicts a curve of $\Sigma_{\mathrm{abs}} = \Sigma^2_{\mathrm{abs}}$ inside C' and hence inside C. Indeed, continuing the branches of $\nu(\lambda)$ for $\lambda \in \mathbb{R}$ there is an interval of Σ_{abs} on the real line confined within isola C; see Figures 9 and 10. In addition, isola C contains an interval of Σ^3_{abs}. Figures 10(a) and (b) show that two real distinct Floquet exponents meet in branch points and become a complex conjugate pair at the endpoints of these intervals. Concerning other Morse indices, Figure 9(b) shows a branch



(a) (b)

Figure 11. *We plot two curves of* Σ^1_{abs} *that cross within isola B at* B_{bif}. (a) *Spectrum in* λ*-plane.* (b) *Location of the crossing point at a local extremum of* $\Im(\nu)$ *for* $\lambda \in \mathbb{R}$.

point in $\Sigma^1_{\mathrm{abs}} \cap \mathbb{R}$ between the points C_1 and C_2 in isola C'. The attached interval of Σ^1_{abs} crosses the self-intersection point of isola C as predicted by Lemma 4.2 and extends into isola B; see Figure 11(a). The curves of imaginary parts in Figure 11(b) illustrate again how two Floquet exponents become a complex conjugate pair at the endpoints of the interval.

Without including a figure we remark that isola A contains an interval of $\Sigma^1_{\mathrm{abs}} \cap \mathbb{R}$ as predicted by Theorem 4.5. In fact, more intervals and crossing curves of generalized absolute spectrum on the real line occur as the branches of Floquet exponents repeatedly cross at the dotted arrows in Figure 9(b).

Finally, we expect the bifurcation of a curve of Σ^j_{abs} into the complex plane at a critical point of the imaginary part $\Im(\nu(\lambda))$ in any interval of $\Sigma^j_{\mathrm{abs}} \in \mathbb{R}$ bounded by branch points; see [11, section 4.2.3]. More precisely, we expect a curve of Σ^*_{abs} crosses vertically at a point where $\frac{\mathrm{d}}{\mathrm{d}\lambda}\Im(\nu(\lambda)) = 0$. Indeed, we can numerically locate such bifurcation points and switch to the bifurcating branch. An example for this in isola B is plotted in Figure 11.

Acknowledgments. The author is grateful to Arnd Scheel and Björn Sandstede for their precious advise, and to Bernold Fiedler, Harald Engel, and their groups for their hospitality. The author thanks Michael Ward and Ralf Wittenberg for their support.

REFERENCES

[1] L. V. Ahlfors, *Complex Analysis*, McGraw–Hill, New York, 1953.
[2] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *Lapack Users' Guide*, SIAM, Philadelphia, 1995.
[3] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw–Hill, New York, 1955.
[4] W. A. Coppel, *Dichotomies in Stability Theory*, Lecture Notes Math. 629, Springer-Verlag, Berlin, 1978.
[5] E. Doedel, R. C. Paffenroth, A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. E. Oldeman, B. Sandstede, and X. Wang, *AUTO2000: Continuation and Bifurcation Software for Ordinary Differential Equations (with HOMCONT)*, Technical report, Concordia University, Montreal, Canada, 2002.
[6] R. A. Gardner, *On the structure of the spectra of periodic traveling waves*, J. Math. Pures Appl. (9), 72 (1993), pp. 415–439.

[7] R. A. Gardner, *Spectral analysis of long wavelength periodic waves and applications*, J. Reine Angew. Math., 491 (1997), pp. 149–181.
[8] D. Henry, *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Math. 840, Springer-Verlag, Berlin, 1981.
[9] S. G. Krantz, *Function Theory of Several Complex Variables*, 2nd ed., AMS Chelsea Publishing, Providence, RI, 2001.
[10] J. D. M. Rademacher, *Homoclinic Bifurcation from Heteroclinic Cycles with Periodic Orbits and Trace-firing of Pulses*, Ph.D. thesis, University of Minnesota, Minneapolis, MN, 2004.
[11] J. D. M. Rademacher, B. Sandstede, and A. Scheel, *Computing Absolute and Essential Spectra Using Continuation*, IMA preprint 2054, IMA, Minneapolis, MN, 2005.
[12] B. Sandstede and A. Scheel, *Gluing unstable fronts and backs together can produce stable pulses*, Nonlinearity, 13 (2000), pp. 1465–1482.
[13] B. Sandstede and A. Scheel, *Absolute and convective instabilities of waves on unbounded and large bounded domains*, Phys. D, 145 (2000), pp. 233–277.
[14] B. Sandstede and A. Scheel, *On the stability of periodic traveling waves with large spatial period*, J. Differential Equations, 172 (2001), pp. 134–188.
[15] J. Schnakenberg, *Simple chemical reaction systems with limit cycle behaviour*, J. Theor. Biol., 81 (1979), pp. 389–400.
[16] G. Schneider, *Diffusive stability of spatial periodic solutions of the Swift-Hohenberg equation*, Comm. Math. Phys., 178 (1996), pp. 679–702.
[17] T. Valent, *Boundary Value Problems of Finite Elasticity. Local Theorems on Existence, Uniqueness, and Analytic Dependence on Data*, Springer Tracts in Natural Philosophy 31, Springer-Verlag, New York, 1988.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 650–669

Motion Planning for an Articulated Body in a Perfect Planar Fluid*

Juan B. Melli†, Clarence W. Rowley†, and Dzhelil S. Rufat†

Abstract. Understanding fish-like locomotion as a result of internal shape changes may result in improved underwater propulsion mechanisms. We use a geometric framework to consider the simplified problem of an articulated two-dimensional body in a potential flow. This paper builds upon the current geometric theory by showing that although the group of Euclidean transformations is non-Abelian, certain tools available for Abelian groups may still be exploited, making use of the semidirect-product structure of this group. In particular, the holonomy in the rotation component may be explicitly computed as a function of the area enclosed by a path in shape space. We use this tool to develop open-loop gaits for an articulated body with two shape variables, using plots of the curvature of the mechanical connection, which relates motion in the shape space to motion of the overall body. Results from numerical computations of the mechanical connection are compared to theoretical results assuming the joints are hydrodynamically decoupled. Finally, we consider a simple method for trajectory tracking in the plane, using a one-parameter family of gaits.

Key words. geometric phase, locomotion, connection, principal bundle

AMS subject classifications. 81Q70, 53C29, 55R10

DOI. 10.1137/060649884

1. Introduction.

In this paper, we study articulated bodies in a potential flow and their locomotion as a result of internal shape changes. In particular, we seek to answer the following question: *Given a desired motion, what are the changes in shape that will achieve such a motion?*

Fish swimming has been studied for many decades in order to understand the physical mechanisms involved and ultimately to improve upon conventional methods of underwater propulsion. The interaction of solid bodies with an ideal fluid was first studied by Kirchhoff in the 1870's [7, 9]. In the 1960's, analytical models by Wu examined the swimming of a two-dimensional flexible plate with a small amplitude traveling wave [24, 25], while Lighthill considered the propulsion of slender-body fishes [12, 13, 14, 15]. More recently, Kelly [5, 6] studied the self-propulsion of fish-like and amoeba-like swimmers in potential and Stokes flow from a geometric point of view, by viewing the configuration space as a principal fiber bundle, in which the base space consists of the internal shape variables, and the fibers are the group of rigid motions of the overall body. The motion along the fiber is related to the motion in the base space by a principal connection, called the *hydromechanical connection*, a term which comes from the geometric mechanics community [19]. These techniques were used in [8] to study geometric phases in articulated free rigid bodies, and similar techniques

*Received by the editors January 13, 2006; accepted for publication (in revised form) by L. Fauci August 17, 2006; published electronically November 22, 2006. This work was supported by the National Science Foundation (award CMS-0347239, with support from an REU supplement) and an NDSEG Graduate Research Fellowship.
http://www.siam.org/journals/siads/5-4/64988.html
†Mechanical Engineering Department, Princeton University, Princeton, NJ 08544 (jmelli@princeton.edu, cwrowley@princeton.edu, drufat@princeton.edu).

were used in [21, 23] to study the motion of deformable bodies and articulated rigid bodies in a fluid, assuming small amplitude motions, and assuming that the individual links were *hydrodynamically decoupled*; that is, that the added inertia of each link is independent of the position or velocity of the other links. Kanso et al. [4] later formulated the potential flow problem explicitly as a reduction by stages [1] (although such a formulation was already implicit in [5]) and accurately computed the added inertia terms using a boundary element method. Tools from control theory and geometric mechanics have also been used to generate gaits for locomotion of underwater vehicles [11, 21, 22].

Benefits of the geometric formulation. One of the advantages of formulating the equations using the structure of a fiber bundle is that one may use the hydromechanical connection to design swimming gaits to achieve desired motions. In this paper, we use the formulation of [5, 4] to develop open-loop swimming gaits to achieve arbitrary rotations in two dimensions. The technique we use is well known for Abelian groups; we show that they may also be used for the rotation component of the Euclidean group, exploiting the semidirect product structure $SE(2) = SO(2) \circledS \mathbb{R}^2$, since $SO(2)$ is Abelian. Whereas some previous results were limited to small amplitude gaits or assumed hydrodynamically decoupled links, our results allow us to develop finite amplitude gaits and accurately compute the motion. By numerically computing the curvature of the hydromechanical connection, it becomes possible to develop such gaits by inspection.

It is well known that fish propulsion relies in part on the shedding of vorticity, for instance, as demonstrated in the experiments of [3]. Although we ignore the effects of viscosity and do not model vortex shedding, our model captures the inertial forces associated with fish-like swimming, which are most important during quick maneuvers. In other words, our model is not valid as representative of real fish swimming unless maneuvers are performed quickly. We consider this to be a first step at developing fish-like gaits and plan to incorporate the effects of vorticity in future models.

This paper is organized as follows. In section 2 we introduce the problem of a three-link fish-like articulated swimmer in a potential flow. Section 3 includes the geometric framework necessary to study the present problem and includes a formula for computing the holonomy in a semidirect product group $S = G \circledS V$, where G is an Abelian group and V is a vector space. This theory is applied in section 4 to find open-loop gaits from plots of the curvature of the local connection to achieve a desired net rotation in the fish-like body. For the three-link geometry, it is found that the most efficient turning gaits involve motions where the body is "C"-shaped, rather than "S"-shaped. Further, we investigate different shape geometries for maneuvering, including the effect of removing the middle link. The curvature plots are compared to the case where the links are assumed to be hydrodynamically decoupled. We then show how the curvature plots can be used to generate small-amplitude gaits for the non-Abelian group components. Finally, in section 5 we consider a family of gaits and present a method for following a trajectory in the plane.

2. Problem formulation.

Consider a fish-like body modeled as a set of articulated rigid links connected by hinged joints. The body is assumed to be two-dimensional and neutrally buoyant. The example we will consider here is a three-link body with two joints, as in Figure 1, which begins at rest, and is immersed in a perfect fluid. The formulation we use is the same

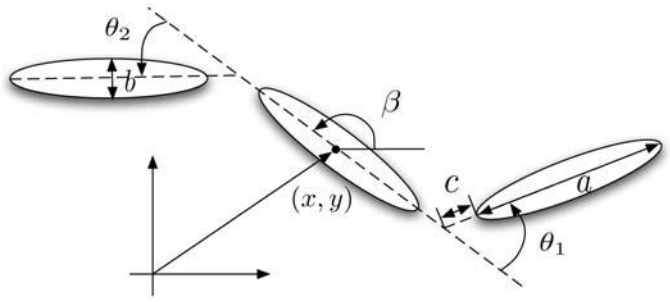


Figure 1. *A three-link swimmer.*

as that in [4], but here we review the main ideas and define the notation used.

Kinematics. The overall motion of the body is represented by the position and orientation of the middle body relative to an inertial frame of reference and is given by an element $s = (x, y, \beta) \in SE(2)$, the special Euclidean group. As noted earlier, we will exploit the semidirect product structure of this group. The shape of the body is determined by the angles θ_1 and θ_2 between the joints, which we call the shape variables. In computations, it will be useful to express the positions of the outer links relative to this middle link by elements $x_1, x_2 \in SE(2)$, where we label the outer links as 1 and 2 as in Figure 1. These elements x_1 and x_2 are of course determined completely by the angles θ_1 and θ_2.

Dynamics. The lack of external forces or torques acting on the body or the fluid leads to the conservation of a momentum-like quantity of the body plus fluid system. We assume the fluid is inviscid, incompressible, and because there is no mechanism for the generation of vorticity, irrotational for all time. Thus the velocity field can be represented as the gradient of a scalar potential,

$$\mathbf{u} = \nabla\phi. \tag{2.1}$$

To determine the equations of motion, we use the Lagrangian formulation, as in [9, 5, 4]. The Lagrangian of the system is equal to the total kinetic energy of the body plus the fluid,

$$L = T_{\mathcal{B}} + T_{\mathcal{F}}, \tag{2.2}$$

where the fluid kinetic energy $T_{\mathcal{F}}$ is given by

$$T_{\mathcal{F}} = \frac{1}{2}\int_{\mathcal{F}} \rho_{\mathcal{F}} |\mathbf{u}|^2 \, \mathrm{d}V, \tag{2.3}$$

where $\rho_{\mathcal{F}}$ is the fluid density. For an ideal fluid, it is well known that the motion of the fluid is determined completely by the motion of the solid bodies [9]; that is, the total kinetic energy may be expressed solely in terms of the velocities of the bodies, with the role of the fluid appearing only as added inertias that of course depend on the configuration of the solid bodies. This formulation is originally due to Kirchhoff [7, 9] and may also be understood as an explicit reduction by the particle-relabeling symmetry [4].

The potential ϕ may therefore be written as a sum of translational and rotational velocity potentials corresponding to each link. The Neumann boundary condition of no penetration is imposed along the surface, requiring that the normal velocity of the fluid, $\nabla\phi \cdot \mathbf{n} = \frac{\partial\phi}{\partial\mathbf{n}}$, at a point along the body surface be equal to the normal velocity of the body at that point.

Since the fluid has infinite extent, the total momentum of the system is indeterminate. However, we may consider the system "impulse," a momentum-like quantity defined by Lord Kelvin [9]. The impulse is a product of the total link inertias (actual plus added inertias) and the body velocities of each link, and may be expressed as in [4] as

$$h_s = \sum_{j=1}^{3} \Big(\mathrm{Ad}^T_{x_1^{-1}}(\mathbb{I}_{1j}\xi_j) + \mathrm{Ad}^T_{x_2^{-1}}(\mathbb{I}_{2j}\xi_j) + \mathbb{I}_{3j}\xi_j \Big), \tag{2.4}$$

where $\xi_j \in \mathbb{R}^3 \cong \mathfrak{se}(2)$ is the velocity of link j with respect to a reference frame attached to that link, where links $1, 2$ are the outer links as before, and link 3 is the middle link. The matrices $\mathbb{I}_{ij}$ are the total inertia terms such that the total kinetic energy of the body plus fluid system is

$$T = \frac{1}{2} \sum_{i,j=1}^{3} \xi_i^T \mathbb{I}_{ij} \xi_j \tag{2.5}$$

and $\mathrm{Ad}^T_{x_j^{-1}}$ are operators that map the momenta term corresponding to body j from the j body-fixed frame to the body-fixed frame of the middle body (see [17, 4] for more discussion of the Ad operator and background on Lie groups). Defining $\zeta_j = x_j^{-1}\dot{x}_j$ as the velocity of link j relative to the middle link, expressed with respect to a frame fixed to body j, we then can write the velocities ξ_j as

$$\xi_j = \zeta_j + \mathrm{Ad}_{x_j^{-1}}\xi_3, \quad j = 1, 2. \tag{2.6}$$

The system impulse may also be expressed with respect to an inertial frame, in which case it will remain unchanged due to the lack of external forces and moments. As currently defined with respect to the body frame, in general, the value will change in time. However, when the system begins at rest, the system impulse h_s remains zero for all time. In this case, (2.4) combined with (2.6) can be rewritten as

$$\mathbb{I}_{\mathrm{loc}}\,\xi_3 + \sum_{\alpha,\beta=1}^{2} \big(\mathrm{Ad}^T_{x_\alpha^{-1}} \mathbb{I}_{\alpha\beta} + \mathbb{I}_{3\beta} \big) \zeta_\beta \;=\; 0, \tag{2.7}$$

where

$$\mathbb{I}_{\mathrm{loc}} \;=\; \sum_{\alpha,\beta=1}^{2} \Big(\mathrm{Ad}^T_{x_\alpha^{-1}} \mathbb{I}_{\alpha\beta} \mathrm{Ad}_{x_\beta^{-1}} + \mathrm{Ad}^T_{x_\alpha^{-1}} \mathbb{I}_{\alpha 3} + \mathbb{I}_{3\beta} \mathrm{Ad}_{x_\beta^{-1}} \Big) + \mathbb{I}_{33} \tag{2.8}$$

is the locked moment of inertia. We may then write (2.7) as

$$\mathbb{I}_{\mathrm{loc}} \big(\xi_3 + \mathcal{A}(x)\dot{x} \big) = 0, \tag{2.9}$$

where $\mathcal{A}$ is the local form of the mechanical connection, an $\mathfrak{se}(2)$-valued one-form on the shape space. When the locked moment of inertia is nondegenerate, (2.9) simplifies to $\xi_3 + \mathcal{A}(x)\dot{x} = 0$. Due to the constraint on the joint between the links, the rigid motion x_α can be parameterized by the single variable θ_α. Also, considering that $\xi_3 = g^{-1}\dot{g}$, the equations of motion for the system can be written as

$$\dot{s}(t) = -T_e L_{s(t)} \left(\mathcal{A}(\theta)\dot{\theta}(t) \right) \tag{2.10}$$

or in shorthand notation

$$\dot{s} = -s\mathcal{A}(\theta)\dot{\theta}, \tag{2.11}$$

where $s \in S$, $\mathfrak{s} = T_e S$, and $\mathcal{A} : TQ \to \mathfrak{s}$ is the (local) *mechanical connection*, a Lie algebra-valued one form on the shape space. It maps a prescribed shape velocity element $\dot{\theta} = (\dot{\theta}_1, \dot{\theta}_2)$ to the negative of the body velocity of the middle body, an element of $\mathfrak{se}(2)$. The connection is a function of the added inertias and is computed numerically at each time step. The motion of the fish can be solved by prescribing a path for θ in shape space and integrating equation (2.10) in time. Trial and error was used in [4] to determine gaits which generate forward and turning motions. Here, we ask the reverse question: *Given a desired motion, what is the closed path in shape space required to achieve this motion?*

3. Holonomy in semidirect product groups. If the group S is Abelian, then for a given path in shape space given by $\theta(t)$, $t \in [0, 1]$, the solution to (2.11) with initial condition $s(0) = e$ is simply

$$s(1) = \exp\left(-\int_0^1 \mathcal{A}(\theta)\dot{\theta}(t)\, dt \right). \tag{3.1}$$

If $\theta(t)$ is a closed path for $t \in [0, 1]$, then this result is equivalent to integrating the connection along this path in shape space, or by the Stokes theorem, to integrating the curvature of the connection over the area enclosed by the path,

$$\exp\left(-\int_0^1 \mathcal{A}(\theta)\dot{\theta}(t)\, dt \right) = \exp\left(-\int_{\partial C} \mathcal{A}(\theta)\, \mathrm{d}\theta \right) = \exp\left(-\iint_C \mathrm{D}\mathcal{A}(\theta)\, \mathrm{d}A \right), \tag{3.2}$$

where the curvature $\mathrm{D}\mathcal{A} : T_qQ \times T_qQ \to \mathfrak{s}$ is the covariant derivative of the connection $\mathcal{A}$, given by

$$\mathrm{D}\mathcal{A}(X, Y) = d\mathcal{A}(X, Y) - [\mathcal{A}(X), \mathcal{A}(Y)], \tag{3.3}$$

where X, Y are vector fields on Q [20]. This is a powerful result because it greatly simplifies the problem of generating gaits to achieve a desired motion. First, one selects the desired element in the group space and solves for the right-hand side of (3.1) by taking the logarithm of both sides. Then, using the equality in (3.2), one finds an area in shape space that encloses the corresponding volume of the curvature of the connection.

Unfortunately, the case when S is non-Abelian has no explicit solution. This is the case for motion in the plane, where the Lie group is $S = SE(2)$. However, $SE(2)$ can be expressed

as the semidirect product $SO(2)\circledS\mathbb{R}^2$. When the Lie group component (G) of the semidirect product group is Abelian, it can be shown that the solution for this component of the semidirect product group is the same as that for the Abelian case (3.1).

Semidirect products. Recall that if G is a Lie group that acts on a vector space V, then one can define the semidirect product $G\circledS V$ as the usual product, with the group operation $(g_1, v_1)(g_2, v_2) = (g_1g_2, g_1v_2+v_1)$, where $g_1, g_2 \in G$, $v_1, v_2 \in V$. The Lie algebras corresponding to the various groups are $\mathfrak{g} = T_eG$, $V = T_eV$, and $\mathfrak{s} = T_eS$.

Theorem 3.1. *Let $S = G\circledS V$, where G is Abelian. Consider a closed curve $\theta(t) \in Q$ for $t \in [0, 1]$, and let $\mathcal{A} : TQ \to \mathfrak{s}$ be a principal connection with components $\mathcal{A}_\mathfrak{g} : TQ \to \mathfrak{g}$ and $\mathcal{A}_V : TQ \to V$. Then if $s(t) = (g(t), v(t)) \in S$ satisfies*

$$\dot{s} = -s\mathcal{A}(\theta)\dot{\theta} \tag{3.4}$$

with $s(0) = (e, 0)$, then

$$\text{holonomy}_G := g(1) = \exp\left(-\int_{\partial C} \mathcal{A}_\mathfrak{g}\, d\theta\right) = \exp\left(-\iint_C \mathrm{D}\mathcal{A}_\mathfrak{g}\, dA\right). \tag{3.5}$$

Proof. Since $\theta(t)$ is given, we may rewrite (3.4) as $\dot{s} = -s\xi(t)$, where $\xi = \mathcal{A}(\theta)\dot{\theta}$ is given. Writing $\xi = (\xi_g, \xi_v)$, where $\xi_g \in \mathfrak{g}$ and $\xi_v \in V$, this becomes (see [18])

$$\dot{g} = -T_eL_g\xi_g, \tag{3.6}$$

$$\dot{v} = -\rho(g)\xi_v. \tag{3.7}$$

Clearly, (3.6) is decoupled from (3.7), and since G is Abelian, the solution is

$$g(t) = \exp\left(-\int_0^t \xi_g(\tau)d\tau\right). \tag{3.8}$$

The holonomy of S in the G component is then

$$\text{holonomy}_G = g(1) = \exp\left(-\int_{\partial C} \mathcal{A}_\mathfrak{g}\, d\theta\right) = \exp\left(-\iint_C \mathrm{D}\mathcal{A}_\mathfrak{g}\, dA\right),$$

where the last equality is by the Stokes theorem. ∎

This result is used to generate gaits for the Abelian component of the Lie group which correspond to turning maneuvers for the fish. As demonstrated in Figure 6(a), this result does not apply in the general case, and we cannot use this method to generate finite amplitude translational gaits.

4. Results.

For the fish-like body depicted in Figure 1, the shape space Q is parameterized by (θ_1, θ_2), so the local connection $\mathcal{A}(\theta_1, \theta_1) : T_{(\theta_1,\theta_2)}Q \to \mathfrak{se}(2)$ is a Lie-algebra-valued one-form on Q, which may be written as

$$\mathcal{A}(\theta_1, \theta_2) = f(\theta_1, \theta_2)\mathrm{d}\theta_1 + g(\theta_1, \theta_2)\mathrm{d}\theta_2, \tag{4.1}$$

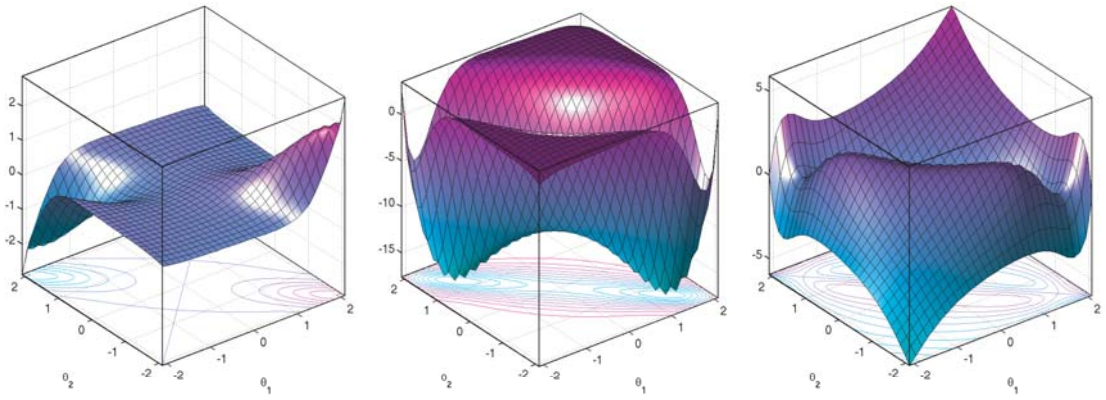
where $f, g : Q \to \mathfrak{se}(2)$ depend on the added inertia terms, which depend on the shape of the body and are computed from (2.9). A boundary element method using source distributions is

employed to compute the added inertia terms at each time step. For details of the computation, see [4]. The curvature $D\mathcal{A}$ is then a Lie-algebra-valued two-form on Q, computed from (3.3) as

$$D\mathcal{A} = \left[\left(\frac{\partial g}{\partial \theta_1} - \frac{\partial f}{\partial \theta_2}\right) - [f, g]\right] \mathrm{d}\theta_1 \wedge \mathrm{d}\theta_2 \tag{4.2}$$

$$= \gamma(\theta_1, \theta_2)\mathrm{d}\theta_1 \wedge \mathrm{d}\theta_2, \tag{4.3}$$

where $\gamma : Q \to \mathfrak{se}(2)$ is the bracketed term in (4.2). The curvature has three components: one corresponding to rotational body velocity and two corresponding to translational body velocity terms. The translational components, denoted u and v, correspond to motion parallel to the direction of the major and minor axes of the middle link, respectively. The three components of γ are computed numerically on a grid in the (θ_1, θ_2) plane and plotted in Figure 2. The plots correspond to a fish-like body with three equal size elliptical links where the dimensions (as in Figure 1) are $a = 20$, $c = 2$ and the aspect ratio of the ellipses is 10.



(a) (b) (c)

Figure 2. (a) ω, (b) u, *and* (c) v *components of curvature.*

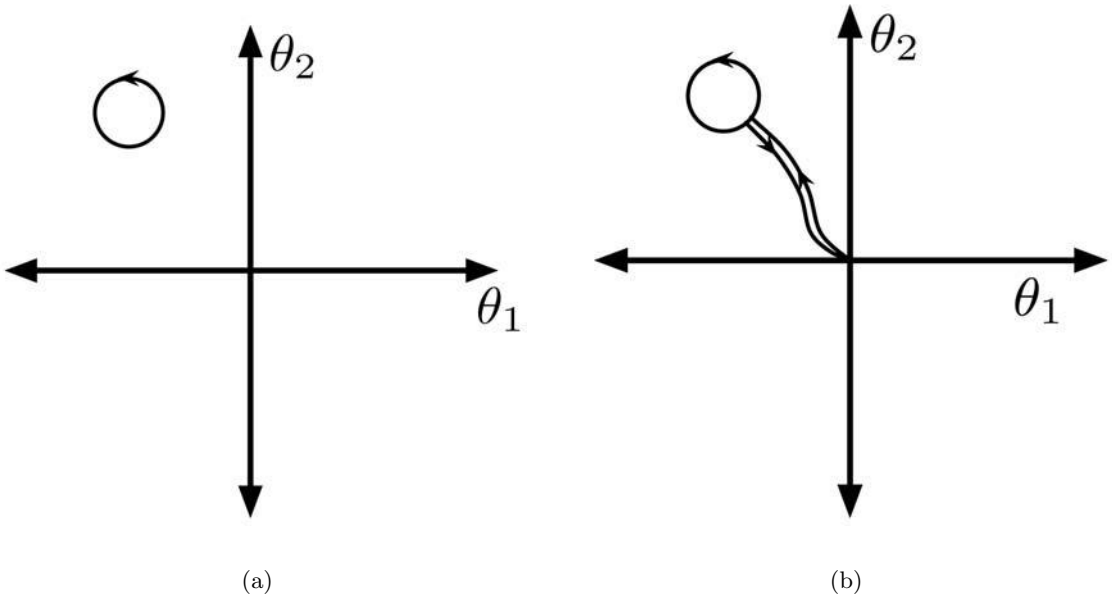
4.1. Using the curvature for gait generation. Since $SO(2)$ is Abelian and $SE(2) = SO(2) \circledS \mathbb{R}^2$, one can use the ω component of the curvature plot to develop finite-amplitude turning gaits, using the theory in section 3. The ω-component plot is shown enlarged in Figure 4. Note that the regions of largest curvature occur in two opposite corners of the shape space. These regions correspond to the fish configuration where the joints are bent in a "C"-shape. The most efficient turning gaits will enclose these regions of high curvature. Intuitively, this makes sense as one would expect a turning fish to coil its body into a "C"-shape to minimize the inertial resistance of the fluid as it turns. Likewise, a fish in an extended or "S"-shape configuration trying to rotate would encounter larger inertial forces and would not be expected to turn easily.

The procedure for gait-generation for the Abelian subgroup component is straightforward:

1. Choose the desired Abelian Lie group element (here, the desired net rotation).

2. Determine the corresponding Lie algebra element by applying the logarithm map.
3. Find a path in shape space that encloses a volume equal to the negative value of the Lie algebra element found in the previous step.

Note that, since the net rotation depends only on the enclosed area, the initial shape configuration need not be a point on the path, as shown in Figure 3.



(a) (b)

Figure 3. *Two paths resulting in the same holonomy in the Abelian subgroup component, but starting and ending at different points in shape space.*

Exponential map. The exponential map takes elements of the Lie algebra and maps them to the Lie group. For $SE(2)$, the exponential map is given by $\exp(u, v, \omega) = (x, y, \beta)$, where

$$\beta = \omega, \tag{4.4}$$

$$(x, y) = \begin{cases} (u, v), & \omega = 0, \\ \frac{1}{\omega}(u \sin\omega + v(1 - \cos\omega), u(\cos\omega - 1) + v \sin\omega), & \omega \neq 0. \end{cases} \tag{4.5}$$

For the purposes of this paper, we will require only the component of this map given by (4.4), so for this component, the exponential map, and its inverse, the log map, are just the identity.

As an example of gait-generation, we choose a desired net rotation of $\beta = \pi/4$. Thus, $\omega = \pi/4$ and we seek a closed path in (θ_1, θ_2) space that encloses a volume equal to $-\pi/4$. Figure 4 illustrates one such path, given by

$$\begin{aligned} \theta_1(t) &= -1.5 + .46\cos(t), \\ \theta_2(t) &= 1.5 - .46\sin(t), \end{aligned} \tag{4.6}$$

where $t \in [0, 2\pi]$. Note that the holonomy is independent of the speed of travel along this path.

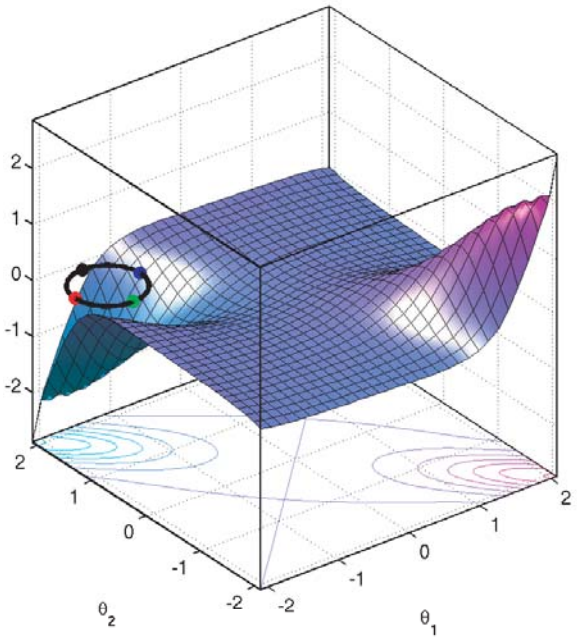


Figure 4. *Path in shape space resulting in net rotation of* $\frac{\pi}{4}$ *radians:* $\theta_1(t) = -1.5 + .46\cos(t)$*;* $\theta_2(t) = 1.5 - .46\sin(t)$.

The fish middle link begins at a zero degree angle from the horizontal and, after one shape space loop, has rotated 45 degrees counterclockwise, as in Figure 5. The angles of the joints are the same in the original and final configuration.

The start and end points of the closed path have no effect on the net rotation, as implied by (3.5). However, the choice for the start and end points will affect the net holonomy in the translational component of motion. Figure 6 shows the trajectory of the center of the fish for various starting points in the prescribed path. The starting points are labeled in Figure 4 in the same color as the corresponding paths in Figures 6(a) and 6(b). In each case, the path is in the clockwise direction. Because the net holonomy in these directions depends on the starting point of the path, it is clear that for the translational component in $SE(2)$, as is the case for general non-Abelian groups, a formula analogous to (3.5) is not possible, as the holonomy cannot depend only on the area enclosed by the path.

As a second example, we present a gait to achieve a net rotation of $\pi/2$, while starting and ending with the fish in a straight configuration (unlike the previous example, where the fish starts and ends in a "C"-shape). In order to begin and end with a straightened out configuration, we choose the gait shape space path to begin and end at the origin, and numerically find an area in shape space that encloses a volume of $-\pi/2$. The path selected is shown in red in Figure 7 along with the initial and final configurations of the swimmer. A video animation of the turning gait is available.

4.2. Determining degree of maneuverability of fish shape via numerical experiments. A numerical study was performed to determine how the body geometry affects the rotational

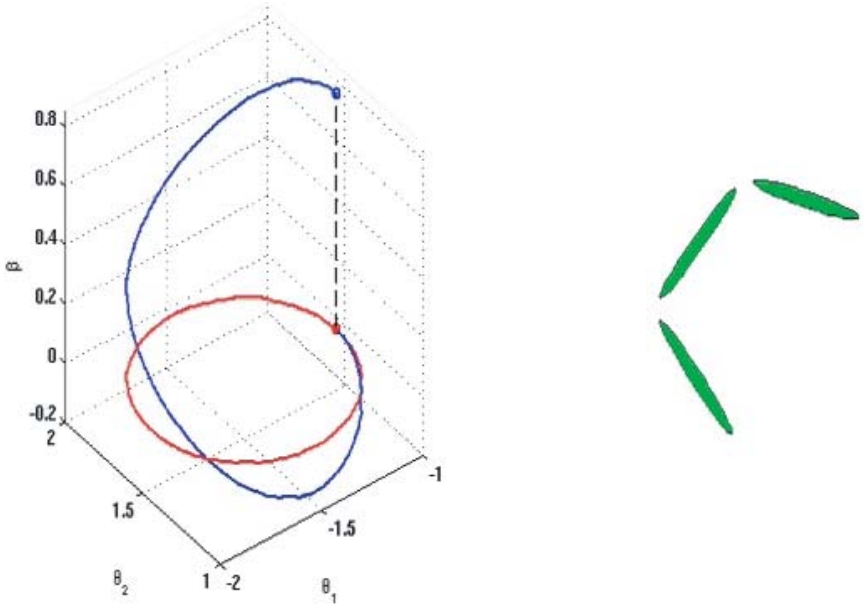
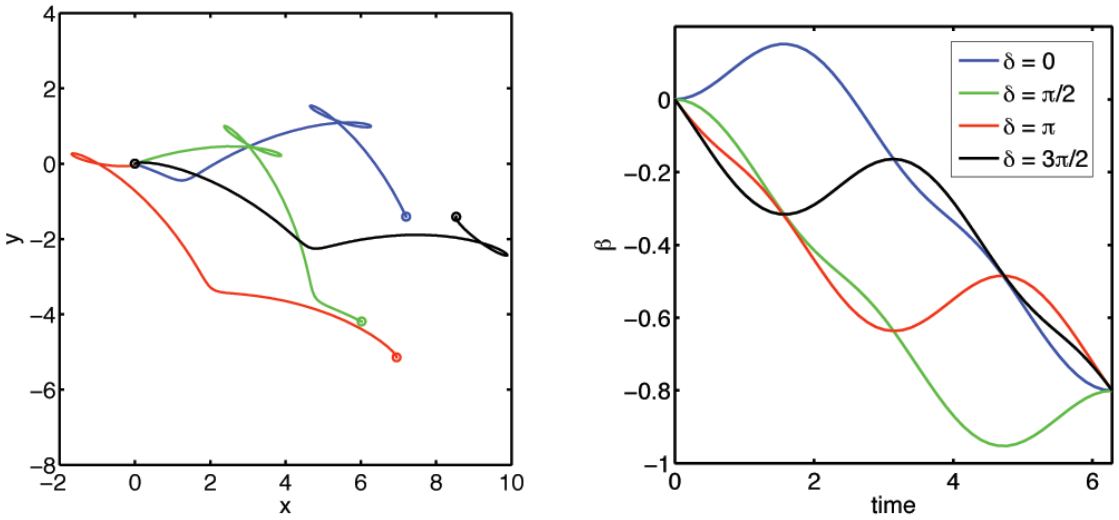


Figure 5. *Turning gait path in shape space in red with the rotation holonomy in blue along with the final configuration for a three-link swimmer after one turning gait. See also the accompanying animation* (64988_01.gif [190KB]).

maneuverability of the fish. This was done by examining the ω component of the curvature plots for various shape geometries along the line $\theta_1 = -\theta_2$, corresponding to "C"-shapes, which have been shown to be the most efficient configurations for turning gaits. The variables examined were the aspect ratio of each individual ellipse as well as the gap between the joints. The number of bodies was held fixed at three and the sizes of all three bodies are assumed to be equal. It was found (see Figure 8(b)) that a larger aspect ratio cross section results in a larger curvature for a given shape configuration. Also, as illustrated in Figure 8(a), increasing the gap between joints improves the maneuverability as well, though the effect saturates for larger gap values. Thus for improved turning, a slender fish with large gaps between joints is preferred to a rounder body with the individual links close together.

4.3. Hydrodynamically coupled versus decoupled. The swimming of an articulated body in a potential flow has been studied by Radford [22], in which it was assumed that the bodies were dynamically coupled, but hydrodynamically decoupled; that is, the added inertias for each body component are determined assuming it is hydrodynamically isolated from all the other bodies. Our current numerical scheme removes this approximation and accurately computes the added inertias of the system, and thus provides a way of validating the hydrodynamically decoupled assumption.

We compare the curvature plots for both the hydrodynamically coupled and decoupled cases. Figure 9 is a plot of the ω component of the curvature for both cases. The plots are qualitatively similar in many respects: both show relatively small curvature in the areas of shape space corresponding to the fish in an "S" configuration. Likewise, the decoupled case accurately predicts that the curvature is greatest when the fish is in a "C" configuration.



(a) (b)

Figure 6. *Group motion:* (a) *Translation and* (b) *rotation resulting from the prescribed paths in shape space:* $\theta_1(t) = -1.5 + .46\cos(t+\delta)$; $\theta_2(t) = 1.5 - .46\sin(t+\delta)$ *for various values of* δ. *The start and end points of the closed path have no effect on the net rotation but do affect the net holonomy in the translational component. Because the net translational holonomy depends on the starting point of the path, a formula analogous to* (3.5) *is not possible, as the holonomy cannot depend only on the area enclosed by the path.*

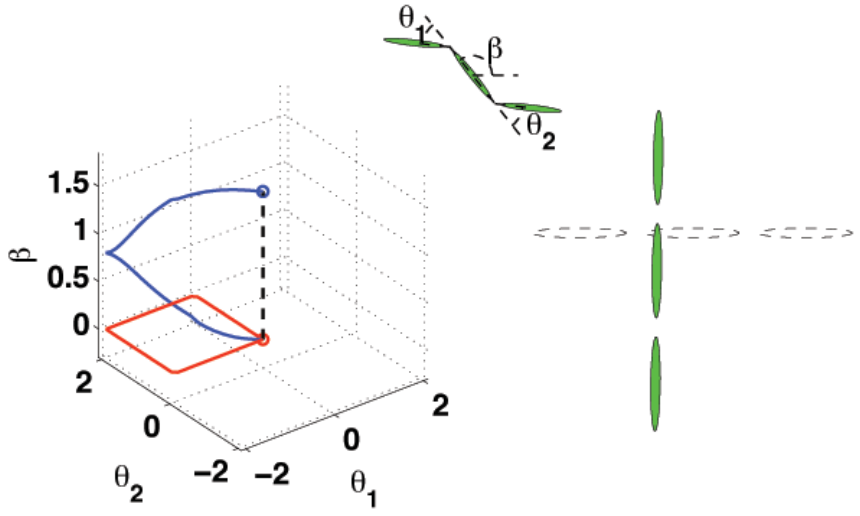
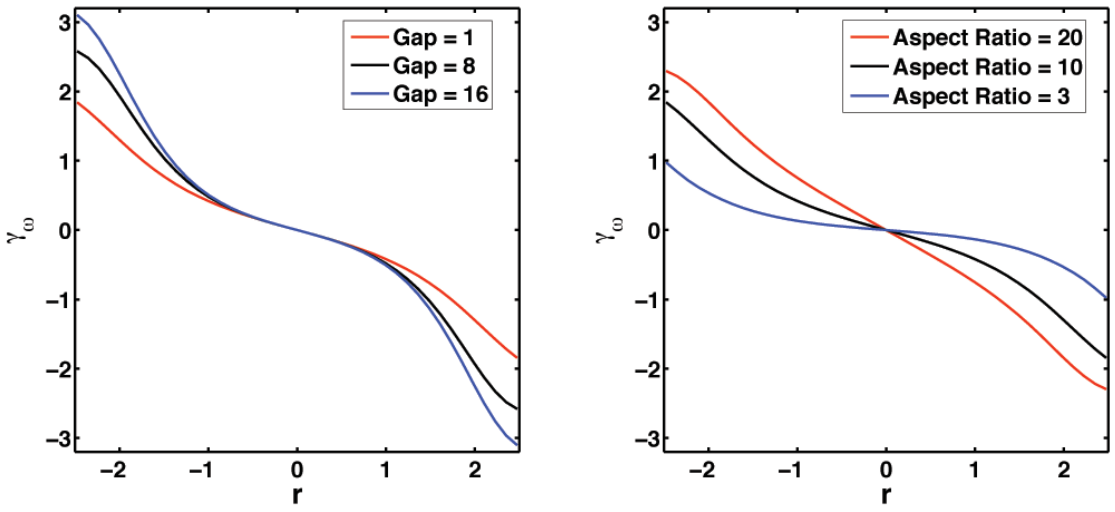


Figure 7. *Turning gait path in shape space in red with the rotation holonomy in blue along with the initial (dotted) and final (solid) configurations for a three-link swimmer after one turning gait. See also the accompanying animation* (64988_02.gif [138KB]).



(a) (b)

Figure 8. *Parametric study of various fish geometries. r is the distance from the origin along the line $\theta_1 = -\theta_2$:* (a) *The gap between links is varied, while the aspect ratio is held fixed at a value of* 10. (b) *The aspect ratio of each link is fixed. The gap and overall length of the fish are held fixed.*

However, the decoupled assumption incorrectly predicts a maximum curvature when the joints are at approximately $\pm\pi/2$. Also, the decoupled assumption underpredicts the curvature in the highest curvature region. However, overall the hydrodynamically decoupled assumption is surprisingly accurate at predicting curvature of the connection, and thus predicting the overall rotation for any given gait.

4.4. Removing the middle link. It has been observed that carangiform fish have developed a narrow necking region in the peduncle, the region anterior of the tail (see Figure 10). Lighthill has proposed that the reduced added mass in this region minimizes the recoil forces experienced by a fish [13]. We crudely model this necking region and reduction in size by removing the middle link and studying the effect on the maneuverability of the fish.

Consider the same three-link, two hinged swimmer as before, but now with the middle link removed, as in Figure 11. Here, the front link can be thought of as representing the dorsal fin, and the rear link as representing the caudal fin in a carangiform fish. (The interactions of these fins in real fishes were considered in the experiments of [2]. For more on undulatory propulsion in carangiform fishes, see [10].) The curvature plots for the two-link swimmer are shown in Figure 12. They are qualitatively similar to those of the three-link swimmer; however they are quantitatively considerably different. The ω-component plot, Figure 12(a), has some characteristics similar to those of the three-link ω curvature plot—most notably a relatively flat surface, except in the two shape space regions corresponding to a "C" type configuration in the three-link body. Unlike the three-link plot, however, the two-link plot has a local maximum in the high curvature region. This is similar to what was found in the case of

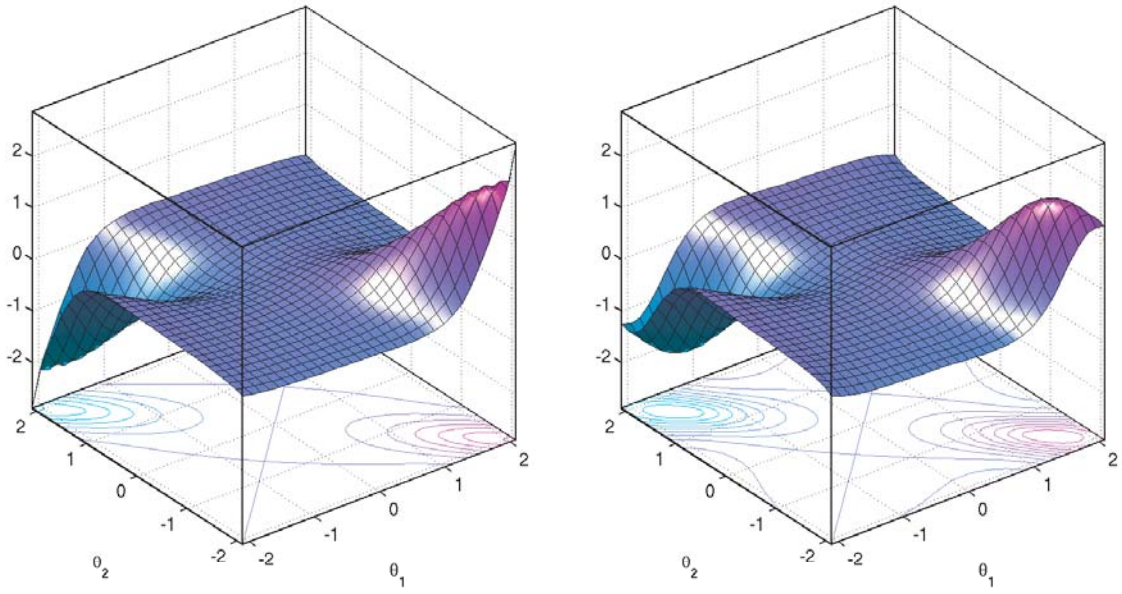
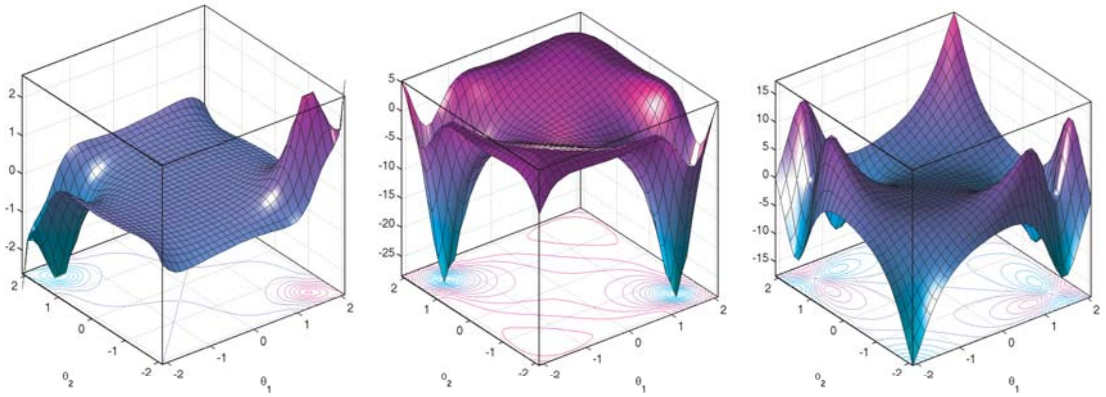


Figure 9. (a) *Hydrodynamically coupled and* (b) *hydrodynamically decoupled* ω *component of curvature plot.*

Figure 10. *Bluegill Sunfish (Lepomis macrochirus), a carangiform mode swimmer. Notice the narrow necking region between the body and tail. (Photo courtesy George Lauder, Harvard University.)*

the hydrodynamically decoupled three-link body. Perhaps the lack of a middle link and thus increased distance between links improves the assumption of hydrodynamic isolation which results in the similarity between these two figures. As a clearer example of this effect, consider the ω-component curvature plots in Figure 13 for the two-link body under the assumption of

Figure 11. *A two-link swimmer. The two links may be viewed as representing the dorsal and caudal fins of a carangiform swimmer.*



(a) (b) (c)

Figure 12. (a) ω, (b) u, *and* (c) v *components of curvature for a two-link, two hinge body assuming hydrodynamically coupled links.*

both hydrodynamically coupled and hydrodynamically decoupled links. The curvature plots for both cases are qualitatively and quantitatively very similar.

Finally, for a more direct comparison, Figure 14 shows the ω curvature component for the three-link body versus that of the two-link body (both under the assumption of hydrodynamically coupled links) along the cross section defined by the line $\theta_1 = -\theta_2$. The curvature value is plotted as a function of distance r from the origin. We see that for a body that is restricted to small angular deformations, the three-link configuration is preferred to the two-link configuration. A body that is able to achieve larger angular displacements of its joints may prefer the two-link configuration over the three-link configuration, as it allows for increased rotational maneuverability for some gaits. For example, for the same path prescribed in (4.6), the net holonomy for the two-link body is 0.99 radians, compared to 0.79 radians for the three-link body.

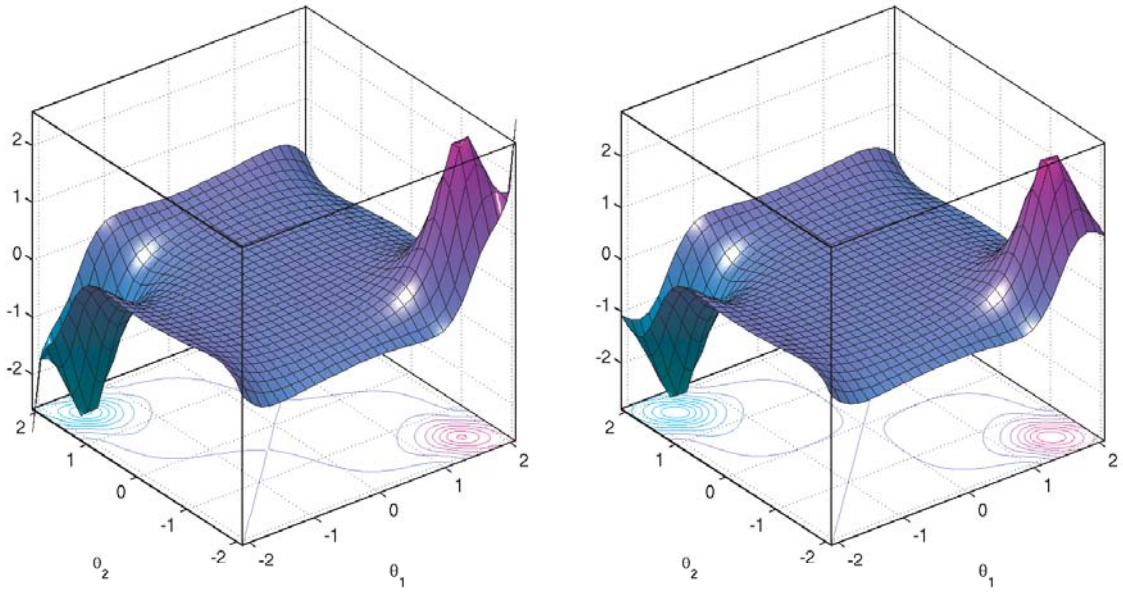


Figure 13. (a) *Hydrodynamically coupled and* (b) *hydrodynamically decoupled* ω *component of curvature for a two-link body.*

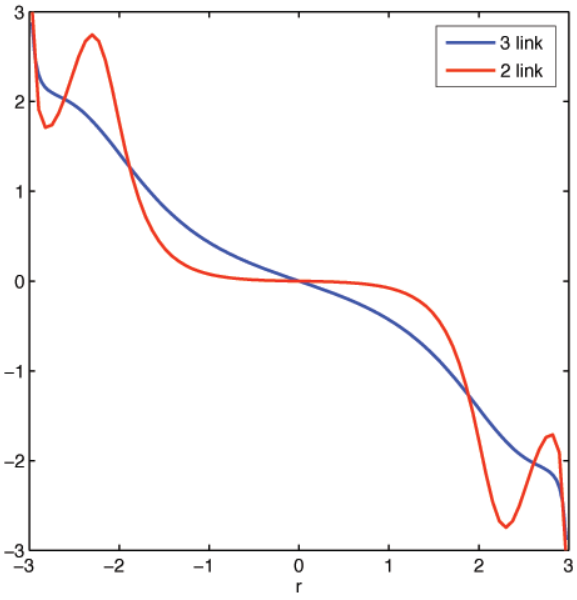


Figure 14. *Comparison of the* ω *curvature component versus distance from the origin along a line defined by* $\theta_1 = -\theta_2$ *for a two-link and three-link body.*

4.5. Trajectory generation. As noted in section 3 and illustrated in Figure 6(a), Theorem 3.1 cannot be applied to the vector space component of a semidirect product group, so curvature plots cannot be used to develop large-amplitude gaits for translation in the plane. However, the curvature plots can be used to develop small-amplitude gaits to achieve local trajectory tracking.

Using the Magnus expansion [16], Radford [23] obtained a local expansion of the group displacement resulting from a closed path α in shape space for systems in the form of (2.11) where the connection $\mathcal{A}$ is a function of the shape variables. The expansion, to third order, is

$$z(\alpha) = -\frac{1}{2}F_{ij}\int_\alpha d\theta^i d\theta^j + \frac{1}{3}(F_{ij,k} - [\mathcal{A}_i, F_{jk}])\int_\alpha d\theta^i d\theta^j d\theta^k + \dots, \tag{4.7}$$

where F_{ij} is the curvature of $\mathcal{A}$,

$$F_{ij} \equiv \mathcal{A}_{j,i} - \mathcal{A}_{i,j} - [\mathcal{A}_i, \mathcal{A}_j], \tag{4.8}$$

which is equivalent to (4.3), and the subscript j represents differentiation with respect to θ_j. F and $\mathcal{A}$ are evaluated at the starting point on the path. The coefficient of the first term is proportional to the local curvature value. Thus, for small-amplitude loops in shape space, the curvature plots can be used to develop gaits. The coefficient of the next term is a function of the local connection and curvature values as well as a higher order derivative of the curvature. This term is nonzero for 2:1 Lissajous-type paths in shape space. Assuming the system is controllable with up to second order bracket motions, (4.7) can be used to determine 1:1 and 2:1 small-amplitude Lissajous-type gaits to achieve the desired trajectory in the group.

Note also that small-time local controllability for the system can be shown from the Lie algebra rank condition. One checks numerically that the vector fields of the local connection plus iterated Lie brackets of those vector fields span the Lie algebra $\mathfrak{se}(2)$ at each point in shape space. (For our three-link swimmer, only the first Lie bracket is required for most points in shape space, and then one checks that second-order brackets span the remaining directions of $\mathfrak{se}(2)$ at degenerate points.) It is then possible to design small-amplitude gaits to achieve group trajectory tracking following a constructive procedure similar to that in [11].

As an example of an application of the curvature plots for non-Abelian Lie groups, we develop a gait to achieve pure net lateral motion, that is, so that the fish swims "sideways." Consider the contour plot of the v component of curvature in Figure 15. Superimposed on this plot are the zero-value contours of both the ω and u curvature components, in blue and magenta, respectively. At the intersection of these two zero-value contours, only the v curvature component is nonzero. A closed path taken about this point, as depicted with a black curve, results in nearly pure net sideways motion of the fish. The net motion will approach a pure lateral direction as the area enclosed by the path approaches zero.

5. Motion planning. Motivated by cross-sectional plots of Figure 14, we consider a family of gaits to develop a method for tracking a desired trajectory. In particular, we focus on gaits in which the displacement per gait period is small compared to the overall desired trajectory displacements. We find that a circular path in shape space about the origin results in pure net forward motion. From the ω curvature plot in Figure 4 and from the cross-sectional plot in Figure 14, we see that shifting the gait away from the origin along the line $\theta_1 = -\theta_2$ increases

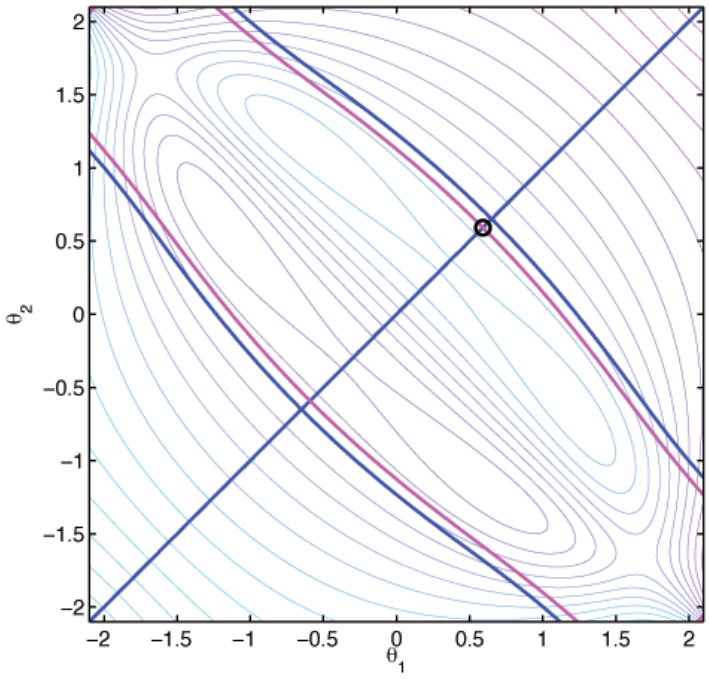


Figure 15. *Contour lines of the v curvature component, with the zero-value contours of the ω and u curvature components in blue and magenta, respectively. A gait represented by a black curve at the intersection of the two zero-value contours results in nearly pure net lateral motion.*

the net rotation. From numerical experiments we find that the net motion is a rotation plus a displacement.

The gait $(\theta_1(t), \theta_2(t))$ is given by the rapidly oscillating complex-valued function

$$\theta(t) = \theta_1(t) + i\theta_2(t) = Ae^{i\omega t} + B(t)(1-i). \tag{5.1}$$

The slowly varying offset B corresponds to shifting the circular trajectory about the line $\theta_1 = -\theta_2$. Here we choose $B(t)$ so that the body follows a desired trajectory in the plane. We hold A, the amplitude of the path, fixed. We perform numerical experiments to compute the displacement δ and rotation β from the forward direction (which is defined as the direction in which the body moves for $B = 0$) over a single period T of oscillation, as functions of B. The average velocity and curvature are given by

$$\langle v\rangle(B) = \frac{\delta}{T} = \frac{\omega\delta(B)}{2\pi}, \qquad \langle \kappa\rangle(B) = \frac{2\sin\left(\frac{\beta(B)}{2}\right)}{\delta(B)}. \tag{5.2}$$

For a fixed value of A, we compute $\langle v\rangle$ and $\langle\kappa\rangle$ for several values of B, a polynomial fit is applied, and finally the inverse function, $B = P(\langle\kappa\rangle)$, is found. We want to follow a specified trajectory, $r(t) = (x(t), y(t))$, where the velocity and curvature along the path are given by

$$v(t) = \sqrt{\dot{x}^2 + \dot{y}^2}, \qquad \kappa(t) = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}. \tag{5.3}$$

Since $B(t)$ is slowly varying (sometimes referred to as an adiabatic approximation), we ignore the contribution of $\dot{B}$ to $\dot{\theta}$. We match $\langle v \rangle(\omega, B) = v(t)$ and $\langle \kappa \rangle(B) = \kappa(t)$ and solve for ω and B as functions of time:

$$B(t) = P(\kappa), \qquad \omega(t) = 2\pi \frac{v}{\delta(B)}. \tag{5.4}$$

Finally, the prescribed path in shape space to approximately track the desired trajectory is

$$\theta(t) = Ae^{i\int_0^t \omega(t')dt'} + B(t)(1-i). \tag{5.5}$$

This method works well only for smoothly varying trajectories where the displacement per gait period is small compared to the dimension of the trajectory. Figure 16 shows the desired path in black along with the actual path in blue due to prescribing open loop controls. An animated video shows the shape deformations as the body follows the desired path.

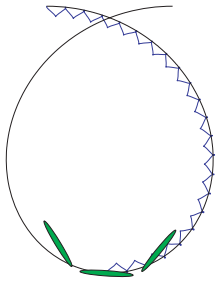


Figure 16. *The desired and actual trajectories in black and blue, respectively, resulting from open-loop controls based on the adiabatic approximation. See also the accompanying animation* (64988_03.gif [841KB]).

6. Conclusions. We have considered the swimming of an articulated fish-like swimming mechanism in a two-dimensional potential flow. We have extended the geometric theory to apply previously known results for systems on Abelian groups to the rotation component of $SE(2)$, exploiting the semidirect product structure. Using plots of the curvature of the local connection form, we developed gaits to achieve a desired rotation for a three-link swimmer. It was found that the most efficient turning gaits correspond to "C"-shaped configurations of the body. A parametric study was performed to assess the degree of maneuverability of different geometries, and it was found that large aspect ratio links with large gaps between the joints result in larger curvature values and are better suited for rotational maneuvering. The curvature plots were used to validate the hydrodynamically decoupled assumption used in previous works and to study the difference between two-link and three-link swimmers. It was shown that removing the middle link can improve the turning gait motion for some gaits. Although the main result is not applicable to non-Abelian groups in general, the curvature plots can be used to develop small-amplitude gaits to follow a desired trajectory. Finally, motivated by the cross section of the ω curvature plot, we considered a family of finite-amplitude gaits and developed a method for tracking a desired trajectory in the plane.

Acknowledgments. We wish to thank Scott Kelly, Jerry Marsden, and Eva Kanso for their insights and conversations on the subject of this paper, and George Lauder for biological inspiration and for providing the photo in Figure 10. We also thank the anonymous referee for the many helpful remarks.

REFERENCES

[1] H. Cendra, J. E. Marsden, and T. S. Ratiu, *Lagrangian reduction by stages*, Mem. Amer. Math. Soc., 152 (2001).
[2] E. G. Drucker and G. V. Lauder, *Locomotor function of the dorsal fin in teleost fishes: Experimental analysis of wake forces in sunfish*, J. Exp. Biol., 204 (2001), pp. 2943–2958.
[3] E. G. Drucker and G. V. Lauder, *Wake dynamics and fluid forces of turning maneuvers in sunfish*, J. Exp. Biol., 204 (2001), pp. 431–442.
[4] E. Kanso, J. E. Marsden, C. W. Rowley, and J. B. Melli-Huber, *Locomotion of articulated bodies in a perfect fluid*, J. Nonlinear Sci., 15 (2005), pp. 255–289.
[5] S. D. Kelly, *The Mechanics and Control of Robotic Locomotion with Applications to Aquatic Vehicles*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 1998.
[6] S. D. Kelly and R. M. Murray, *Modelling efficient pisciform swimming for control*, Internat. J. Robust Nonlinear Control, 10 (2000), pp. 217–241.
[7] G. Kirchhoff, *Ueber die bewegung eines rotationskörpers in einer flüssigkeit*, Crelle, 71 (1869), p. 237.
[8] P. S. Krishnaprasad and R. Yang, *Geometric phases, anholonomy, and optimal movement*, in Proceedings of the IEEE Conference Robotics and Automation, Volume 3, IEEE Robotics and Automation Society, Sacramento, CA, 1991, pp. 2185–2189.
[9] H. Lamb, *Hydrodynamics*, Dover, New York, 1945.
[10] G. V. Lauder and E. D. Tytell, *Hydrodynamics of undulatory propulsion*, in Fish Biomechanics, Fish Physiology 23, R. E. Shadwick and G. V. Lauder, eds., Academic Press, San Diego, 2006, pp. 425–468.
[11] N. E. Leonard and P. S. Krishnaprasad, *Motion control of drift-free, left-invariant systems on Lie groups*, IEEE Trans. Automat. Control, 40 (1995), pp. 1539–1554.
[12] M. J. Lighthill, *Note on the swimming of slender fish*, J. Fluid Mech., 9 (1960), pp. 305–317.
[13] M. J. Lighthill, *Hydrodynamics of aquatic animal propulsion*, Ann. Rev. Fluid Mech., 1 (1969), pp. 413–446.
[14] M. J. Lighthill, *Aquatic animal propulsion of high hydromechanical efficiency*, J. Fluid Mech., 44 (1970), pp. 265–301.
[15] M. J. Lighthill, *Large-amplitude elongated-body theory of fish locomotion*, Proc. Roy. Soc. London B, 179 (1971), pp. 125–138.
[16] W. Magnus, *On the exponential solutions of differential equations for a linear operator*, Comm. Pure Appl. Math., 7 (1954), pp. 649–673.
[17] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Texts Appl. Math. 17, 2nd ed., Springer-Verlag, New York, 1999.
[18] J. E. Marsden, T. Ratiu, and A. Weinstein, *Semidirect products and reduction in mechanics*, Trans. Amer. Math. Soc., 281 (1984), pp. 147–177.
[19] J. E. Marsden, *Lectures on Mechanics*, London Math. Soc. Lecture Note Ser. 174, Cambridge University Press, Cambridge, UK, 1990.
[20] J. E. Marsden, R. Montgomery, and T. S. Ratiu, *Reduction, symmetry, and phases in mechanics*, Mem. Amer. Math. Soc., 88 (436) (1990).
[21] R. Mason and J. W. Burdick, *Propulsion and control of deformable bodies in an ideal fluid*, in Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, 1999, pp. 773–780.
[22] J. E. Radford, *Symmetry, Reduction and Swimming in a Perfect Fluid*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2003.

[23] J. E. Radford and J. W. Burdick, *Local motion planning for nonholonomic control systems evolving on principal bundles*, in Proceedings of the International Symposium on Mathematical Theory of Networks and Systems, Padova, Italy, 1998.
[24] T. Y. Wu, *Swimming of a waving plate*, J. Fluid Mech., 10 (1961), pp. 321–344.
[25] T. Y. Wu, *Hydrodynamics of swimming propulsion. Part* 1. *Swimming of a two-dimensional flexible plate at variable forward speeds in an inviscid fluid*, J. Fluid Mech., 46 (1971), pp. 337–355.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 670–698

Neural Fields with Distributed Transmission Speeds and Long-Range Feedback Delays*

Fatihcan M. Atay† and Axel Hutt‡

Abstract. We introduce distributed axonal transmission speeds and a long-range constant feedback loop into the standard neural field model. We analyze the stability of spatially homogeneous equilibrium solutions for general connectivity kernels. By studying reduced models based on the assumption of small delays, we determine the effects of the delays on the stability and bifurcations. We show in a reduced model that delayed excitatory feedback generally facilitates stationary bifurcations and Turing patterns, while suppressing the bifurcation of periodic solutions and traveling waves. The reverse conclusion holds for inhibitory feedback. In case of oscillatory bifurcations, the variance of the distributed propagation and feedback delays affects the frequency of periodic solutions and the phase speed of traveling waves. Moreover, we give a nonlinear analysis of traveling fronts and find that distributed transmission speeds can maximize the front speed.

Key words. synaptic networks, nonlocal interaction, distributed delays, spatio-temporal patterns, traveling fronts

AMS subject classifications. 34K99, 37N25, 92C20

DOI. 10.1137/050629367

1. Introduction.

Several models of coupled neurons have recently attracted much attention, including topological neural networks [25, 34, 42, 47, 55] and networks involving spatial structures [1, 6, 18, 23, 27, 31, 59, 61]. This paper considers the latter networks describing a continuous synaptically coupled neural field extended in space, whose dynamics are governed by an integro-differential equation. We expand this classical model by incorporating two novel elements suggested by experimental evidence, namely a distribution of signal transmission speeds within the field, and a long-range feedback term involving distributed time delays. These elements constitute two sources of temporal delays with different character and have significant consequences for the dynamics of the neural ensemble. The present work studies their effects on the dynamics through the stability of equilibria, the bifurcations leading to spatial patterns and oscillations, and the analysis of traveling fronts. The measured activity of the neural system indicates that such spatio-temporal dynamics are intimately related to various brain functions and cognition. We mention the space-time instabilities during several types of hallucinations [5] and epileptic seizures [40, 41], the stimulus-evoked traveling waves in turtle visual cortex [48], and the evoked traveling pulses in somatosensory brain slices [46].

*Received by the editors April 17, 2005; accepted for publication (in revised form) by B. Ermentrout May 28, 2006; published electronically December 1, 2006.
http://www.siam.org/journals/siads/5-4/62936.html
†Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany (atay@member.ams.org).
‡Humboldt University at Berlin, Newtonstr. 15, 12489 Berlin, Germany (axel.hutt@physik.hu-berlin.de). This author has obtained financial support from the Deutsche Forschungsgemeinschaft (grant SFB-555).

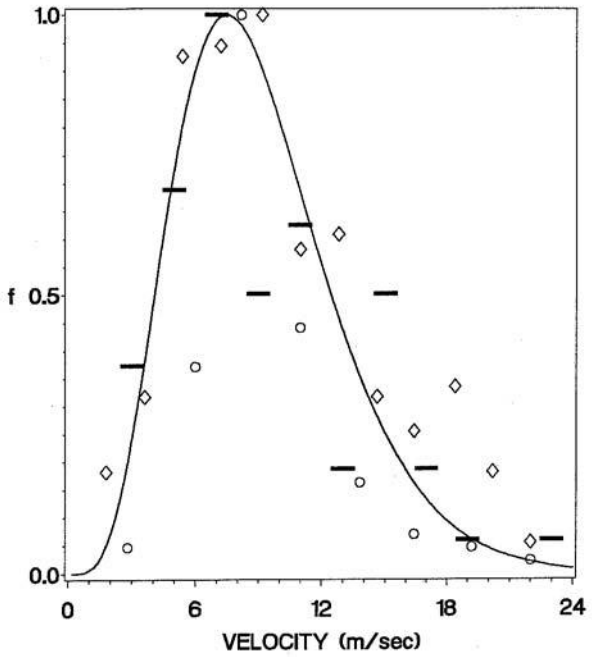


Figure 1. *Estimated distribution of propagation velocity in myelinated cortico-cortical fibers (after Nunez* [44, Figure 2-6, p. 85]*, reprinted by permission of Oxford University Press, Inc.; data originally reviewed in* [35]*). Different symbols refer to data taken from different parts of the brain.*

The mathematical analysis is thus an important step in understanding the relation between neural processes and the model parameters.

The spatial structure in most cortical areas is not known in detail; therefore, the spatial field is usually assumed to be homogeneous, with an underlying arbitrary spatial connectivity. On the other hand, the temporal structure of neural activity is well studied. Multiple time scales are present in real neural areas, which may originate from the delayed impulse response of chemical synapses [20], the finite axonal speed [44], or the presence of feedback loops [56]. To be more specific, finite axonal transmission speed results in a space-dependent delay between two distant locations; i.e., the neural field shows retarded interaction. Moreover, experimental studies reveal not a single axonal speed but statistically distributed speeds in cortico-cortical connections in rats [44] and in intracortical connections in visual cortex of cats and monkeys [7, 16]. In all studies, the histogram of axonal speeds follows a gamma-distribution with maxima between 5 m/s and 12 m/s in rats and at about 0.2 m/s in the cat and monkey brain (Figure 1). To our best knowledge, such distributions of propagation speeds have not been considered yet in dynamical neural models.

In addition, feedback loops play an important role in real neural processing [53]. For instance, detailed studies of interhemispheric connections [17], connections of the visual areas of different functional order [8, 43], and reciprocal thalamic connections [14, 45, 57] indicate interareal feedback loops. In sum, most experimental findings indicate a network of interacting neural areas [52], which allows the brain to accomplish complex tasks, such as motor

coordination or visual feature binding [54]. Due to this network structure, some theoretical studies examined the effects of delayed self-interaction and the interactions of different neural areas [32]. In this context, a well-known example is corticothalamic feedback connections [45], which have been modeled recently by a simple constant feedback delay [49, 51]. The present work extends these previous studies by introducing arbitrary nonlocal feedback kernels; in other words, we consider the case in which the distance between the initial and the final spatial locations of the loop obeys an arbitrary probability density distribution. Furthermore, the time delays in the loop are also described by a general distribution function.

In the following sections, we study the stability of equilibrium solutions of neural fields subject to two sources of temporal delays, namely the transmission delays and the delays in the nonlocal feedback loops. These two delay types turn out to have quite different characters, which are reflected in their dynamical effects. Our aim is to draw general conclusions about the dynamics of the model without reference to specific choices of the connectivity, transmission speed, or feedback delay distributions. The model is introduced in section 2, and the stability of spatially homogeneous equilibria is analyzed in section 3 for arbitrary kernels. In section 4 we introduce approximate models under the assumption of large but finite axonal speeds and small feedback delays. Based on such a model, in section 4.2 we are able to give further details of the effects of the delays on stability and bifurcations. The theoretical analysis is supplemented by numerical simulations in section 5 for some commonly used choices of the connectivity and delay distribution functions. Section 6 treats traveling fronts. The discussion in section 7 closes the work.

2. Distributed speeds and nonlocal feedback. This section aims to motivate the model equation [2, 30]. A detailed review of the derivation of the basic model can be found in, e.g., [6, 28]. Here we additionally introduce the distribution of transmission speeds and the delayed long-range feedback loop.

An essential ingredient of neural activity is the input-output behavior of synapses which convert incoming pulses to postsynaptic potentials. In the coarse-grained population model, at time t and some point x in the field, ensembles of excitatory (e) and inhibitory (i) chemical synapses respond to incoming pulse activity $P_{e,i}(x,t)$ by temporal convolution with some impulse response function $h(t)$. Specifically,

$$\begin{aligned} V_{e,i}(x,t) &= g_{e,i} \int_{-\infty}^{t} h(t-t')\ P_{e,i}(x,t')\,dt' \\ &= g_{e,i}\mathcal{I} \circ P_{e,i}(x,t), \end{aligned} \tag{2.1}$$

where $g_{e,i}$ denotes the efficacy of excitatory and inhibitory synapses, $\mathcal{I}$ is the integral operator, and $V_{e,i}$ denotes the excitatory and inhibitory postsynaptic potential. The presynaptic pulse activity $P_{e,i}(x,t)$ originates from the somatic pulse activity $P_k^s(x,t)$ of distant neural ensembles of neuron type k and is given by

$$\begin{aligned} P_{e,i}(x,t) = g_0 \int_0^\infty g(v) \int_\Omega K_{e,i}(x-x') P_k^s \left(x', t - \frac{|x-x'|}{v} \right) dx' dv \\ + \bar{\mu}_{e,i} \int_0^\infty f(\tau) \int_\Omega F_{e,i}(x-x') P_k^s(x', t-\tau)\,dx' d\tau + E_{e,i}(x,t). \end{aligned} \tag{2.2}$$

Here, Ω represents a spatial domain, which is taken to be the real line in this paper. The kernels $K_e(x)$ and $K_i(x)$ represent the spatial distributions for excitatory and inhibitory synaptic connections, respectively, while $F_e(x)$ and $F_i(x)$ give the distributions of excitatory and inhibitory long-range delayed feedback connections, respectively. The factors $\bar{\mu}_e, \bar{\mu}_i$ weight the corresponding feedback connections, and $g_0 > 0$ weights the interareal contribution towards the delayed feedback distribution. Moreover, the somatic conversion of the effective membrane potential $V = V_e - V_i$ to ensemble pulse activity P^s can be modeled by a sigmoidal function $P^s_k(x,t) = S(V(x,t) - V_0)$ with the mean firing threshold V_0 [21, 28]. The present model does not distinguish different neuron types and thus presumes $P^s_k(x,t) = P^s(x,t)$; though the relaxation of this constraint may give additional insights into neural dynamics [37], its discussion would exceed the aim of the present work. The sigmoidal function $S : \mathbb{R} \to \mathbb{R}^+$ is differentiable and monotone increasing. The terms E_e and E_i represent the external stimuli terminating at excitatory and inhibitory synapses, respectively. Equation (2.2) incorporates several realistic mechanisms as sources of time delays in the neural field, including the signal propagation delays within the field and the delays in the long-range feedback loop. The finite speed v of signal propagation within the field gives rise to distance-dependent delays; such delays have been discussed in some recent studies [10, 13, 30, 32, 46]. As a novel aspect, in this paper we relax the constraint of a single transmission speed v and consider more naturally a distribution function of speeds $g(v)$. As mentioned in the introduction, the motivation for distributed transmission speeds originates from experimental findings. In addition, we take into account the feedback delays τ in the feedback loop, whose distribution is given by the function f. In contrast to the transmission speeds, only a few details on the distribution of feedback delays are known from experiments; we mention the corticothalamic feedback delay [14], the feedback delay between visual areas V1 and V2 in monkeys [24], and the delayed inhibitory response in the retina of the horseshoe crab *Limulus polyphemus* [26]. The distribution functions g and f are probability densities on $[0,\infty)$, i.e., nonnegative functions satisfying

$$\int_0^\infty g(v)\,dv = \int_0^\infty f(\tau)\,d\tau = 1. \tag{2.3}$$

Combining (2.1) and (2.2), we find the equation

$$\begin{aligned} V(x,t) = \int_{-\infty}^{t} h(t-t') \Bigg[\bar{\alpha} \int_0^\infty g(v) \int_{-\infty}^\infty K(z) S(V(x+z, t' - |z|/v))\,dz\,dv \\ + \bar{\beta} \int_0^\infty f(\tau) \int_{-\infty}^\infty F(z) S(V(x+z, t'-\tau))\,dz\,d\tau + E(x,t') \Bigg] dt' \end{aligned} \tag{2.4}$$

with $K = a_e K_e - a_i K_i$, $F = \mu_e F_e - \mu_i F_i$, $a_{e,i} = g_0 g_{e,i}$, $\mu_{e,i} = g_{e,i}\bar{\mu}_{e,i}$ and the external input $E = E_e - E_i$. The kernels $K, F : \mathbb{R} \to \mathbb{R}$ are continuous, integrable, and even; that is, $K(-z) = K(z)$, $F(-z) = F(z)$ for all $z \in \mathbb{R}$. They can have different signs for small and large values of $|z|$, showing different local and lateral contributions of excitation and inhibition. Furthermore, with the help of the nonnegative scaling factors $\bar{\alpha}$ and $\bar{\beta}$, they can be assumed to satisfy the normalization conditions

$$\int_{-\infty}^\infty |K(z)|\,dz = \int_{-\infty}^\infty |F(z)|\,dz = 1. \tag{2.5}$$

By introducing a temporal differentiation operator $L = \mathcal{I}^{-1}$, where $Lh(t) = \delta(t)$ and δ is the Dirac delta function, (2.4) can be written as an integro-differential equation. Thus, h plays the role of the Green's function corresponding to L. In this way we obtain our main equation

$$L\left(\frac{\partial}{\partial t}\right) V(x,t) = \bar{\alpha} \int_0^\infty g(v) \int_{-\infty}^\infty K(z) S(V(x+z, t-|z|/v))\, dz\, dv + \bar{\beta} \int_0^\infty f(\tau) \int_{-\infty}^\infty F(z) S(V(x+z, t-\tau))\, dz\, d\tau + E(x,t). \tag{2.6}$$

As an example for Green's function h, we mention the form suggested by experiments [15], namely $h(t) = (\exp(-t/T_1) - \exp(-t/T_2))/(T_1 - T_2)$ with synaptic time constants $T_1, T_2 > 0$. Letting $T_1 \to T_2 = T$ yields the well-studied case $h(t) = t \exp(-t/T)/T$ (see, e.g., [50]), corresponding to a second order differentiation operator with time scale T. We shall consider this case in sections 4.2 and 5, as well as a first order operator in section 6. However, the general formulation will be given for an arbitrary order operator L, often with the assumption that L is stable, i.e., that all its characteristic values have negative real parts.

We close this section with several comments about the model (2.6). To begin with, the unbounded spatial domain $(-\infty, \infty)$ is physically unrealistic; however, it gives a good approximation when boundary effects are negligible inside the domain. Our results also hold when the domain is a circle, corresponding to a bounded region with periodic boundary conditions, and we use the circle as the spatial domain when we do numerical simulations in section 5. Since an unbounded domain is merely a mathematical convenience, the kernels K and F will typically have compact support, especially when they are experimentally determined, although the model allows arbitrary kernels defined on $(-\infty, \infty)$. In a similar vein, it may be more desirable to formulate the model in terms of v^{-1} instead of v, because the probability density for v^{-1} will have a compact support. Moreover, in terms of v^{-1} it makes sense to discuss the perturbation of the distribution about the Dirac delta at zero, corresponding to instantaneous information transmission. We keep v in (2.6) mainly because it allows slightly more intuitive and less cumbersome notation, and because the experimental measurements of the distribution are given in terms of v. Similarly, the probability density f for the feedback delays will have compact support, making all delays in (2.6) bounded. In summary, the unbounded spatial and temporal domains appearing in the model are mostly for convenience of notation and reduce to finite ranges for physically reasonable choices of distributions. A main point of interest is the different nature of the two delay types in the two terms on the right-hand side of (2.6), which we study in the subsequent sections.

3. Equilibrium solutions. In this section we consider the basics of a stability analysis for spatially uniform equilibria of the model (2.6). The material here is essentially a brief version of sections 2 and 3 of [2], with the purpose of indicating the differences introduced by distributed speeds and nonlocal feedback.

For a constant input $E(x,t) \equiv E^*$, a spatially uniform equilibrium solution $V(x,t) \equiv V^*$ of (2.6) satisfies $V^* = \kappa S(V^*) + E^*$, where

$$\kappa = \bar{\alpha} \int_{-\infty}^\infty K(z)\, dz + \bar{\beta} \int_{-\infty}^\infty F(z)\, dz.$$

It is easy to see that if S is a continuous and increasing function (such as a sigmoid) and $\kappa \leq 0$, then V^* is unique for any $E^* \in \mathbb{R}$, whereas if $\kappa > 0$, then there may be multiple equilibria V^* [2]. The stability of the equilibrium solution V^* is determined by the linear variational equation

$$L\left(\frac{\partial}{\partial t}\right) u(x,t) = \alpha \int_0^\infty g(v) \int_{-\infty}^\infty K(z) u(x+z, t-|z|/v)\, dz\, dv + \beta \int_0^\infty f(\tau) \int_{-\infty}^\infty F(z) u(x+z, t-\tau)\, dz\, d\tau, \tag{3.1}$$

where $u(x,t) = V(x,t) - V^*$ and

$$\alpha = \bar{\alpha} S'(V^*) \geq 0, \quad \beta = \bar{\beta} S'(V^*) \geq 0. \tag{3.2}$$

Note that the values of α and β depend implicitly on the external input E^* through the value of V^*. Using the ansatz $u(x,t) = e^{\lambda t} e^{ikx}$ in (3.1), where $\lambda \in \mathbb{C}$ and $k \in \mathbb{R}$, we obtain the dispersion relation between the temporal and spatial modes as

$$L(\lambda) = \alpha \int_0^\infty g(v) \int_{-\infty}^\infty K(z) \exp(-\lambda |z|/v) \exp(-ikz)\, dz\, dv + \beta \int_0^\infty f(\tau) e^{-\lambda \tau} d\tau \int_{-\infty}^\infty F(z) \exp(-ikz)\, dz. \tag{3.3}$$

Notice that the last term in (3.3) factors as the product of the Laplace transform of f and the Fourier transform of F. By contrast, such a factorization does not arise for the term multiplying α. This difference accounts for the different character of the two delay sources, whose effects are further studied in section 4.2. For now we make some elementary observations regarding the stability of the equilibrium solution. When the characteristic values of the operator L have negative real parts, the zero solution of (3.1) is stable for $\alpha = \beta = 0$. It follows that V^* is stable for small α, β, as quantified by the next theorem.

Theorem 3.1. *Suppose that L is a stable polynomial. If*

$$\alpha + \beta < \min_{\omega \in \mathbb{R}} |L(i\omega)|, \tag{3.4}$$

then V^* is asymptotically stable. In particular, if $L(\lambda) = \lambda + \rho$, $\rho > 0$, then the condition

$$\alpha + \beta < \rho \tag{3.5}$$

is sufficient for the asymptotic stability of V^*. If $L(\gamma) = \lambda^2 + \gamma\lambda + \rho$ with $\gamma, \rho > 0$, then V^* is asymptotically stable, provided that the condition

$$\frac{\gamma^2}{2} > \rho - \sqrt{\rho^2 - (\alpha+\beta)^2} \tag{3.6}$$

holds in addition to (3.5).

Proof. In the ansatz $u(x,t) = e^{\lambda t}e^{ikx}$ let $\lambda = \sigma + i\omega$, where σ and ω are real numbers. We will prove that $\sigma < 0$ if (3.4) holds. Suppose by way of contradiction that (3.4) holds but $\sigma \geq 0$. From the dispersion relation (3.3),

$$\begin{aligned}
|L(\sigma+i\omega)| &= \Big|\alpha \int g(v) \int_{-\infty}^{\infty} K(z)\exp(-(\sigma+i\omega)|z|/v)\exp(-ikz)\,dz\,dv \\
&\qquad + \beta \int_0^{\infty} f(\tau)\exp(-(\sigma+i\omega)\tau)\,d\tau \int_{-\infty}^{\infty} F(z)\exp(-ikz)\,dz\Big| \\
&\leq \alpha \int g(v) \int_{-\infty}^{\infty} |K(z)|\,|\exp(-\sigma|z|/v)|\,dz\,dv + \beta \int_0^{\infty} f(\tau)d\tau \int_{-\infty}^{\infty} |F(z)|\,dz \\
&\leq \alpha \int_{-\infty}^{\infty} |K(z)|dz + \beta \int |F(z)|\,dz \\
&= \alpha + \beta, \qquad (3.7)
\end{aligned}$$

where we have used (2.3) and (2.5). On the other hand, by Lemma 2.2 in [2],

$$|L(i\omega)| \leq |L(\sigma+i\omega)|$$

for all $\sigma \geq 0$ and $\omega \in \mathbb{R}$, which together with (3.7) implies

$$|L(i\omega)| \leq \alpha + \beta.$$

This, however, contradicts (3.4). Thus $\sigma > 0$, and the first statement of the theorem is proved. The remainder is similar to the proof of Theorem 2.1 in [2]. ∎

The change of stability as the parameters α, β are increased beyond the quantity given in (3.4) is characterized by the existence of an eigenvalue $\lambda = i\omega$, $\omega \in \mathbb{R}$. The dispersion relation (3.3) then implies that

$$\begin{aligned}
L(i\omega) &= \alpha \int_0^{\infty} g(v) \int_{-\infty}^{\infty} K(z)\exp(-i(kz+\omega|z|/v))\,dz\,dv \\
&\qquad + \beta \int_0^{\infty} f(\tau)e^{-i\omega\tau}d\tau \int_{-\infty}^{\infty} F(z)\exp(-ikz)\,dz. \qquad (3.8)
\end{aligned}$$

The bifurcating solutions near such critical cases can be qualitatively classified as stationary or oscillatory, depending on whether $\omega = 0$ or $\omega \neq 0$, respectively, and as spatially homogeneous or inhomogeneous, depending on whether $k = 0$ or $k \neq 0$, respectively, as summarized in Table 1.

Table 1
Classification of possible local bifurcations of equilibria.

	Stationary ($\omega = 0$)	Oscillatory ($\omega \neq 0$)
Homogeneous ($k = 0$)	Spatially constant equilibrium	Global oscillations
Inhomogeneous ($k \neq 0$)	Spatial patterns	Traveling waves

For stationary bifurcations which are spatially homogeneous ($\omega = k = 0$), the bifurcating solution is also an equilibrium. From (3.8), this case occurs when

$$L(0) = \alpha \int_{-\infty}^{\infty} K(z)\, dz + \beta \int_{-\infty}^{\infty} F(z)\, dz. \tag{3.9}$$

Suppose $L(0) > 0$. If $\beta = 0$, (3.9) can be satisfied only when $\int_{-\infty}^{\infty} K(z)\, dz > 0$, that is, when the connectivity is dominantly excitatory. On the other hand, if $\beta > 0$, such a bifurcation can occur even with inhibitory connectivity kernels, provided that $\int_{-\infty}^{\infty} F(z)\, dz > 0$. Hence, excitatory nonlocal feedback can induce this type of bifurcation even when inhibitory connections dominate the field. Conversely, inhibitory feedback can prevent this bifurcation from occurring.

The stationary bifurcations which are spatially inhomogeneous ($\omega = 0$ and $k \neq 0$) yield spatial patterns, also called Turing modes. The condition for their appearance is given by (3.8) as

$$\begin{aligned} L(0) &= \alpha \int_{-\infty}^{\infty} K(z) \exp(-ikz)\, dz + \beta \int_{-\infty}^{\infty} F(z) \exp(-ikz)\, dz \\ &= \alpha \hat{K}(k) + \beta \hat{F}(k), \quad k \neq 0, \end{aligned} \tag{3.10}$$

where $\hat{K}$ and $\hat{F}$ denote the Fourier transforms of K and F, respectively. Reasoning as in the above paragraph, we conclude that excitatory nonlocal feedback can induce stationary bifurcations and patterns, whereas inhibitory feedback can stabilize the equilibrium against such bifurcations.

By (3.9) and (3.10), stationary bifurcations are independent of the propagation speeds and the feedback delays, as well as the temporal differentiation operator L (except through $L(0)$). The oscillatory bifurcations, however, are very much influenced by the delays and may even be absent if delay effects are small. Indeed, in terms of the mean feedback delay $E[\tau] = \int_0^{\infty} \tau f(\tau)\, d\tau$ and the mean propagation delay τ_p, defined by

$$\tau_p = \underbrace{\int_0^{\infty} \frac{g(v)}{v}\, dv}_{1/\bar{v}} \underbrace{\int_{-\infty}^{\infty} |zK(z)|\, dz}_{\xi} \tag{3.11}$$

(where the first integral is the mean value of v^{-1} and ξ is a characteristic interaction distance within the field), the following estimate can be derived.

Theorem 3.2. *Suppose $L(\lambda) = \eta\lambda^2 + \gamma\lambda + \rho$, where $\eta \geq 0$ and $\gamma, \rho \in \mathbb{R}$. If*

$$\alpha\tau_p + \beta E[\tau] < |\gamma|, \tag{3.12}$$

then (3.1) has no solutions of the form $u(x,t) = \exp i(\omega t + kx)$ with ω real and nonzero.

Proof. From (3.8),

$$\begin{aligned} L(\lambda) = \alpha \int_0^{\infty} g(v) \int_{-\infty}^{\infty} K(z) \exp(-\lambda|z|/v) \cos kz\, dz\, dv \\ + \beta \int_0^{\infty} f(\tau) e^{-\lambda\tau} d\tau \int_{-\infty}^{\infty} F(z) \cos(kz)\, dz \end{aligned}$$

since the functions $F(z)$ and $K(z)\exp(-\lambda|z|/v)$ are even in z. The imaginary part of the above expression at the value $\lambda = i\omega$ is

$$\begin{aligned}\operatorname{Im} L(i\omega) = &-\alpha \int_0^\infty \int_{-\infty}^\infty g(v)K(z)\sin(\omega|z|/v)\cos(kz)\,dz \\ &-\beta \int_0^\infty f(\tau)\sin(\omega\tau)\,d\tau \int_{-\infty}^\infty F(z)\cos(kz)\,dz. \end{aligned} \tag{3.13}$$

If $L(\lambda) = \eta\lambda^2 + \gamma\lambda + \rho$, then $\operatorname{Im} L(i\omega) = \gamma\omega$, and comparing the magnitudes in (3.13) gives

$$\begin{aligned}|\gamma\omega| &\le \alpha \int_0^\infty \int_{-\infty}^\infty g(v)|K(z)\sin(\omega z/v)|\,dz\,dv + \beta \int_0^\infty f(\tau)|\sin\omega\tau|\,d\tau \int_{-\infty}^\infty |F(z)|\,dz \\ &\le \alpha \int_0^\infty \int_{-\infty}^\infty g(v)|K(z)\omega z/v|\,dz\,dv + \beta|\omega| \int_0^\infty \tau f(\tau)\,d\tau,\end{aligned}$$

where we have used the normalization (2.5) and the estimate $|\sin(x)| \le |x|$, $x \in \mathbb{R}$. If $\omega \neq 0$, then ω may be canceled to yield

$$|\gamma| \le \alpha \int_0^\infty \frac{g(v)}{v}\,dv \int_{-\infty}^\infty |zK(z)|\,dz + \beta \int_0^\infty \tau f(\tau)\,d\tau = \alpha\tau_p + \beta E[\tau].$$

This, however, contradicts the assumption (3.12). Hence $\omega = 0$, which now proves the theorem. ∎

We note that the mean propagation delay τ_p can be estimated for cortico-cortical connections in humans [35, 58], and we find $\bar{v} \approx 8$ ms^{-1}, $\xi \approx 0.10$ m, leading to $\tau_p \approx 12$ ms. Similarly, for the corticothalamic loop in humans $E[\tau] \approx 120$ ms [14], while the feedback loop between V1 and V2 exhibits $E[\tau] \approx 2$ ms [24]. Such estimates together with Theorem 3.2 give a practical means of determining the possibility of oscillatory bifurcations of spatially uniform equilibria. The importance of oscillatory bifurcations stems from the prevalence of oscillatory activity in neural systems. For instance, a spatially homogeneous and oscillatory bifurcation ($\omega \neq 0$ and $k = 0$) corresponds to spatially uniform, or synchronous, oscillations, whereas the spatially inhomogeneous case ($\omega \neq 0$ and $k \neq 0$) corresponds to traveling waves, with wave speed given by ω/k. The analysis of these bifurcations by solving (3.8), however, is not straightforward. Obtaining general results without making specific assumptions about the kernels is particularly difficult. In the next section we will use an approximation scheme to draw qualitative conclusions for general classes of connectivity and feedback kernels and delay distributions. Based on this analysis, we will show in section 4.2 that the stabilizing effect of inhibitory connectivity and feedback against stationary bifurcations is accompanied by their tendency to induce oscillatory bifurcations when delays are present. Similarly, excitatory feedback, which can induce stationary bifurcations, can also prevent oscillatory bifurcations. We are thus able to connect the effects of field connectivity and nonlocal feedback with the effects of delays.

4. Reduced models.

We now investigate the bifurcations of equilibria in reference to the two novel aspects of the neural field introduced in section 2—namely, distributed propagation speeds and feedback delays. Our aim is to obtain an analytical understanding of the general

dynamical features of the model, without recourse to specific forms for the connectivity and feedback kernels as well as speed and delay distributions. To make the problem tractable under such generalities, we study an approximate model for which stability and bifurcations can be analytically calculated and the effects of delays can be determined. Based on this analysis, we make qualitative predictions and obtain parameter values which serve as the starting point for numerical investigation of the full model. We return to the original model in section 5 to numerically verify the predictions for particular choices of kernels.

4.1. Series approximation. To investigate the stability and bifurcations of the equilibrium solution, we introduce the power series expansions

$$u(x+z,t-|z|/v)=\sum_{m=0}^{\infty}\frac{(-|z|/v)^m}{m!}\frac{\partial^m}{\partial t^m}u(x+z,t),$$
$$u(x+z,t-\tau)=\sum_{m=0}^{\infty}\frac{(-\tau)^m}{m!}\frac{\partial^m}{\partial t^m}u(x+z,t)$$

into the linearized equation (3.1) and obtain

$$\begin{aligned}L\left(\frac{\partial}{\partial t}\right)u(x,t)=\alpha\int_{-\infty}^{\infty}\sum_{m=0}^{\infty}\frac{(-1)^m E[v^{-m}]}{m!}|z|^m K(z)\frac{\partial^m}{\partial t^m}u(x+z,t)\,dz\\ +\beta\int_{-\infty}^{\infty}\sum_{m=0}^{\infty}\frac{(-1)^m E[\tau^m]}{m!}F(z)\frac{\partial^m}{\partial t^m}u(x+z,t)\,dz,\end{aligned}\tag{4.1}$$

where $E[\tau^m]=\int_0^\infty \tau^m f(\tau)\,d\tau$ and $E[v^{-m}]=\int_0^\infty v^{-m}g(v)\,dv$ are the expected values of τ^m and v^{-m}, respectively. Now consider the case when the above equation can be well approximated by using the first $N+1$ terms of the infinite series. This assumption is justified when the delays arising from signal propagation and nonlocal feedback are not too large and their distributions are sufficiently concentrated near their mean values, and when the connection kernel K decays sufficiently fast away from the origin [2]. For example, K is typically of exponential order; that is, there exist positive constants κ_1 and κ_2 such that

$$|K(z)|\le\kappa_1\exp(-\kappa_2|z|),\quad z\in\mathbb{R}.$$

Then, using the fact that K is even,

$$\frac{1}{m!}\int_{-\infty}^{\infty}|z^m K(z)|\,dz\le\frac{2\kappa_1}{m!}\int_0^\infty z^m\exp(-\kappa_2 z)\,dz$$
$$=2\kappa_1\kappa_2^{-(m+1)}\frac{\Gamma(m+1)}{m!}=2\kappa_1\kappa_2^{-(m+1)}.$$

So if $\kappa_2>1$, the contribution of the terms $|z|^m K(z)/m!$ to the integral in (4.1) becomes negligible for large m.

We thus have the reduced model for the neural field near the equilibrium solution

$$L\left(\frac{\partial}{\partial t}\right)u(x,t) = \alpha\int_{-\infty}^{\infty}\sum_{m=0}^{N}\frac{(-1)^m E[v^{-m}]}{m!}|z|^m K(z)\frac{\partial^m}{\partial t^m}u(x+z,t)\,dz$$
$$+\beta\int_{-\infty}^{\infty}\sum_{m=0}^{N}\frac{(-1)^m E[\tau^m]}{m!}F(z)\frac{\partial^m}{\partial t^m}u(x+z,t)\,dz. \tag{4.2}$$

The corresponding dispersion relation has the form

$$L(\lambda) = \sum_{m=0}^{N}\frac{(-1)^m\lambda^m}{m!}\left(\alpha E[v^{-m}]\hat{K}_m(k) + \beta E[\tau^m]\hat{F}(k)\right), \tag{4.3}$$

where

$$\hat{K}_m(k) = \int_{-\infty}^{\infty}|z|^m K(z)e^{-ikz}\,dz$$

denotes the Fourier transforms of the moments of K. We also use the usual notation $\hat{K}$ for $\hat{K}_0$. All the transforms $\hat{K}_m$ as well as $\hat{F}$ are real-valued since the kernels K and F are even functions. The next theorem gives conditions for the stability of the zero solution of (4.2). Recall that the leading principal minors of a $q\times q$ real matrix $A=[a_{ij}]$ are the q real numbers

$$a_{11}, \det\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}, \det\begin{pmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}, \ldots, \det(A).$$

Theorem 4.1. *Let $L(\lambda)=\sum_{m=0}^{n} l_m\lambda^m$, and suppose that the pair (λ,k) satisfies (4.3). Let*

$$c_m(k) = l_m - b_m(k),$$

where

$$b_m(k) = \begin{cases} \frac{(-1)^m}{m!}\left(\alpha E[v^{-m}]\hat{K}_m(k) + \beta E[\tau^m]\hat{F}(k)\right), & m=0,1,\ldots,N,\\ 0, & m>N.\end{cases}$$

Let

$$q = \max\{m : c_m(k)\neq 0\},$$

and define the $q\times q$ real matrix

$$Q(k) = \operatorname{sign}(c_q)\cdot\begin{pmatrix} c_{q-1} & c_q & 0 & \cdots & 0\\ c_{q-3} & c_{q-2} & c_{q-1} & \cdots & 0\\ c_{q-5} & c_{q-4} & c_{q-3} & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & 0\\ 0 & 0 & \cdots & c_1 & c_0\end{pmatrix}.$$

Then $\operatorname{Re}\lambda \geq 0$ *if and only if some leading principal minor of* $Q(k)$ *is nonpositive. Consequently, the zero solution of* (4.2) *is asymptotically stable if and only if all leading principal minors of* $Q(k)$ *are positive for all* k.

Proof. The relation (4.3) can be viewed as a qth order polynomial equation in λ, where the coefficients depend on k. We thus write (4.3) as

$$p_k(\lambda) = \sum_{m=0}^{q} c_m(k)\lambda^m = 0,$$

where $c_m(k)$ are as in the statement of the theorem. The stability of $p_k(\lambda)$ is given by the Routh–Hurwitz criteria, which can be expressed as the leading principal minors of $Q(k)$ being positive; see, e.g., [3]. ∎

Theorem 4.1 allows us to determine the temporal stability of the spatial mode corresponding to a particular wave number k. In this way, it is possible to systematically investigate the bifurcation structure and its dependence on the parameters. As an application, we next analyze a special case where the stability and bifurcations can be completely characterized and a detailed description of the effects of delays can be given.

4.2. Effects of delays in a second order model. The reduced model (4.2) with $N = 0$ is equivalent to neglecting all delays in the original field model. Taking $N = 1$, on the other hand, gives an equation that depends only on the mean values of the delays and not on their distributions. Hence, the lowest order model which exhibits the effects of distributed speeds and feedback delays is obtained with $N = 2$. We now consider the case $N = 2$ in some detail to gain insight into the effects of speed and delay distributions. For definiteness, we focus on the second order temporal differential operator

$$L\left(\frac{\partial}{\partial t}\right) = \frac{\partial^2}{\partial t^2} + \gamma\frac{\partial}{\partial t} + \rho, \quad \gamma, \rho > 0, \tag{4.4}$$

which has been considered by several authors in models of neural field dynamics, e.g., [6, 28, 30, 50, 60]. As we will make use of the approximation scheme introduced above, the results here hold for sufficiently large transmission speeds. The analysis is based on the following result, which essentially follows from Theorem 4.1.

Theorem 4.2. *Consider* (4.2) *with* $N = 2$ *and* L *given by* (4.4). *Then the following hold:*

1. *(stability) Let* (λ, k) *be a solution of the dispersion relation* (4.3) *such that*

$$\alpha E[v^{-2}]\hat{K}_2(k) + \beta E[\tau^2]\hat{F}(k) < 2. \tag{4.5}$$

Then $\operatorname{Re}(\lambda) < 0$ *if and only if the conditions*

$$\rho - \alpha\hat{K}(k) - \beta\hat{F}(k) > 0 \tag{4.6}$$

and

$$\gamma + \alpha E[v^{-1}]\hat{K}_1(k) + \beta E[\tau]\hat{F}(k) > 0 \tag{4.7}$$

are satisfied. The conclusion remains valid if all inequalities in (4.5)–(4.7) *are simultaneously reversed. When* $\alpha E[v^{-2}]\hat{K}_2(k) + \beta E[\tau^2]\hat{F}(k) = 2$, $\operatorname{Re}(\lambda) < 0$ *if and only if the left-hand sides of* (4.6) *and* (4.7) *have the same (nonzero) sign.*

2. *(stationary bifurcations) The pair* (λ, k) *with* $\lambda = 0$ *satisfies* (4.3) *if and only if*

$$\rho - \alpha\hat{K}(k) - \beta\hat{F}(k) = 0. \tag{4.8}$$

3. *(oscillatory bifurcations) The pair* (λ, k) *with* $\lambda = i\omega$, $\omega \in \mathbb{R}\backslash\{0\}$, *satisfies* (4.3) *if and only if*

$$\gamma + \alpha E[v^{-1}]\hat{K}_1(k) + \beta E[\tau]\hat{F}(k) = 0 \tag{4.9}$$

and

$$\omega^2 = \frac{\rho - \alpha\hat{K}(k) + \beta\hat{F}(k)}{2 - \alpha E[v^{-2}]\hat{K}_2(k) - \beta E[\tau^2]\hat{F}(k)} > 0. \tag{4.10}$$

Proof. With $N = 2$ and L as given in the statement of the theorem, (4.3) is a quadratic equation in λ:

$$p_k(\lambda) = c_2(k)\lambda^2 + c_1(k)\lambda + c_0(k) = 0, \tag{4.11}$$

where

$$\begin{aligned} c_0(k) &= \rho - \alpha\hat{K}(k) - \beta\hat{F}(k), \\ c_1(k) &= \gamma + \alpha E[v^{-1}]\hat{K}_1(k) + \beta E[\tau]\hat{F}(k), \\ c_2(k) &= 1 - \frac{1}{2}\alpha E[v^{-2}]\hat{K}_2(k) - \frac{1}{2}\beta E[\tau^2]\hat{F}(k). \end{aligned}$$

It is easy to see that, in case $c_2(k) \neq 0$, the roots of p_k have negative real parts if and only if all the coefficients $c_m(k)$ are nonzero and have the same sign. Furthermore, $p(0) = 0$ if and only if $c_0(k) = 0$, and $p(i\omega) = 0$ if and only if $c_1(k) = 0$ and $\omega^2 = c_0(k)/c_2(k) > 0$. In case $c_2(k) = 0$, p_k has the unique root $\lambda = -c_0(k)/c_1(k)$. ∎

We note that Theorem 4.2 actually holds without the assumption that γ and ρ are positive.

When $\alpha = \beta = 0$ (corresponding to the "uncoupled" dynamics), the equilibrium solution is asymptotically stable since γ and ρ are positive in (4.4). By (4.6) and (4.7), stability persists for sufficiently small α, β, in agreement with Theorem 3.1. As α or β is further increased, stability can be lost through a stationary or oscillatory bifurcation, characterized by the conditions (4.8) and (4.9)–(4.10), respectively. Note that the stationary bifurcation condition (4.8) depends only on the connectivity kernels and not on the delays. Furthermore, in case of instantaneous signal transmission, $E[\tau] = E[v^{-1}] = 0$, and by (4.9) oscillatory bifurcations cannot occur, which agrees with Theorem 3.2. For nonzero delays both types of bifurcations become possible. In case of small delays, one can assume that (4.5) holds throughout some parameter range. Then studying the validity of conditions (4.6) and (4.7) while increasing α or β from zero yields an important result: the condition that is first violated determines the type of bifurcation that actually occurs, i.e., whether a stationary or oscillatory bifurcation takes place, as either (4.8) or (4.9) is satisfied. We now give a qualitative account of the effects of delays on these bifurcations. Under the assumption of small delays as stated in section 4, the conclusions drawn will also hold for the full model (2.6) at nearby parameter values. Numerical solution of (2.6) will be used in section 5 to confirm the results for specific choices of kernels.

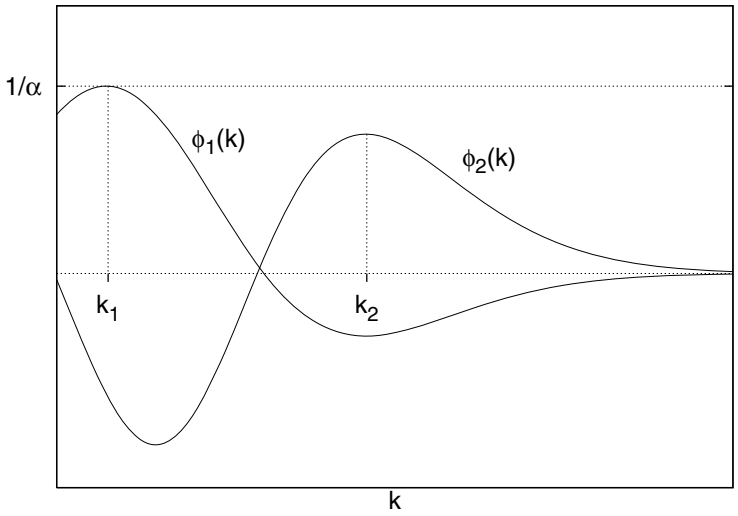


Figure 2. *Schematic diagram of the competition between stationary and oscillatory bifurcations. The curves ϕ_1 and ϕ_2 denote the first and the second terms inside the braces in (4.12). As the bifurcation parameter α is increased from zero, the horizontal line $1/\alpha$ becomes tangent to one of the curves. The tangency with ϕ_1 is equivalent to a stationary bifurcation, while the tangency with ϕ_2 is equivalent to an oscillatory bifurcation. Since these tangencies depend only on the maxima of the curves, located at k_1 and k_2, respectively, the type of bifurcation can be determined by comparing $\phi_1(k_1)$ and $\phi_2(k_2)$.*

Propagation delays. We first focus on the effect of propagation delays by setting $\beta = 0$, i.e., by ignoring the contribution of feedback. From (4.8) and (4.9) we have the conditions

$$\frac{1}{\alpha} = \frac{\hat{K}(k)}{\rho}$$

and

$$\frac{1}{\alpha} = -\frac{E[v^{-1}]\hat{K}_1(k)}{\gamma}$$

for stationary and oscillatory bifurcations, respectively. As α is increased from zero, the horizontal line $1/\alpha$ can become tangent to the graph of $\hat{K}(k)/\rho$ or $-\hat{K}_1(k)E[v^{-1}]/\gamma$, at some value $k = k^*$. Whichever tangency occurs first (i.e., for the smaller value of α) determines the type of bifurcation as the equilibrium solution loses its stability. In other words, k^* denotes the point where the function

$$\phi(k) := \max\left\{\frac{\hat{K}(k)}{\rho}, \frac{-E[v^{-1}]\hat{K}_1(k)}{\gamma}\right\} \tag{4.12}$$

assumes its maximum, which exists if ϕ is not strictly negative, and is generically unique. If $\phi(k^*) = \hat{K}(k^*)/\rho$, then the bifurcation is stationary; otherwise it is oscillatory (see Figure 2 for a schematic picture). Hence, higher values of $E[v^{-1}]$ (slower signal propagation) will facilitate

oscillatory bifurcations by increasing the magnitude of the second term inside the braces in (4.12), provided that $\hat{K}_1$ is not a nonnegative function. The oscillation frequency of such bifurcations is given by (4.10) as

$$(\omega^*)^2 = \frac{\rho - \alpha\hat{K}(k^*)}{2 - \alpha E[v^{-2}]\hat{K}_2(k^*)}.$$

Furthermore, the bifurcating solution is spatially homogeneous if and only if $k^* = 0$. If $k^* \neq 0$, then ω^*/k^* is the phase speed of the traveling waves that bifurcate. Thus, the main effect of propagation delays is on oscillatory bifurcations. However, this effect is intimately related to the connectivity kernel through $\hat{K}_1$. Note that even if K is sign definite, such as a purely excitatory or purely inhibitory connection, $\hat{K}_1$ can in general take both positive and negative values.

As noted in [2], if a sufficiently general class of kernels is considered, the maximum of ϕ in (4.12) is likely to occur at some nonzero k rather than at the precise value $k = 0$. Consequently, one expects to see traveling waves (respectively, Turing patterns) when signal propagation is slow (respectively, fast), as the dominant mode of bifurcation in the absence of the feedback term.[1]

Feedback delays. We now include the effect of feedback delays by allowing β to be nonzero. Similar to above, the wave number k^* of the bifurcating solution is the point where the function

$$\Phi(k) = \max\left\{\frac{\alpha\hat{K}(k) + \beta\hat{F}(k)}{\rho}, \frac{-\alpha E[v^{-1}]\hat{K}_1(k) - \beta E[\tau]\hat{F}(k)}{\gamma}\right\} \tag{4.13}$$

assumes its maximum value. The type of bifurcation is again determined by the relative magnitudes of the two terms inside the braces, which can be studied by plotting them on the same graph. Thus, when $\hat{F}$ is a positive function (which is typical for most common forms of excitatory feedback), the first term increases while the second one decreases for increasing β. We conclude that the presence of excitatory feedback connections generally facilitates stationary bifurcations and suppresses oscillatory ones if $E[\tau]$ is nonzero. The reverse conclusion holds for inhibitory feedback connections. The feedback connections thus show a definite preference on the type of bifurcations, given by the sign of $\hat{F}$. Furthermore, their effect on oscillatory bifurcations increases with the mean feedback delay $E[\tau]$.

A particular observation which may have significance in applications concerns the case when the feedback connections are highly nonlocal, that is, the kernel F has a large variance. It is a well-known fact from Fourier theory that the corresponding transform $\hat{F}$ then has a small variance and thus is highly concentrated near the origin. It follows that increasing β from zero will cause Φ to have its maximum near $k^* = 0$; in fact, if k is restricted to discrete values (e.g., when the spatial domain is a circle), then the maximum will occur at $k^* = 0$. Consequently, nonlocal feedback which has a sufficiently large spatial range tends to enhance spatially *homogeneous* bifurcations, which are stationary if $\hat{F}$ is a positive function or

[1] Of course, in applications which dictate a specific form for the connectivities, this expectation may not always be realized.

oscillatory if $\hat{F}$ is negative and $E[\tau]$ is sufficiently large. This is in contrast with the case $\beta = 0$ noted above, where spatially *inhomogeneous* bifurcations might be the more likely scenario.

Distributed delays and speeds. Finally, we consider the effects of distributed transmission speeds and feedback delays on the bifurcations. To this end, we keep the mean values $E[\tau]$ and $E[v^{-1}]$ fixed and increase the variances $\mathrm{Var}[\tau]$ and $\mathrm{Var}[v^{-1}]$, assuming (4.5) throughout as before. It then follows from the conditions (4.6) and (4.7) that the stability of the equilibrium solution is unaffected by the variances, and the bifurcations occur at the same parameter values. This is also apparent from the above analysis, which shows that the wave number k^*, in particular, does not depend on $\mathrm{Var}[\tau]$ or $\mathrm{Var}[v^{-1}]$. The main effect of the variances is on the frequency ω^* of oscillatory bifurcations. By (4.10),

$$(\omega^*)^2 = \frac{\rho - \alpha\hat{K}(k^*) + \beta\hat{F}(k^*)}{2 - \alpha\hat{K}_2(k^*)(\mathrm{Var}[v^{-1}] + E^2[v^{-1}]) - \beta\hat{F}(k^*)(\mathrm{Var}[\tau] + E^2[\tau])}. \tag{4.14}$$

Suppose that the variance $\mathrm{Var}[\tau]$ is changed while the mean value $E[\tau]$ is kept fixed. If $\hat{F}(k^*)$ is positive (resp., negative), then ω^* will increase (resp., decrease) with increasing variance of the feedback delays τ. In case of bifurcating traveling waves, an increase in ω^* corresponds to an increase in the speed ω^*/k^* of the waves, since k^* is unaffected by the variance of delays and speeds. Thus, increasing the variance of delays will typically increase the frequency of bifurcating oscillatory solutions or the speed of traveling waves for excitatory feedback connections and decrease it for inhibitory feedback connections. Similarly, if $\hat{K}_2(k^*)$ is positive (resp., negative), then ω^* will increase (resp., decrease) with increasing variance $\mathrm{Var}[v^{-1}]$ of propagation speeds. Figure 7 in section 5 illustrates the dependence of the wave speed on the variance of the propagation speed for a particular choice of connectivity and transmission speed distribution.

5. Numerical results. In this section we present simulation results confirming the analysis of previous sections with specific choices for the connectivities and transmission speed distributions. For numerical calculations we take the spatial domain Ω to be the circle with circumference C (or equivalently the interval $[0, C]$ with periodic boundary conditions) and assume that the feedback connections are uniformly distributed over this domain (global feedback). Thus,

$$F(z) \equiv \frac{1}{C}, \qquad \int_0^C F(z)\,dz = 1,$$

with the Fourier transform

$$\hat{F}(k) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k = \frac{2\pi n}{C},\ n \in \mathbb{Z}; \end{cases}$$

thus, we denote $\hat{F}(k) = \delta(k)$. We further assume that L is the second order operator (4.4) and take $\rho = L(0) = 1$.

In this setting, the condition (3.9) for temporally stationary and spatially homogeneous bifurcation takes the form

$$1 - \beta = \alpha \int_{-\infty}^{\infty} K(z)\,dz.$$

The condition (3.10) for Turing patterns also simplifies to

$$1 = \alpha \hat{K}(k), \quad k \neq 0,$$

which depends only on the distribution K of intra-areal connections. We choose K to be

$$K(z) = \frac{a_e}{2} e^{-|z|} - \frac{a_i}{2} r e^{-r|z|}, \tag{5.1}$$

where r denotes the relation of excitatory and inhibitory spatial ranges and a_e and a_i represent excitatory and inhibitory synaptic weights [30]. For instance, in case of a single propagation speed and $r > a_e/a_i$, the neural field exhibits local inhibition and lateral excitation and thus facilitates traveling waves [30].

5.1. Feedback delays. We now present simulation results displaying the role of global feedback delay. We apply an explicit Euler-forward algorithm for the time integration, which stores past activity according to the distance-dependent propagation delays and the feedback delays. The initial conditions are chosen randomly from a uniform distribution; however, see figure captions for more details. At each temporal iteration step, both spatial integrals in (2.6) have to be computed. In previous studies, we applied modified Riemannian sums [2, 30], which need a large number of spatial grid points to obtain small numerical errors. More advanced integration rules take into account the kernel properties. Here we use the VEGAS algorithm [22, 39], which is a Monte-Carlo integration algorithm in combination with importance sampling. In principle, the algorithm samples points from the probability function to be integrated, so that the points are concentrated in the regions that make the largest contribution to the integral. This approach can be applied to (2.6), as the function S is bounded and the kernels define the contribution to the integral. In the simulations, 2000 calls are applied for a single Monte-Carlo integration. The integration is repeated five times, and the average result is considered. Other variants of Monte-Carlo integration algorithms, such as plain Monte-Carlo or the MISER algorithm [22], showed worse performance. We remark that our criteria for the best algorithm have been the speed and the error at computing the norm of both connectivity kernels. The final parameters allow the computation of the kernel norms to an error of 10^{-5}. Since (2.6) necessitates integral computations for each value of x, we discretize the spatial domain into N intervals, i.e., $x_n = Cn/N$ with $n = 0, \ldots, N$. The sigmoidal transfer function has been chosen to be the logistic function $S(V) = 10/(1 + \exp(-1.8(V - 3.0)))$.

According to section 4.2, global inhibitory feedback with delays can destabilize the equilibrium solution and induce periodic oscillations. In the simulation, we applied a single feedback delay with $f(\tau) = \delta(\tau - \tau_0)$, i.e., $E[\tau] = \tau_0$ and $\mathrm{Var}[\tau] = 0$, and chose a single transmission speed, i.e., $g(v) = \delta(v - v_0)$. The function Φ given in (4.12) now has the form

$$\Phi(k) = \max\left\{ \frac{\alpha \hat{K}(k) + \beta \delta(k)}{\rho}, \frac{-\alpha \hat{K}_1(k)/v_0 - \beta \tau_0 \delta(k)}{\gamma} \right\},$$

and will have a maximum at $k^* = 0$ if β and τ_0 are sufficiently large. Therefore, as the equilibrium loses its stability, spatially constant oscillations emerge. Decreasing either β or τ_0 will decrease the effect of global feedback, and the maximum of Φ might jump to some $k^* \neq 0$, yielding a spatially nonhomogeneous bifurcation. Figure 3 illustrates a jump from constant

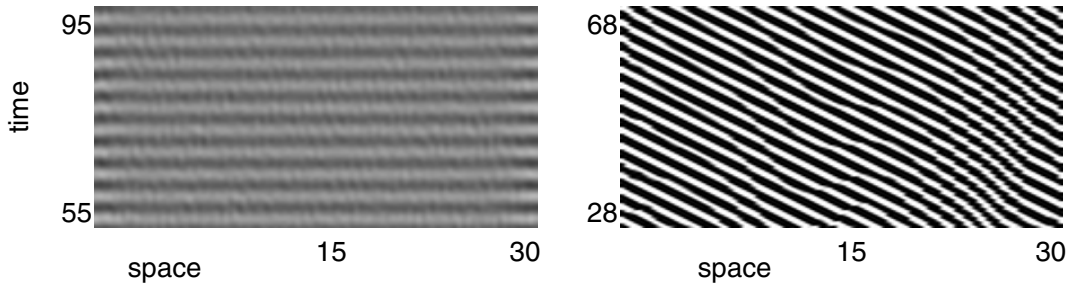


Figure 3. *Two space-time plots of the field activity with inhibitory feedback near a Hopf bifurcation. Left panel: constant oscillations with strong nonlocal inhibition* $\bar{\mu}_i = 3.5$ ($\beta = -4.93$) ($E^* = 4.6$, $\alpha = 8.1$). *Right panel: traveling waves with* $k^* = 3.4$ *and weaker nonlocal inhibition* $\bar{\mu}_i = 1.0$ ($\beta = -3.82$) ($E^* = 5.4$, $\alpha = 22.31$). *In both cases the other parameters are* $C = 30$, $N = 100$, $v_0 = 2.5$, $\tau_0 = 2.5$, $r = 6.0$, $a_e = 5.0$, $a_i = 4.9$, $\bar{\mu}_e = 0$, $dt = 0.08$, *and the numerical wave number shows good agreement with the analytical critical wave number* k^*. *The greyscale encodes the activity deviation from the stationary state and is fixed if the activity magnitude exceeds a fixed threshold. This clipping procedure has been introduced for illustration reasons and results in artificial sharp edges in the plots.*

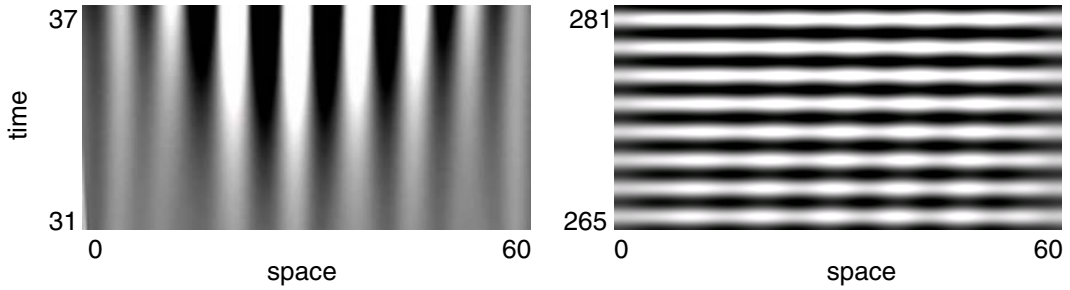


Figure 4. *Space-time plots of the field activity with inhibitory feedback at bifurcation thresholds. Left panel: Turing instability with wave number* $k^* = 0.73$ *and vanishing delay time* $\tau_0 = 0$ *(parameters are* $E^* = 1.93$, $\alpha = 1.77$, $\beta = -1.16$, $dt = 0.06$*). Right panel: oscillatory global instability with* $k^* = 0$ *and delay time* $\tau_0 = 2.5$ ($E^* = 1.6$, $\alpha = 1.335$, $\beta = -0.873$, $dt = 0.02$). *In both cases the other parameters are* $C = 60$, $N = 200$, $v = 2.0$, $\bar{\mu}_i = 1.5$, $\bar{\mu}_e = 0$, $r = 0.5$, $a_e = 5.0$, $a_i = 4.9$, *and the numerical wave number shows good agreement with the analytical critical wave number* k^*. *The greyscale encoding follows the same rules as in Figure 3.*

oscillations to traveling waves induced by changing the inhibitory feedback.

Figure 4 reveals further feedback effects on the stability of neural fields. For inhibitory feedback and vanishing feedback delay time τ_0, there is a stationary Turing bifurcation for certain parameters (Figure 4, left panel). That is, the first term in (4.13) is larger than the second one. Now increasing the delay time τ_0 makes the second term exceed the first term (since $\beta < 0$), which destabilizes the field towards an oscillatory bifurcation. Hence, by the special choice of the feedback kernel, global oscillations occur (Figure 4, right panel).

5.2. Distributed propagation speeds. Motivated by experimental findings [7, 16, 44] (see Figure 1), we model the distribution of propagation speeds by the (truncated) gamma family of densities

$$g(v) = \begin{cases} \frac{N_{p,q}}{q^p \Gamma(p)} v^{p-1} \exp(-v/q) & \text{if } v \in (v_l, v_h), \\ 0 & \text{otherwise,} \end{cases} \tag{5.2}$$

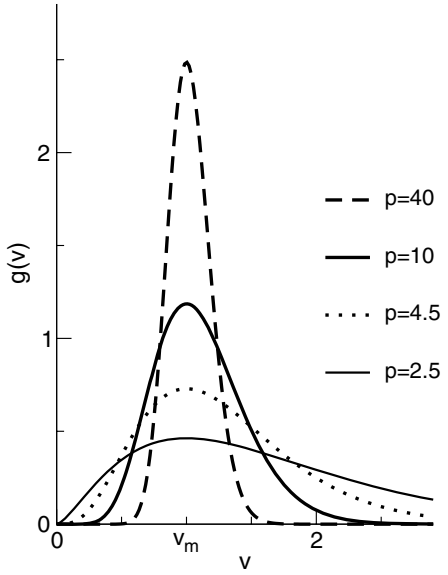


Figure 5. *Gamma-distributed transmission speeds.*

with parameters $p > 2$ and $q > 0$. The constant $N_{p,q}$ arises from the normalization condition (2.3) if one stipulates that g should be zero outside some interval $v \in (v_l, v_h)$, where v_l and v_h are physiologically dictated lower and upper bounds for the possible propagation speeds. In the limit as $v_l \to 0$ and $v_h \to \infty$ one obtains the usual gamma density with $N_{p,q} = 1$. The maximum of $g(v)$ (i.e., the mode of the distribution) occurs at $v_m = q(p-1)$. Figure 5 shows the shape of the (untruncated) distribution for various p and some fixed value of v_m. Furthermore,

$$E[v^{-1}] = \frac{N_{p,q}}{N_{p-1,q}} \frac{1}{q(p-1)} = \frac{N_{p,q}}{N_{p-1,q}} \frac{1}{v_m}, \qquad E[v^{-2}] = \frac{N_{p,q}}{N_{p-2,q}} \frac{p-1}{p-2} \frac{1}{v_m^2},$$

$$\mathrm{Var}[v^{-1}] = \frac{N_{p,q}}{v_m^2 N_{p-2,q} N_{p-1,q}^2 (p-2)} \left(N_{p-1,q}^2 (p-1) - N_{p,q} N_{p-2,q} (p-2) \right).$$

Figure 6 shows the dependence of $E[1/v^2]$ and $\mathrm{Var}[v^{-1}]$ on v_m and p. For a fixed value of v_m, a fast decay to a constant value is observed by increasing p.

With this choice of g, we determine the phase speed ω/k of bifurcating waves from (4.14), where for simplicity we neglect feedback effects by setting $\beta = 0$. In Figure 7 the resulting phase speed is plotted with respect to the parameters v_m and p of the gamma-distribution. The phase speed is seen to be lower than v_m for all p, which affirms causality, and shows accordance to previous results obtained for single transmission speeds [2, 10, 30]. Furthermore, for lower values of p (broader speed distribution) the phase speed is smaller, an effect which is more significant at lower values of p. For the chosen parameter set, we calculate $\hat{K}_2(k^*) < 0$; so, as mentioned in section 4.2, a broader transmission speed distribution yields a lower phase speed for traveling waves.

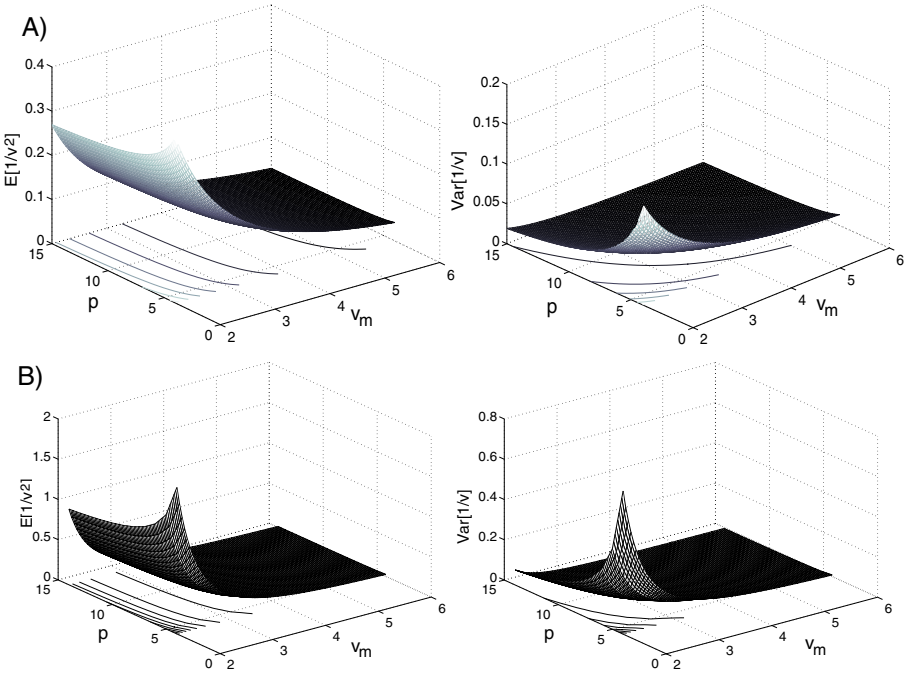


Figure 6. *The dependence of* $E[1/v^2]$ *and* $\mathrm{Var}[1/v]$ *on* v_m *and* p *for* (A) *untruncated gamma-distribution and* (B) *truncated gamma-distribution with* $v_l = 5$, $v_h = 50$. *The value of* q *is given by* $q = v_m/(p-1)$.

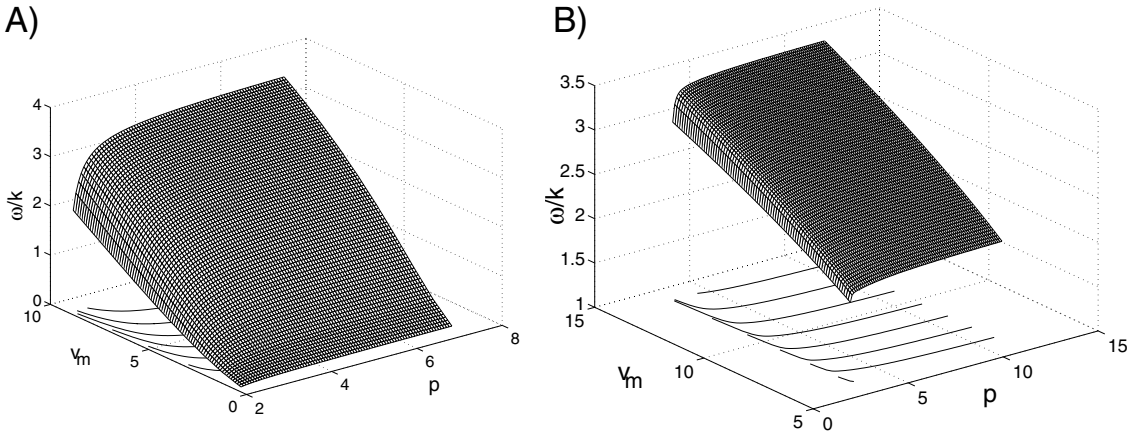


Figure 7. *Propagation speed of bifurcating traveling waves with respect to the parameters of the speed distribution for* (A) *the untruncated gamma-distribution and* (B) *the truncated gamma-distribution with* $v_l = 4$, $v_h = 100$. *Other parameters are* $a_e = 100$, $a_i = 99$, $r = 3$, $\mu = 0$.

6. Traveling fronts.

We now give an analysis of the effects of distributed propagation speeds and feedback delays on traveling fronts. Here, we take the function S in (2.6) to be the Heaviside function Θ with a fixed firing threshold u_0. In addition, we take the external input E to be zero, which introduces no loss of generality since a constant external input amounts to

shifting the firing threshold.[2] For definiteness, we consider a first order temporal differential operator

$$L\left(\frac{\partial}{\partial t}\right) = \frac{\partial}{\partial t} + 1. \tag{6.1}$$

Then the evolution equation (2.6) reads

$$\begin{aligned}\frac{\partial V(x,t)}{\partial t} &= -V(x,t) + \int_0^\infty g(v) \int_{-\infty}^\infty K(x-y)\Theta(V(y,t-|x-y|/v)-u_0)dy\,dv \\ &\quad + \int_0^\infty f(\tau) \int_{-\infty}^\infty F(x-y)\Theta(V(y,t-\tau)-u_0)dy\,d\tau,\end{aligned} \tag{6.2}$$

where for simplicity of notation we have subsumed the scaling factors $\bar{\alpha}$ and $\bar{\beta}$ into the definitions of K and F, respectively.

6.1. The stationary front solution. We switch to a frame moving with speed c by the change of variables $V(x,t) = \bar{u}(x-ct) = \bar{u}(z)$, which transforms (6.2) into

$$\begin{aligned}\frac{d\bar{u}(z)}{dz} &= \frac{1}{c}\bar{u}(x,t) - \frac{1}{c}\int_0^\infty g(v) \int_{-\infty}^\infty K(y)\Theta\left(u\left(z-y+\frac{c|y|}{v}\right)-u_0\right) dy\,dv \\ &\quad - \frac{1}{c}\int_0^\infty f(\tau) \int_{-\infty}^\infty F(y)\Theta(u(z-y+c\tau)-u_0)dy\,d\tau.\end{aligned} \tag{6.3}$$

The boundary conditions are set to $\lim_{z\to-\infty}\bar{u}(z) = u_\infty > 0$, $\lim_{z\to\infty}\bar{u}(z) = 0$, and $\bar{u}(0) = u_0$, with $0 < u_0 < u_\infty$. A simple calculation yields $u_\infty = \eta_1 + \eta_2$, where $\eta_1 = \int_{-\infty}^\infty K(x)\,dx$ and $\eta_2 = \int_{-\infty}^\infty F(x)\,dx$. Then we obtain

$$\begin{aligned}z \geq 0 : \frac{d\bar{u}(z)}{dz} &= \frac{1}{c}\bar{u}(z) \\ &\quad - \frac{1}{c}\int_{v_l}^{v_h} g(v) \int_{\gamma_- z}^\infty K(y)\,dy\,dv - \frac{1}{c}\int_0^\infty f(\tau) \int_{z+c\tau}^\infty F(y)\,dy\,d\tau, \\ z < 0 : \frac{d\bar{u}(z)}{dz} &= \frac{1}{c}\bar{u}(z) - \frac{1}{c}\int_{v_l}^{v_h} g(v) \int_{\gamma_+ z}^\infty K(y)\,dy\,dv - \frac{1}{c}\int_0^\infty f(\tau) \int_{z+c\tau}^\infty F(y)\,dy\,d\tau,\end{aligned} \tag{6.4}$$

with $\gamma_\pm = v/(v\pm c)$. It is not hard to check that (6.4) has finite solutions for $z\to\infty$, provided that

$$\begin{aligned}u_0 &= \frac{\eta_1}{2} + \eta_2 - R(c\tau) \\ &\quad - \int_0^\infty e^{-z'/c} E_g\left[\gamma_- K(\gamma_- z')\right] dz' - \int_0^\infty e^{-z'/c} E_f\left[F(z'+c\tau)\right] dz'\end{aligned} \tag{6.5}$$

holds with the antiderivative $R(x)$ of $F(x)$, i.e., $dR(x)/dx = F(x)$. Here we have used the expectation operators E_g and E_f with respect to the densities g and f, namely,

$$E_g[r] = \int_0^\infty g(v)r(v)\,dv, \qquad E_f[r] = \int_0^\infty f(\tau)r(\tau)\,d\tau.$$

[2] A nonconstant input may yield additional effects, whose study would exceed the aim of the present work.

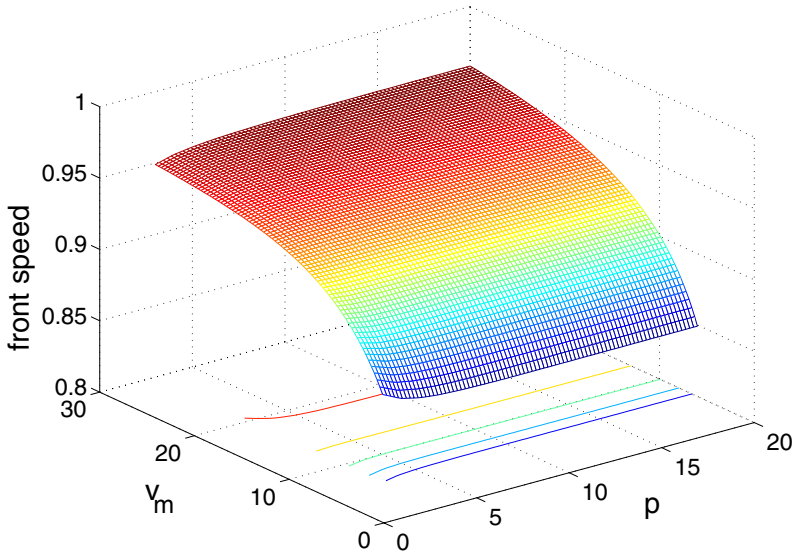


Figure 8. *The traveling front speed c for excitatory fields and vanishing feedback, plotted with respect to the parameters v_m and p of the transmission speed distribution. Further parameters are $a_e = 4$, $a_i = 0$, $u_0 = 0.01$, $v_l = 2$, $v_h = 50$, and $q = v_m/(p-1)$.*

The traveling front solution can now be explicitly calculated as

$$
\begin{aligned}
z \geq 0 : \bar{u}(z) &= \eta_1 + \eta_2 - E_g[W(\gamma_- z)] - E_f[R(z + c\tau)] \\
&\quad - \int_z^\infty e^{(z-z')/c} E_g[\gamma_- K(\gamma_- z')]\, dz' - \int_z^\infty e^{(z-z')/c} E_f[F(z' + c\tau)]\, dz', \qquad (6.6) \\
z < 0 : \bar{u}(z) &= \eta_1 + \eta_2 - E_g[W(\gamma_+ z)] - E_f[R(z + c\tau)] \\
&\quad - \int_z^0 e^{(z-z')/c} E_g[\gamma_+ K(\gamma_+ z')]\, dz' - \int_0^\infty e^{(z-z')/c} E_g[\gamma_- K(\gamma_- z')]\, dz' \\
&\quad - \int_z^\infty e^{(z-z')/c} E_f[F(z' + c\tau)]\, dz', \qquad (6.7)
\end{aligned}
$$

where $W(x)$ is the antiderivative of $K(x)$, i.e., $dW(x)/dx = K(x)$.

Equation (6.5) implicitly defines the front speed c in terms of the kernels K and F, the threshold u_0, the transmission speed distribution $g(v)$, and the feedback delay distribution $f(\tau)$. To illustrate the dependence of the front speed on distributed delays, Figure 8 presents the solutions of (6.5) for the synaptic excitatory kernel taken from (5.1) and various speed distributions (5.2) for vanishing feedback delay. The relation of c and v_m shows accordance to previous results for single propagation speeds [27, 46]. In addition, for low values of v_m an increase in the speed distribution width, i.e., a decrease of p, yields an increase in the front speed. In contrast, large values of v_m yield a decreasing front speed for an increase of the speed distribution width, similar to the observations of sections 4.2 and 5 obtained under the assumption of large propagation speeds. Hence, the shape of the transmission speed distributions affects the front speed. A subtler effect is discovered when the front speed is plotted against the variance instead of the parameter p (Figure 9). It is seen that the front speed can

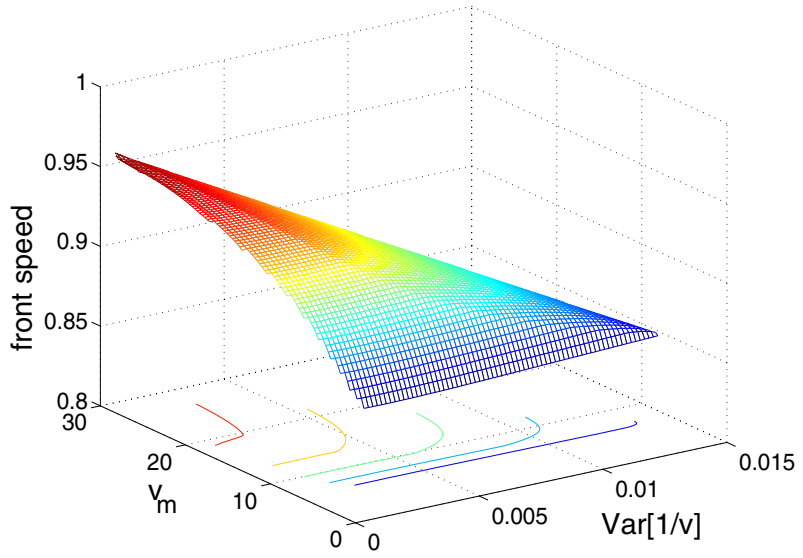


Figure 9. *The traveling front speed c plotted against the mode v_m and the variance* $\mathrm{Var}[1/v]$ *of the speed distribution for excitatory fields and vanishing feedback. Parameter values are $a_e = 4$, $a_i = 0$, $u_0 = 0.01$, $v_l = 2$, $v_h = 50$, and $q = v_m/(p-1)$.*

be maximized by a positive variance of speed distributions. This might indicate an interesting biological principle for maximizing signal propagation in fields of nonlocal interaction.

6.2. Stability of fronts. Considering small deviations $s(x-ct,t) = u(x,t) - \bar{u}(x-ct)$ from the stationary solution, (6.2) yields, in linear order in s,

$$\begin{aligned} -c\frac{\partial s(z,t)}{\partial z} + \frac{\partial s(z,t)}{\partial t} &= -s(z,t) \\ &+ \int_{-\infty}^{\infty} K(y) E_g \left[\delta\left(\bar{u}\left(z - y + \frac{c|y|}{v} \right) - u_0 \right) s\left(z - y + \frac{c|y|}{v}, t - \frac{|y|}{v} \right) \right] dy \\ &+ \int_{-\infty}^{\infty} F(y) E_f \left[\delta(\bar{u}\,(z - y + c\tau) - u_0)\, s(z - y + c\tau, t - \tau) \right] dy, \end{aligned} \tag{6.8}$$

where $\delta(x)$ denotes the Dirac delta distribution. Subsequently, we obtain

$$\begin{aligned} z > 0 : -c\frac{\partial s(z,t)}{\partial z} + \frac{\partial s(z,t)}{\partial t} &= -s(z,t) \\ &+ E_g \left[\frac{\gamma_- K(\gamma_- z)}{|\bar{u}'(0)|} s\left(0, t - \frac{z}{v - c} \right) \right] + E_f \left[\frac{F(z + c\tau)}{|\bar{u}'(0)|} s(0, t - \tau) \right], \end{aligned} \tag{6.9}$$

$$\begin{aligned} z < 0 : -c\frac{\partial s(z,t)}{\partial z} + \frac{\partial s(z,t)}{\partial t} &= -s(z,t) \\ &+ E_g \left[\frac{\gamma_+ K(\gamma_+ z)}{|\bar{u}'(0)|} s\left(0, t + \frac{z}{v + c} \right) \right] + E_f \left[\frac{F(z + c\tau)}{|\bar{u}'(0)|} s(0, t - \tau) \right], \end{aligned} \tag{6.10}$$

with $\bar{u}'(0) = \partial\bar{u}/\partial z$ computed at $z = 0$. We observe that $s(z,t)$ is continuous but not continuously differentiable at $z = 0$, which results from the applied Heaviside function Θ.

Now inserting the ansatz $s(z,t) = \phi(z)e^{\lambda t}$, we find

$$
\begin{aligned}
z > 0 : \frac{\partial \phi(z)}{\partial z} &= \frac{1+\lambda}{c}\phi(z) \\
&\quad - E_g\left[\frac{\gamma_- K(\gamma_- z)}{|\bar{u}'(0)|c} e^{-z\lambda/(v-c)}\right]\phi(0) - E_f\left[\frac{F(z+c\tau)}{|\bar{u}'(0)|c}e^{-\lambda\tau}\right]\phi(0), \\
z < 0 : \frac{\partial \phi(z)}{\partial z} &= \frac{1+\lambda}{c}\phi(z) \\
&\quad - E_g\left[\frac{\gamma_+ K(\gamma_+ z)}{|\bar{u}'(0)|c} e^{z\lambda/(v+c)}\right]\phi(0) - E_f\left[\frac{F(z+c\tau)}{|\bar{u}'(0)|c}e^{-\lambda\tau}\right]\phi(0),
\end{aligned}
$$

whose solutions are

$$
\begin{aligned}
z > 0 : \phi(z) = \phi(0)e^{(1+\lambda)z/c}&\left(1 - \frac{1}{|\bar{u}'(0)|c}\int_0^z E_g\left[\gamma_- e^{-z'(\gamma_-\lambda/v+(1+\lambda)/c)}K(\gamma_- z')\right]dz'\right. \\
&\left. - \frac{1}{|\bar{u}'(0)|c}\int_0^z E_f\left[e^{-\lambda\tau}e^{-z'(1+\lambda)/c}F(z'+c\tau)\right]dz'\right), \qquad (6.11) \\
z < 0 : \phi(z) = \phi(0)e^{(1+\lambda)z/c}&\left(1 + \frac{1}{|\bar{u}'(0)|c}\int_z^0 E_g\left[\gamma_+ e^{-z'(-\gamma_+\lambda/v+(1+\lambda)/c)}K(\gamma_+ z')\right]dz'\right. \\
&\left. + \frac{1}{|\bar{u}'(0)|c}\int_z^0 E_f\left[e^{-\lambda\tau}e^{-z'(1+\lambda)/c}F(z'+c\tau)\right]dz'\right). \qquad (6.12)
\end{aligned}
$$

Here, we used the continuity of $\phi(z)$ at $z = 0$, i.e., $\lim_{z\to+0}\phi(z) = \lim_{z\to-0}\phi(z) = \phi(0)$. Equations (6.11) and (6.12) are equivalent to the solutions found in [12] for traveling fronts involving a single transmission speed and without the constant feedback delay.

It can be seen that finite solutions $\phi(z)$ exist for $z \to \infty$, provided that the equation

$$
\begin{aligned}
0 = 1 &- \frac{1}{|\bar{u}'(0)|c}E_g\left[\mathcal{L}_0\left(\frac{1+\lambda}{c} - \frac{1}{v}\right)\right] \\
&- \frac{1}{|\bar{u}'(0)|c}E_f\left[e^{-\lambda\tau}\int_0^\infty e^{-z'(1+\lambda)/c}F(z'+c\tau)dz'\right] \qquad (6.13)
\end{aligned}
$$

holds for $\mathrm{Re}(\lambda) > -1$, while the solutions $\phi(z)$ are finite for $z \to -\infty$ if

$$
|\bar{u}'(0)|c = -E_g\left[\mathcal{L}_0\left(-\frac{1+\lambda}{c} - \frac{1}{v}\right)\right] - E_f\left[e^{-\lambda\tau}\int_0^\infty e^{-z'|1+\lambda|/c}F(z'-c\tau)dz'\right] \qquad (6.14)
$$

holds for $\mathrm{Re}(\lambda) \le -1$. Here, the symmetry of both kernels has been used implicitly. The equations (6.13) and (6.14) determine λ implicitly. We observe that $\lambda = 0$ represents a solution of (6.13), which reflects the translation symmetry of the moving fronts. Since (6.13) may yield $\mathrm{Re}(\lambda) > 0$, it represents a stability condition of the stationary front. The foregoing analysis is equivalent to the method of Evans functions for nonlocal interactions [12, 33, 62], which has been used to obtain a single implicit condition, namely the Evans function, for determining the stability of the field in the absence of speed distributions and feedback delays.

A numerical simulation of the evolution equation (6.2) is presented in Figure 10. The

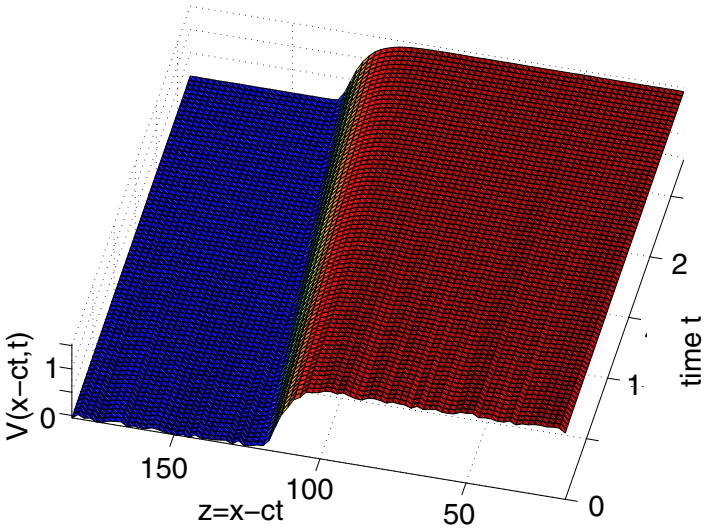


Figure 10. *The simulated traveling front for excitatory fields, distributed transmission speeds, and vanishing feedback delay. Parameter values are* $a_e = 1$, $a_i = 0$, $r = 0.5$, $u_0 = 0.1$, $v_m = 4$, $v_l = 2.5$, $v_h = 6$, $p = 3.15$, *and* $c = 1.97$.

initial conditions are chosen as $V(x) = \bar{u}(x) + \Gamma(x)$, where $\Gamma(x) \in [-0.05, 0.05]$ is a uniformly distributed random variable and $\bar{u}(x)$ is given by (6.6)–(6.7). It turns out that the random activity is damped out, and the solution converges to the stationary front. In other words, the stationary front $\bar{u}(x)$ is stable for the chosen parameter values.

7. Conclusion. We have introduced transmission speed distributions and a nonlocal feedback loop with distributed delays into the standard neural field model, and studied the stability and bifurcations of spatially homogeneous equilibrium solutions, as well as traveling fronts. As expected, the relation between the connectivities and delays plays an important role in the analysis of bifurcations. We have investigated this relation for general field connectivities and feedback kernels as well as speed and feedback delay distributions. The results have significance in the context of understanding the basic mechanisms of neural activity, since distributed delays and speeds arise naturally in real neural systems. To the best of our knowledge, these effects have been neglected in the literature until now. For the numerical solution of the field equation, we have introduced a Monte-Carlo approach. In contrast to conventional integration procedures such as the trapezoidal rule [46], this approach allows the computation of the interaction integral with a divergent kernel. The numerical results illustrate the effects of nonlocal feedback and the corresponding delays, and confirm the analytical findings. Finally, we have examined stationary traveling fronts involving distributed speeds and feedback delays.

One of the highlights of the presented analysis is the different nature of the two delay sources, namely, the delays caused by the finite speed of signal transmission in the field, and the delays in the long-range feedback loop. The former delays vary continuously, depending on the distance between locations, and so are always distributed, even for a single transmission speed. This continuous dependence on distance interacts in a nontrivial way with the connectivity

kernel. On the other hand, the delays in the feedback loop have a simpler dependence on the feedback connectivity. The difference is made clear by the particular way the terms enter the dispersion relation (3.3). To overcome the analytical difficulties, we have considered in section 4 the case when the delays are not too large, and studied approximate models based on series expansion. The difference of the two delay types is still apparent in reduced models at all orders. For instance, the statement of Theorem 4.1 involves higher moments $\hat{K}_m$ of the field connectivity kernel, showing their intimate relation with the speed distribution, whereas the feedback terms contribute only by a simple Fourier transform $\hat{F}$. One can see the physical effects in a second-order approximate model studied in section 4.2. Because the feedback connectivity enters (4.13) only through $\hat{F}$ in both terms, its role essentially depends simply on the relative magnitudes of α and β, i.e., the relative strengths of field and feedback connectivities. In contrast, the field connectivity K appears through $\hat{K}$ in the first term and $\hat{K}_1$ in the second. Since $\hat{K}_1$ emphasizes the role of distant interaction more, the role of the involved propagation delays is accentuated, which can facilitate oscillatory bifurcations. Based on similar arguments involving Fourier transforms, we have concluded that feedback connections with large variance tend to induce spatially homogeneous bifurcations, whereas general field connectivity kernels might favor spatially inhomogeneous bifurcations.

The presented study shows that distributed propagation speeds and feedback delays affect the speed of bifurcating waves and traveling fronts. The quantitative changes can be significant for sufficiently broad distributions; otherwise, they may be relatively small. The precise effects depend on the parameters of a particular biological application, which we do not further pursue here. More interestingly, one can observe qualitative changes by the introduction of distributions, such as obtaining a maximum front speed at a positive variance of the transmission speed distribution. (Whether this represents a biological optimization principle is a tempting conjecture beyond the purpose of the present paper.) Moreover, the stability analysis of stationary fronts involving distributed propagation speeds and feedback delays yields a stability condition, which represents the Evans function of the system.

The present work studies only a limited range of dynamical behavior. Nevertheless, the effects of delays are much richer in this setting of spatially extended systems with nonlocal interaction than in simple feedback systems with distributed delays considered in, e.g., [4]. Further research can be expected to discover other interesting features in more global dynamics away from equilibria, for instance, in the evoked response of the field as studied in [29], when distributed speeds and feedback loops are introduced into the model. Moreover, the existence and stability study of bumps [9, 38], spirals [36], or breathers [11, 19] in case of distributed propagation speeds and nonlocal feedback delays promises a wide range of novel phenomena and will yield new insight into neural information processing.

Acknowledgments. The authors are grateful to Peter Mathe for valuable discussions on Monte-Carlo methods, and to Jens Rademacher and Matthias Wolfrum for valuable discussions on traveling fronts.

REFERENCES

[1] S. Amari, *Dynamics of pattern formation in lateral-inhibition type neural fields*, Biol. Cybernet., 27 (1977), pp. 77–87.

[2] F. M. Atay and A. Hutt, *Stability and bifurcations in neural fields with finite propagation speed and general connectivity*, SIAM J. Appl. Math., 65 (2005), pp. 644–666.
[3] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed., McGraw-Hill, New York, 1970.
[4] S. Bernard, J. Beláir, and M. C. Mackey, *Sufficient conditions for stability of linear differential equations with distributed delay*, Discrete Contin. Dyn. Syst. Ser. B, 1 (2001), pp. 233–256.
[5] J. R. Brasic, *Hallucinations*, Percep. Motor Skills, 86 (1998), pp. 851–877.
[6] P. C. Bressloff and S. Coombes, *Physics of the extended neuron*, Int. J. Mod. Phys. B, 11 (1997), pp. 2343–2392.
[7] V. Bringuier, F. Chavane, L. Glaeser, and Y. Fregnac, *Horizontal propagation of visual activity in the synaptic integration field of area* 17 *neurons*, Science, 283 (1999), pp. 695–699.
[8] C. Cavada and P. S. Goldman-Rakic, *Multiple visual areas in the posterior parietal cortex of primates*, Progr. Brain Res., 95 (1993), pp. 123–137.
[9] S. Coombes, *Waves and bumps in neural field theories*, Biol. Cybernet., 93 (2005), pp. 91–108.
[10] S. Coombes, G. J. Lord, and M. R. Owen, *Waves and bumps in neuronal networks with axo-dendritic synaptic interactions*, Phys. D, 178 (2003), pp. 219–241.
[11] S. Coombes and M. Owen, *Bumps, breathers, and waves in a neural network with spike frequency adaptation*, Phys. Rev. Lett., 94 (2005), paper 148102.
[12] S. Coombes and M. R. Owen, *Evans functions for integral neural field equations with Heaviside firing rate function*, SIAM J. Appl. Dyn. Syst., 3 (2004), pp. 574–600.
[13] S. M. Crook, G. B. Ermentrout, M. C. Vanier, and J. M. Bower, *The role of axonal delays in the synchronization of networks of coupled cortical oscillators*, J. Comput. Neurosci., 4 (1997), pp. 161–172.
[14] F. H. L. da Silva, J. E. Vos, J. Mooibroek, and A. van Rotterdam, *Relative contribution of intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis*, Electroencephalogr. Clin. Neurophysiol., 50 (1980), pp. 449–456.
[15] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, Cambridge, MA, 2001.
[16] R. Eckhorn, A. M. Gail, A. Bruns, A. Gabriel, B. Al-Shaikhli, and M. Saam, *Different types of signal coupling in the visual cortex related to neural mechanisms of associative processing and perception*, IEEE Trans. Neural Networks, 15 (2004), pp. 1039–1052.
[17] A. K. Engel, O. Koenig, A. K. Kreiter, and W. Singer, *Interhemispheric synchronization of oscillatory neuronal response in cat visual cortex*, Science, 252 (1991), pp. 1177–1179.
[18] G. B. Ermentrout and J. D. Cowan, *A mathematical theory of visual hallucination patterns*, Biol. Cybernet., 34 (1979), pp. 137–150.
[19] S. Folias and P. Bressloff, *Breathers in two-dimensional excitable neural media*, Phys. Rev. Lett., 95 (2005), paper 208107.
[20] W. J. Freeman, *Mass Action in the Nervous System*, Academic Press, New York, 1975.
[21] W. J. Freeman, *Tutorial on neurobiology: From single neurons to brain chaos*, Internat. J. Bifur. Chaos, 2 (1992), pp. 451–482.
[22] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossia, *GNU Scientific Library Reference Manual*, 2nd ed., Network Theory, Bristol, U.K., 2003.
[23] W. Gerstner, *Time structure of the activity in neural network models*, Phys. Rev. E, 51 (1995), pp. 738–758.
[24] P. Girard, J. M. Hupe, and J. Bullier, *Feedforward and feedback connections between areas* $V1$ *and* $V2$ *of the monkey have similar rapid conduction velocities*, J. Neurophysiol., 85 (2001), pp. 1328–1331.
[25] H. Haken, *Brain Dynamics*, Springer, Berlin, 2002.
[26] H. K. Hartline and F. Ratliff, *Inhibitory interaction in the retina of Limulus*, in Handbook of Sensory Physiology, Springer-Verlag, New York, 1971, Vol. VII/2, pp. 381–447.
[27] A. Hutt, *Effects of nonlocal feedback on traveling fronts in neural fields subject to transmission delay*, Phys. Rev. E, 70 (2004), paper 052902.
[28] A. Hutt and F. M. Atay, *Analysis of nonlocal neural fields for both general and gamma-distributed connectivities*, Phys. D, 203 (2005), pp. 30–54.
[29] A. Hutt and F. M. Atay, *Spontaneous and evoked activity in extended neural populations with gamma-distributed spatial interactions and transmission delay*, Chaos Solitons Fractals, (2007), to appear.

[30] A. Hutt, M. Bestehorn, and T. Wennekers, *Pattern formation in intracortical neuronal fields*, Network: Comput. Neural Syst., 14 (2003), pp. 351–368.
[31] V. K. Jirsa, *Connectivity and dynamics of neural information processing*, Neuroinformatics, 2 (2004), pp. 183–204.
[32] V. K. Jirsa and H. Haken, *Field theory of electromagnetic brain activity*, Phys. Rev. Lett., 7 (1996), pp. 960–963.
[33] T. Kapitula, N. Kutz, and B. Sandstede, *The Evans function for nonlocal equations*, Indiana Univ. Math. J., 53 (2004), pp. 1095–1126.
[34] J. Karbowski and N. Kopell, *Multispikes and synchronization in a large-scale neural network with delays*, Neural Comput., 12 (2000), pp. 1573–1606.
[35] R. D. Katznelson, *Normal modes of the brain: Neuroanatomic basis and a physiologic theoretical model*, in Electric Fields of the Brain: The Neurophysics of EEG, P. Nunez, ed., Oxford University Press, New York, 1981, pp. 401–442.
[36] C. Laing, *Spiral waves in nonlocal equations*, SIAM J. Appl. Dyn. Syst., 4 (2005), pp. 588–606.
[37] C. Laing and S. Coombes, *The importance of different timings of excitatory and inhibitory models*, Network: Comput. Neural Syst., 17 (2006), pp. 151–172.
[38] C. R. Laing and W. C. Troy, *PDE methods for nonlocal models*, SIAM J. Appl. Dyn. Syst., 2 (2003), pp. 487–516.
[39] G. P. Lepage, *A new algorithm for adaptive multidimensional integration*, J. Comput. Phys., 27 (1978), pp. 192–203.
[40] F. H. Lopes da Silva, W. Blanes, S. N. Kalitzin, J. Parra, P. Suffczynski, and D. N. Velis, *Epilepsies as dynamical diseases of brain systems: Basic models of the transition between normal and epileptic activity*, Epilepsia, 44 (2003), pp. 72–83.
[41] J. Milton and P. Jung, eds., *Epilepsy as a Dynamic Disease*, Springer-Verlag, New York, 2002.
[42] L. G. Morelli, G. Abramson, and M. N. Kuperman, *Associative memory on a small-world neural network*, Eur. Phys. J. B Condens. Matter Phys., 38 (2004), pp. 495–500.
[43] S. Neuschwander, M. Castelo-Branco, J. Baron, and W. Singer, *Feed-forward synchronization: Propagation of temporal patterns along the retinothalamocortical pathway*, Phil. Trans. R. Soc. London, 357 (2002), pp. 1869–1876.
[44] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms*, Oxford University Press, New York, Oxford, 1995.
[45] D. Pinault, *The thalamic reticular nucleus: Structure, function and concept*, Brain Res. Rev., 46 (2004), pp. 1–31.
[46] D. J. Pinto and G. B. Ermentrout, *Spatially structured activity in synaptically coupled neuronal networks:* I. *Traveling fronts and pulses*, SIAM J. Appl. Math., 62 (2001), pp. 206–225.
[47] T. Prager and L. Schimansky-Geier, *Stochastic resonance in a non-Markovian discrete state model for excitable systems*, Phys. Rev. Lett., 91 (2003), paper 230601.
[48] J. C. Prechtl, L. B. Cohen, B. Pesaran, P. P. Mitra, and D. Kleinfeld, *Visual stimuli induce waves of electrical activity in turtle cortex*, Proc. Natl. Acad. Sci. USA, 94 (1997), pp. 7621–7626.
[49] P. A. Robinson, P. N. Loxley, S. C. O'Connor, and C. J. Rennie, *Modal analysis of corticothalamic dynamics, electroencephalographic spectra and evoked potentials*, Phys. Rev. E, 63 (2001), paper 041909.
[50] P. A. Robinson, C. J. Rennie, and J. J. Wright, *Propagation and stability of waves of electrical activity in the cerebral cortex*, Phys. Rev. E, 56 (1997), pp. 826–840.
[51] P. A. Robinson, R. W. Whitehouse, and C. J. Rennie, *Nonuniform corticothalamic continuum model of encephalographic spectra with application to split-alpha peaks*, Phys. Rev. E, 68 (2003), paper 021922.
[52] E. Salinas and T. J. Sejnowski, *Correlated neuronal activity and the flow of neural information*, Nat. Rev. Neurosci., 2 (2001), pp. 539–550.
[53] W. Singer, *Neural synchrony: A versatile code for the definition of relations?*, Neuron, 24 (1999), pp. 49–65.
[54] W. Singer and C. M. Gray, *Visual feature integration and the temporal correlation hypothesis*, Ann. Rev. Neurosci., 18 (1995), pp. 555–586.
[55] M. Spiridon and W. Gerstner, *Effect of lateral connections on the accuracy of the population code for a network of spiking neurons*, Network: Comput. Neural Syst., 12 (2001), pp. 409–421.

[56] M. Steriade, E. G. Jones, and R. R. Llinas, *Thalamic Oscillations and Signalling*, Wiley, New York, 1990.
[57] M. Steriade, D. A. McCormick, and T. J. Sejnowski, *Thalamocortical oscillations in the sleeping and aroused brain*, Science, 262 (1993), pp. 679–685.
[58] V. V. Braitenberg and A. Schüz, *Anatomy of the Cortex*, Springer, Berlin, 1991.
[59] H. R. Wilson and J. D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophys. J., 12 (1972), pp. 1–24.
[60] J. J. Wright, *Simulation of EEG: Dynamic changes in synaptic efficacy, cerebral rhythms, and dissipative and generative activity in cortex*, Biol. Cybernet., 81 (1999), pp. 131–147.
[61] J. J. Wright and R. R. Kydd, *The electroencephalogram and cortical neural networks*, Network, 3 (1992), pp. 341–362.
[62] L. Zhang, *On stability of traveling wave solutions in synaptically coupled neuronal networks*, Differential Integral Equations, 16 (2003), pp. 513–536.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 699–725

Delayed-Mutual Coupling Dynamics of Lasers: Scaling Laws and Resonances*

T. W. Carr†, I. B. Schwartz‡, Min-Young Kim§, and Rajarshi Roy§

Abstract. We consider a model for two lasers that are mutually coupled optoelectronically by modulating the pump of one laser with the intensity deviations of the other. Signal propagation time in the optoelectronic loop causes a significant delay leading to the onset of oscillatory output. Multiscale perturbation methods are used to describe the amplitude and period of oscillations as a function of the coupling strength and delay time. For weak coupling the oscillations have the laser's relaxation period, and the amplitude varies as the one-fourth power of the parameter deviations from the bifurcation point. For order-one coupling strength the period is determined as multiples of the delay time, and the amplitude varies with a square-root power law. Because we allow for independent control of the individual coupling constants, for certain parameter values there is an atypical amplitude-resonance phenomena. Finally, our theoretical results are consistent with recent experimental observations when the inclusion of a low-pass filter in the coupling loop is taken into account.

Key words. coupled lasers, delay, Hopf bifurcation, resonance

AMS subject classifications. 34D15, 37G15, 39A11, 78A60

DOI. 10.1137/050647918

1. Introduction.

In recent work, we presented experimental and simulation results for two mutually coupled lasers [1] with time-delayed asymmetric coupling. The light emitted from one laser propagates through fiber-optic cables to a photodetector that generates an electronic signal proportional to the light-intensity deviations from steady state. The electronic signal may be attenuated or amplified before it modulates the pump current of the other laser. The propagation time of the signal in the optoelectronic path introduces a significant time delay, and the coupling strength in each direction can be controlled separately. Our work in [1] investigated how the time delay and asymmetric coupling led to oscillatory and pulsating laser output. In this paper, we present a more theoretical exploration of the dynamics that result from this coupling configuration.

In [2] we investigated this same configuration (asymmetric pump coupling) but without including the effect of delay. In addition, the coupling constant from laser-2 to laser-1 was negative, while the coupling constant from laser-1 to laser-2 was positive. That is, we assumed that the electronic coupling signal from laser-2 was inverted before applying it to the pump of laser-1. For purely harmonic signals, having opposite-sign coupling constants is equivalent

*Received by the editors December 18, 2005; accepted for publication (in revised form) by B. Krauskopf July 31, 2006; published electronically December 5, 2006.
http://www.siam.org/journals/siads/5-4/64791.html
†Department of Mathematics, Southern Methodist University, Dallas, TX 75275-0156 (tcarr@smu.edu).
‡Nonlinear Dynamical Systems Section, Code 6792, Naval Research Laboratory, Washington, DC 20375 (schwartz@nlschaos.nrl.navy.mil).
§Department of Physics, University of Maryland, College Park, MD 20742 (mmyykim@cnd.mcgill.ca, rroy@glue.umd.edu).

to time-delayed coupling when the delay is half of the period. We found that as the coupling constant from laser-2 to laser-1 is increased in magnitude, (i) there is a Hopf bifurcation to oscillatory output, (ii) for certain parameter values there exists a small-coupling resonance such that the amplitudes of both lasers are pulsating, and (iii) for large coupling, laser-2 acts as a small-amplitude, nearly harmonic modulation to laser-1. Laser-1 exhibits period-doubling bifurcations to chaos and complex subharmonic resonances.

In this paper we include the effect of the delayed coupling, which results in a coupled set of delay-differential equations (DDEs). With delay there is again a Hopf bifurcation to oscillatory output. However, the delay allows for periodic oscillations not just at the laser's relaxation period as in [2], but at periods that are integer multiples of the delay; we refer to the former oscillatory solutions as "internal modes" and to the latter as "external modes." We show that as the coupling strength is increased, the first instability is a Hopf bifurcation to the internal mode. Our nonlinear analysis then shows that the amplitude of the internal mode varies as the one-fourth root of the bifurcation parameter's deviation from the Hopf bifurcation point.

However, in the experiments of [1] the output was not the internal mode but an external mode. The period of the oscillatory output was a multiple of the delay, and the amplitude of the oscillations varied as the square root of the bifurcation parameter. The discrepancy is due to the fact that the optoelectronic coupling contains an intrinsic low-pass filter that attenuates the internal mode. We show that the filter selects the external mode with the greatest possible frequency passed by the low-pass filter. A nonlinear analysis then shows that the amplitude of the external modes does indeed vary with a square-root power law.

For the coupled lasers without delay in [1], we found that the parameters could be tuned to cause a resonance-type effect with respect to the coupling; more specifically, the coupling could be tuned to maximize the amplitude of the oscillatory output. We find that this resonance effect also occurs with the inclusion of the delay. However, there are some dramatic qualitative differences as the coupling parameters are varied. In the system with delay, the branch of periodic solutions in the bifurcation diagram can be smoothly folded to form parameter intervals of bistability. Then there is a critical value of the same coupling parameter beyond which the bistability disappears suddenly and nonsmoothly.

Our coupling scheme is an example of "incoherent coupling" [3] because it depends only on the laser's intensity and not on the complex electric field. This is because the intensity of one laser affects the other only indirectly; in our case, the intensity of one laser modulates the pump current of the other. In contrast, "coherent coupling" refers to when the optical field of one laser is directly injected into the cavity of the other laser. Analogously, a single laser with delayed and reinjected self-feedback would be called "coherent" feedback. Semiconductor lasers with delayed coherent feedback have been extensively studied because of the their widespread application in electronics and communication systems; there the laser's output signal may be reflected off external surfaces back into the laser. The Lang–Kobayashi DDEs [4] are the canonical model for semiconductor lasers with delayed self-feedback and have been used to investigate phenomena ranging from the onset of instabilities to the development of chaotic output referred to as "coherence collapse" (see [5] for a review). Coupled sets of Lang–Kobayashi or related DDEs are most often used to described coherently coupled semiconductor lasers with delay. Two recent studies [6, 7] are in the same spirit as this paper as they track

the number of and properties of oscillatory solutions that appear as the coupling strength is increased.

The main simplification that results from considering incoherent coupling between the lasers is that neither the laser-frequency detuning nor the electric-field phase differences affect the dynamics of the coupled system [8]. Pieroux, Erneux, and Otsuka [9] have shown that the electro-optical coupling we consider in this paper leads to an equivalent dynamical model as incoherent coupling. There have recently been a number of other investigations of lasers with incoherent coupling. The main focus of many of these works was on chaotic synchronization (see [10] and included references). Two recent papers by Vicente et al. [11, 12] consider an implementation of optoelectronic coupling similar to that we consider here; i.e., the intensity of one laser modulates the pump current of the other. Their investigation of the oscillatory solutions is mainly numerical and they discuss interesting phenomena such as amplitude quenching, frequency locking, and routes to chaos.

Delay-coupled relaxation and limit-cycle oscillators have been the subject of many investigations (see [13, 14, 15, 16, 17, 18] and their included references). These systems are self-oscillatory and the amplitude is often fixed by the intrinsic properties of the oscillator. In many cases the system may be reduced to a phase equation, or a system of phase equations if the oscillators are coupled [19].

In contrast, for laser systems the amplitude is an important dynamical variable. This is because the lasers we consider are weakly damped, nearly conservative systems such that the steady state is the only asymptotically stable solution. Oscillations must be induced by external mechanisms such as modulation, injection, or coupling [20]. Thus, the amplitude of the oscillations is highly dependent upon the external mechanism, in our case the coupling, rather than the individual laser. We should mention that for coupled limit-cycle oscillators there is the amplitude instability referred to as "amplitude death," where for specific values of the coupling (and oscillator parameter values) the amplitude becomes zero.

After presenting our model, we show that the linear-stability analysis predicts a critical value of the coupling constant (that depends upon the delay time) such that there is Hopf bifurcation to the internal mode, i.e., oscillations with the intrinsic relaxation period. We then use multiple-scale perturbation methods, modified to account for time delays [21], to analyze the long-time evolution of the internal mode. The results are a pair of complex Stuart–Landau DDEs for the oscillation amplitudes that include a slow-time delay term. We analyze the amplitude equations to determine the bifurcation properties of the internal mode. Of particular note is that we allow for independent control of the coupling constants; most other studies of coupled lasers and oscillators consider symmetric coupling, where the coupling is the same across all of the elements. The independent control is important because, as we will show, it allows for a singularity in the bifurcation equations that marks the "resonance" of large amplitude solutions.

The Hopf bifurcation to the internal mode occurs for small coupling. However, as the coupling is increased, the external modes appear via Hopf bifurcations. We have extended our multiple-scale analysis to be able to describe the bifurcation of the external modes that occur for $O(1)$ coupling. The analytical challenge is that the delay terms remain a part of the leading-order problem. Describing the bifurcation of the external modes is important for comparing our results to those in experiments [22], because the experimental system contains

low-pass filters that attenuate the internal mode originating from the first Hopf bifurcation.

An alternative to our multiple-scale method for deriving slow-time amplitude equations is to apply center-manifold theory with averaging [23, 14, 15]. However, the averaging method does not retain the delay in the slow time; that is, the amplitude equations are ordinary differential equations—not DDEs. The distinction is not important for their or our investigation because we look for steady-state solutions of the amplitude equations. However, time-varying amplitudes may require consideration of the slow-time delay.

Reddy, Sen, and Johnston [16, 24] considered delay-coupled Stuart–Landau DDEs similar to the amplitude equations we derive here. However, the complex coefficients they use are appropriate for limit-cycle oscillators and not the lasers that we consider. They focus their work on the properties of synchronization and amplitude death.

In the next section, we nondimensionalize the model for two-coupled lasers, with the result being the focus of the rest of the paper. We perform the linear-stability analysis in section 3. Slow-time evolution equations for small and $O(1)$ coupling are derived in section 4 and section 5, respectively. We close with a discussion of the results.

2. Class-B model.

We consider two class-B lasers [20, 25] modeled by rate equations as

$$\begin{aligned} \frac{dI_j}{dt} &= (D_j - 1)I_j, \quad j = 1, 2, \\ \frac{dD_j}{dt} &= \epsilon_j^2[A_j(t) - (1 + I_j)D_j], \end{aligned} \tag{2.1}$$

where I_j and D_j are the scaled intensity and population inversion, respectively, of each laser. Scaled time t is measured with respect to the cavity-decay constant κ, $t = (2\kappa)\, t_r$, where t_r measures real time. $\epsilon^2 = \gamma_\parallel/(2\kappa)$ is a ratio of the inversion-decay rate, $\gamma_\parallel$, to the cavity-decay rate, κ, and measures the relative lifetime of photons to excited electrons. $A(t)$ is the dimensionless pump rate and corresponds to the energy input to the laser by an external source, e.g., another laser, an incoherent light source, or an electronic current. The mass-action coupling term ID models "stimulated emission"; a photon passing through the laser cavity stimulates an excited electron to drop to the lower energy level, resulting in one less excited electron and one more photon. Equations (2.1) may be derived as a reduced model from semiclassical Maxwell–Bloch equations [20, 25, 26]; the latter use Maxwell's equations to describe the laser's electromagnetic fields coupled with the Bloch equations from quantum mechanics for the amplifying media.

If the pump rate is a constant, $A_j(t) = A_{j0}$, then the laser relaxes to the steady state $D_{j0} = 1$, $I_{j0} = A_{j0} - 1$. To facilitate further analysis, we define new variables for the deviations from the nonzero steady state [27] as

$$I_j = I_{j0}(1 + y_j), \quad D_j = 1 + \epsilon_j \sqrt{I_{j0}}\, x_j, \quad s = \epsilon_1 \sqrt{I_{10}}\, t. \tag{2.2}$$

Our goal is to investigate the effects of coupling through the pump. In addition, we account for the effect of a delay when the signal from one laser takes a nonnegligible time before affecting the other. Thus, the pump coupling is taken to be

$$A_j(t) = A_{j0} - I_{j0}\delta_k y_k(t - \tau_k). \tag{2.3}$$

Thus, we feed the intensity *deviations* $y_k = (I_k - I_{k0})/I_{k0}$ from the steady-state output of laser k to the pump of laser j; the strength of the coupling is controlled by δ_k. The pump-coupling scheme allows for easy electronic control of the coupling signal. The negative coupling results from the configuration of the electronic coupling circuits in [1]. Finally, we assume that the decay constants of the two lasers are related by $\epsilon_2 = \epsilon_1 \frac{\sqrt{I_{10}}}{\sqrt{I_{20}}}\beta$. The new rescaled equations are

$$\begin{aligned}
\frac{dy_1}{dt} &= x_1(1+y_1),\\
\frac{dx_1}{dt} &= -y_1 - \epsilon x_1(a_1 + by_1) - \delta_2 y_2(t-\tau_2),\\
\frac{dy_2}{dt} &= \beta x_2(1+y_2),\\
\frac{dx_2}{dt} &= \beta[-y_2 - \epsilon\beta x_2(a_2 + by_2) - \delta_1 y_1(t-\tau_1)],
\end{aligned} \tag{2.4}$$

where

$$a_1 = \frac{1+I_{10}}{\sqrt{I_{10}}}, \quad a_2 = \frac{\sqrt{I_{10}}(1+I_{20})}{I_{20}}, \quad \text{and } b = \sqrt{I_{10}}. \tag{2.5}$$

For notational convenience we have let $s \to t$ and dropped the subscript on ϵ_1 ($\epsilon_1 \to \epsilon$). Rogister et al. [28] have considered a model very similar to ours; however, in their case the cross-coupling term is instantaneous, while the delay appears through self-feedback of each laser's own intensity.

3. Linear stability of the steady state.

3.1. Characteristic equation. In the new variables, the steady state is given by $x_j = y_j = 0$. The linear stability of the steady state is determined by studying the evolution of small perturbations, for which we obtain the characteristic equation

$$[\lambda(\lambda + \epsilon a_1) + 1][\lambda(\lambda + \epsilon a_2\beta^2) + \beta^2] - \beta^2\delta_1\delta_2 e^{-\lambda(\tau_1+\tau_2)} = 0. \tag{3.1}$$

The delay results in the exponential term $\exp(-\lambda(\tau_1 + \tau_2))$. The transcendental form of the characteristic equation and, hence, the possibility of an infinite number of roots is typical for DDEs.

If either $\delta_j = 0$, then there are only decaying oscillations, which indicates that a single uncoupled laser is a weakly damped oscillator; this is the general characteristic of "class-B" lasers [25]. For $\delta_j \neq 0$ numerical simulations indicate there is a Hopf bifurcation as the δ_j are increased. To identify the Hopf bifurcation point, we let $\lambda = i\omega$ and separate the characteristic equation into real and imaginary parts. After some algebra we obtain a single equation for the frequency and another equation for the value of $\Delta = \delta_1\delta_2$ at the bifurcation point. The equations are

$$\begin{aligned}
0 = \tan(\omega\tau_s)&[(1-\omega^2)(\beta^2-\omega^2) - \epsilon^2 a_1 a_2 \beta^2\omega^2]\\
&+ \epsilon\omega(1-\omega^2)(a_1 + a_2\beta^2),
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
\beta^2\Delta_H^2 = &[(1-\omega^2)(\beta^2-\omega^2)\\
&- \epsilon^2 a_1 a_2 \beta^2\omega^2]^2 + [\epsilon\omega(1-\omega^2)(a_1 + a_2\beta^2)]^2,
\end{aligned} \tag{3.3}$$

where $\tau_s = \tau_1 + \tau_2$ is the "round-trip" delay time. The transcendental equation for ω can be solved numerically and its solution is then substituted into (3.3) to determine the value of the coupling at the bifurcation point $\Delta = \Delta_H$.

To simplify further discussion we consider $\beta = 1$. This implies $\epsilon_2 = \epsilon_1\sqrt{I_{10}/I_{20}}$, which is a specific relationship between the lasers' decay constants, $\gamma_{\|,j}$ and κ_j, and the pump rate, A_j0; if $\beta = 1$ and $A_{10} = A_{20}$ then the lasers are identical. The results are qualitatively the same for nearly identical lasers with $\beta \approx 1$. To simplify notation, we define $c_1 = a_1 + a_2$ and $c_2 = a_1a_2$ and have

$$0 = \tan(\omega\tau_s)[(1-\omega^2)^2 - \epsilon^2 c_2\omega^2] + \epsilon\omega(1-\omega^2)c_1, \tag{3.4}$$

$$\Delta_H^2 = [(1-\omega^2)^2 - \epsilon^2 c_2\omega^2]^2 + [\epsilon\omega(1-\omega^2)c_1]^2. \tag{3.5}$$

For $\epsilon \ll 1$ the leading approximation to the frequency is given by $\omega = \pm 1$ or $\omega = m\pi/\tau_s$, m an integer. We will refer to the former as the *internal mode* because this is the scaled laser-relaxation frequency; more specifically, an oscillatory solution of (2.4) with period $\omega = 1$ is called the internal mode. Similarly, oscillatory solutions with period $\omega = m\pi/\tau_s$, m an integer, are called *external modes* because their periods are determined by the delay. For any fixed value of the delay τ_s there is the internal mode and an infinite number of external modes.

From (3.5) we see that for almost all values of the delay, as the coupling is increased from zero, the bifurcation to the internal mode occurs at a lower value of the Δ_H than the external mode (this is true except when $\tau_s = n\pi$, n even, which we will discuss in later paragraphs). More specifically, for the internal mode with $\omega \approx 1$ the Hopf bifurcation occurs when $\Delta_H = O(\epsilon^2) \ll 1$. For the external mode with $\omega = O(1)$ the value of the coupling at the Hopf bifurcation is $\Delta_H = O(1)$. This is illustrated in Figure 1, where we have plotted (3.5). We will determine more precise approximations for these bifurcation points below; however, we will first make additional qualitative observations.

Let us fix the delay to be a multiple of π, i.e., $\tau_s = n\pi$, n an integer. Then $\omega = \pm 1$, corresponding to the internal mode, is an exact solution to the frequency equation (3.4) and $\Delta_H = \pm\epsilon^2 c_2$. The $+$ is taken if n is odd and the $-$ is taken if n is even (the sign association comes from the real and imaginary parts of the characteristic equation before ω and Δ were separated). In the present paper, we will consider $\delta_j > 0$ and $\Delta > 0$; results for $\Delta < 0$ are qualitatively the same. Thus, we have the exact solution $\omega = \pm 1$, $\Delta_H = +\epsilon^2 c_2$ when $\tau_s = n\pi$, n odd. An analysis of (3.5) shows this occurs at a minimum in the curve $\Delta_H(\omega)$ as illustrated in Figure 1.

If τ_s is *not* a multiple of π, say, $n\pi < \tau_s < (n+1)\pi$, the Hopf bifurcation to the internal mode still occurs before the bifurcation to any of the external modes. The frequency of the internal mode is $\omega \approx 1$ (instead of $\omega = 1$), and $\Delta_H > \epsilon^2 c_2$ but still $O(\epsilon^2)$. Thus, if we plot Δ_H as a function of the delay τ_s, the *minimum* of the curves $\Delta_H(\tau_s)$ occurs when $\tau_s = n\pi$, n odd. This is illustrated in Figure 3 and will be discussed further below.

Now consider the external modes with $\omega \approx m\pi/\tau_s$. For negative couplings when $\delta_j > 0$ (see (2.3)) and $\Delta > 0$ an analysis of the characteristic equation shows that m *must be even.* Thus, the external modes have frequencies $\omega \approx 2\pi/\tau_s$, $4\pi/\tau_s$, etc. In Figure 1 we illustrate three cases when $\tau_s = n\pi$, $n = 1$, 5, and 53. The external modes then have $\omega = m/n$, m even.

Finally, we consider the case when $\tau_s = n\pi$ for n *even.* As discussed in the previous paragraph, for negative couplings the external modes with frequency $\omega = m\pi/\tau_s$ require that

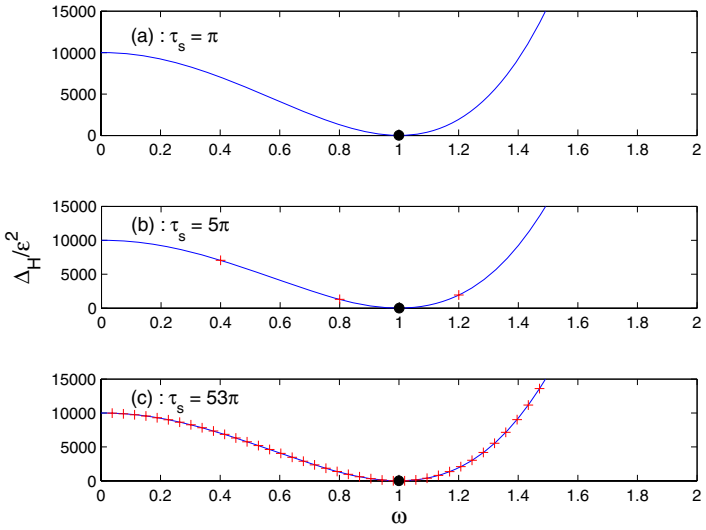


Figure 1. *Neutral stability curve $\Delta_H(\omega)$ (solid curve) for Hopf bifurcations described by (3.5). The solid circle at $\omega \approx 1$ indicates the internal mode. The crosses (+) at $\omega \approx m\pi/\tau_s$ indicate the external modes. Because $\tau_s = n\pi$, then $\omega \approx m/n$ for the external modes ($\epsilon = 0.01$, $a_1 = a_2 = 2$, $b = 1$, $\beta = 1$).*

$m = 2, 4, 6 \ldots$. Thus, for each external mode there is a value of $\tau_s = n\pi$, n even, such that $\omega \approx 1$ similar to the internal mode. This is illustrated in the numerical data of Figure 3a for $\tau_s = 2\pi$, while in Figure 3b we see that the bifurcations to these two modes occur at nearly the same value of the coupling. Referencing Figure 1, the two modes would have frequencies on opposite sides of $\omega = 1$ such that Δ_H is the same for each. This is referred to as a "double Hopf" [21, 29] and may lead to more complicated bifurcation scenarios that we will not pursue in the present paper. We also note that at the double Hopf bifurcation, the designation of "internal mode" versus "external mode" is ambiguous because both have $\omega \approx 1$. Indeed, following the numerical data in Figure 3a, as τ_s is varied through 2π, we see that the $n = 1$ internal mode becomes the $m = 2$ external mode, while the $m = 2$ external mode becomes the $n = 3$ internal mode.

3.2. Approximations to internal and external modes. As discussed above, for any delay τ_s when the coupling is increased, the first periodic solutions to appear correspond to the internal mode and will have frequency $\omega \approx 1$. For arbitrary delay we can find an approximation for the frequency by letting $\omega = 1 + \epsilon\omega_1 + O(\epsilon^2)$ in (3.4) to find that ω_1 satisfies

$$\tan(\tau_s)[4\omega_1^2 - c_2] - 2c_1\omega_1 = 0, \tag{3.6}$$

which is a simple quadratic for ω_1. The bifurcation point is then approximated as

$$\begin{aligned} \Delta_H^2 &= \epsilon^4[16\omega_1^4 + 4(c_1^2 - 2c_2)\omega_1^2 + c_2^2] \\ &= \epsilon^4(4\omega_1^2 + a_1^2)(4\omega_1^2 + a_2^2). \end{aligned} \tag{3.7}$$

We can obtain simpler results than (3.6) and (3.7) if we require that $\tau_s \approx n\pi$, n odd; that is, in Figure 2 we examine locally to the critical values $\omega(\tau_s) = 1$ and the minimum of the

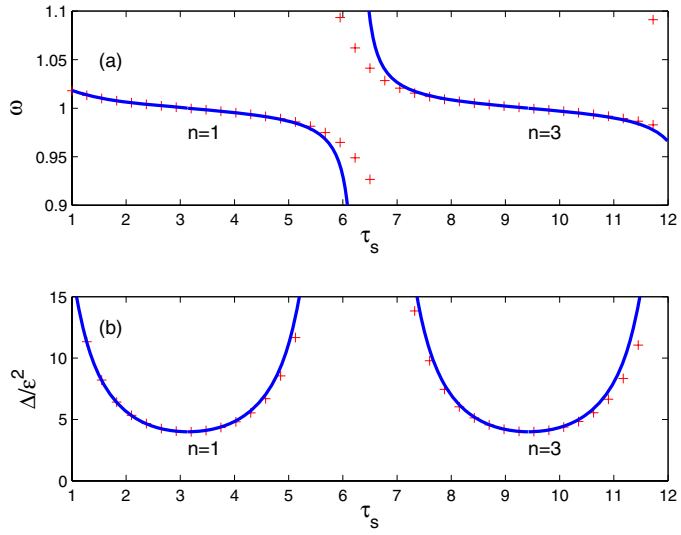


Figure 2. *Frequency ω and coupling $\Delta = \delta_1\delta_2$ at the Hopf bifurcation point as a function of the delay $\tau_s = \tau_1 + \tau_2$. The solid curve is the asymptotic results of (3.6) and (3.7), while the + are the result of numerically evaluating (3.4) and (3.5) ($\epsilon = 0.01$, $a_1 = a_2 = 2$, $\beta = 1$).*

curves $\Delta_H(\tau_s)$. If $\tau_s = n\pi + \epsilon\tau_{s1}$, then

$$\omega = 1 - \epsilon^2 \frac{c_2}{2c_1}\tau_s + O(\epsilon^2), \quad \Delta_H^2 = \epsilon^4 c_2^2 + O(\epsilon^5). \tag{3.8}$$

For the external modes we let $\omega = \omega_0 + \epsilon\omega_1 + O(\epsilon^2)$ with $\omega_0 = m\pi/\tau_s$, m even, and find that

$$\omega = \omega_0 - \epsilon \frac{\omega_0 c_1}{(1-\omega_0^2)\tau_s} + \dots \tag{3.9}$$

and the Hopf bifurcation point is approximated as

$$\Delta_H^2 = (1-\omega_0^2)^4 + \epsilon \frac{8c_1}{\tau_s}(1-\omega_0^2)^2\omega_0^2 + \dots. \tag{3.10}$$

The external mode results are valid for τ_s such that $\omega_0 \neq 1$; thus, τ_s must be away from the double Hopf point when $\tau_s = m\pi$.

In Figures 2 and 3 we plot the linear stability given by (i) numerically evaluating (3.4) and (3.5) (marked with +) and (ii) the asymptotic approximations from (3.6)–(3.7) and (3.9)–(3.10) (solid curves). Figure 2 focuses on the bifurcation of the internal mode. For any fixed value of the delay time τ_s, we increase Δ until the curve is crossed at $\Delta = \Delta_H$; the frequency is $\omega \approx 1$. In Figure 3 we pan out so that both the internal and external modes are visible. From here we see that as Δ is increased, the bifurcation to the internal mode always occurs first. As Δ is increased further, we cross secondary curves indicating the bifurcations to the external modes. More specifically, suppose that $\tau = \pi$ ($n = 1$). After the internal mode, the external

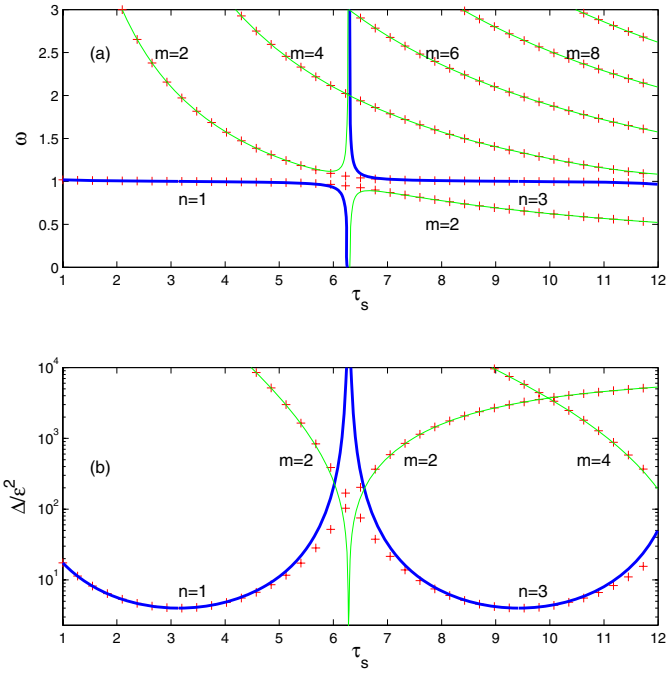


Figure 3. *Same as Figure 2 but with a wider range in ω and Δ. The + indicate the numerical results. The thick curves are the analytical approximations of the internal modes (same as Figure 2). The thin curves are the approximations of the external modes.*

modes with frequencies $\omega = m/n = m/1$, $m = 2, 4, \ldots$, appear in sequence. For longer delay times, the order in which the external modes appear will be different; this can be seen in Figure 3 for $\tau_s \approx 11$, where the bifurcation to the external mode with $m = 4$ occurs before the mode with $m = 2$. Alternatively, in Figure 1b when $\tau_s = 5\pi$, as Δ is increased the bifurcation to the periodic solutions occurs in the order $\omega \approx 1, \approx 0.8, \approx 1.2, \approx 0.4$, which corresponds to the internal mode, followed by the external modes $\omega = m/n$, $n = 5$, $m = 4, 6, 2, \ldots$.

4. Hopf bifurcation of the internal mode.

4.1. Two-time scales. We use the method of multiple scales to analyze the oscillatory solutions that appear at the Hopf bifurcation points. For the uncoupled lasers, oscillations decay on an $O(\epsilon)$-time scale, which suggests that we introduce the slow time $T = \epsilon t$; time derivatives become $\frac{d}{dt} = \frac{\partial}{\partial t} + \epsilon \frac{\partial}{\partial T}$. We analyze the nonlinear problem using perturbation expansions in powers of $\epsilon^{1/2}$, e.g., $x_j(t) = \epsilon^{1/2} x_{j1}(t, T) + \epsilon x_{j2}(t, T) + \ldots$; the relevant nonlinear terms and the slow-time derivative then balance at $O(\epsilon^{3/2})$. Finally, for simplicity we set $\beta = 1$ indicating identical lasers; however, relaxing this assumption to nearly identical lasers ($\beta = 1 + O(\epsilon)$) does not qualitatively change the results.

We now consider the effect of the two-time scale assumption on the delay term. With the

additional slow time, the delay term becomes

$$y_j(t-\tau) \to y(t-\tau, T-\epsilon\tau). \tag{4.1}$$

If $\delta_j \ll 1$, then consideration of the delay term is postponed until higher order and appears as part of the solvability condition for the slowly varying amplitude. However, if $\delta_j = O(1)$, then the leading-order problem will contain the delay term and analytical progress is much more difficult.

If $\tau \gg O(1/\epsilon)$ (large delay), then $\epsilon\tau = O(1)$ and the delay term must be retained in the slow argument. However, if $\tau = O(1)$, then the slow argument can be expanded as [21]

$$y_j(t-\tau) = y_j(t-\tau, T) - \epsilon\tau \frac{\partial}{\partial T} y_j(t-\tau, T) + \dots. \tag{4.2}$$

Now the delay of the slow argument is postponed to higher order. We shall see that this leads to simpler slow-evolution equations for $\delta \ll 1$ (see (4.5)) and is necessary to make any progress at all for $\delta = O(1)$ (see section 5).

It should be noted that care must be taken when using a series expansion of a delay term in a differential equation. The Taylor series may itself be justified but using it can change the stability of the differential equation. A simple example is given in [30], while [31, 32] provide more theoretical discussions concerning restrictions on the size of the delay. In our presentation we will check the validity of our approximations by comparing our analytical and numerical results.

4.2. Bifurcation equation. From the linear-stability analysis, we know that the first bifurcation will be to the internal mode with $\omega \approx 1$ when $\Delta = \delta_1\delta_2 = O(\epsilon^2)$ (we assume that $\tau_s \neq n\pi$, n even, and thus do not consider the case of the double-Hopf bifurcation). We present the case that both coupling constants are of the same order $\delta_j = \epsilon d_j$. However, as long as $\Delta = O(\epsilon^2)$ the bifurcation results are qualitatively the same.

Proceeding with the multiple-scale analysis, we find that at the leading order, $O(\epsilon^{1/2})$, we obtain the solutions

$$y_{j1}(t,T) = B_j(T)e^{it} + c.c., \quad x_{j1}(t,T) = iB_j(T)e^{it} + c.c., \tag{4.3}$$

which exhibit oscillations with radial frequency 1 on the t time scale. To find the slow evolution of $B_j(T)$, we must continue the analysis to $O(\epsilon^{3/2})$. Then, to prevent the appearance of unbounded secular terms, we determine "solvability conditions" for the $B_j(T)$. Due to the scalings of δ_j and y_j, the delay terms $\delta_j y_j(t-\tau_j, T-\tau_{\epsilon,j})$ are $O(\epsilon^{3/2})$ and contribute to the solvability condition. The final result is

$$\frac{\partial B_j}{\partial T} = -\frac{a_j}{2}B_j(T) - \frac{i}{6}|B_j(T)|^2 B_j(T) + \frac{i}{2}d_k B_k(T-\tau_{\epsilon,k})e^{-i\tau_k}, \tag{4.4}$$

$j,k = 1,2$, $j \neq k$, and where $\tau_{\epsilon,k} = \epsilon\tau_k$. The effect of the delay in the slow time appears explicitly in the delay terms $B_k(T-\tau_{\epsilon_k})$. The delay in the fast time has resulted in the exponential terms $e^{-i\tau_k}$. Equation (4.4) is valid for arbitrary values of the delay (we have not simplified the delay term $B_k(T-\tau_{\epsilon,k})$). However, according to (4.2), if $\tau = O(1)$, then

$\tau_{\epsilon,k} = O(\epsilon)$ and delay in the slow time is postponed until higher order. This results in a simpler solvability condition where the delay in the slow time does not appear.

$$\frac{\partial B_j}{\partial T} = -\frac{a_j}{2}B_j - \frac{i}{6}|B_j|^2 B_j + \frac{i}{2}d_k B_k e^{-i\tau_k}, \quad j,k = 1,2, \quad j \neq k. \tag{4.5}$$

Periodic solutions to the original laser equations, (2.4), correspond to T-independent solutions to the solvability conditions. The conditions are the same for both (4.4) and (4.5) because the delay terms in (4.4) become constants. The full T-dependent solvability conditions, including delay, are required only to analyze the stability of the periodic solutions.

To determine the amplitude and phase of the periodic solutions, we let $B_j(T) = R_j(T)e^{i\theta_j(T)}$, define the phase difference $\psi = \theta_2 - \theta_1$, and set the time derivatives to zero. We obtain

$$\begin{aligned} 0 &= -a_1 R_1 - d_2 R_2 \sin(\psi - \tau_2), \\ 0 &= -a_2 R_2 + d_1 R_1 \sin(\psi + \tau_1), \\ 0 &= -\frac{1}{3}(R_2^2 - R_1^2) + d_1 \frac{R_1}{R_2}\cos(\psi + \tau_1) - d_2 \frac{R_2}{R_1}\cos(\psi - \tau_2). \end{aligned} \tag{4.6}$$

We find that

$$\tan(\psi) = \frac{a_2 d_2 S_2 \sin(\tau_2) - a_1 d_1 S_1 \sin(\tau_1)}{a_2 d_2 S_2 \cos(\tau_2) + a_1 d_1 S_1 \cos(\tau_1)}, \tag{4.7}$$

where $S_1 = R_1^2$ and $S_2 = R_2^2$ and the amplitudes are found from the implicit equations

$$\begin{aligned} a_2^2 d_2^2 S_2^2 + a_1^2 d_1^2 S_1^2 + 2a_1 a_2 d_1 d_2 \cos(\tau_s) S_1 S_2 - d_1^2 d_2^2 \sin^2(\tau_s) S_1 S_2 = 0, \\ d_1 d_2 \sin(\tau_s) S_2 (S_1 - S_2) + 3 d_1 d_2 \cos(\tau_s)(a_2 - a_1) S_1 S_2 \\ + 3(a_1 d_1^2 S_1^2 - a_2 d_2^2 S_2^2) = 0. \end{aligned} \tag{4.8}$$

Of particular note is that just as in the linear-stability analysis, the amplitudes of the periodic solutions depend on the sum of the delays $\tau_s = \tau_1 + \tau_2$. This corresponds to the effective round-trip time of information from laser-j returning to laser-j.

In Appendix A we present explicit solutions of (4.8) that specify $S_j = R_j^2$ as a function of the parameters and the delay τ_s. The expressions are rather complicated but are easily evaluated numerically. To obtain simpler expressions, we consider $\tau_s = (n + \xi)\pi$, where $\xi \ll 1$ and n an odd integer. That is, we tune the delay time τ_s to be near one of the minimums of the neutral stability curves in Figure 2. The bifurcation equation is

$$R_2^2 = 3(a_1 + a_2)\frac{\sqrt{D_1}}{D_2} + \xi Z_{21} + O(\xi^2), \tag{4.9}$$

where

$$D_1 = \frac{d_1 d_2}{a_1 a_2} - 1, \quad D_2 = \frac{a_2 d_2}{a_1 d_1} - 1, \tag{4.10}$$

and Z_{21} is specified in Appendix A.

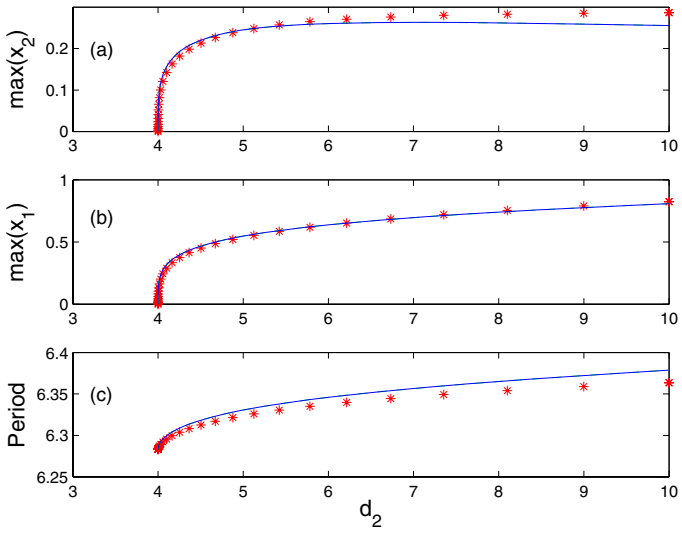


Figure 4. *Internal mode, $n = 5$ with $\tau_1 = \tau_2 = n\pi/2$, and $\epsilon = 0.01$. Numerical data points indicated by $*$. The solid curve is the bifurcation equation determined by (A.6)–(A.8) for $\xi = 0$, or (A.1)–(A.5). Because $\tau_s = \tau_1 + \tau_2 = n\pi$, $\xi = 0$ so the bifurcation equations are equivalent ($d_1 = 1$, $a_1 = a_2 = 2$, $b = 1$, $\beta = 1$).*

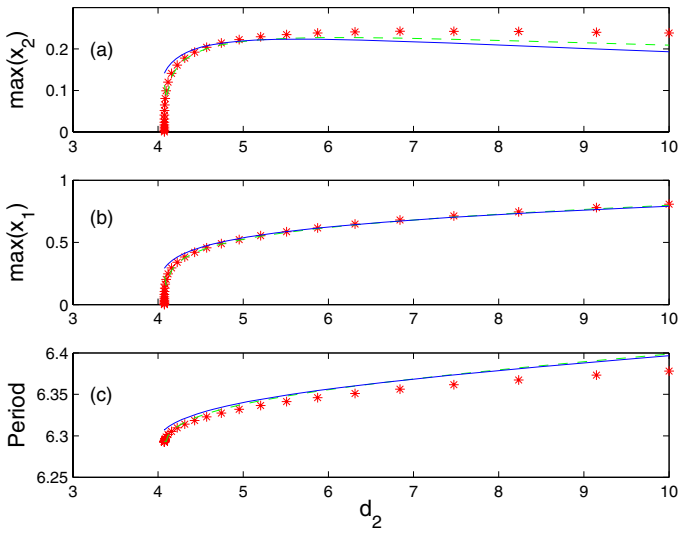


Figure 5. *Same as Figure 4 ($n = 5$) except that $\tau_1 = \tau_2 = (n + 0.1)\pi/2$. Thus, $\tau_s = (n + 0.1)\pi$, $\xi = 0.1\pi$, and the delay is tuned to the right of the minimum of the neutral-stability curve. The solid curves are the approximate bifurcation equations based on $\xi \ll 1$ given by (A.6)–(A.8). The dashed curves are the more general bifurcation equations given by (A.1)–(A.5). All other parameters are the same as in Figure 4.*

We compare the results of numerically computed bifurcation diagrams [33] and our analytical results in Figures 4–6. In Figures 4 and 6, $\tau_s = 5\pi$ and 53π, respectively, and hence

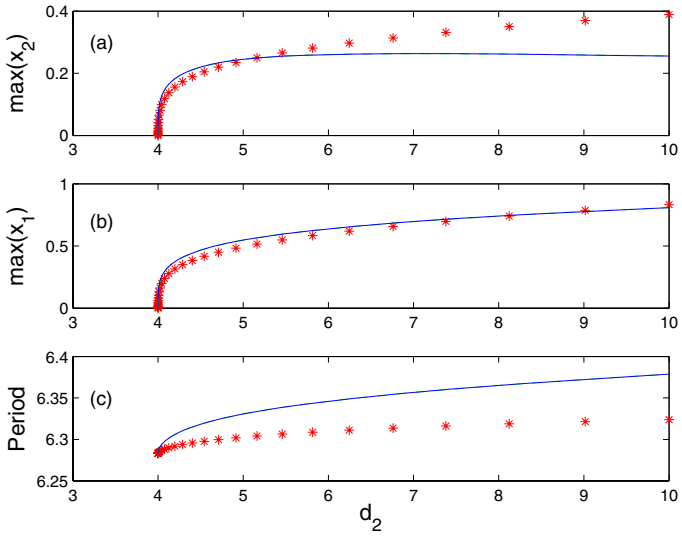


Figure 6. *Same as Figure 4 except that* $n = 53$.

$\xi = 0$ in (4.9). The analytical bifurcation equation describes the numerical results quite well even for d_2 far away from the bifurcation point. An expansion of (4.9) near the bifurcation point shows that

$$R_2 \sim (\Delta - \Delta_H)^{1/4}, \quad \Delta = \delta_1 \delta_2. \tag{4.11}$$

We also note that (4.9) is identical to the bifurcation equation we derived in [2] for two coupled lasers with opposite sign coupling. The coupled laser equations in [2] are identical to (2.4) except that in [2] (i) there is no delay such that the coupling is instantaneous, and (ii) we allowed the sign of one of the coupling constants to be positive. Near the bifurcation point, the small-amplitude periodic oscillations are nearly harmonic as indicated by (4.3). Thus, a positive coupling constant corresponds to an effective phase shift of half the period or, more generally, a phase shift of $\tau_s = n\pi$ (see Appendix B). Thus, we expect that when τ_s is tuned to the minimum of the neutral stability curves in Figure 2, the bifurcation equations of the two cases should be equivalent.

In Figure 5 we show the case when $\xi \neq 0$. Here we have tuned the delay τ_s to be to the right of the minimum with $\tau_s = (5 + 0.1)\pi$. The dashed curve is the result of numerically evaluating the full bifurcation equations, (4.8). The solid curve is the approximation (4.9) and it shows very good agreement with the numerical results. In general, we maintain good agreement as we tune τ_s away from the minimum up to $\tau_s = (n \pm 0.15)\pi$, n odd. The full bifurcation equation, (4.8), where we have not made any assumption on the delay, maintains good agreement out to at least $\tau_s = (n \pm 0.4)\pi$; beyond that, however, it loses fidelity as we approach the double-Hopf bifurcations at $\tau_s = (n \pm 1)\pi$ ($\tau_s = n\pi$, n even).

Finally, we note that the first correction to the bifurcation equation (4.9) is linear in ξ. Thus, the bifurcation equation is not symmetric about the minimums of the neutral-stability

curves; that is, the amplitude will be larger on one side of the minimum than the other for a given perturbation of Δ above the minimum.

4.3. Small-coupling resonance. The bifurcation equation (4.9) is singular when its denominator is zero ($D_2 = 0$), or

$$\Delta = \Delta_S \equiv \frac{a_1}{a_2}\delta_1^2 \tag{4.12}$$

($\Delta = \delta_1\delta_2$). If the parameters are such that the singular point is before the Hopf point, $\Delta_S < \Delta_H$, then the singularity can be ignored because the vanishing denominator will not occur when periodic solutions exist for $\Delta > \Delta_H$. However, if $\Delta_S > \Delta_H$, the bifurcation equation will exhibit the pole-type singularity. Near the singular point, the bifurcation equation predicts that the amplitude of the oscillations will become large corresponding to a resonance. The condition that $\Delta_S > \Delta_H$ corresponds to $\delta_1^2 > \epsilon^2 a_2^2$. Thus, in laser-2 the increase in the population inversion provided by pump coupling is greater than the effective losses. In [2], when the lasers were coupled without delay, we investigated this resonance in detail and the bifurcation equation matched the result of simulations very well. For the present case with delay coupling, however, we find that there are some dramatic differences.

In Figure 7a, we compare the analytical and numerical bifurcation branches. For $d_1 = 2.3$, the bifurcation equation predicts a strong resonance peak. The numerical data for $d_1 = 2.3$ ($*$) exhibits a small resonance peak and there is good agreement between the analytical and numerical curves both before and after the singularity. In Figure 7b we increase d_1 from 3.0 to 4.5 to follow the deformation of the bifurcation branch from an isolated resonance peak to a curve that forms a loop before continuing to higher values of d_1; this folding of the bifurcation branch is new to the system with delay and was not observed in [2]. We want to make clear that the bifurcation branch is *not* intersecting itself. The apparent intersection results from projecting the full bifurcation curve of d_1 versus (x_1, y_1, x_2, y_2), which does not intersect, onto the (d_1, x_2) plane.

Because the multiple-scale method is a local analysis and we assumed small-amplitude solutions, it is not unexpected that the analytical and numerical results match well near the bifurcation point. Similarly, in the vicinity of the resonance peak, the amplitude of the solutions becomes $O(1)$ and so it is not surprising that the analytical results fail to match the numerical results. That said, in our previous work with coupled lasers without delay [2], the analytically determined bifurcation equation matched the numerically computed resonance peak throughout the parameter range, i.e., even for $O(1)$ amplitudes; that was a bit surprising. We can surmise that in the present case it is the folding of the bifurcation branch that causes the mismatch. It may be that if we continued our analysis to higher order and derived additional nonlinear corrections to the bifurcation equation, we might better describe the folded branch. This would be a nontrivial and tedious calculation; but without doing so our supposition remains speculation. In [2] we used a different asymptotic method better suited to analyzing periodic solutions with $O(1)$ amplitudes with only partial success. Its application to the present problem with delay would be quite complicated and we have not made the attempt.

In Figure 7b we increase d_1 to 5.0 and see that there is qualitative change in the bifurcation branch; there is no longer a loop, and the bifurcation branch continues to negative values of

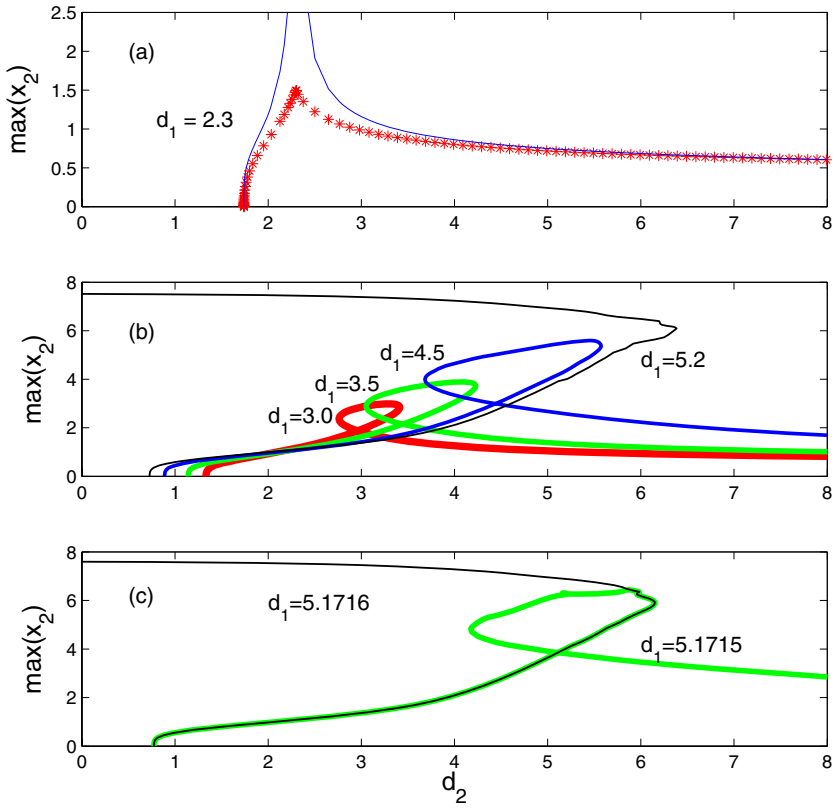


Figure 7. *Internal mode, $n = 1$ with $\tau_s = n\pi$.* (a) *The solid curve is the predicted bifurcation branch from* (A.6) *when d_1 is tuned such that the singularity in the bifurcation equation occurs after the Hopf bifurcation. The $*$ are the numerical bifurcation data for $d_1 = 2.3$.* (b) *Numerical bifurcation branches as they deform between $d_1 = 3.0$ and $d_1 = 5.2$.* (c) *There is a critical value of d_1, where on either side of the critical value the bifurcation branch has a loop and proceeds to higher values of d_1, or bends back without a loop to negative values of d_1. All other parameters are the same as in Figure 4.*

d_1. In Figure 7b we see that the change in the nature of the bifurcation curve appears quite abruptly at a critical value of d_1. Typically, one expects smooth folding and unfolding of bifurcation branches as a parameter is changed [34]. In the present case, we have not been able to isolate an interval of d_1 over which such an unfolding might occur. To our knowledge, such a discontinuous unfolding of a branch of solutions in a bifurcation diagram has not been previously described.

Finally, if d_1 is increased further the bifurcation branch folds back and continues for positive d_2; there also appear to be secondary bifurcations but we have not explored these in any detail. We mention that as the delay τ_s increases, or as ϵ increases, the effect of the singularity is diminished. Conversely, with reduced values of ϵ, the bifurcation branch may exhibit more complicated turns and folds. It is intuitively understandable that increasing the damping (ϵ) will diminish any type of resonance phenomena. The role of the longer delay is unclear.

5. Hopf bifurcation of the external modes. In this section, we describe the external modes that emerge via Hopf bifurcations as the coupling is increased. We fix $\delta_1 = O(1)$ and use δ_2 as the bifurcation parameter with $\delta_2 = \delta_{20} + \epsilon\delta_{22} + \dots$. We again have the slow time $T = \epsilon t$ and the perturbation expansion of x and y in powers of $\epsilon^{1/2}$. The multiple-scale analysis is more complicated because the $O(1)$ couplings result in a leading-order problem that contains the delay terms. Due to the two time scales, the delay terms are of the form $\delta_j y_j(t - \tau_j, T - \epsilon\tau_j)$. To make analytical progress we need to remove the slow delay from the leading-order problem. This requires that we assume $\epsilon\tau_j \ll 1$ such that by using (4.2) the slow delay is postponed to higher order. We then have a restriction on the size of the delay such that $\tau = o(1/\epsilon)$. Thus, we find that our results fit well with the external modes corresponding to the case $\tau_s = 5\pi$, Figure 1b, but are inaccurate in the case $\tau_s = 53\pi$, Figure 1c. Finally, to simplify the presentation, we consider the case of equal delays $\tau_1 = \tau_2 = \tau$, such that $\tau_s = 2\tau$; analysis of the general case with unequal delays can be carried out in the same way.

5.1. Leading order. The leading order $\mathrm{O}(\epsilon^{1/2})$ problem is

$$\frac{\partial}{\partial t} X_1(t,T) = L \cdot X_1(t,T) - D \cdot X_1(t-\tau,T), \tag{5.1}$$

where

$$L = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 & 0 & \delta_{20} \\ 0 & 0 & 0 & 0 \\ 0 & \delta_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{5.2}$$

$$\text{and } X_1(t,T) = \begin{pmatrix} x_{11}(t,T) \\ y_{11}(t,T) \\ x_{21}(t,T) \\ y_{21}(t,T) \end{pmatrix}.$$

We look for oscillatory solutions of the form $X_1(t,T) = U_1 B(T)\exp(i\omega t)$, where $B(T)$ is a slowly varying scaler amplitude (to be determined from a solvability condition at $O(\epsilon^{3/2})$). U_1 is a vector that is determined by substituting our ansatz into (5.1) to obtain

$$0 = J \cdot U_1, \quad \text{where} \quad J = \begin{pmatrix} -i\omega & -1 & 0 & -\delta_{20}e^{-i\omega\tau} \\ 1 & -i\omega & 0 & 0 \\ 0 & -\delta_1 e^{-i\omega\tau} & -i\omega & -1 \\ 0 & 0 & 1 & -i\omega \end{pmatrix}. \tag{5.3}$$

For a nonzero solution U_1, we require $\det J = 0$. This results in the same condition obtained from the leading-order linear-stability problem; specifically,

$$(1-\omega^2)^2 - \delta_1\delta_{20}e^{-i2\omega\tau} = 0. \tag{5.4}$$

Thus, we have that

$$\omega = \frac{m\pi}{2\tau}, \quad m = \text{even, positive integer.} \tag{5.5}$$

For later reference it will be useful to note that

$$e^{-i2\omega\tau} = 1, \quad \text{and} \quad e^{-i\omega\tau} = \pm 1 = \nu_m, \tag{5.6}$$

where $\nu_m = +1$ if $m/2$ is odd (e.g., $m = 2, 6, \dots$) and $\nu_m = -1$ if $m/2$ (e.g., $m = 4, 8, \dots$) is even. Finally, we find that

$$U_1 = \begin{pmatrix} i\omega \\ 1 \\ i\omega u_1 \\ u_1 \end{pmatrix}, \quad \text{where} \quad u_1 = \nu_m \sqrt{\frac{\delta_1}{\delta_{20}}}. \tag{5.7}$$

5.2. Second order. At $O(\epsilon)$ the problem is

$$\frac{\partial}{\partial t} X_2(t,T) = L \cdot X_2(t,T) - D \cdot X_2(t-\tau,T) + F_2, \tag{5.8}$$

$$\text{where} \quad F_2 = \begin{pmatrix} 0 \\ x_{11}y_{11} \\ 0 \\ x_{21}y_{21} \end{pmatrix}.$$

Because the homogeneous problem is the same as the $O(\epsilon^{1/2})$ problem we can, without loss of generality, set the homogeneous solution to 0. The inhomogeneous term F is proportional to $\exp(i2\omega t)$ so that the solution is

$$X_2(t,T) = B(T)^2 U_2 e^{i2\omega t} + c.c., \tag{5.9}$$

where U_2 is specified in Appendix C.

5.3. Third order. At $O(\epsilon^{3/2})$ we find the solvability condition that determines the slow-evolution equation for $B(T)$. The $O(\epsilon^{3/2})$ problem is

$$\frac{\partial}{\partial t} X_3(t,T) = L \cdot X_3(t,T) - D \cdot X_3(t-\tau,T) + F_3, \tag{5.10}$$

where

$$F_3 = \begin{pmatrix} -a_1 x_{11} - \delta_{22} y_{21}(t-\tau,T) + \delta_{20}\tau \frac{\partial}{\partial T} y_{21}(t-\tau,T) - \frac{\partial}{\partial T} x_{11} \\ x_{12}y_{11} + x_{11}y_{12} - \frac{\partial}{\partial T} y_{11} \\ -a_2 x_{21} + \delta_1 \tau \frac{\partial}{\partial T} y_{11}(t-\tau,T) - \frac{\partial}{\partial T} x_{21} \\ x_{22}y_{21} + x_{21}y_{22} - \frac{\partial}{\partial T} y_{21} \end{pmatrix}. \tag{5.11}$$

The vector F_3 has terms proportional to $\exp(i\omega t)$ and $\exp(i2\omega t)$ and the former will lead to solutions of the form $(U_3 + V_3 t)\exp(i\omega t)$. The secular terms $V_3 t$ must be eliminated to prevent unbounded solutions for large t, which implies that a solvability condition must be imposed on F_3. The solvability condition is formulated as follows: We look for a solution to (5.10) of

the form $X_3 = U\exp(i\omega t)$ and at the same time identify the terms F in F_3 proportional to $\exp(i\omega t)$. We then obtain an algebraic system of equations for the vector U as

$$0 = J \cdot U + F, \tag{5.12}$$

where

$$F = \begin{pmatrix} (-i\omega a_1 - \delta_{22}u_1\nu_m)B - (i\omega - \delta_{20}\tau u_1\nu_m)B_T \\ i\omega(u_{22}-1)|B|^2B - B_T \\ -i\omega a_2 u_1 B - (i\omega - \delta_1\tau\nu_m)B_T \\ i\omega u_1(u_{24}-u_1^2)|B|^2B - u_1 B_T \end{pmatrix}. \tag{5.13}$$

For U to have a nonzero solution, the Fredholm alternative requires that $V^H \cdot F = 0$, where V is the solution to $J^H \cdot V = 0$ (the superscript H refers to Hermitian). We find that $V^H = (u_1, i\omega u_1, 1, i\omega)$, and the resulting condition for the amplitude $B(T)$ is

$$\frac{\partial B}{\partial T} = (p_l + iq_l)B + (p_n + iq_n)B|B|^2, \tag{5.14}$$

where $p_{l,n}$ and $q_{l,n}$ are given in Appendix C.

5.4. Bifurcation equation. To analyze the solvability condition given by (5.14), we let $B(T) = R(T)e^{i\theta(T)}$ to obtain

$$\frac{\partial R}{\partial T} = (p_l + p_n R^2)R, \tag{5.15}$$

$$\frac{\partial \theta}{\partial T} = q_l + q_n R^2. \tag{5.16}$$

The equation for θ determines the frequency correction as a function of the amplitude. The bifurcation equation is determined by considering steady-state solutions to the equation for R, and we find that

$$R^2 = -\frac{p_l}{p_n} = -\left(\frac{|\delta_1|}{r_2\omega^2|1-\omega^2|}\right)\left[\delta_1\delta_{22} - \frac{2\omega^2(a_1+a_2)}{\tau}\right]. \tag{5.17}$$

The onset of oscillations occurs when $R = 0$ and determines δ_{22} at the Hopf bifurcation point; this result matches exactly that obtained in the linear-stability analysis.

In Figures 8 and 9, we compare the bifurcation equation (5.17) to the numerically computed result. In each figure we have $\tau_s = 5\pi = O(1)$ (see Figures 4 and 5 for analysis of the internal mode). In Figure 8 the external mode corresponds to $m = 4$ and its direction of bifurcation is subcritical. In Figure 9 we have $m = 6$ and the bifurcation is supercritical. In each case, there is good agreement between the numerical and analytical results local to the bifurcation point, where the multiple-scale analysis's validity is strongest.

The direction of bifurcation (super- or subcritical) in (5.17) is controlled by the sign of the constant r_2, which is given in (C.5). Analysis of r_2 shows that $r_2 > 0$ in the interval $\sqrt{2/5} < \omega < \omega_z(d_1)$, where $\omega_z(d_1)$ is shown in Figure 10 (the lower bound is a zero of the denominator of r_2, while the upper bound is the sole real zero of the numerator). Thus, for

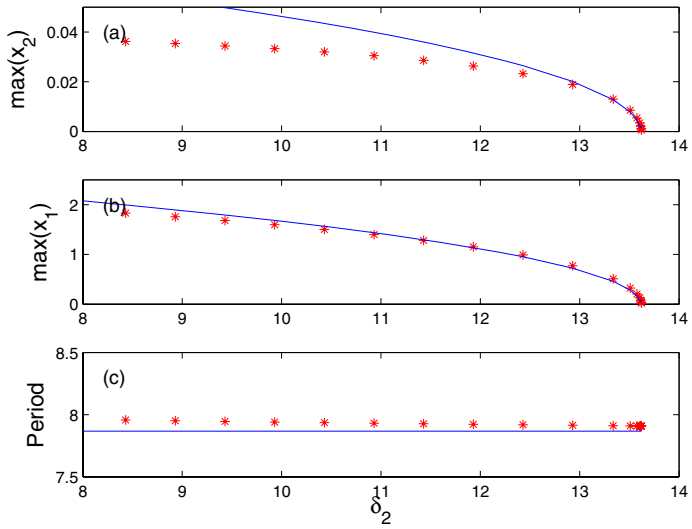


Figure 8. *External mode, $n = 5$, $m = 4$ with $\tau_1 = \tau_2 = n\pi/2$, and $\epsilon = 0.01$. Numerical data points are indicated by $*$. The solid curve is the asymptotic approximation based on (5.17) ($d_1 = 1$, $a_1 = a_2 = 2$, $b = 1$, $\beta = 1$).*

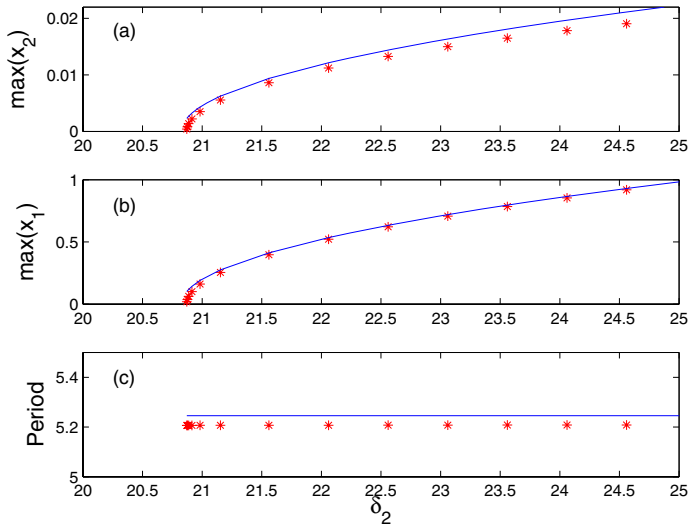


Figure 9. *Same as Figure 8 except $m = 6$. All other parameter values are the same.*

external modes with frequencies within this interval, the bifurcation will be subcritical; this is the case for the external mode in Figure 8. For all other modes $r_2 < 0$ and the direction of the bifurcation will be supercritical.

For reference, in Figure 11 we show the bifurcation diagrams of the internal mode ($\omega \approx 1$) and the external modes ($\omega = m/n$, $m = 4$ and $m = 6$) for the case $n = 5$; that is, we have

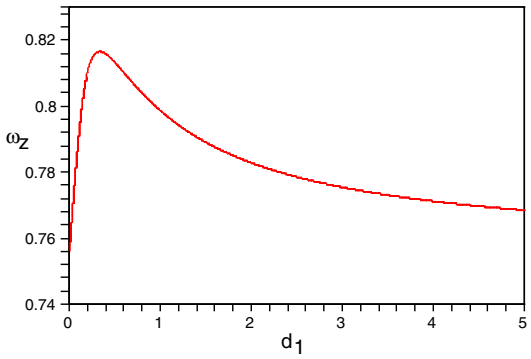


Figure 10. *Upper bound on the value of ω as a function of d_1 giving a sign change for parameter r_2 in (C.5).*

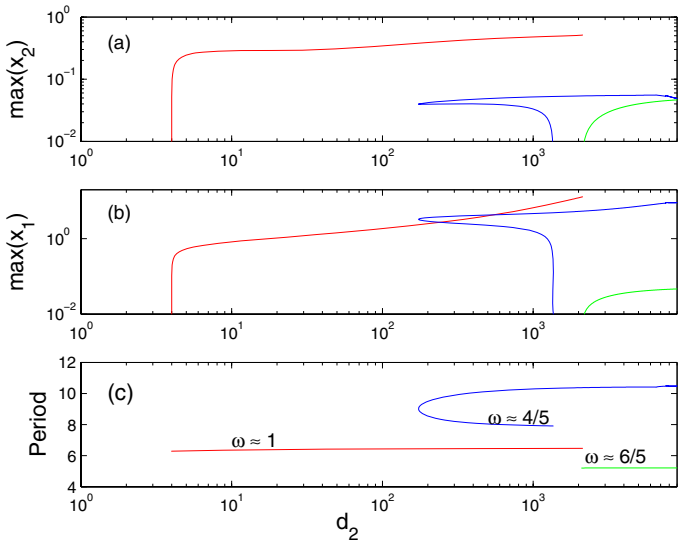


Figure 11. *Numerically computed bifurcation diagrams, both internal and external modes, for $n = 5$ and $\tau_s = n\pi$. The left-most branch corresponds to the internal mode shown in Figure 4; periodic solutions are stable until $d_2 \approx 59$, when there is a period-doubling bifurcation. The middle and right-most branches are the continuation of the external modes shown in Figure 8 ($m = 4$) and Figure 9 ($m = 6$), respectively; periodic solutions are unstable along these branches ($\epsilon = 0.01$, $d_1 = 1$, $a_1 = a_2 = 2$, $b = 1$, $\beta = 1$).*

combined the bifurcation diagrams of Figures 4, 8, and 9. The subcritical bifurcation for the external mode $m = 4$ folds to provide an interval of hysteresis. The bifurcation branches are projections onto planes so that the intersections of the curves are not relevant. The periodic solutions corresponding to the external modes are unstable as they bifurcate from the unstable branch of steady-state solutions. The internal mode is stable until $d_2 \approx 59$, when a period-doubling sequence to chaos begins.

Finally, we note that for the external modes

$$R \sim [\Delta - \Delta_H]^{1/2}. \tag{5.18}$$

The amplitude of the external modes varies as the square root of the distance from the bifurcation point. In contrast, the internal mode varies as one-fourth the power of the deviation.

6. Discussion. We have analyzed the output of two mutually coupled lasers, where the light intensity deviations from steady state modulated the pump of the other laser. The coupling strength in one direction is held fixed, while we examine the effect of increasing the coupling strength in the other direction. The signal-propagation time through the optical fiber and the electronic circuit causes a delay leading to a model that is a system of DDEs. Linear-stability analysis finds that the steady-state solution becomes unstable at a Hopf bifurcation; we call the resulting periodic solution the internal mode because it oscillates at the laser's relaxation-oscillation frequency. As the coupling is increased, subsequent instabilities occur with frequencies determined by the round-trip delay time $\omega_e = m\pi/\tau_s$, $m = 2, 4, 6, \ldots$; we call these periodic solutions external modes.

Using a multiple-scale analysis, we derive bifurcation equations for both the internal and external modes. We find that the amplitude of the internal mode increases with the 1/4-root of the deviation from the bifurcation point and is supercritical; i.e., $R \sim +(\Delta - \Delta_H)^{1/4}$. The amplitude of the external modes increase with the 1/2-root of the deviation and may be super- or subcritical; i.e., $R \sim \pm(\Delta - \Delta_H)^{1/2}$. Both the initial instability and the bifurcation results depend on the product of the coupling constants $\Delta = \delta_1\delta_2$. For the analysis of the internal mode, we assumed that both coupling constants were of the same relative size, $\delta_j = O(\epsilon)$. However, other scalings satisfy the Hopf condition, e.g., $\delta_1 = O(\epsilon^{1/2})$ and $\delta_2 = O(\epsilon^{3/2})$. We have found that this does not change the qualitative properties of the bifurcation.

We have focused our analysis on the onset of oscillatory instabilities and just beyond. We have not considered the stability of the internal and external modes as the coupling is increased further. However, numerical simulations indicated period-doubling bifurcations of individual modes as well as multimode behavior.

With independent control of the individual coupling strengths, we have observed an atypical resonance phenomena. As in our previous work on two mutually coupled lasers without delay [2], there is an interval of the coupling parameter δ_2 over which the amplitude of the oscillations becomes large; we have referred to this effect as a "resonance." The bifurcation equation, (4.9), can describe the amplitude of the oscillations for δ_2 above and below resonance. A singularity in the bifurcation equation indicates the existence of the resonance, but (4.9) does not describe the amplitude within the resonance well. The parametric form of the singularity indicates that physically, the coupling term provides an effective negative-damping that cancels with the lasers' self-damping and hence provides a resonance effect.

What is different from the results in [2] is that the bifurcation branch folds with a change in the coupling parameter δ_1; in Figure 7 we see that the resonance peak forms a loop in the (d_1, x_2) plane. We also find that instead of a smooth unfolding of the loop, there appears to be a critical value of the parameter d_1 where the folded portion of the bifurcation branch abruptly disappears. To our knowledge, such a discontinuous unfolding of a branch of solutions in a

bifurcation diagram has not been previously described. We do not understand the mechanism or manner by which the abrupt change occurs, but this will be the focus of future work.

The application of multiple-scale perturbation techniques to DDEs is, to our knowledge, a relatively recent development [21]. For our analysis of the internal modes, the delay terms appeared only as part of the solvability condition; thus, the multiple-scale analysis was relatively straightforward. In contrast, for the external modes that bifurcate when the coupling constants are $O(1)$, the delay terms are included in the leading-order problem. However, by looking for periodic solutions we are able to continue the analysis and formulate a solvability condition. Finally, as mentioned in the text, multiple-scale expansion of the delay term in (4.2) amounts to a Taylor series expansion that, while it may be justified in and of itself, can lead to erroneous results for the DDE [30]. Thus, it is important to compare our analytical results to those from numerical simulations as an important check of the work.

As discussed in the introduction, the method of averaging has also been used to analyze the weakly nonlinear characteristics of delay problems. Averaging and the multiple-scale technique will lead to a similar slow-time evolution equation for the amplitudes. However, the multiple-scale technique accounts for the delay in the slow time where averaging does not. Because we focused on the existence of periodic solutions, the slow-time delay is removed and both averaging and multiple scales give equivalent results. However, stability of the periodic solutions, or the investigation of more complicated phenomena such as quasiperiodicity, would require the slow delay from the multiple-scale analysis.

Finally, we finish with a discussion relating the results of our analysis to experimental results observed in [1] and [22]. The coupling circuit in the experimental system has two important characteristics. First, the signal is inverted, which results in the negative signs that appear in front of the coupling constants in (2.4). Second, the circuit acts as a low-pass filter that suppresses coupling of the relaxation oscillations. We discuss the effect of each of these below.

The signal inversion of the coupling circuit results in negative coupling constants in (2.4). However, both the linear-stability and the leading-order bifurcation results depend on the product of the coupling constants $\Delta = \delta_1\delta_2$. Thus, local to the Hopf bifurcation it does not matter if both δ_j are negative or both are positive. More generally, our results depend only on whether Δ is positive or negative—not on the signs of the individual coupling constants. This is effectively a symmetry result, because for small amplitudes the oscillations are nearly harmonic. This means a sign change is merely a phase shift. However, we have observed in numerical simulations that when the amplitudes become larger such that the intensity is pulsating, the symmetry is lost such that positive coupling results in different system output from negative coupling.

As mentioned above, the optoelectronic coupling circuit in [1] and [22] acts as a low-pass filter on the coupling signal. We do not account for this in our model of the system (recently, Illing and Gauthier [35] have analyzed a DDE where they explicitly account for the bandlimited response of their feedback system). However, because our results address the linear and nonlinear dynamics of both the internal and external modes, we can make a comparison between theory and experiment.

The low-pass filter characteristic of the experimental coupling circuit attenuates the high-frequency relaxation oscillations, corresponding to the internal mode. The result is that the

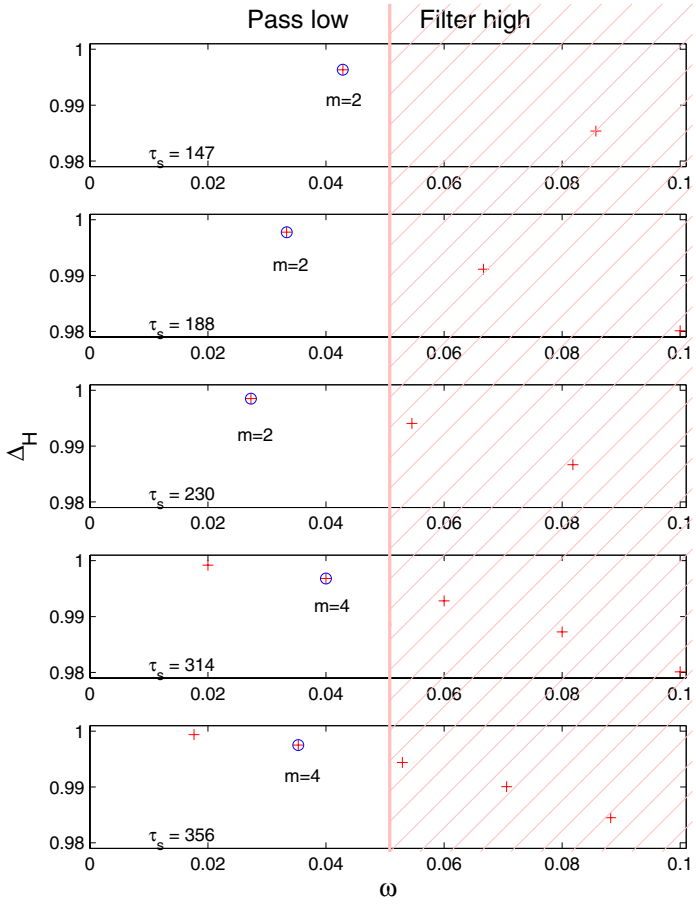


Figure 12. *For each value of delay, linear stability predicts a set of external modes with frequencies $\omega_e = m\pi/\tau_s$, m even, indicated by $+$. The frequencies of the external modes have been plotted on the linear-stability curve $\Delta_H(\omega)$ as in Figure 1. External modes with higher frequency bifurcate first because $\Delta_H(\omega)$ is decreasing. For each value of the delay, the $(\circ)$ indicates the specific external mode observed in the experiment. All of the observed modes are less than a cutoff frequency due to the low-pass filter characteristic of the coupling circuit.*

experimental system oscillates at one of the low-frequency external modes. This is illustrated in Figure 12, where we indicate the experimentally observed external mode with a $(\circ)$ for different values of the delay. The $(+)$ indicate the theoretical value of $\Delta_H(\omega)$ for a bifurcation to an external mode. Linear stability predicts that as the frequency increases the coupling strength, Δ_H, required for a Hopf bifurcation decreases (see also Figure 1 for small ω). However, there is a filter-cutoff frequency ω_c such that the relaxation oscillations and higher-frequency external modes are attenuated and only the external modes with frequencies $\omega_e < \omega_c$ are observed.

For example, in [1] we observed oscillations with period equal to the round-trip delay

time; specifically, $\omega = 2\pi/T = 2\pi/\tau_s$ corresponding to the external mode with $m = 2$. This indicates that the filter-cutoff frequency, ω_c, is such that $2\pi/\tau_s < \omega_c < 4\pi/\tau_s$. If ω_c were greater than $4\pi/\tau_s$, then we would have expected to observe the external mode with $m = 4$ because the value of the coupling at the Hopf bifurcation point decreases as ω_e increases; i.e., it will bifurcate at a lower value of the coupling.

More recent experiments [22] confirm that with longer delays, external modes with $m > 2$ are exhibited; this is shown in the lower two plots in Figure 12. That is, as the delay increases, the external modes with larger m will be below the cutoff frequency, $\omega_e = m\pi/\tau_s < \omega_c$. And it is always the mode with the largest frequency, but still below cutoff, that is exhibited because $\Delta_H(\omega)$ is least for that frequency. Taking all five plots together, the experimental data suggest that the low-pass filter cutoff frequency is $\omega_c \approx 0.05$ because all of the observed modes have frequency less than ω_c. Finally, we add that it was observed in [1] that the amplitude of the oscillations followed a square-root power law as a function of the coupling. This is exactly as predicted by the bifurcation equation (5.17) for the external modes.

Our comparison between the theory and the experiment does have limitations. We do not know the detailed filter characteristics of the optoelectronic coupling circuit. It is known that the filter profile is most certainly not a step function but is instead frequency dependent; it depends on the properties of the optical cable, the electronic coupling circuit, and the frequency response of the laser to the modulated pump. In addition, for large values of the delay the frequency difference between the external modes becomes very small and many external modes are excited at nearly the same value of the coupling. For these cases the linear stability theory may be insufficient to identify the mode that is observed in the experiments.

Appendix A. Amplitudes of periodic solutions.

$$R_2^2 = \frac{N}{D}, \tag{A.1}$$

where

$$\begin{aligned} N = {} & 3Q\left[a_1 d_1^2 d_2 \sin^2\tau_s - a_1^2 d_1 (a_1 + a_2)\cos\tau_s\right] \\ & - 3\,a_1 d_1^3 d_2^2 \sin^3\tau_s + 3a_1^2 d_1^2 d_2 (3a_2 + a_1)\sin\tau_s\cos\tau_s \\ & + 6a_1^3 a_2 d_1 (a_1 + a_2)\sin\tau_s, \end{aligned} \tag{A.2}$$

$$\begin{aligned} D = {} & Q\sin\tau_s\left[\,a_1^2 d_1 - \,d_1 d_2^2 \sin^2\tau_s + 2\,a_1 a_2 d_2\cos\tau_s\right] \\ & + d_1^2 d_2^3 \sin^4\tau_s - 4\,a_1 a_2 d_1 d_2^2 \cos\tau_s \sin^2\tau_s - 4\,a_1^2 a_2^2 d_2 \sin^2\tau_s \\ & + 2\,a_1^3 a_2 d_1 \cos\tau_s - a_1^2 d_1^2 d_2 \sin^2\tau_s + 2\,a_1^2 a_2^2 d_2, \end{aligned} \tag{A.3}$$

$$Q = \sqrt{d_1^2 d_2^2 \sin^2\tau_s - 4\,a_1 a_2 d_1 d_2 \cos\tau_s - 4\,a_1^2 a_2^2}, \tag{A.4}$$

and

$$R_1^2 = \frac{d_2}{2a_1^2 d_1}\left(d_1 d_2 \sin^2\tau_s - 2\,a_2 a_1 \cos\tau_s - Q\sin\tau_s\right) R_2^2. \tag{A.5}$$

When the delay is tuned to be near one of the minimums of the neutral stability curves,

i.e., $\tau_s = m\pi + \xi$, $\xi \ll 1$, we have

$$
\begin{aligned}
R_2^2 &= Z_{20} + \xi Z_{21} + O(\xi^2),\\
Z_{20} &= 3(a_1 + a_2)\frac{\sqrt{D_1}}{D_2},\\
Z_{21} &= \frac{3(D_1+1)}{2d_1^2 D_2^2}\left[D_1(a_2^3 + 3a_2^2 a_1) + (a_2 - a_1)(d_1^2 - a_2^2)\right],\\
D_1 &= \frac{d_1 d_2}{a_1 a_2} - 1,\\
D_2 &= \frac{a_2 d_2}{a_1 d_1} - 1,
\end{aligned}
\tag{A.6}
$$

$$
R_1^2 = \left(\frac{a_2}{d_1}\right)^2 (D_1+1)\left[Z_{20} + \xi(Z_{20}\sqrt{D_1} + Z_{21})\right] + O(\xi^2),
\tag{A.7}
$$

$$
\omega = 1 - \epsilon\left(\frac{1}{6}Z_{20} + \frac{1}{2}a_2\sqrt{D_1}\right) - \xi\epsilon\left[\frac{1}{6}Z_{21} + \frac{1}{4}a_2(D_1+1)\right] + O(\xi^2).
\tag{A.8}
$$

Appendix B. Delay and opposite-sign coupling. Consider a negative coupling constant such that

$$
\frac{dx_1}{dt} \sim +\delta_2 y_2(t).
\tag{B.1}
$$

If y_2 is harmonic with frequency $\omega = 1$ and given by $y_2 = A\cos(t)$, then

$$
\begin{aligned}
\frac{dx_1}{dt} \sim\ & +\delta_2 A\cos(t),\\
& -\delta_2 A\cos(t - n\pi), \quad n \text{ an odd integer},\\
& -\delta_2 A\cos(t - \tau_s),
\end{aligned}
\tag{B.2}
$$

where $\tau_s = n\pi$ is the total "round-trip" or system delay because the other coupling constant, δ_1, is negative.

Appendix C. External modes: Coefficients.

C.1. Second order: $O(\epsilon)$.

$$
U_2 = \begin{pmatrix} i\omega(2u_{22} - 1)\\ u_{22}\\ i\omega(2u_{24} - u_1^2)\\ u_{24}\end{pmatrix}
\tag{C.1}
$$

and

$$
u_{22} = -\frac{2\omega^2[1 - 4\omega^2 - \delta_1]}{(1-4\omega^2)^2 - \delta_1\delta_{20}}, \quad u_{24} = -\frac{2\omega^2[u_1^2(1-4\omega^2) - \delta_1]}{(1-4\omega^2)^2 - \delta_1\delta_{20}}.
\tag{C.2}
$$

C.2. Third order: $O(\epsilon^{3/2})$.

$$p_l = r_1[\delta_1\delta_{22}\tau - 2\omega^2(a_1+a_2)], \quad q_l = r_1\left\{\nu_m u_1\omega\left[(a_1+a_2)\frac{\tau}{\delta_{20}} + 2\delta_{22}\right]\right\}, \tag{C.3}$$

$$p_n = r_1 r_2 \tau \nu_m \frac{\omega^2}{u_1}\delta_1, \quad q_n = r_1 r_2 2\omega^3, \tag{C.4}$$

$$r_1 = \frac{1}{2(\delta_1\delta_{20}\tau^2 + 4\omega^2)}, \quad r_2 = \frac{\delta_1(\delta_1+\delta_{20}+4\omega^2) - (1+u_1^2)(1-2\omega^2)(1-4\omega^2)}{(1-4\omega^2)^2 - \delta_1\delta_{20}}. \tag{C.5}$$

Acknowledgment. I. B. Schwartz acknowledges the support of the Office of Naval Research.

REFERENCES

[1] M.-Y. Kim, R. Roy, J. L. Aron, T. W. Carr, and I. B. Schwartz, *Scaling behavior of laser population dynamics with time-delayed coupling: Theory and experiment*, Phys. Rev. Lett., 94 (2005), 088101.
[2] T. W. Carr, M. L. Taylor, and I. B. Schwartz, *Negative-coupling resonances in pump-coupled lasers*, Phys. D, 215 (2006), pp. 152–163.
[3] H. Yaska and H. Kawaguchi, *Linewidth reduction and optical frequency stabilization of a distributed feedback laser by incoherent optical negative feedback*, Appl. Phys. Lett., 53 (1988), pp. 1360–1362.
[4] R. Lang and K. Kobayashi, *External optical feedback effects on semiconductor injection laser properties*, IEEE J. Quantum Electron., 16 (1980), pp. 347–355.
[5] G. H. M. van Tartwijk and D. Lenstra, *Semiconductor lasers with optical injection and feedback*, Quantum Semiclass. Opt., 7 (1995), pp. 87–143.
[6] J. Javaloyes, P. Mandel, and D. Pieroux, *Dynamical properties of lasers coupled face to face*, Phys. Rev. E, 67 (2003), 036201.
[7] H. Erzgräber, B. Krauskopf, and D. Lenstra, *Compound laser modes of mutually delay-coupled lasers*, SIAM J. Appl. Dyn. Syst., 5 (2006), pp. 30–65.
[8] R. D. Li, P. Mandel, and T. Erneux, *Periodic and quasiperiodic regimes in self-coupled lasers*, Phys. Rev. A, 41 (1990), pp. 5117–5126.
[9] D. Pieroux, T. Erneux, and K. Otsuka, *Minimal model of a class-b laser with delayed feedback: Cascading branching of periodic solutions and period-doubling bifurcation*, Phys. Rev. E, 50 (1994), pp. 1822–1829.
[10] D. W. Sukow, A. Gavrielides, T. Erneux, M. J. Baracco, Z. A. Parmenter, and K. L. Blackburn, *Two-field description of chaos synchronization in diode lasers with incoherent optical feedback and injection*, in Proceedings of SPIE, Proc. SPIE 5722, SPIE, Bellingham, WA, 2005, pp. 256–258.
[11] R. Vicente, S. Tang, J. Mulet, C. R. Mirasso, and J. M. Liu, *Dynamics of semiconductor lasers with bidirectional optoelectronic coupling: Stability, route to chaos and entrainment*, Phys. Rev. E, 70 (2004), 046216.
[12] R. Vicente, S. Tang, J. Mulet, C. R. Mirasso, and J. M. Liu, *Synchronization properties of two self-oscillating semiconductor lasers subject to delayed optoelectronic mutual coupling*, Phys. Rev. E, 73 (2006), 047201.
[13] D. V. R. Reddy, A. Sen, and G. L. Johnston, *Driven response of time delay coupled limit cycle oscillators*, Commun. Nonlinear Sci. Numer. Simul., 8 (2003), pp. 493–518.
[14] A. K. Sen and R. H. Rand, *A numerical investigation of the dynamics of a system of two time-delay coupled relaxation oscillators*, Comm. Pure Appl. Math., 2 (2003), pp. 567–577.
[15] S. Wirkus and R. Rand, *The dynamics of two coupled Van der Pol oscillators with delay coupling*, Nonlinear Dynam., 30 (2002), pp. 205–221.
[16] D. V. R. Reddy, A. Sen, and G. L. Johnston, *Time delay induced death in coupled limit cycle oscillators*, Phys. Rev. Lett., 80 (1998), pp. 5109–5112.

[17] S. R. Campbell and D. L. Wang, *Relaxation oscillators with time delay coupling*, Phys. D, 111 (1998), pp. 151–178.
[18] S. G. Lee, S. Kim, and H. Kook, *Synchrony and clustering in two and three synaptically coupled Hodgkin-Huxley neurons with delay*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 7 (1997), pp. 889–895.
[19] M. K. S. Yeung and S. H. Strogatz, *Time delay in the Kuromoto model of coupled oscillators*, Phys. Rev. Lett., 82 (1999), pp. 648–651.
[20] N. B. Abraham, P. Mandel, and L. M. Narducci, *Dynamical instabilities and pulsations in lasers*, Prog. Opt., 25 (1988), pp. 3–190.
[21] D. Pieroux, T. Erneux, A. Gavrielides, and V. Kovanis, *Hopf bifurcation subject to a large delay in a laser system*, SIAM J. Appl. Math., 61 (2000), pp. 966–982.
[22] M.-Y. Kim, *Delay Induced Instabilities in Coupled Semiconductor Lasers and Mack-Glass Electronic Circuits*, Ph.D. thesis, University of Maryland, College Park, MD, 2005.
[23] C. Chicone and Z. C. Feng, *Synchronization phenomena for coupled delay-line oscillators*, Phys. D, 198 (2004), pp. 212–230.
[24] D. V. R. Reddy, A. Sen, and G. L. Johnston, *Time delay effects on coupled limit cycle oscillators at Hopf bifurcation*, Phys. D, 129 (1999), pp. 15–34.
[25] F. T. Arecchi, G. L. Lippi, G. P. Puccioni, and J. R. Tredicce, *Deterministic chaos in laser with injected signal*, Optics Comm., 51 (1984), pp. 308–314.
[26] P. Mandel, *Theoretical Problems in Cavity Nonlinear Optics*, Cambridge Studies in Modern Optics, Cambridge University Press, New York, 1997.
[27] I. B. Schwartz and T. Erneux, *Subharmonic hysteresis and period doubling bifurcations for a periodically driven laser*, SIAM J. Appl. Math., 54 (1994), pp. 1083–1100.
[28] F. Rogister, D. Pieroux, M. Sciamanna, P. Megret, and M. Blondel, *Anticipating synchronization of two chaotic laser diodes by incoherent optical coupling and its application to secure communications*, Opt. Commun., 207 (2002), pp. 295–306.
[29] W. Govaerts, J. Guckenheimer, and A. Khibnik, *Defining functions for multiple Hopf bifurcations*, SIAM J. Numer. Anal., 34 (1997), pp. 1269–1288.
[30] R. D. Driver, *Ordinary and Delay Differential Equations*, Springer-Verlag, New York, 1977.
[31] R. D. Driver, D. W. Sasser, and M. L. Slater, *The equation $x'(t) = ax(t) + b(t-\tau)$ with "small" delay*, Amer. Math. Monthly, 80 (1973), pp. 990–995.
[32] L. E. El'sgol'ts and S. B. Norkin, *Introduction to the Theory and Application of Differential Equations with Deviating Arguments*, Math. Sci. Engrg. 105, Academic Press, New York, 1973.
[33] K. Engelborghs, T. Luzyanina, and G. Samaey, *DDE-BIFTOOL v.*2.00 *User Manual: A MATLAB Package for Bifurcation Analysis of Delay Differential Equations*, Technical report TW-330, Department of Computer Science, K. U. Leuven, Leuven, Belgium, 2001.
[34] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems*, Springer-Verlag, New York, 1990.
[35] L. Illing and D. J. Gauthier, *Hopf bifurcations in time-delay systems with band-limited feedback*, Phys. D, 210 (2005), pp. 180–202.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 726–758

The Kelvin–Helmholtz Instability of Momentum Sheets in the Euler Equations for Planar Diffeomorphisms*

Robert I. McLachlan† and Stephen R. Marsland‡

Abstract. The Euler equations that describe geodesics on the group of diffeomorphisms of the plane admit singular solutions in which the momentum is concentrated on curves, the so-called momentum sheets analogous to vortex sheets in the Euler fluid equations. We study the stability of straight and circular momentum sheets for a large family of metrics. We prove that straight sheets moving normally to themselves under an H^1 metric, corresponding to peakons for the one-dimensional (1D) Camassa–Holm equation, are linearly stable in Eulerian coordinates, suffering only a weak instability of Lagrangian particle paths, while most other cases are unstable but well-posed. Expanding circular sheets are algebraically unstable for all metrics. The evolution of the instabilities are followed numerically, illustrating several typical dynamical phenomena of momentum sheets.

Key words. Euler equations, planar diffemorphisms, momentum sheets, Kelvin–Helmholtz instability

AMS subject classifications. 37K65, 37M15

DOI. 10.1137/060655808

1. Introduction.

The Kelvin–Helmholtz instability of shear layers, which occurs when one fluid layer slides over another of the same or different density, is one of the most famous and ubiquitous instabilities in fluid mechanics. Although studied intensively since its discovery by Kelvin in 1868 and Helmholtz in 1871, it is not yet fully understood. One key idealization is to consider the motion of a vortex sheet, which represents the limit of a thin shear layer in an incompressible, inviscid, two-dimensional (2D) fluid. Early simulations showing the instability developing into a chain of spiral vortices were confounded when Derek Moore showed in 1979 [20] that the sheet in fact develops an infinite curvature singularity at a finite time, a time even before the first vortices appear. Subsequent studies [10, 13, 14] have focused on regularized versions of the equations of motion and their relationship to weak solutions of the Euler fluid equations.

In this paper we study the analogous, Kelvin–Helmholtz-like instability for a related family of equations, the Euler equations for diffeomorphisms:

$$\dot{m} + u \cdot \nabla m + \nabla u^T \cdot m + m(\operatorname{div} u) = 0, \tag{1.1}$$

where $\dot{m}$ denotes differentiation with respect to time, $u(x,t)$ $(u, x \in \mathbb{R}^n,\ t \in \mathbb{R})$ is a velocity

*Received by the editors March 31, 2006; accepted for publication (in revised form) by J. Marsden August 28, 2006; published electronically December 5, 2006. This research was supported by the Marsden Fund and the New Zealand Institute of Mathematics and Its Applications.
http://www.siam.org/journals/siads/5-4/65580.html
†Mathematics Department, Massey University, Palmerston North, New Zealand (r.mclachlan@massey.ac.nz).
‡Institute of Information Science, Massey University, Palmerston North, New Zealand (s.r.marsland@massey.ac.nz).

field, and $m(x,t)$ its associated momentum. The velocity u and momentum m are related by

$$m = \mathcal{A}u, \tag{1.2}$$

where $\mathcal{A}$ is an elliptic operator (e.g., $\mathcal{A} = (1-\nabla^2)^k$) called the inertia operator.

It was Arnold's celebrated discovery [1] that the Euler fluid equations describe geodesics on the group of volume-preserving (i.e., incompressible) diffeomorphisms. Similarly, (1.1), (1.2) describe geodesics on the group of all diffeomorphisms. There has been much recent interest in the Euler equations (1.1), (1.2) because they arise in several different fields, including computer vision and fluid dynamics. In computer vision they appear in two and three dimensions in the field of image registration, such as in the averaged template matching equations [19, 22] and the geodesic interpolating clamped-plate spline [16], while in fluid dynamics they coincide with the Camassa–Holm wave equation for particular choices of $\mathcal{A}$ [5].

Euler equations such as (1.1) have a natural geometric origin, which we sketch here although these details are not needed in the paper [2, 12, 15]. Let $\mathfrak{G}$ be a Lie or diffeomorphism group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{G}$ be equipped with a left- or right-invariant metric that restricts to a metric $\langle\langle , \rangle\rangle$ on $\mathfrak{g}$. Typically, this metric is defined by a linear *inertia operator* $\mathcal{A} : \mathfrak{g} \to \mathfrak{g}^*$ via

$$\langle\langle u, v\rangle\rangle := \langle u, \mathcal{A}v\rangle. \tag{1.3}$$

The geodesic equation on $T\mathfrak{G}$ can be reduced to give a noncanonical Hamiltonian evolution equation on $\mathfrak{g}$ called the *Euler–Poincaré* equation, or transferred (via the Legendre transform, $\mathfrak{g} \mapsto \mathfrak{g}^*$, which in this case is $u \to m := \mathcal{A}u$) to a Lie–Poisson system on $\mathfrak{g}^*$ called the *Euler* equations

$$\dot{m} = \pm \mathrm{ad}^*_{\mathcal{A}^{-1}m} m, \tag{1.4}$$

where the sign is $+$ for left- and $-$ for right-invariant metrics. The variable $m \in \mathfrak{g}^*$ is called the *momentum* and $u \in \mathfrak{g}$ the *velocity*.

The most famous Euler equation on a Lie group is the equation of motion of a free rigid body ($\mathfrak{G} = SO(3)$, m = body angular momentum). An infinite-dimensional example is the Landau–Lifshitz equation on the loop group $\mathfrak{G} = C^\infty(\mathbb{R}, SO(3))$. Examples on diffeomorphism groups include the Euler fluid equations on the group $\mathfrak{G} = \mathrm{Diff}_{\mathrm{vol}}(\mathbb{R}^n)$ of volume-preserving diffeomorphisms with respect to the L^2 metric (m = vorticity) [2], the Camassa–Holm equation ($\mathfrak{G} = \mathrm{Diff}(\mathbb{S}^1)$, H^1 metric) [5], and the second-grade fluid equations ($\mathfrak{G} = \mathrm{Diff}_{\mathrm{vol}}(\mathbb{R}^n)$, H^1 metric) [24].

For $\mathfrak{G} = \mathrm{Diff}(\mathbb{R}^n)$ the Euler equations are given by (1.1), (1.2) (see [2, 5, 22] for further details). The inverse of the inertia operator $\mathcal{A}$ is given by convolution with the Green's function $\mathbf{G}$ of $\mathcal{A}$, i.e., $u = \mathbf{G} * m$, where $*$ denotes convolution and $\mathcal{A}\mathbf{G}(x,x') = \delta(x-x')$ for $x, x' \in \mathbb{R}^n$. We shall only consider Euclidean-invariant and diagonal $\mathcal{A}$ (i.e., $\mathcal{A} = a(-\nabla^2)I$ for some function a); in this case $\mathbf{G}(x,x') = G(\|x-x'\|)$ for a scalar function G, which we call the *kernel*. Examples in the literature include Gaussian [7] and Bessel function kernels [24] associated with the H^1 metric and a family of H^k metrics [17]. We will study (1.1), (1.2) for arbitrary Euclidean-invariant operators $\mathcal{A}$.

A striking feature of Euler equations on diffeomorphism groups is that they admit (formally, at least) exact solutions in which the momentum is concentrated on a submanifold of dimension $l < n$ [5]. The submanifold M is embedded in $\mathbb{R}^n$, say, $q : M \times \mathbb{R} \to \mathbb{R}^n$, and the momentum field is given by

$$m(x,t) = \int_M \delta(x - q(s,t))p(s,t)\,\mathrm{d}s, \tag{1.5}$$

where $p(s,t)$ is the "strength" of the momentum (vorticity in the fluid case) on the submanifold. For the Euler fluid equations these solutions are point vortices for $l = 0$, vortex sheets for $l = n - 1$, and vortex filaments for $l = 1$, $n = 3$. These are widely studied both in their own right and as a means of approximating the evolution of smooth or other vorticities [14].

For the 2D and three-dimensional (3D) Euler fluid equations, convergence of the point vortex solutions to solutions for smooth initial data has been established [14]. The speed of convergence can be improved by smoothing out the point vortices to *vortex blobs*, even though the (e.g., Gaussian) blobs are no longer exact solutions of the Euler equations. Instead, their evolution can be regarded as that of delta-functions under a slightly different inertia operator. Altering the shape of the blob (which corresponds to the kernel) is equivalent to altering the metric. In this way we are led to consider a wide class of metrics.

Very little is known about the Euler equations (1.1), (1.2). Most studies have taken $\mathcal{A} = (1 - \alpha^2\nabla^2)$, i.e., an H^1 metric. In one dimension this gives the Camassa–Holm equation, which is completely integrable. Here the codimension-1 singular solutions are known to play a special role: they are solitons (m a sum of delta functions, u a sum of shifted and scaled "peakons" $e^{-|x|}$) and are stable in the sense that the long-time solution for initial data that decays to zero at infinity contains only such solitons [5, 8]. Extensive simulations which have been performed with the same metric in dimensions 2 and 3 [9] suggest that bump-like initial data also evolves toward codimension-1 singular solutions, i.e., momentum sheets. These near-sheets interact with each other and can even undergo reconnection events. The sheets appear to be stable objects, even asymptotically stable, attracting a substantial part of phase space. These numerical results suggest that to understand the behavior of (1.1), (1.2) it is crucial to first understand the behavior of their momentum sheet solutions.

From now on we consider momentum sheets in two dimensions, i.e., $n = 2$ and $\dim M = 1$.

For such momentum-sheet data the Euler equations form a canonical Hamiltonian system with configuration space the space of embeddings of M in $\mathbb{R}^2$, i.e., $Q = \mathrm{Emb}(M, \mathbb{R}^2)$ and phase space T^*Q [5]. In coordinates, the positions $q(s,t) \in \mathbb{R}^2$ and conjugate momenta $p(s,t) \in \mathbb{R}^2$, where the parameter $s \in M$ labels points on the sheets. For $n = 2$, $l = 1$, M is a union of open or closed curves. The Hamiltonian is the kinetic energy

$$H = \frac{1}{2}\iint_{M\times M} p(s)^T \mathbf{G}(q(s) - q(t))p(u)\,\mathrm{d}s\,\mathrm{d}u \tag{1.6}$$

and the equations of motion are

$$\begin{aligned} \dot{q} &= \int_M \mathbf{G}(q(s) - q(u))p(u)\,\mathrm{d}u, \\ \dot{p} &= -\int_M p(s)^T \nabla\mathbf{G}(q(s) - q(u))p(u)\,\mathrm{d}u \end{aligned} \tag{1.7}$$

or, for Euclidean-invariant and diagonal $\mathbf{G}$,

$$
\begin{aligned}
H &= \frac{1}{2}\iint_{M\times M} G(r)p(s)\cdot p(u)\,\mathrm{d}s\,\mathrm{d}u,\\
\dot{q} &= \int_M G(r)p(u)\,\mathrm{d}u,\\
\dot{p} &= -\int_M p(s)\cdot p(u)G'(r)\frac{\mathbf{r}}{r}\,\mathrm{d}u,
\end{aligned}
\tag{1.8}
$$

where

$$
\mathbf{r} = q(s) - q(u), \quad r = \|\mathbf{r}\|. \tag{1.9}
$$

If the metric is Euclidean-invariant then (1.1), (1.2), and (1.8) are too, and one can, via a symmetry reduction, study the behavior of translation or rotation-invariant initial data. This is carried out for sets of parallel straight and concentric circular momentum sheets in [6]. Our goal here is to investigate the stability of these solutions in the simplest settings of either (i) a single straight sheet, moving at a certain angle to itself, or (ii) a single expanding circular sheet.

The motion of straight vortex sheets for the Euler fluid equations, subject to the classical Kelvin–Helmholtz instability, is known to be ill-posed [14]. The growth rate of Fourier mode ω is $\mathcal{O}(\omega)$, so that even for analytic initial data, eventually analyticity is lost and the (classical) solution terminates in a singularity. The growth rates can be moderated by incorporating either a physical regularizing effect, such as surface tension [10] or altering the metric from L^2 to H^1 to obtain the Euler-α or second-grade fluid equations [11], or a nonphysical regularization, such as smoothing the kernel.

We shall see that straight momentum sheets are also unstable for the diffeomorphism case, and because of the close analogy in the geometric structure of the equations and of the initial data, we shall call it a Kelvin–Helmholtz instability. However, there are several key differences between the fluid and diffeomorphism cases. The latter has no volume-preserving constraint and more degrees of freedom (the sheet strength is constant for fluids, while $p(x,t) \in \mathbb{R}^2$ for diffeomorphisms); these make the diffeomorphism case more unstable. But it has more filtering of the high modes in its inertia operator ($\mathcal{A} = 1$ for the fluid case, while typically $\mathcal{A} = (1-\nabla^2)^k$ for diffeomorphisms), which makes the diffeomorphism case more stable. The stability and well-posedness of the diffeomorphism case is therefore difficult to predict.

Stretched circular vortex sheets, on the other hand, are stable if the stretching is sufficiently strong. (A single vortex sheet does not expand by itself but can be made to expand by placing a line source at the origin.) They are well-posed but unstable under slow expansion, and stable under fast expansion [21]. We expect to see similar behavior here, especially because the numerical solutions for smooth initial data tend to show curved sheets expanding radially.

In section 2 we describe the particle-relabeling symmetry of (1.7) and its associated conserved quantity, the potential vorticity $q_s\cdot p$. In section 3.1 we linearize the equations of motion (1.8), giving a system of four linear integro-differential equations. Remarkably, for straight sheets they can be solved, giving a complete and explicit determination of linear stability for all metrics (section 3.2). One case, of a straight sheet moving normally to itself under an H^1

metric (section 3.3), is only algebraically unstable, and we prove in Proposition 3.9 that even this instability is confined to the Lagrangian particle motion; the Eulerian motion of the sheet is linearly stable. In section 3.4 we follow the evolution of the unstable cases numerically. For circular sheets (section 4) the linearized equations are even worse, having nonconstant coefficients in space and time. The time variation can only be dealt with by taking the limit of large radius; in this case the dependence on the metric drops out and the resulting PDE can be solved explicitly. We find that radially expanding circular sheets are algebraically unstable for all metrics.

2. Symmetries and conserved quantities. The equations of motion for momentum sheets, (1.7), have a family of conserved quantities arising from a particle-relabeling symmetry. More precisely, the symmetry group $\mathfrak{G} = \mathrm{Diff}(M)$ acts on sheet positions in $Q = \mathrm{Emb}(M, \mathbb{R}^2)$ on the right by composition,

$$\mathfrak{G} \times Q \to Q : (\varphi, q(s)) \mapsto q(\varphi(s)), \tag{2.1}$$

and on T^*Q by cotangent lifts,

$$\mathfrak{G} \times T^*Q \to T^*Q : (\varphi, (q(s), p(s))) \mapsto (q(\varphi(s)), p(\varphi(s))\varphi'(s)). \tag{2.2}$$

Equation (2.2) is a symplectic map, as can be seen by the fact that it is generated by the generating function $S(q, P)$ as $(q(s), p(s)) \mapsto (Q(s), P(s))$ with

$$S(q, P) = \int_M q(\varphi(s))P(s)\,\mathrm{d}s = \int_M q(s)P(\varphi^{-1}(s))\frac{\mathrm{d}s}{\varphi'(\varphi^{-1}(s))} \tag{2.3}$$

according to

$$\begin{aligned} Q(s) &= \frac{\delta S}{\delta P} = q(\varphi(s)), \\ p(s) &= \frac{\delta S}{\delta q} = P(\varphi^{-1}(s))\frac{1}{\varphi'(\varphi^{-1}(s))} \Rightarrow P(s) = p(\varphi(s))\varphi'(s). \end{aligned} \tag{2.4}$$

The generators of the group action are the vector fields

$$\begin{aligned} \dot{q} &= f(s)q_s, \\ \dot{p} &= (f(s)p)_s, \end{aligned} \tag{2.5}$$

where $f(s) = \dot{\varphi}(s)$. These are Hamiltonian vector fields with Hamiltonian $J_f = \int_M f(s)q_s(s) \cdot p(s)\,\mathrm{d}s$, which is therefore the momentum map for the group action. The Hamiltonian (1.6) is invariant under the group action

$$\begin{aligned} g \cdot H &= \frac{1}{2}\iint_{M\times M} \varphi'(s)p(\varphi(s))^T\mathbf{G}(q(\varphi(s)) - q(\varphi(u)))p(\varphi(u))\varphi'(u)\,\mathrm{d}s\,\mathrm{d}u \\ &= \frac{1}{2}\iint_{M\times M} p(\tilde{s})^T\mathbf{G}(q(\tilde{s}) - q(\tilde{u}))p(\tilde{u})\,\mathrm{d}\tilde{s}\,\mathrm{d}\tilde{u} \\ &= H, \end{aligned} \tag{2.6}$$

where we have made the change of variables $\tilde{s} = \varphi(s)$, $\tilde{u} = \varphi(u)$. Therefore, the momentum map J_f is a conserved quantity of Hamilton's equations for H for any function $f(s)$. Choosing $f(s)$ to be a delta-function, we get the pointwise conserved quantities $J(s) = q_s(s) \cdot p(s)$ for each s, whose conservation can be directly checked from the equations of motion. By analogy with the shallow water equations we call $J(s)$ the *potential vorticity* of the momentum sheet.

It can be seen from (2.2) that the momentum sheet strength $p(s)$ is not a fundamental physical property of the sheet, because it changes when different coordinates are used on the sheet. The fundamental variables of the sheets are the invariants of the group action, namely, the unit tangent vector $q_s(s)/\|q_s(s)\|$ and the *true momentum* $p(s)/\|q_s(s)\|$. The true momentum contains only one independent component because of the conservation of potential vorticity $q_s(s) \cdot p(s)$.

If the metric is Euclidean-invariant the equations have four other conserved quantities: the energy H, the total momentum $\int p(s)\,\mathrm{d}s$ (from translation symmetry), and the angular momentum $\int q(s) \times p(s)\,\mathrm{d}s$ (from rotational symmetry).

3. Stability of straight sheets.

3.1. Analysis of the general case. We first determine the linearized equations of motion for straight sheets, which, because of translation symmetry, can be diagonalized by a Fourier transform. Finding the growth rates requires finding the eigenvalues of 4×4 matrices. However, this is not required in the determination of stability, for which an explicit condition can be found. We give the condition in terms of the critical angle of propagation of the sheet for a disturbance of a given wavenumber to be unstable as a function of the Fourier transform of the Green's function of the metric.

Proposition 3.1. *Consider a straight sheet located at $q(s,t) = tU + (0,s)^{\mathrm{T}}$ with constant momentum $p(s,t) = \rho = (\rho_1, \rho_2)$ and velocity $U = (U_1, U_2)$, where s is the Lagrangian parameter along the sheet. Under a small perturbation to $(q,p) + \varepsilon z(s,t)$, the growth of the perturbation is controlled to leading order by the 4×4 linear systems*

$$\dot{\tilde{z}} = \begin{pmatrix} 0 & a\widetilde{G} & \widetilde{G} & 0 \\ 0 & b\widetilde{G} & 0 & \widetilde{G} \\ c & 0 & 0 & 0 \\ 0 & (a^2+b^2)\widetilde{G} & a\widetilde{G} & b\widetilde{G} \end{pmatrix} \tilde{z}, \tag{3.1}$$

where

$$\begin{aligned} \tilde{z}(\omega,t) &= \int_{-\infty}^{\infty} z(s,t) e^{-2\pi i\omega s}\,\mathrm{d}s, \\ \widetilde{G}(\omega) &= \int_{-\infty}^{\infty} G(\omega) e^{-2\pi i\omega s}\,\mathrm{d}s, \\ c &= \rho^2 \widetilde{F}|_0^{\omega}, \\ a &= -2\pi i\omega\rho_1, \\ b &= -2\pi i\omega\rho_2, \\ \rho^2 &= \rho_1^2 + \rho_2^2, \quad \textit{and} \end{aligned}$$

$$\widetilde{F}|_0^\omega = 4\pi^2 \int_0^\omega \omega' \widetilde{G}(\omega')\,\mathrm{d}\omega'.$$

The system (3.1) has an eigenvalue with positive real part, indicating an exponential instability in wavenumber ω, if and only if

$$\Delta := -1 + (11\cos^2\phi - 8)\mu + (\cos^4\phi + 12\cos^2\phi - 16)\mu^2 + (\cos^4\phi)\mu^3 < 0, \tag{3.2}$$

where

$$\mu = \mu(\omega) = \frac{4\pi^2\omega^2\widetilde{G}(\omega)}{\widetilde{F}|_0^\omega} \tag{3.3}$$

and

$$\rho_1 = \rho\cos\phi, \quad \rho_2 = \rho\sin\phi,$$

or equivalently, if and only if

$$\cos^2\phi < \frac{(5+4\mu)^{3/2} - 11 - 12\mu}{2\mu(1+\mu)}. \tag{3.4}$$

Proof. First note that, for compatibility, $U = G_0\rho$, where $G_0 = \int_{-\infty}^{\infty} G(r)\,\mathrm{d}r$. We make a small perturbation and let

$$\begin{aligned} q(s,t) &= tU + \begin{pmatrix} 0 \\ s \end{pmatrix} + \varepsilon u, \\ p(s,t) &= \rho + \varepsilon v. \end{aligned} \tag{3.5}$$

The distance d between two points with parameters s and σ on the curve is given by

$$\begin{aligned} \|q(s) - q(\sigma)\|_2 = \Big[&((U_1 t + \varepsilon u_1(s)) - (U_1 t + \varepsilon u_1(\sigma)))^2 \\ &+((U_2 t + s + \varepsilon u_2(s)) - (U_2 t + \sigma + \varepsilon u_2(\sigma)))^2\Big]^{1/2} \\ = {}& s - \sigma + \varepsilon(u_2(s) - u_2(\sigma)) + \mathcal{O}(\varepsilon^2), \end{aligned} \tag{3.6}$$

where $s - \sigma$ is held fixed. The first equation of motion

$$\dot{q} = \int_{-\infty}^{\infty} G(\|q(s) - q(t)\|_2) p(\sigma)\,\mathrm{d}\sigma \tag{3.7}$$

linearizes to

$$\begin{aligned} \dot{u} &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_{-\infty}^{\infty} G(s - \sigma + \varepsilon(u_2(s) - u_2(\sigma)))(\rho + \varepsilon v(\sigma))\,\mathrm{d}\sigma \\ &= \int_{-\infty}^{\infty} \big(G(s-\sigma)v(\sigma) + \rho G'(s-\sigma)(u_2(s) - u_2(\sigma))\big)\,\mathrm{d}\sigma \\ &= \int_{-\infty}^{\infty} \big(G(s-\sigma)v(\sigma) - \rho G'(s-\sigma)u_2(\sigma)\big)\,\mathrm{d}\sigma, \end{aligned} \tag{3.8}$$

where in the last line we have used that $\int_{-\infty}^{\infty} G'(\sigma)\,\mathrm{d}\sigma = 0$ because $G(\sigma)$ is an even function. As the right-hand side is a convolution, it is diagonalized by the Fourier transform, giving

$$\dot{\tilde{u}} = \widetilde{G}(\tilde{v} - 2\pi i\omega \tilde{u}_2 \rho). \tag{3.9}$$

(We are using the Fourier transform pair $\tilde{u}(\omega) = \int_{-\infty}^{\infty} u(x)e^{-2\pi i\omega x}\,\mathrm{d}x$, $u(x) = \int_{-\infty}^{\infty} \tilde{u}(\omega)e^{2\pi i\omega x}\,\mathrm{d}\omega$.) Introducing $F(r) = G'(r)/r$, the momentum equation is

$$\begin{aligned} \dot{p} &= -\int_{-\infty}^{\infty} (p(s)\cdot p(\sigma))\nabla G(\|q(s)-q(\sigma)\|_2)\,\mathrm{d}\sigma \\ &= -\int_{-\infty}^{\infty} (p(s)\cdot p(\sigma))F(\|q(s)-q(\sigma)\|_2)(q(s)-q(\sigma))\,\mathrm{d}\sigma, \end{aligned} \tag{3.10}$$

which linearizes to

$$\begin{aligned} \dot{v} &= -\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_{-\infty}^{\infty} [(\rho_1+\varepsilon v_1(s))(\rho_1+\varepsilon v_1(\sigma)) + (\rho_2+\varepsilon v_2(s))(\rho_2+\varepsilon v_2(\sigma))] \\ &\qquad F(s-\sigma+\varepsilon(u_2(s)-u_2(\sigma)))\begin{pmatrix} \varepsilon(u_1(s)-u_1(\sigma)) \\ s-\sigma+\varepsilon(u_2(s)-u_2(\sigma)) \end{pmatrix}\mathrm{d}\sigma \\ &= -\int_{-\infty}^{\infty} \begin{pmatrix} \rho^2 F(s-\sigma)(u_1(s)-u_1(\sigma)) \\ (\rho\cdot v(\sigma))G'(s-\sigma) + \rho^2 G''(s-\sigma)(u_2(s)-u_2(\sigma)) \end{pmatrix}\mathrm{d}\sigma, \end{aligned} \tag{3.11}$$

which is diagonalized by the Fourier transform to become

$$\begin{aligned} \dot{\tilde{v}}_1 &= \rho^2 \widetilde{F}|_0^{\omega} \tilde{u}_1, \\ \dot{\tilde{v}}_2 &= -(2\pi i\omega(\rho\cdot\tilde{v}) + (2\pi\omega)^2\rho^2\tilde{u}_2)\widetilde{G}. \end{aligned} \tag{3.12}$$

Here we have used, for example, $\int_{-\infty}^{\infty} F(\sigma)\,\mathrm{d}\sigma = \widetilde{F}(0)$. The diagonalized, linearized equations of motion (3.9), (3.12) are given collectively by the 4×4 linear systems in (3.1). The expression for $\widetilde{F}|_0^{\omega}$ in terms of $\widetilde{G}(\omega)$ is easily established by manipulating the Fourier transforms.

The characteristic polynomial of the coefficient matrix of (3.1) is

$$\begin{aligned} P(\lambda) &= \lambda\left((c\widetilde{G}-\lambda^2)(2bG-\lambda) - a^2\widetilde{G}^2\lambda\right) \\ &= \lambda\left((\rho^2\widetilde{F}|_0^{\omega}\widetilde{G}-\lambda^2)(-4\pi i\omega\rho_2\widetilde{G}-\lambda) + 4\pi^2\omega^2\rho_1^2\widetilde{G}^2\lambda\right). \end{aligned} \tag{3.13}$$

Although $P(\lambda)$ has complex coefficients, $iP(i\lambda)$ has real coefficients. If the cubic $iP(i\lambda)/\lambda$ has one real and two complex conjugate roots, then $P(\lambda)$ has precisely one root with a positive real part, indicating instability. This is true if and only if the discriminant of the cubic is negative. This discriminant, after removing the positive factor (see Proposition 3.5 and Corollary 3.6 below) $4(\widetilde{G}\widetilde{F}|_0^{\omega})^3$ and expressing it in terms of the new variable $\mu(\omega)$, is Δ in (3.2).

The stability criterion (3.4) is then determined simply by solving a quadratic equation. ■

Note that (3.1) is a *complex* Hamiltonian system, with conjugate variables $\tilde{u}(\omega)$ and $\bar{\tilde{v}}(\omega)$, so its eigenvalues do not obey the usual $(\lambda, -\lambda, \bar{\lambda}, -\bar{\lambda})$ symmetry. When written as an 8×8 real Hamiltonian system, all the eigenvalues given below are paired with their complex conjugates.

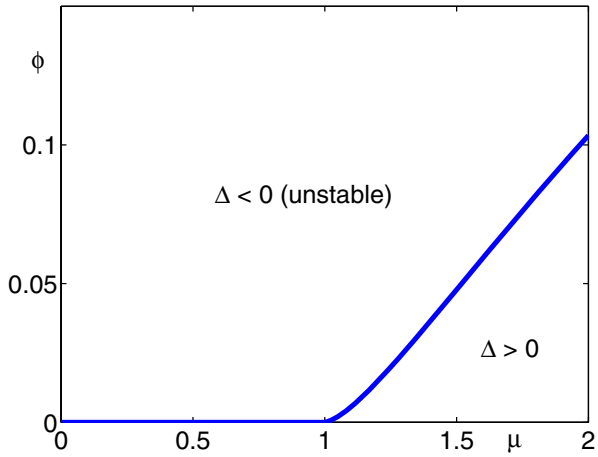


Figure 1. *The stability region for straight sheets in terms of ϕ, the angle of propagation of the sheet to its normal (i.e., $\rho_1 = \rho\cos\phi$, $\rho_2 = \rho\sin\phi$), and $\mu(\omega) = \omega^2\widetilde{G}(\omega)/\int_0^\omega w'\widetilde{G}(\omega')\,\mathrm{d}\omega'$. Here ω is the wavenumber of the disturbance and $\widetilde{G}$ is the Fourier transform of the Green's function of the metric. Since typically $\mu \le 2$, only sheets moving nearly normal to themselves can be stable at any wavenumber.*

Note from (3.13) that $\lambda = 0$ is always an eigenvalue. This corresponds to the particle-relabeling symmetry and its associated conserved quantity, the potential vorticity $J(s)$.

The instability region in terms of ϕ and μ is shown in Figure 1. For the metrics we will consider, it is always the case that $0 < \mu(\omega) \le 2$ for all ω.

Corollary 3.2. *$\Delta = -1$ at $\mu = 0$ for all ϕ, indicating instability when μ is sufficiently small.*

(We shall see below that for all H^k metrics with $k > \frac{3}{2}$, $\mu(\infty) = 0$; i.e., all such straight sheets are unstable at high wavenumbers.)

Corollary 3.3. *Stability can be lost, but never gained, with increasing angle of propagation ϕ.*

Proof. The coefficients of μ in Δ (3.2) are all nonincreasing functions of $\phi \in [0, \pi/2]$. ∎

Corollary 3.4. *The eigenvalues can be reduced to the solution of a quadratic in two cases, when $\rho_1 = 0$ and when $\rho_2 = 0$.*

When $\rho_2 = 0$ the sheet is moving perpendicularly to itself and the eigenvalues are $(0, 0, \lambda, -\lambda)$, where

$$\lambda^2 = \widetilde{G}(c + a^2\widetilde{G}) = \widetilde{G}(\widetilde{F}|_0^\omega - (2\pi\omega)^2\widetilde{G}) = \widetilde{G}\widetilde{F}|_0^\omega(1-\mu). \tag{3.14}$$

The double zero eigenvalue has a single eigenvector and hence generates an algebraic $\mathcal{O}(t)$ instability for all G. Exponential stability, on the other hand, is determined by μ and depends on a subtle balance between the Fourier transforms of $G(r)$ and $G'(r)/r$.

When $\rho_1 = 0$ the sheet is moving tangentially to itself and the eigenvalues are

$$(0, -4\pi i\rho_2\omega\widetilde{G}, \pm\rho_2\sqrt{\widetilde{F}|_0^\omega\widetilde{G}}).$$

We will consider Euclidean-invariant metrics of the form $\mathcal{A} = a(-\nabla^2)$, where $a(\eta^2)$ is the Fourier symbol of the operator. For example, for the H^1 metric $1-\nabla^2$, we have $a(\eta) = 1+\eta$. For such metrics we can easily calculate both the Green's function and its Fourier transform.

Proposition 3.5. *The Green's function $G(\|x\|)$ of the operator $\mathcal{A} = a(-\nabla^2)$ and its Fourier transform $\widetilde{G}(\omega)$ are given by*

$$G(r) = \frac{1}{2\pi}\int_0^\infty \frac{1}{a(\eta^2)} J_0(\eta r)\eta \,\mathrm{d}\eta \tag{3.15}$$

(where J_0 is the zero order Bessel function of the first kind) and

$$\widetilde{G}(\omega) = \frac{1}{\pi}\int_0^\infty \frac{\mathrm{d}\nu}{a(\nu^2 + (2\pi\omega)^2)}, \tag{3.16}$$

respectively.

Proof. A Hankel transform of the radially symmetric PDE $\mathcal{A}G = \delta(\|x\|)$ gives (3.15) (see, e.g., [25]). We then have

$$\begin{aligned}
\widetilde{G}(\omega) &= \int_{-\infty}^{\infty} G(r) e^{-2\pi i\omega r}\,\mathrm{d}r \\
&= 2\int_0^\infty G(r)\cos(2\pi\omega r)\,\mathrm{d}r \\
&= \frac{1}{\pi}\int_0^\infty \mathrm{d}r \int_0^\infty \mathrm{d}\eta J_0(\eta r)\eta\cos(2\pi\omega r)\frac{1}{a(\eta^2)} \\
&= \frac{1}{\pi}\int_{2\pi\omega}^\infty \frac{\eta\,\mathrm{d}\eta}{a(\eta^2)\sqrt{\eta^2 - (2\pi\omega)^2}} \\
&= \frac{1}{\pi}\int_0^\infty \frac{\mathrm{d}\nu}{a(\nu^2 + (2\pi\omega)^2)}
\end{aligned} \tag{3.17}$$

because

$$\int_0^\infty J_0(\eta r)\cos(2\pi\omega r)\,\mathrm{d}r = \begin{cases} \frac{1}{\sqrt{\eta^2 - (2\pi\omega)^2}}, & 0 < 2\pi\omega < \eta, \\ 0, & \eta < 2\pi\omega. \end{cases} \tag{3.18}$$

Here $\nu^2 = \eta^2 - (2\pi\omega)^2$. ∎

Apart from being an easy way to calculate $\widetilde{G}$, (3.16) has the following immediate consequences.

Corollary 3.6. *For any Euclidean-invariant, diagonal metric, straight sheets moving tangentially to themselves are linearly unstable at all positive wavenumbers.*

Proof. From positive-definiteness of the metric, $a(\nu) > 0$ for all ν and hence $\widetilde{G}(\omega) > 0$. Therefore $\widetilde{F}|_0^\omega > 0$ for all $\omega > 0$ and the mode with eigenvalue $|\rho_2|\sqrt{\widetilde{F}|_0^\omega \widetilde{G}}$ is unstable. ∎

Corollary 3.7. *For any H^k metric, $a(\nu) = \mathcal{O}(\nu^k)$ as $n \to \infty$ and hence $\widetilde{G}$ exists and is $\mathcal{O}(\omega^{1-2k})$ as $\omega \to \infty$ if $k > \frac{1}{2}$. Similarly, $\widetilde{F}|_0^\omega$ exists and is $\mathcal{O}(\omega^{3-2k})$ for $\frac{1}{2} < k < \frac{3}{2}$ and $\mathcal{O}(1)$ for $k > \frac{3}{2}$. Consequently, the unstable eigenvalue for tangentially moving sheets is $\mathcal{O}(\omega^{2-2k})$ for $\frac{1}{2} < k < \frac{3}{2}$ and $\mathcal{O}(\omega^{(1-2k)/2})$ for $k > \frac{3}{2}$. Therefore* (i) *for $\frac{1}{2} < k < 1$ the equation is linearly ill-posed for initial data in any fixed Sobolev space;* (ii) *for $k \geq 1$ the growth rates are bounded for large ω and hence do not lead to linear ill-posedness.*

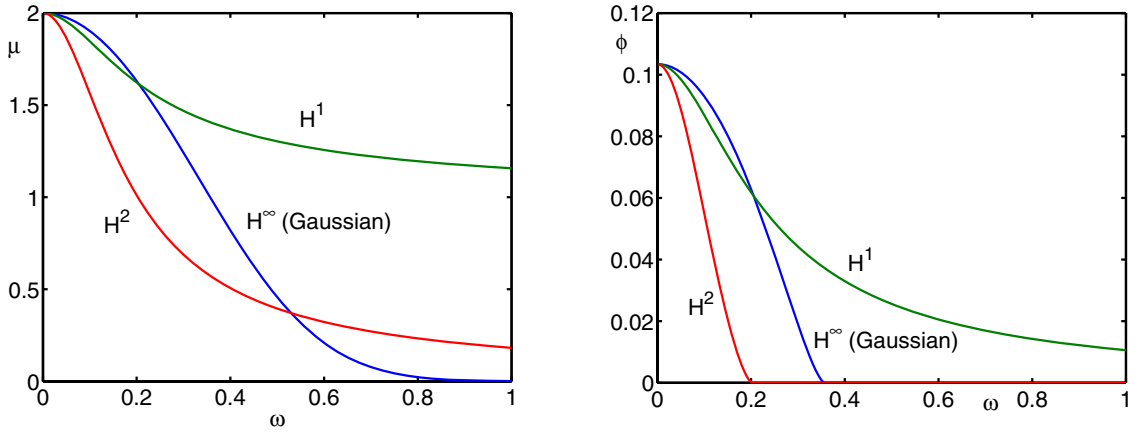


Figure 2. *Left: Each kernel $G(r)$ determines a function $\mu(\omega)$ which determines the stability of a disturbance with wavenumber ω (see (3.3), (3.4), and Figure 1). Right: The resulting stability regions for each metric. Here $\phi(\omega)$ is the critical angle of propagation of a straight sheet to its normal above which a disturbance with wavenumber ω is unstable.*

Note that the critical case $k = 1$ is interesting, however, because then the eigenvalue tends to a constant as $\omega \to \infty$. For $k > 1$ the instability will manifest itself at the frequency with the maximum eigenvalue and then follow an essentially unique nonlinear evolution, forgetting the initial perturbation, while for $k = 1$ all high frequencies grow at the same rate and the nonlinear evolution depends on the initial perturbation.

Corollary 3.8. *For any H^k metric, as $\omega \to \infty$ we have $\mu = \mathcal{O}(1)$ for $\frac{1}{2} < k < \frac{3}{2}$ and $\mu = \mathcal{O}(\omega^{3-2k})$ for $k > \frac{3}{2}$. Therefore, for sheets moving normally to themselves, all sufficiently large wavenumbers are unstable for all H^k metrics with $k > \frac{3}{2}$.*

3.2. Stability under various metrics.

Case 1. Gaussian metric. We consider the H^∞ metric $\mathcal{A} = \exp(-\nabla^2/4)$, with Fourier symbol $a(\nu^2) = \exp(\nu^2/4)$, for which an application of (3.15) and (3.16) gives the Gaussian kernel

$$G(r) = \frac{1}{\pi} e^{-r^2} \tag{3.19}$$

and

$$\begin{aligned} \widetilde{G} &= \frac{1}{\sqrt{\pi}} e^{-\pi^2\omega^2}, \\ \widetilde{F}|_0^\omega &= \frac{2}{\sqrt{\pi}} (1 - e^{-\pi^2\omega^2}) \end{aligned} \tag{3.20}$$

(see Figure 2). The growth rates for a range of propagation angles ϕ are shown in Figure 3. When $\phi = 0$, wavenumbers with $0 < \omega < 0.3568$ (i.e., wavelength more than $1/0.3568 = 2.8027$) have zero real part and are neutrally stable, while wavenumbers with $\omega > 0.3568$ are unstable. The exponential smoothing of the high frequencies by the Green's function is not enough to stabilize these perturbations. Although arbitrarily high frequencies are unstable,

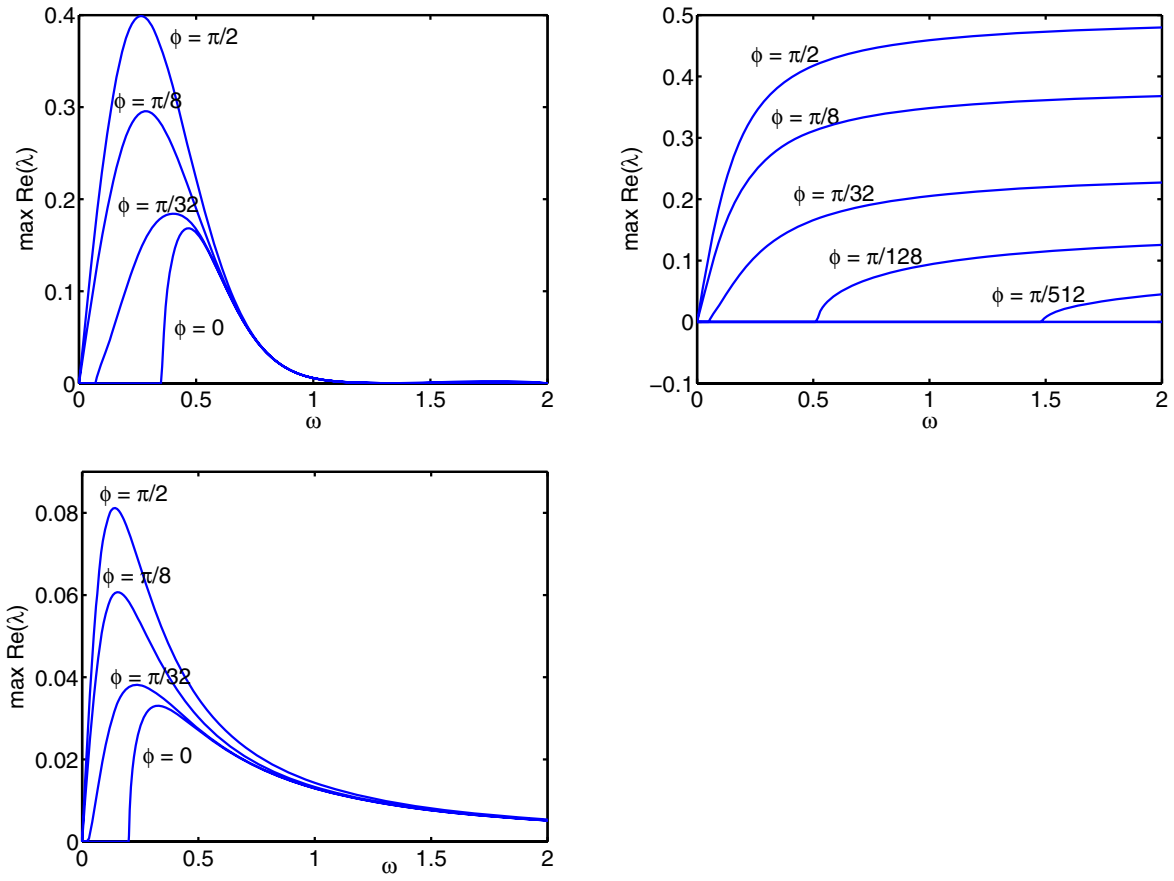


Figure 3. *The growth rate of disturbances with wavenumber ω to a straight sheet with angle of propagation ϕ to its normal. Top left: H^∞ (Gaussian) metric. Top right: H^1 metric. Bottom: H^2 metric.*

the exponential smoothing of the high frequencies is reflected in the stability analysis by their exponentially small growth rate. The growth rate is strongly peaked with a maximum of $\lambda = 0.1684$ at $\omega = 0.4671$.

When $\phi = \pi/2$, all wavenumbers are unstable, with the growth rate strongly peaked with a maximum of $\lambda = 0.3989$ at $\omega = 0.2650$.

Case 2. H^1 metric. For the H^1 metric $\mathcal{A} = 1 - \nabla^2$, $a(\nu) = 1 + \nu$ and an application of (3.15) and (3.16) give (where K_0 is the zero order modified Bessel function of the second kind)

$$\begin{aligned} G(r) &= K_0(r)/(2\pi), \\ \widetilde{G} &= \frac{1}{2\sqrt{1 + (2\pi\omega)^2}}, \\ \widetilde{F}|_0^\omega &= \frac{1}{2}\left(\sqrt{1 + (2\pi\omega)^2} - 1\right). \end{aligned} \tag{3.21}$$

In the eigenvalue calculations, the dominant terms in $\widetilde{G}$ and $\widetilde{F}|_0^\omega$ cancel out with the following results. When $\phi = 0$, the real part of all eigenvalues is zero for all ω. There is no exponential instability—only an algebraic instability due to the double zero eigenvalue. We examine this case in detail in section 3.3, proving a theorem on linear stability. Curiously, although the H^1 metric applies less smoothing to the momentum than the H^∞ metric does, it is in a sense more stable. For $\phi > 0$ there is a band of instability in (ω^*, ∞), where ω^* decreases from ∞ to 0 as ϕ increases from 0 to 0.1034. The growth rates tend to a positive constant as $\omega \to \infty$, indicating that the nonlinear evolution of the instability depends on the high-frequency part of the initial perturbation. For $\phi > 0.1034$ all wavenumbers are unstable. The maximum growth rate of $\lambda \to \frac{1}{2}$ occurs for $\phi = \pi/2$, $\omega \to \infty$.

Case 3. H^2 metric. For $\mathcal{A} = (1 - \nabla^2)^2$ an application of (3.15) and (3.16) gives (where K_1 is the first order modified Bessel function of the second kind)

$$\begin{aligned} G(r) &= \frac{1}{4\pi} r K_1(r), \\ \widetilde{G} &= \frac{1}{4\left(4\pi^2\omega^2 + 1\right)^{3/2}}, \text{ and} \\ \widetilde{F}|_0^\omega &= \frac{1}{4}\left(1 - \frac{1}{\sqrt{4\pi^2\omega^2 + 1}}\right). \end{aligned} \tag{3.22}$$

The extra smoothing of high wavenumbers in the H^2 compared to the H^1 metric is already enough to make the behavior of straight sheets qualitatively the same as the H^∞ (Gaussian) case, as the general theory of Corollaries 3.7 and 3.8 indicates. They are exponentially unstable for all ϕ. For $\phi = 0$ all wavenumbers $\omega > 0.2024$ are unstable, with the growth rate decaying slowly, $\mathcal{O}(\omega^{-3/2})$, as $\omega \to \infty$. The maximum growth rate of only $\mathrm{Re}(\lambda) = 0.03305$ comes at $\omega = 0.3266$. For $\phi = \pi/2$, the maximum growth rate is $\lambda = 0.08119$ at $\omega = 0.1404$.

3.3. Stability of straight sheets in an H^1 metric. The case $\phi = 0$ of a straight sheet moving normally to itself corresponds to the motion of a single peakon in the 1D Camassa–Holm equation. This case was shown in section 3.2, Case 2, to have no eigenvalues with positive real parts. The perturbation grows as $\mathcal{O}(t)$ due to the double zero eigenvalue. However, we now show that the algebraic instability applies only to the Lagrangian motion; the Eulerian motion, in which one identifies two sheets related by the particle-relabeling symmetry (2.2), is in fact linearly stable. The only effect of the instability is to push particles along the sheet. To study this, in the following proposition we introduce a norm that measures the perturbation to the true momentum and to the x-component of the tangent to the sheet. (Only the x-component is relevant to the reduced motion, because $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\big|_{\varepsilon=0} q_s/\|q_s\| = (u_{1s}, 0)^T$.)

Some care is required because although the Fourier modes are bounded in time, they are not uniformly bounded with respect to ω. It would be possible for the solution itself to grow in time. These considerations dictate the choice of norm in the following proposition. We use the Sobolev norms defined by $\|u\|_{H^n}^2 = \int (1 + |\omega|^{2n})|\tilde{u}|^2 \,\mathrm{d}\omega$.

Proposition 3.9. *Let $\mathcal{A} = (1 - \nabla^2)$ and consider the motion of a straight sheet moving normally to itself: let $q(s,0) = (0,s)^T + \varepsilon u(s,0)$, $p(s,0) = (1,0)^T + \varepsilon v(s,0)$, $z = (u,v)^T$. Let $j(s) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\big|_{\varepsilon=0} J(s)$ be the perturbation to the potential vorticity $J(s) = q_s \cdot p$ and $k(s,t) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\big|_{\varepsilon=0} p(s,t)/\|q_s(s,t)\|$ be the perturbation to the true momentum. Then for any $n \geq 1$ the*

motion is linearly stable in the norm

$$\|z\|_n = \|u_{1s}\|_{H^n} + \|k\|_{H^{n+1/2}} + \|j\|_{H^{n+1}}; \tag{3.23}$$

i.e., for all $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that* $\|z(\cdot,t)\|_n < \varepsilon$ *for all* $t > 0$ *and for all* $z(\cdot,0)$ *such that* $\|z(\cdot,0)\|_n < \delta$.

Proof. We know that the perturbation $z(s,t) = \mathcal{O}(t)$. First we will show that there are no $\mathcal{O}(t)$ terms in $u_1(s,t)$ and $k(s,t)$ and that $\tilde{u}_1(\omega,t)$ and $\tilde{k}(\omega,t)$ are bounded in time for all ω. However, the bounds are not uniform in ω so more precise estimates will be required to establish the final result.

First, we calculate the perturbations to the potential vorticity and true momentum:

$$\begin{aligned}
j &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} q_s \cdot p \\
&= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} \begin{pmatrix} \varepsilon u_{1s} \\ 1+\varepsilon u_{2s} \end{pmatrix} \cdot \begin{pmatrix} 1+\varepsilon v_1 \\ \varepsilon v_2 \end{pmatrix} \\
&= u_{1s} + v_2, \\
k &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} \frac{p(s,t)}{\|q_s(s,t)\|} \\
&= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} \frac{\begin{pmatrix} 1+\varepsilon v_1 \\ \varepsilon v_2 \end{pmatrix}}{\sqrt{(\varepsilon u_{1s})^2 + (1+\varepsilon u_{2s})^2)}} \\
&= \begin{pmatrix} v_1 - u_{2s} \\ v_2 \end{pmatrix}.
\end{aligned} \tag{3.24}$$

Of course, since J is conserved by the nonlinear motion, j is conserved by the linearized equations of motion. We express the linear system (3.1) in Jordan canonical form for $\omega \neq 0$ as

$$\begin{aligned}
\dot{y}_1 &= y_2, \\
\dot{y}_2 &= 0, \\
\dot{y}_3 &= i\lambda y_3, \\
\dot{y}_4 &= -i\lambda y_4,
\end{aligned} \tag{3.25}$$

where $\lambda = \sqrt{\widetilde{G}\widetilde{F}|_0^\omega(\mu-1)}$ and $\tilde{z} = Xy$ with

$$X = \begin{pmatrix} 0 & -a/c & -1 & -1 \\ 1 & 0 & a\sqrt{\widetilde{G}/\beta} & -a\sqrt{\widetilde{G}/\beta} \\ -a & 0 & c/\sqrt{\widetilde{G}\beta} & -c/\sqrt{\widetilde{G}\beta} \\ 0 & 1/\widetilde{G} & -a & -a \end{pmatrix}, \tag{3.26}$$

where $\beta = c - (2\pi\omega)^2\widetilde{G}$. As a check, we calculate that $y_2 = (-a\widetilde{G}c/\beta)(-a\tilde{z}_1 + \tilde{z}_4)$ which is proportional to the Fourier transform of $u_{1s} + v_2$, the perturbed potential vorticity, which is

why $\dot{y}_2 = 0$ (recall that $a = -2\pi i\omega$). Therefore, the solution of (3.1) for $\omega \neq 0$ is

$$\begin{aligned}\widetilde{z} &= Xy \\ &= X \begin{pmatrix} y_2(0)t \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mathcal{O}(1) \\ &= y_2(0)t \begin{pmatrix} 0 \\ 1 \\ 2\pi i\omega \\ 0 \end{pmatrix} + \mathcal{O}(1),\end{aligned} \tag{3.27}$$

from which we conclude that $\tilde{u}_1$, $\tilde{v}_2$, and $\tilde{k} = \tilde{v}_1 - 2\pi i\omega\tilde{u}_2$ are bounded in time for each ω.

To study the motion in more detail we let $w = (u_{1s}, u_2, k, j)^T = Lz$, $\tilde{w} = \widetilde{L}\tilde{z}$, and compute the solution

$$\tilde{w}(t) = \widetilde{L}X \operatorname{diag}(1, 1, \mathrm{e}^{i\lambda t}, \mathrm{e}^{-i\lambda t})X^{-1}\widetilde{L}^{-1}\tilde{w}(0), \tag{3.28}$$

of which the relevant components are

$$\begin{aligned}\widetilde{u_{1s}}(\omega, t) &= \cos(\lambda t)\widetilde{u_{1s}}(\omega, 0) + \sin(\lambda t)\gamma(\omega)\tilde{k}(\omega, 0) - (1 - \cos(\lambda t))\gamma(\omega)^2\tilde{j}(\omega, 0), \\ \tilde{k}(\omega, t) &= \cos(\lambda t)\tilde{k}(\omega, 0) - \sin(\lambda t)\left(\gamma(\omega)^{-1}\widetilde{u_{1s}}(\omega, 0) + \gamma(\omega)\tilde{j}(\omega, 0)\right),\end{aligned} \tag{3.29}$$

where

$$\begin{aligned}\gamma(\omega) &= 2\pi i\omega\sqrt{\widetilde{G}/\beta} \\ &= 2\pi\omega\left(\left(1 + 4\pi^2\omega^2\right)^{1/2} - 1\right)^{-1/2} \\ &= \begin{cases} \sqrt{2}, & \omega \to 0, \\ \mathcal{O}(\omega^{1/2}), & \omega \to \infty. \end{cases}\end{aligned} \tag{3.30}$$

Using $\lambda(0) = 0$ we can check that (3.29) holds for $\omega = 0$ as well as for $\omega \neq 0$. Elementary calculus now establishes the bounds

$$\begin{aligned}|\gamma(\omega)| &\leq \sqrt{2\pi}(1 + |\omega|)^{1/2} \quad \forall\omega, \\ |\gamma(\omega)^{-1}| &\leq \frac{1}{\sqrt{2}}(1 + |\omega|)^{-1/2} \quad \forall\omega.\end{aligned} \tag{3.31}$$

We will also use

$$\begin{aligned}(1 + |\omega|)^l(1 + |\omega|^{2n}) &\leq \sqrt{2}(1 + |\omega|^{2n+l}), \\ (1 + |\omega|)^{-l}(1 + |\omega|^{2n}) &\leq (1 + |\omega|^{2n-l})\end{aligned} \tag{3.32}$$

for all ω and for all $n, l > 0$. We then have

$$\begin{aligned}
\|u_{1s}(\cdot,t)\|_{H^n} &\le \|u_{1s}(\cdot,0)\|_{H^n} + \|\mathcal{F}^{-1}\gamma(\omega)\tilde{k}(\omega,0)\|_{H^n} + 2\|\mathcal{F}^{-1}\gamma(\omega)^2)\tilde{j}(\omega,0)\|_{H^n} \\
&= \|u_{1s}(\cdot,0)\|_{H^n} + \left(\int (1+|\omega|)^{2n}|\gamma(\omega)\tilde{k}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&\quad + \left(\int (1+|\omega|)^{2n}|\gamma(\omega)^2\tilde{j}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&\le \|u_{1s}(\cdot,0)\|_{H^n} + \left(2\sqrt{2}\pi\int (1+|\omega|^{2n+1})|\tilde{k}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&\quad + 2\left(4\sqrt{2}\pi^2\int (1+|\omega|^{2n+2})|\tilde{j}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&= \|u_{1s}(\cdot,0)\|_{H^n} + \sqrt{2\sqrt{2}\pi}\|k(\cdot,0)\|_{H^{n+1/2}} + 2^{1/4}4\pi\|j(\cdot,0)\|_{H^{n+1}}
\end{aligned} \tag{3.33}$$

and similarly

$$\begin{aligned}
\|k(\cdot,t)\|_{H^{n+1/2}} &\le \|k(\cdot,0)\|_{H^n+1/2} + \|\mathcal{F}^{-1}\gamma(\omega)^{-1}\widetilde{u_{1s}}(t)\|_{H^{n+1/2}} + \|\mathcal{F}^{-1}\gamma(\omega)\tilde{j}(0)\|_{H^{n+1/2}} \\
&= \|k(\cdot,0)\|_{H^n+1/2} + \left(\int (1+|\omega|)^{2n+1}|\gamma(\omega)^{-1}\tilde{k}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&\quad + \left(\int (1+|\omega|)^{2n+1}|\gamma(\omega)\tilde{j}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&\le \|k(\cdot,0)\|_{H^n+1/2} + \left(\frac{1}{2}\int (1+|\omega|^{2n})|\tilde{k}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&\quad + \left(2\sqrt{2}\pi\int (1+|\omega|^{2n+2})|\tilde{j}(\omega,0)|^2\,\mathrm{d}\omega\right)^{1/2} \\
&= \|k(\cdot,0)\|_{H^{n+1/2}} + \frac{1}{\sqrt{2}}\|u_{1s}(\cdot,0)\|_{H^n} + \sqrt{2\sqrt{2}\pi}\|j(\cdot,0)\|)_{H^{n+1}}.
\end{aligned} \tag{3.34}$$

Because j is a conserved quantity, we also have that $\|j(\cdot,t)\|_{H^{n+1}} = \|j(\cdot,0)\|_{H^{n+1}}$. Therefore

$$\begin{aligned}
\|z(\cdot,t)\|_n &= \|u_{1s}(\cdot,t)\|_{H^n} + \|k(\cdot,t)\|_{H^{n+1/2}} + \|j(\cdot,t)\|_{H^{n+1}} \\
&\le \left(1+\frac{1}{\sqrt{2}}\right)\|u_{1s}(\cdot,0)\|_{H^n} + (1+\sqrt{2\sqrt{2}\pi})\|k(\cdot,0)\|_{H^{n+1/2}} + C\|j(\cdot,0)\|_{H^{n+1}} \\
&\le C\|z(\cdot,0)\|_n,
\end{aligned} \tag{3.35}$$

where $C = 1 + \sqrt{2\sqrt{2}\pi} + 2^{1/4}4\pi$. That is, we have the required stability with $\delta = \varepsilon/C$. ∎

Note that stability of the y-component v_2 of the true momentum follows from $v_2 = j - u_{1s}$, which implies that $\|v_2(\cdot,t)\|_{H^n} \le \|j(\cdot,t)\|_{H^n} + \|u_{1s}(\cdot,t)\|_{H^n} \le C\|z(\cdot,0)\|_n$ for some constant C.

It is necessary in Proposition 3.9 for the norm to include u_{1s} and not u_1 itself, the horizontal perturbation, for u_1 is not in fact bounded. To see why, we examine the $\omega = 0$ Fourier mode of the solution. From (3.1), this obeys $\dot{\tilde{u}}(0,t) = \widetilde{G}(0)\tilde{v}(0,t)$, $\dot{\tilde{v}}(0,t) = 0$, with solution

$\tilde{u}(0,t) = \tilde{u}(0,0) + \widetilde{G}(0)\tilde{v}(0,0)t$, $\tilde{v}(0,t) = \tilde{v}(0,0)$. In other words, the average speed of the sheet is increased by $\widetilde{G}(0)\int v(s,0)\,\mathrm{d}s$, leading to a linear drift away from the unperturbed position. However, it can be proved, analogously to the proof of Proposition 3.9, that the motion is *orbitally* stable in a norm including $u_1(\cdot,0)$.

By the Sobolev embedding theorem, these results also imply that the derivatives of the solution up to a suitable order are also bounded by $C\varepsilon$ for some constant C.

3.4. Numerical study of the instability. To follow the nonlinear development of the instability we will solve the Euler equations for momentum sheets, (1.8), numerically. The basic method is to discretize the sheet by a set of Lagrangian particles, or, equivalently, to approximate the integrals in (1.8) by constant-weight quadrature. Let the particles have positions q_i and momenta p_i, $i = 1, \dots, N$, and let the particle spacing be h. The Hamiltonian becomes

$$H = \frac{1}{2}h^2 \sum_{i,j=1}^{N} G(\|q_i - q_j\|) p_i \cdot p_j, \tag{3.36}$$

and the equations of motion are

$$\begin{aligned} \dot{q}_i &= h \sum_{j=1}^{N} G(\|q_i - q_j\|) p_j, \\ \dot{p}_i &= -h \sum_{j=1}^{N} p_i \cdot p_j G'(\|q_i - q_j\|) \frac{q_i - q_j}{\|q_i - q_j\|}. \end{aligned} \tag{3.37}$$

We integrate these with the symplectic midpoint rule [4, 18] $x_{k+1} = x_k + \Delta t f(\bar{x}_k)$, where the midpoint value $\bar{x}_k = (x_k + x_{k+1})/2$. Previous midpoint values are stored and extrapolated to provide a high-order initial guess for $\bar{x}_k$, so that one can solve for x_{k+1} by simple iteration to within the desired tolerance (typically 10^{-12}) in just two or three iterations. Higher-order variants were tested but not needed.

To simulate periodic sheets (say, with period 1 in the y-direction), it is necessary to augment the Hamiltonian to include the effect of the periodic images:

$$H_{\mathrm{ext}} = \frac{1}{2}h^2 \sum_{i,j=1}^{N} \sum_{k=-M}^{M} G(\|q_i - q_j + k(0,1)^T\|) p_i \cdot p_j. \tag{3.38}$$

The parameter M (the number of periodic images) is adjusted so that the contribution of the omitted images is less than 10^{-12}. Since $G(r) = o(e^{-y/\alpha})$ in all the cases we consider, and we use a length-scale $\alpha = 0.2$, $M \le 3$ is sufficient. (In the motion of vortex sheets, governed by the Biot–Savart law, the contribution of all the periodic images can be summed analytically; but we could not achieve this for our choice of kernel.)

For analytic circular (resp., periodic) sheets and analytic G, the discretization (3.36) (resp., (3.38)) corresponds to discretizing an analytic periodic integral by the trapezoidal rule and hence is spectrally accurate. For H^k metrics, however, G is not analytic. In this case we apply the theory of quadrature of weakly singular periodic integrands developed by Sidi and Israeli [26]. (A similar issue arises for axisymmetric vortex sheets [23].)

Theorem 3.10 (from [26]). *If $\phi(u) = |u-s|^\sigma \log|u-s| f(u) + \tilde{f}(u)$, $\sigma > -1$, with $f(u)$ and $\tilde{f}(u)$ $2m$ times differentiable on $[a,b]$, and $\phi(u)$ periodic with period $T = b-a$ and $2m$ times differentiable on $(-\infty,\infty)\backslash\{t+kT\}_{k=-\infty}^{\infty}$, grid points $u_j = a+jh$, $h=(b-a)/n$, $s=u_i$ for some i, ζ the Riemann zeta function, and*

$$Q_n[\phi] = h \sum_{\substack{j=1 \\ u_j \neq s}}^{n} \phi(u_j) + \tilde{f}(s)h + 2[\zeta'(-\sigma) - \zeta(-\sigma)\log h] f(s) h^{\sigma+1}, \tag{3.39}$$

then the error

$$\int_a^b \phi(u)\,\mathrm{d}u - Q_n[\phi] = 2\sum_{\mu=1}^{m-1} [\zeta'(-\sigma-2\mu) - \zeta(-\sigma-2\mu)\log h] \frac{f^{(2\mu)}(s)}{(2\mu)!} h^{2\mu+\sigma+1} + \mathcal{O}(h^{2m}) \tag{3.40}$$

as $h \to 0$.

Corollary 3.11. *The Hamiltonians (3.36) for circular sheets and (3.38) for straight sheets, for Euclidean-invariant H^k metrics with $k \geq 2$, have quadrature errors of $\mathcal{O}(h^{2k-1})$.*

Proof. The form of the singularity in $G(r)$ at $r=0$ is controlled by the highest derivatives ∇^{2k} in the inertia operator $\mathcal{A}$ of the metric. $G(r)$ has the form used in Theorem 3.10 with $\sigma = 2k-2$. For $k \geq 2$, the term $\tilde{g}(t) = G(0)$ is included in the quadrature in (3.36); the term $\zeta(-\sigma)$ drops out because $\zeta(2-2k) = 0$ for $k \geq 2$; but the term $\zeta'(-\sigma)$ is missing from the quadrature (3.36) and hence determines the leading order error of the quadrature. For circular sheets, the integrands are periodic as in Theorem 3.10, while for straight sheets the integrands decay rapidly at infinity, so there are no boundary terms. ∎

The interesting case $k=1$ requires more attention. In this case $G(r) \sim c\log|r|$ as $r \to 0$, i.e., $\sigma = 0$ in (3.39). Simply omitting the singular term from the integrand, i.e., setting $G(0) = 0$, gives a very poor quadrature rule of error $\mathcal{O}(h\log h)$. Consider the integrand $G(\|q(s)-q(u)\|)p(u)$, for which an expansion near $u=s$ gives

$$G(\|q(s)-q(u)\|)p(u) = c\log(|s-u|\|q'(s)\|)p(s) + o(s-u),$$

i.e., $f(s) = cp(s)$ and $\tilde{f}(s) = c\log\|q'(s)\|p(s)$. We could therefore apply the quadrature of (3.39) to get a method of order $\mathcal{O}(h^3)$. However, this would involve approximating the term $\|q'(s)\|$ by, e.g., finite differences, leading to complicated ODEs. Instead, we have started with the quadrature

$$P_n[\phi] = -2cp(s)\zeta(0)h\log h + h\sum_{\substack{j=1 \\ u_j \neq s}}^{n} \phi(u_j), \tag{3.41}$$

which is $\mathcal{O}(h) + \mathcal{O}(h^3)$ accurate. It can be interpreted as a regularization of the kernel with $G(0) = -2c\zeta(0)\log h = c\log h$. One step of Richardson extrapolation, i.e., $2P_n[\phi] - P_{n/2}[\phi]$, eliminates the $\mathcal{O}(h)$ term, establishing the following.

Proposition 3.12. *For kernels obeying $G(r) \sim c \log |r|$ as $r \to 0$, the quadrature error*

$$\int G(\|q(s) - q(u)\|)p(u)\,\mathrm{d}u - h\sum_{j=1}^{n} G_{ij}p_j = \mathcal{O}(h^3), \tag{3.42}$$

where

$$G_{ij} = \begin{cases} G(\|q_i - q_j\|), & j - i \text{ odd}, \\ -2c\log 2, & j = i, \\ 0, & j \neq i,\ j - i \text{ even}. \end{cases} \tag{3.43}$$

The quadrature here is essentially the "every other point" trapezoidal rule (as studied, e.g., as the modified point vortex method in [3]) but with the singular point included with a special weight to account for the singularity.

For the standard H^1 metric with $\mathcal{A} = (1 - \alpha^2\nabla^2)$, we have $G(r) = K_0(r)/(2\pi\alpha^2) \sim -\frac{1}{2\pi\alpha^2}\log|r|$ as $r \to 0$, i.e., $c = -\frac{1}{2\pi\alpha^2}$.

There are two further numerical difficulties with the H^1 kernel. First, recall from section 3.2 that the bounded growth rates of the high frequencies depend delicately on the asymptotic properties of $\widetilde{G}(\omega)$. These properties are destroyed by the numerical quadrature of (3.42), and indeed, a direct solution of Hamilton's equations shows a high frequency instability. We have suppressed this instability using the Krasny filter [13]; during the simulation, any Fourier modes with an amplitude less than a chosen cutoff (we used 10^{-10}) are set to 0. This prevents growth of the instability until the amplitudes of all modes are greater than the cutoff. After this time the evolution of the sheet tends to proceed very quickly until it reaches the point at which all accuracy is lost, typically just before a collision of the sheet with itself, so that the instability does not have time to dominate the solution. Convergence of point vortex methods for vortex sheets using the Krasny filter is established in [3].

The second difficulty occurs only in the simulation of sheets moving transversely to themselves, i.e., in the case $\phi = 0$. Recall that this case has a double zero eigenvalue for each wavenumber. Under the numerical discretization, this double zero splits into a $\pm$ pair of eigenvalues of the order of the square root of the discretization error. This is not a problem for the Gaussian kernel, for which we have spectral accuracy, but it is a problem for the H^1 case. The eigenvalue pair can be real or imaginary depending on h and on the metric. If real, it creates a numerical instability affecting all wavenumbers. The only way we have found to deal with this is to perform careful convergence checks of all our results for this case.

The results for a number of simulations are shown in Figures 4–11 along with some error estimates. (There are animations of the sheet motion available for each of these cases at http://www-ist.massey.ac.nz/smarsland/KelvinHelmholtz.html.) In each case we have computed three approximate solutions, $x^{(1)} = x_{h,\Delta t}$, $x^{(2)} = x_{\frac{1}{2}h,\frac{1}{2}\Delta t}$, and $x^{(3)} = x_{h,\frac{1}{2}\Delta t}$. The global error in $x^{(2)}$ due to the (second-order) time discretization is estimated by $(x^{(3)} - x^{(1)})/3$. The estimation of the global error due to the space discretization depends on the metric. For the H^1 metric, it is estimated for $x^{(2)}$ by $(x^{(2)} - x^{(3)})/7$ (since the spatial error is $\mathcal{O}(h^3)$), and for the Gaussian metric it is estimated for (the less accurate solution) $x^{(3)}$ by $x^{(2)} - x^{(3)}$; the error in $x^{(2)}$ is much smaller. We also compute the error in the potential vorticity $J(s)$. We then take 2-norms with respect to space for all the errors and plot their evolution in time

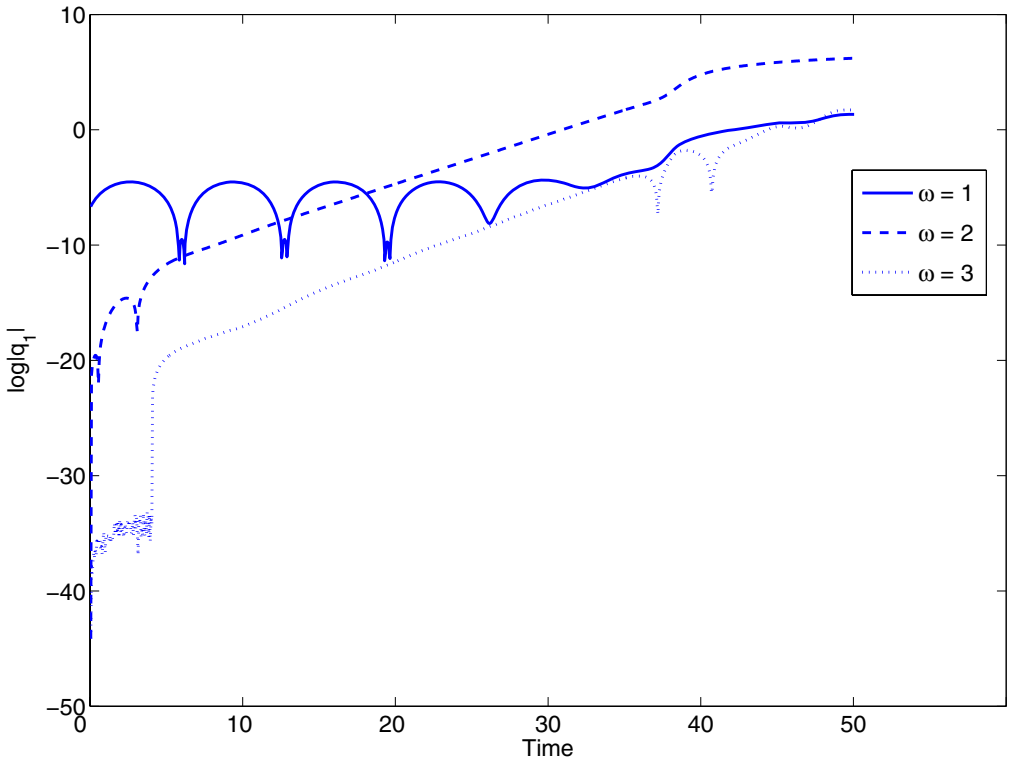


Figure 4. *Evolution of an initially nearly straight momentum sheet under a Gaussian metric with unperturbed motion normal to the sheet ($\phi = 0$). The initial perturbation is confined to the stable mode $\omega = 1$, which oscillates while triggering a nonlinear response in the most unstable mode $\omega = 2$, which entrains a forced response in the less unstable mode $\omega = 3$.*

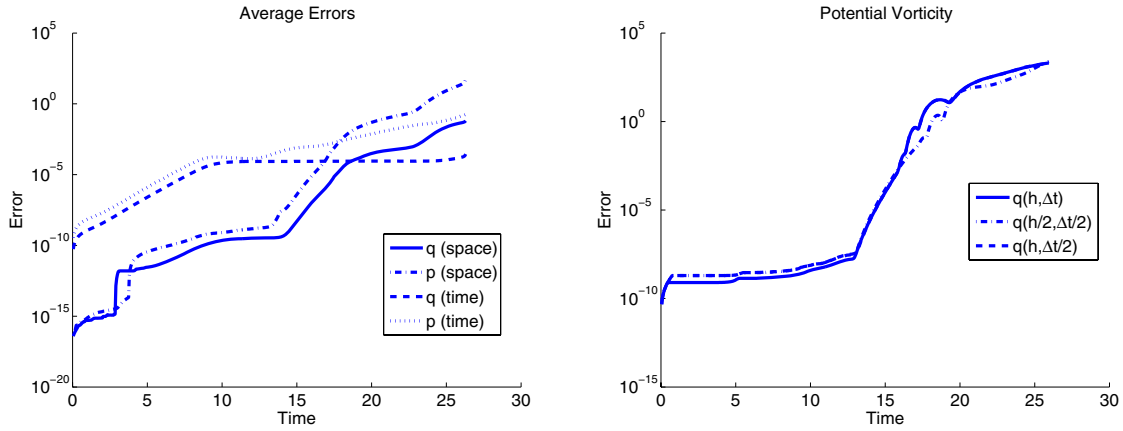


Figure 5. *The spatial and temporal components of the mean global error in q and p (left) and the error in a pseudospectral estimate of the potential vorticity (right) for the simulation of Figure 4.*

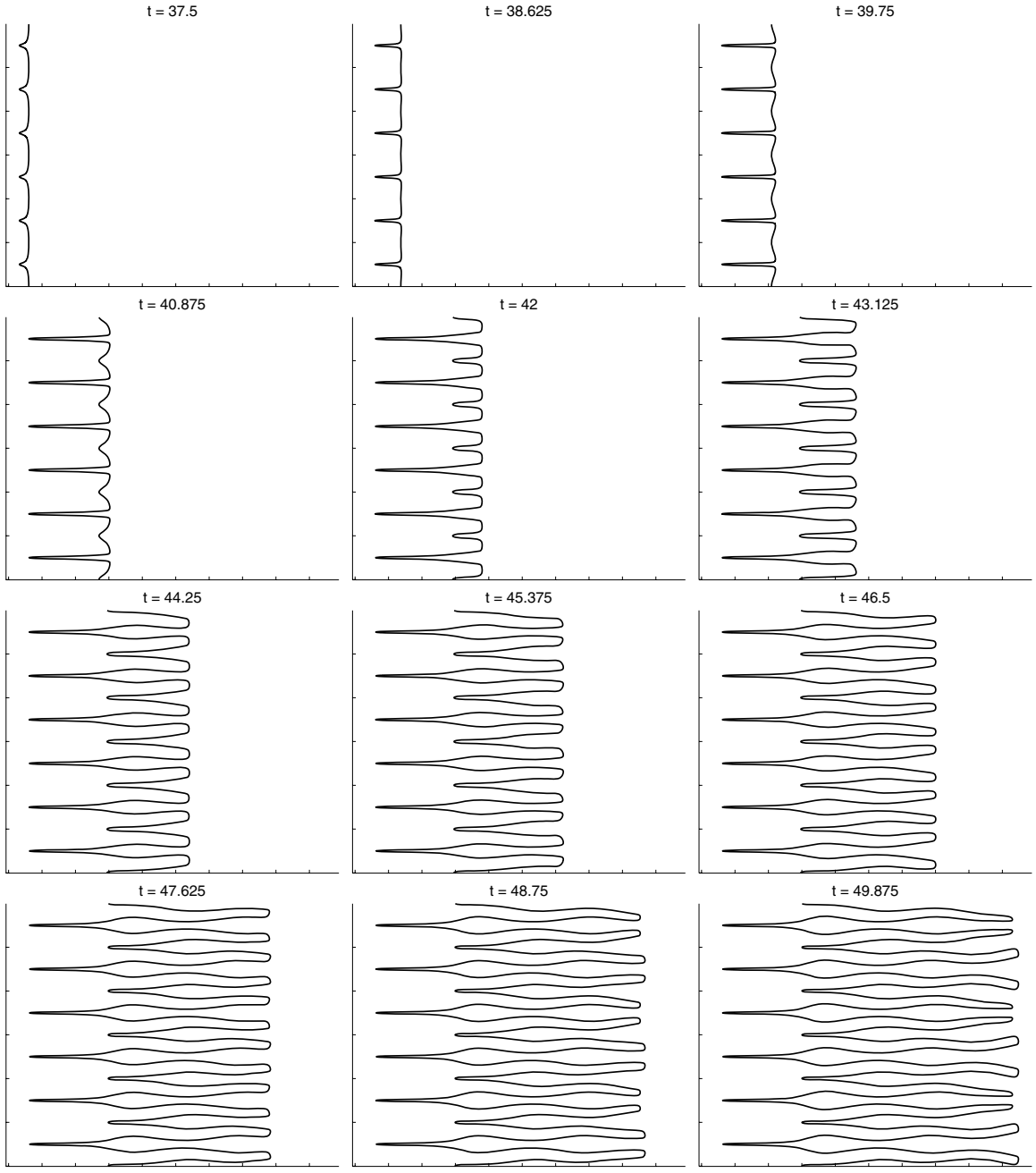


Figure 6. *The motion of the sheet for the simulation of Figure 4, i.e., Gaussian metric and $\phi = 0$. The solutions are 1-periodic in the y-direction and three symmetric images are shown. See text for further details. The x and y axes have the same scaling throughout.*

for the Gaussian kernel in Figure 5. In all cases $h = 1/512$ and $\Delta t = 0.05$, so that for the most accurate solution $x^{(2)}$ there are 1024 grid points and the time step is 0.025. (In fact, the dominant errors can be removed by Richardson extrapolation applied to the final results; this

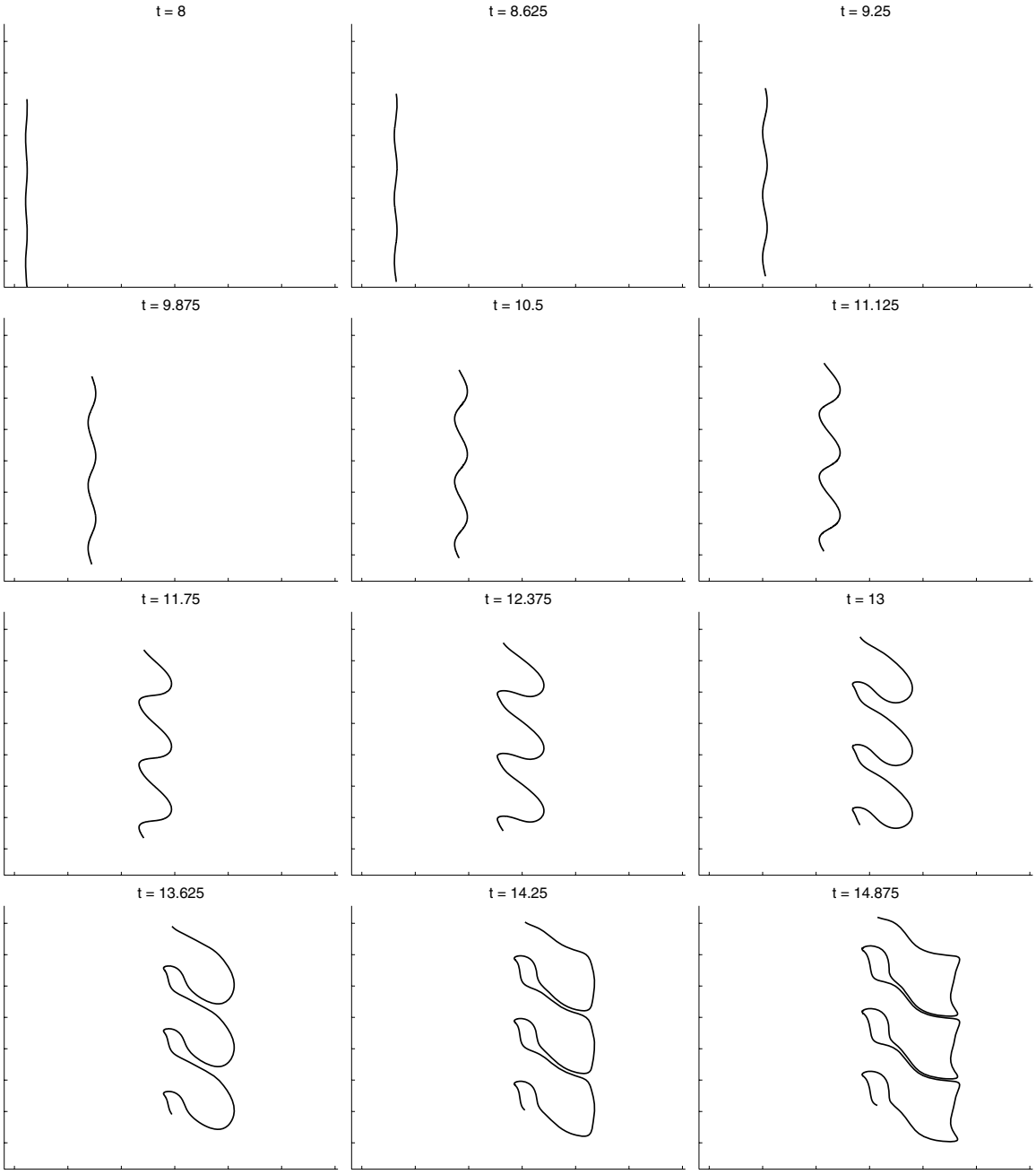


Figure 7. *The motion of the sheet for the Gaussian metric and* $\phi = \pi/8$.

does not change our observations of the results.)

For the major part of all the simulations the errors are reasonably small. However, when the sheets become stretched or develop sharp corners, the errors are large and the results should be interpreted with caution. In addition, it is known [10] that the point particle

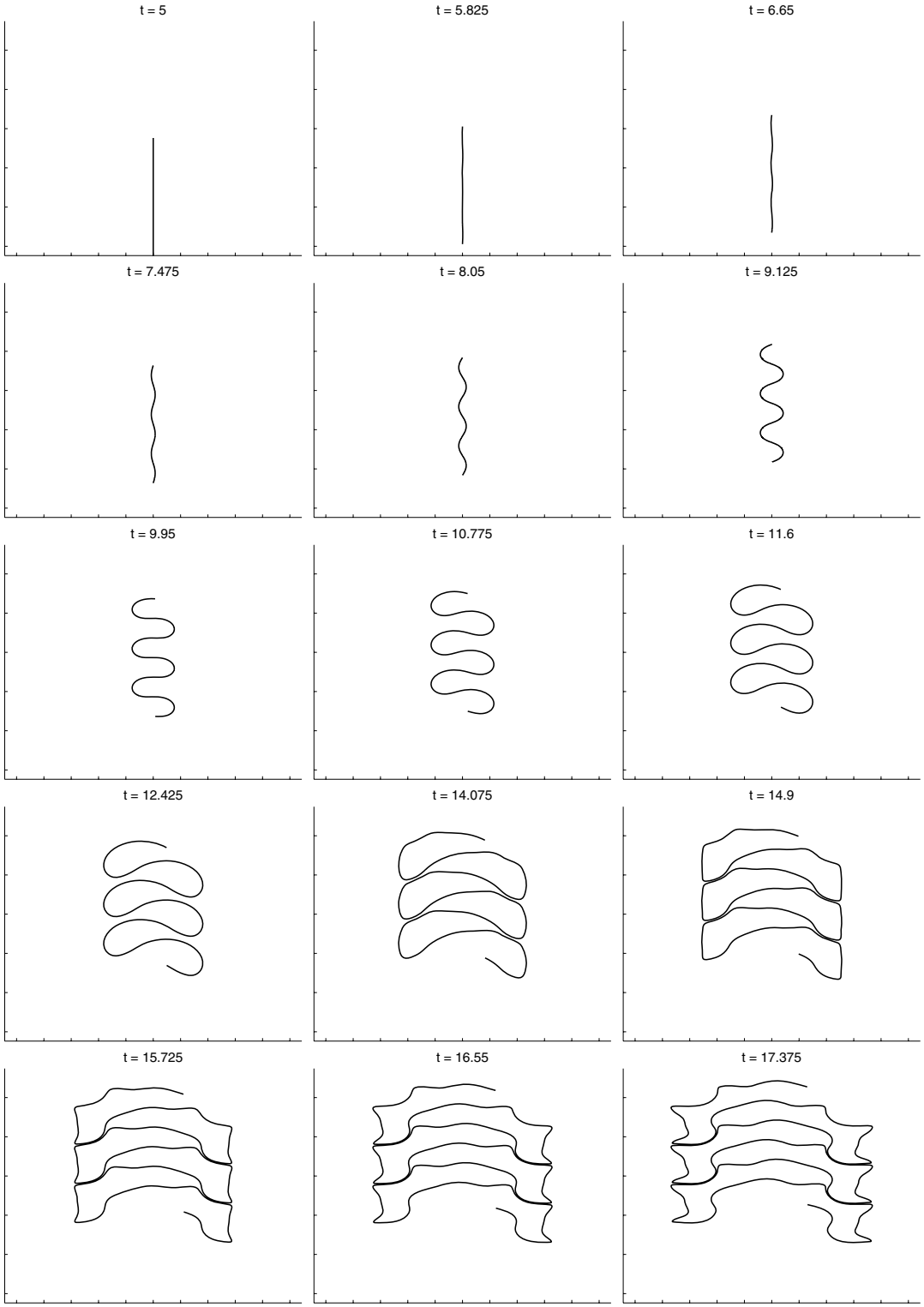


Figure 8. *The motion of the sheet for the Gaussian metric and* $\phi = \pi/2$.

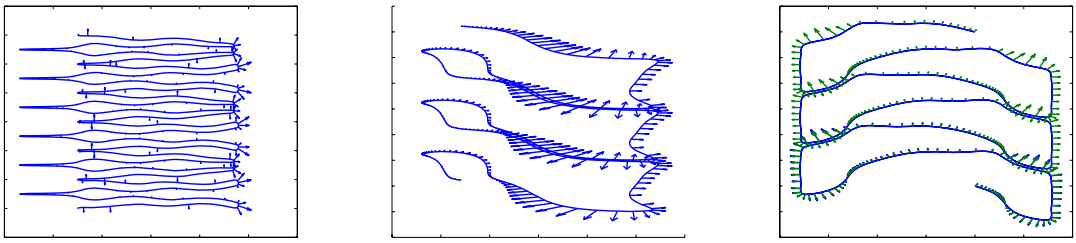


Figure 9. *Close ups of the motion for the Gaussian metric, showing the momentum (at every* 10*th point).* Left: $\phi = 0$*; see Figure* 6. Center: $\phi = \pi/8$*; see Figure* 7. Right: $\phi = \pi/2$*; see Figure* 8.

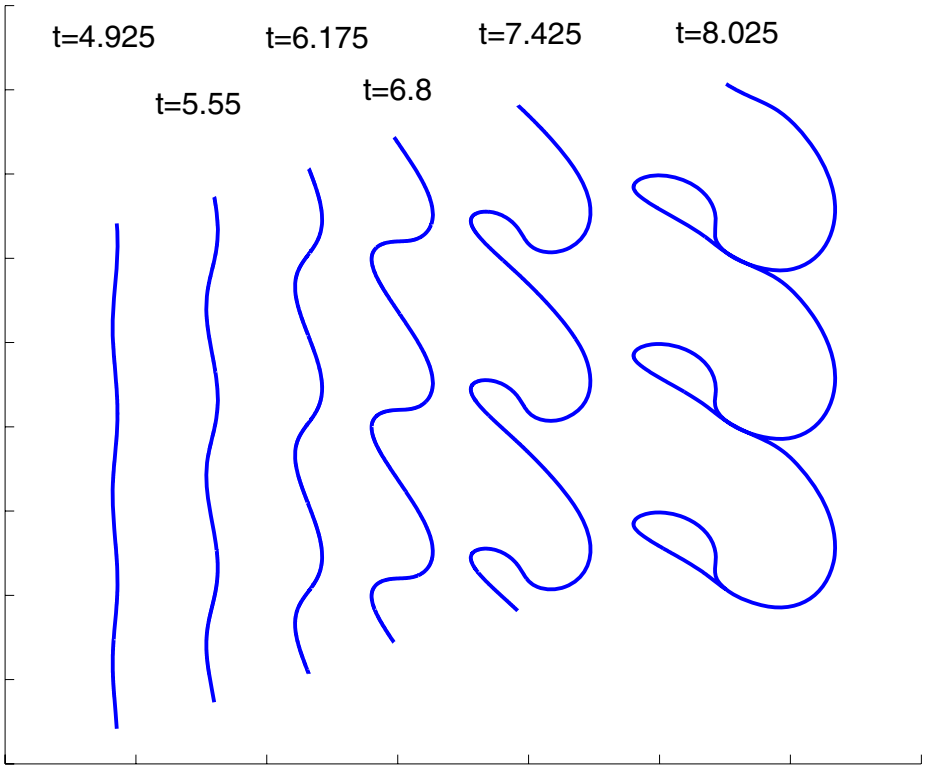


Figure 10. *The motion of the sheet for the* H^1 *metric and* $\phi = \pi/8$.

approximation for sheets becomes inaccurate when two sections of the sheet approach closer than a few times the particle spacing.

For the Gaussian kernel, the numerical results indicate that the sheet remains analytic for all time and does not develop spontaneous singularities or self-intersections. For the H^1 kernel,

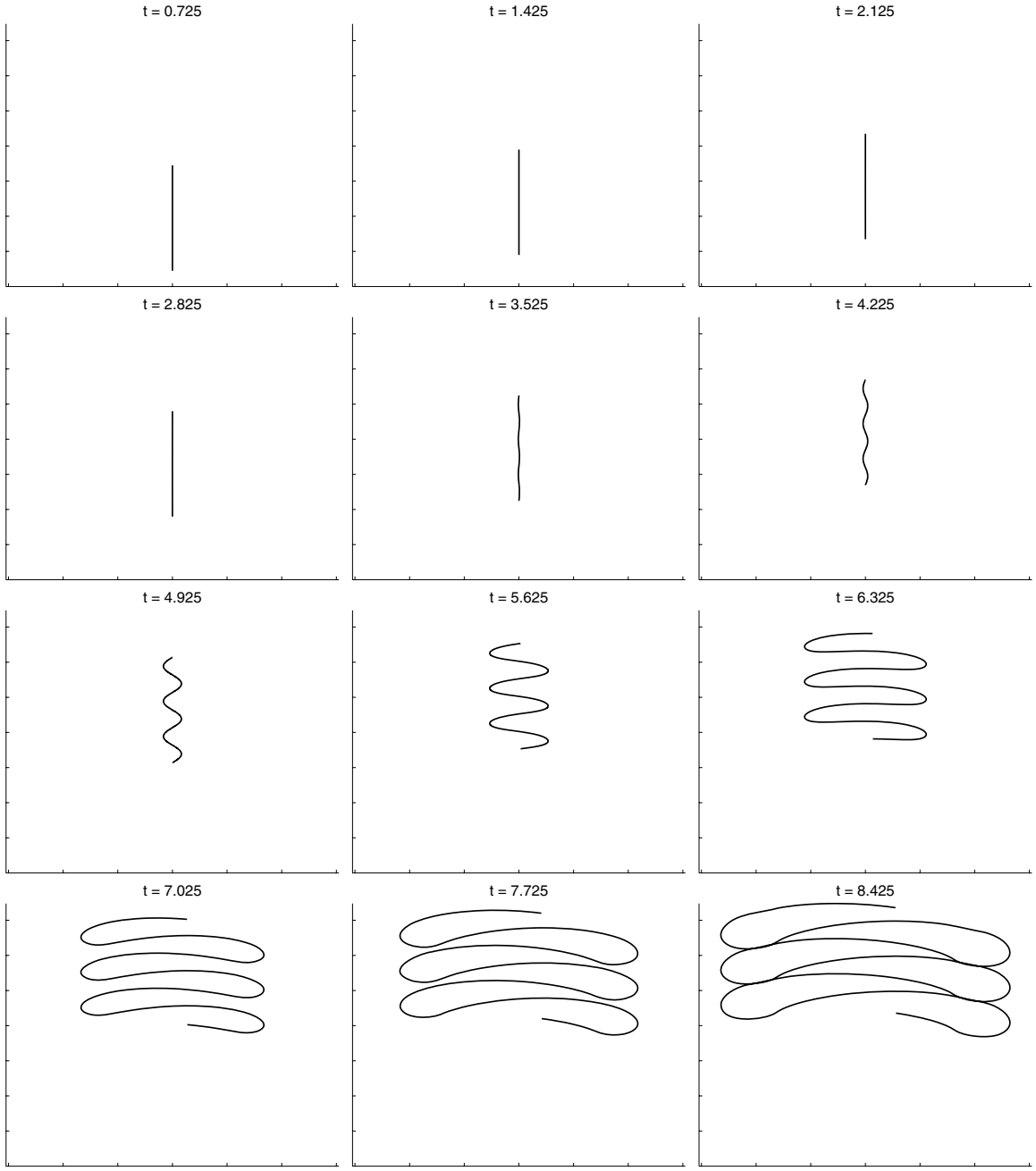


Figure 11. *The motion of the sheet for the H^1 metric and $\phi = \pi/2$.*

the motion can terminate in a finite-time self-intersection. Indeed, this is already known to happen when two 1D Camassa–Holm solitons approach one another, which corresponds here to the collision of two approaching straight sheets. The numerical results indicate, however, that the sheet remains analytic prior to a self-intersection and does not develop spontaneous

self-intersections. In all cases the initial perturbation is confined to wavenumber 1 and has amplitude 5×10^{-6} in q and 2.5×10^{-5} in p.

Three cases are shown for the Gaussian kernel $G(r) = \mathrm{e}^{-(r/\alpha)^2}$ with $\alpha = 0.2$. (Note the different space- and time-scales from those considered in section 3.2. Note also that three symmetric repetitions of the sheet are shown.) For normal motion ($\phi = 0$), from section 3.2, we have that wavenumbers with $\alpha\omega > 0.3568$ are unstable, with the instability strongly peaked at $\alpha\omega = 0.4671$. In the simulations with periodic boundary conditions, only integer wavenumbers are permitted. Therefore, $\omega = 1$ is stable, and $\omega \geq 2$ is unstable, with $\omega = 2$ the most unstable. These results are confirmed in Figure 4, which shows wavenumber $\omega = 1$ undergoing three periodic oscillations before the nonlinear interactions it generates in $\omega = 2$ grow to dominate the solution. $\omega = 3$ is slaved to the most rapidly growing mode $\omega = 2$. The nonlinear evolution of the instability (Figure 6) shows the development of two fingers on length-scale α, the tail of which stays put. The body of the finger consists of two sheets with opposite momentum, which approach each other like the capture of 1D particles [17] and essentially cancel one another out. This is the simplest instance of what we call the "zippering" phenomenon of sheets. The remaining straight part of the sheet then undergoes a second instability, also with $\omega = 2$ but displaced by a quarter period. As these secondary fingers grow they cause the primary fingers to unzip. In the final state shown the motion consists of approximately 1D motion of the horizontal fingers combined with continued evolution of the tips.

For small nonzero ϕ the picture is qualitatively the same, but as ϕ increases, new features of the motion emerge. The motion for $\phi = \pi/8$ is shown in Figure 7. Wavenumber 1 is now unstable, and, in addition, the momentum in the tail of the finger swings around to form an outwardly expanding bubble. The zippered sections of the sheet are now curved and undergo dynamics of their own. Eventually the bubbles and zippers will meet their symmetric images and undergo further zipperings. The secondary instability also occurs but its evolution appears to be constrained by the primary zippers.

As ϕ increases further the bubbles grow larger, until the extreme case (Figure 8) $\phi = \pi/2$ is reached. Now the bubbles are so large that there is no recognizable finger and no zippering at the bubble's tail. Instead, each left- and right-moving bubble expands to meet its periodic image, sections of these pairs zipper up, and the heads of each bubble in a beautiful maneuver reform to form left- and right-moving straight sheets. These, of course, are unstable and begin to finger, causing the main zippers to begin to unzip, while the partly zippered internal parts of the sheet continue to evolve in their own right.

The initial stages of the evolution of the nonlinear stability for the H^1 metric are roughly similar. (See Figure 10 for $\phi = \pi/8$ and Figure 10 for $\phi = \pi/2$.) We know, however, from the 1D case, that sheets can collide in finite time, with the classical solution ending in a singularity. All the simulations do appear to end in this way, although we have not performed a detailed study of the collision process. Intriguingly, however, when two sections of the sheet appear to be on a collision path, they first match their shapes and try to zipper up. Only after a long zippered section has formed do the sheets collide at some point on the zipper. In the case $\phi = \pi/2$ the zippered section even appears to reseparate along its length before colliding at the ends of the zipper. Although these late-stage phenomena are observed in both the high- and low-accuracy simulations, they should still be interpreted with caution as the global errors here are large.

4. Stability of expanding circular sheets. A single circular sheet is described by its radius $r(t)$, radial momentum $p_r(t)$, and rotational momentum $p_\phi(t)$. If the momentum points toward the center of the circle, the circle collapses to a point, ending in a singularity; otherwise the momentum vector is asymptotically outwardly normal to the sheet. Here we study only this limiting case; i.e., we take $p_\phi = 0$.

The linearized equations of motion form a complicated system of integro-differential equations which, as far as we can tell, cannot be solved analytically. However, they can be solved asymptotically for large r (and hence for large t). The following proposition shows that expanding circular sheets are algebraically, but not exponentially, unstable. Moreover, the instability does not depend in a significant way on the initial perturbation, which enters the solution only parametrically. All frequencies grow at the same rate. The growth relative to the size of the expanding sheet is only $\mathcal{O}(t^{1/3})$.

Proposition 4.1. *Consider a circular momentum sheet of radius $r(t)$, radial momentum $p_r(t)$, and rotational momentum $p_\phi(t) = 0$ subject to a small perturbation, so that its position and momentum are described by*

$$q(r,t) = r(t)\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} + \varepsilon u(\theta,t), \quad p(r,t) = p_r(t)\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} + \varepsilon v(\theta,t), \tag{4.1}$$

where θ is the Lagrangian parameter of the sheet. If $p_r(0) > 0$ so that the sheet is expanding, then as $t \to \infty$ we have

$$r(t) = \mathcal{O}(t^{2/3}), \quad p_r(t) = \mathcal{O}(t^{1/3}), \quad u(t) = \mathcal{O}(t), \quad v(t) = \mathcal{O}(t^{2/3})$$

for all kernels $G(r)$.

Proof. For fixed $\theta \neq \alpha$, the distance $d = \|q(\theta) - q(\alpha)\|_2$ between two points on the sheet is given by

$$\begin{aligned} d = {} & 2r\sin\frac{1}{2}|\alpha-\theta| \\ & + \varepsilon\left(2\sin\frac{1}{2}|\alpha-\theta|\right)^{-1}[(v_1(\theta)-v_1(\alpha))(\cos\theta-\cos\alpha) + (v_2(\theta)-v_2(\alpha))(\sin\theta-\sin\alpha)] \\ & + \mathcal{O}(\varepsilon^2). \end{aligned} \tag{4.2}$$

The unperturbed motion is described asymptotically for large r by a Hamiltonian system with $H = (G_0 p_r^2)/(2r) + \mathcal{O}(p_r^2/r^3)$, where $G_0 = \int_{-\infty}^{\infty} G(r)\,dr$; dropping the higher-order terms, the equations of motion are

$$\begin{aligned} \dot r &= G_0\frac{p_r}{r}, \\ \dot p_r &= -\frac{1}{2}G_0\frac{p^2}{r^2} \end{aligned} \tag{4.3}$$

with solution

$$\begin{aligned} r(t) &= \left(\frac{3}{2}\sqrt{2G_0E}\,t + r_0^{3/2}\right)^{2/3} = \mathcal{O}(t^{2/3}), \\ p_r(t) &= \sqrt{2Er(t)/G_0} = \mathcal{O}(t^{1/3}), \end{aligned} \tag{4.4}$$

where E is the value of the energy. Note that the momentum per unit length of the sheet is $\mathcal{O}(t^{-1/3})$ and tends to zero as $t \to \infty$.

Linearizing the equations of motion, making the substitution $\tilde{r} = 2r \sin \frac{1}{2}|\alpha - \theta|$, and expanding the integrands for large r give the linearized system

$$\begin{aligned} u_t &= \frac{G_0}{r} v + \frac{G_0 p_r}{r^2} A^T(\theta) u_\theta, \\ v_t &= \frac{G_0 p_r}{r^2} (A(\theta) v)_\theta \end{aligned} \tag{4.5}$$

with coefficient matrix

$$A(\theta) = \begin{pmatrix} \cos\theta \sin\theta & \sin^2\theta \\ -\cos^2\theta & -\cos\theta \sin\theta \end{pmatrix}. \tag{4.6}$$

Equation (4.5) forms a canonical Hamiltonian system with Hamiltonian

$$H = \int_0^{2\pi} \left(\frac{G_0 p_r}{r^2} u_\theta^T A v + \frac{G_0}{2r} v^T v \right) \mathrm{d}\theta, \tag{4.7}$$

which can also be derived by directly expanding the full Hamiltonian. In general, of course, such nonautonomous, nonconstant-coefficient PDEs cannot be solved explicitly. In this case, however, the matrix A has constant Jordan form,

$$A = XJX^{-1}, \quad X = \begin{pmatrix} -\tan\theta & -\sec^2\theta \\ 1 & 0 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{4.8}$$

and, even though $A(\theta)$ and $A(\alpha)$ do not commute for $\theta \neq \alpha$, the equations are nevertheless dramatically simplified by the canonical change of variables

$$v = X\bar{v}, \quad u = X^{-T}\bar{u} \tag{4.9}$$

in terms of which the new Hamiltonian is

$$\begin{aligned} \bar{H} &= \int_0^{2\pi} \bar{u}^T (X^{-1}AX)\bar{v} + \bar{u}^T (X^{-1})_\theta AX\bar{v} + \frac{1}{2}\bar{v}^T X^T X \bar{v} \, \mathrm{d}\theta \\ &= \int_0^{2\pi} \frac{G_0 p_r}{r^2} \left(\bar{u}_\theta^T J v + \bar{u}^T K \bar{v} \right) + \frac{G_0}{2r} \bar{v}^T L v \, \mathrm{d}\theta, \end{aligned} \tag{4.10}$$

where

$$K = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, \quad L = \sec^2\theta \begin{pmatrix} 1 & \tan\theta \\ \tan\theta & \sec^2\theta \end{pmatrix}, \tag{4.11}$$

with equations of motion

$$\begin{aligned} \dot{\bar{u}}_1 &= \frac{G_0}{r} \sec^2\theta (\bar{v}_1 + \bar{v}_2 \tan\theta), \\ \dot{\bar{u}}_2 &= \frac{G_0 p_r}{r^2} (\bar{u}_{1\theta} - \bar{u}_2) + \frac{G_0}{r} \sec^2\theta (\bar{v}_1 \tan\theta + \bar{v}_2 \sec^2\theta), \\ \dot{\bar{v}}_1 &= \frac{G_0 p_r}{r^2} \bar{v}_{2\theta}, \\ \dot{\bar{v}}_2 &= \frac{G_0 p_r}{r^2} \bar{v}_2. \end{aligned} \tag{4.12}$$

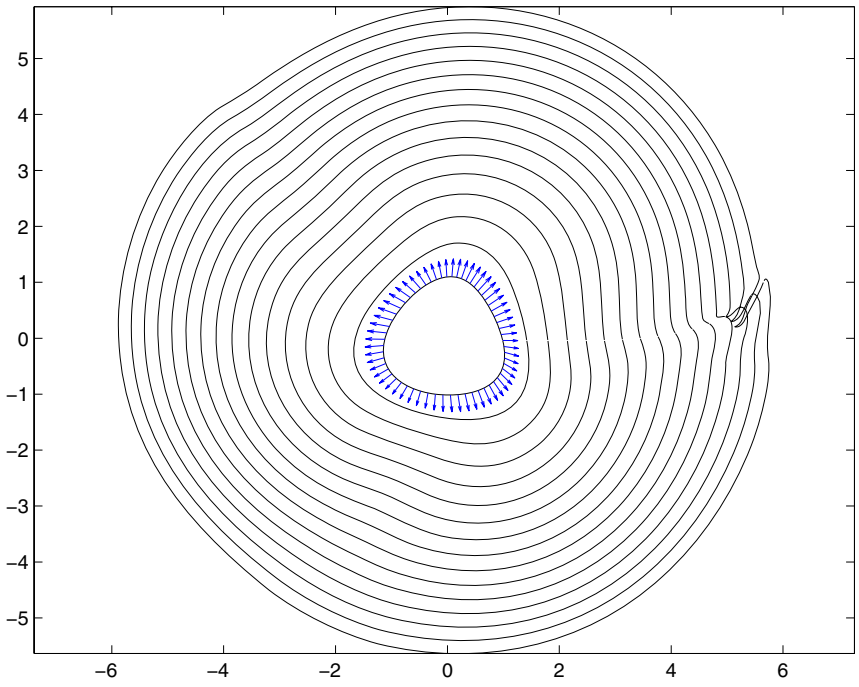


Figure 12. *The nonlinear development of the algebraic ($\mathcal{O}(t^{1/3})$) instability of circular momentum sheets. Here $r(0) = p_r(0) = 1$, $G(r) = \exp(-(r/0.2)^2)$, and a small perturbation is applied to q and p in the 2nd Fourier mode.*

Notice that the change of variables has reduced the PDEs to (effectively) ODEs. Substituting the asymptotic behavior of $r(t)$ and $p_r(t)$ gives

$$\begin{aligned}
\dot{\bar{u}}_1 &= ct^{-2/3}\sec^2\theta(\bar{v}_1 + \bar{v}_2\tan\theta),\\
\dot{\bar{u}}_2 &= \frac{2}{3t}(\bar{u}_{\theta 1} - \bar{u}_2) + ct^{-2/3}\sec^2\theta(\bar{v}_1\tan\theta + \bar{v}_2\sec^2\theta),\\
\dot{\bar{v}}_1 &= \frac{2}{3t}\bar{v}_{2\theta},\\
\dot{\bar{v}}_2 &= \frac{2}{3t}\bar{v}_2.
\end{aligned} \tag{4.13}$$

The solution for the momentum perturbation (as $t \to \infty$) is

$$\bar{v}_2 = t^{2/3}b(\theta), \quad \bar{v}_1 = t^{2/3}b'(\theta), \tag{4.14}$$

where $b(\theta)$ is determined by the initial conditions. Substituting (4.14) into (4.13) for $\bar{u}$ shows that only the $\mathcal{O}(t^{-2/3})$ terms contribute to the leading order behavior of $\bar{u}$. This leading order behavior is more easily determined from the original formulation

$$u_t = ct^{-2/3}v = \alpha d(\theta) \quad (\text{where } d = X(b', b)^T), \tag{4.15}$$

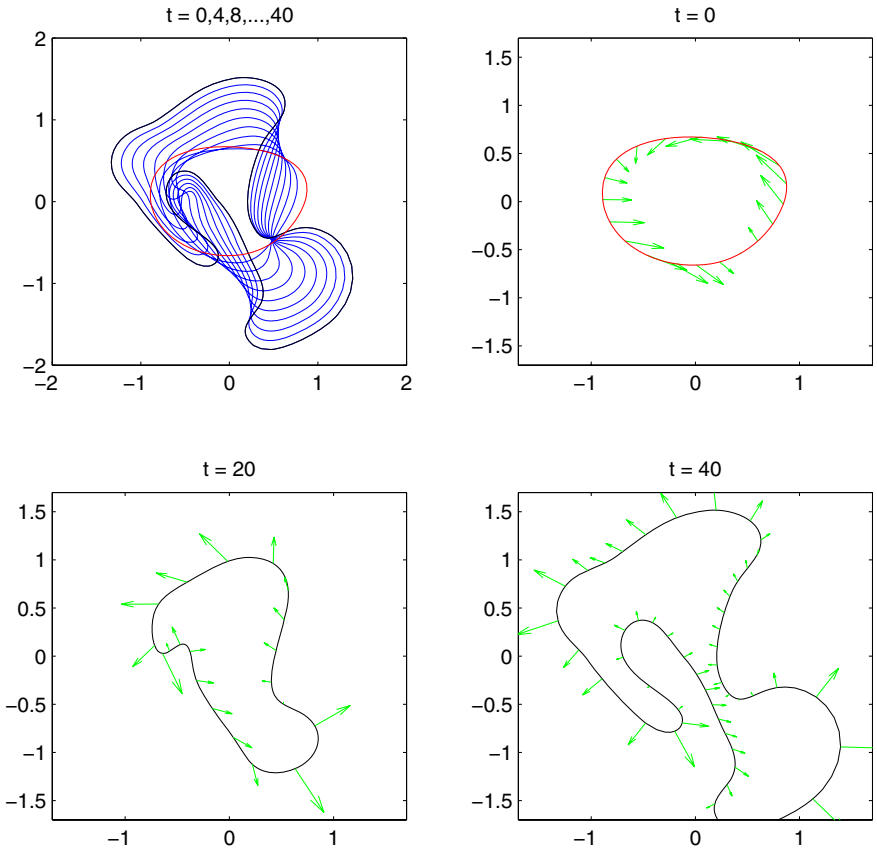


Figure 13. *Motion of a momentum sheet with noncircular initial conditions, Gaussian metric.*

where $\alpha = [(4p_{r0}^2)/(9G_0 r_0)]^{-1/3}$, so

$$u = \gamma t d(\theta). \tag{4.16}$$ ■

Note that the asymptotic behavior of the unperturbed sheet and of the leading order behavior of perturbation depends on the metric only through G_0. Higher-order terms depend on higher moments of G, and these can be found through a regular perturbation expansion proceeding in powers of r^{-2}. Also, note that the $\mathcal{O}(t^{2/3})$ growth of the sheet is critical in determining its stability; any slower, and the growth of perturbations would be bounded; any faster, and they would grow exponentially, not algebraically.

The development of the instability is shown in Figure 12 under the Gaussian metric $G(r) = \exp(-(r/\alpha)^2)$ with $\alpha = 0.2$. Here $r(0) = p_r(0) = 1$ and the initial perturbation is confined to the second Fourier mode. The nonlinear development of the instability shows itself on the length-scale α, followed by rapid fingering.

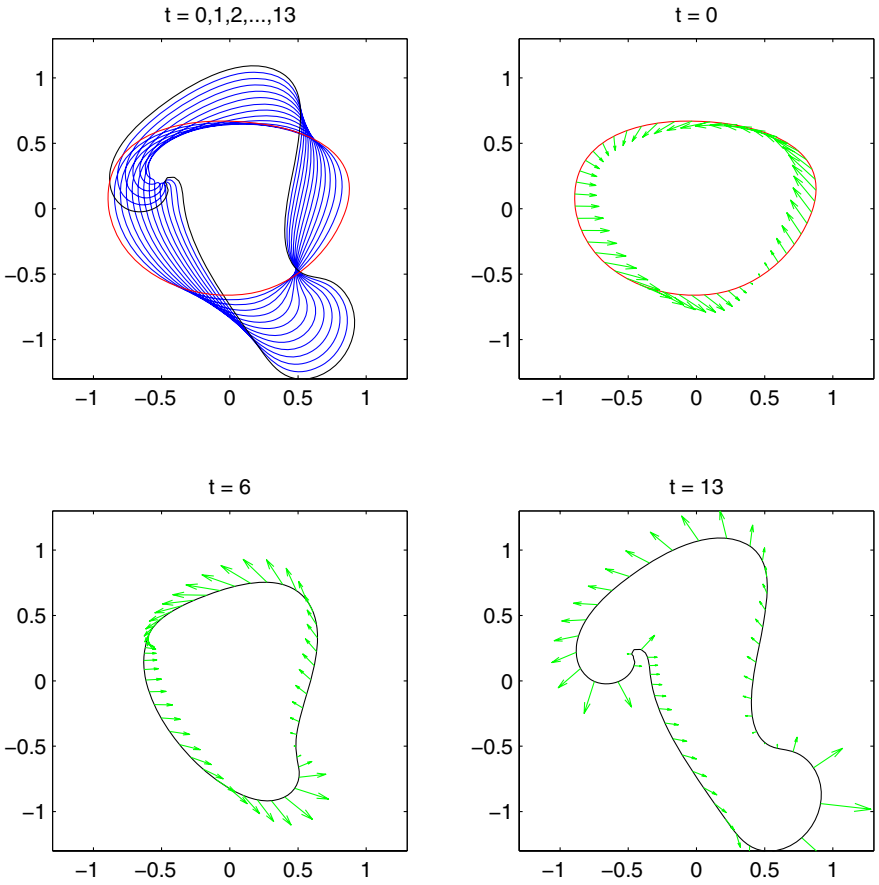


Figure 14. *Motion of a momentum sheet with noncircular initial conditions, H^1 metric. The initial conditions are the same as in Figure 13.*

Some dynamics of momentum sheets for initial conditions far from circular are shown in Figure 13 for the Gaussian kernel $G(r) = \exp(-(r/0.2)^2)$. The momentum in the northwest corner of the sheet, initially pointing inward, swings around to point outward, forming an outwardly expanding bubble, a fairly stable object much like the circular sheets examined above. Another bubble is seen in the southeast corner. However, two other sections of the sheet do not succeed in turning around, and at $t = 40$ are approaching each other (later they will zipper up as in Figure 8). The transition from outward to inward momentum at the southwest corner seems to be unstable and leads to the development of a growing finger.

The initial conditions in Figure 14 are the same as in Figure 13, but we now switch to the H^1 kernel $G(r) = K_0(r/0.2)$. Apart from the time-scale, the motion is broadly similar

until some fine structure emerges on the length-scale of the kernel. Shortly after the final time shown, the motion ends in a self-collision of the sheet.

Acknowledgment. We would like to thank the referees for their comments.

REFERENCES

[1] V. I. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier (Grenoble), 16 (1966), pp. 319–361.

[2] V. I. Arnold and B. A. Khesin, *Topological Methods in Hydrodynamics*, Appl. Math. Sci. 125, Springer-Verlag, New York, 1998.

[3] R. E. Caflisch, T. Y. Hou, and J. Lowengrub, *Almost optimal convergence of the point vortex method for vortex sheets using numerical filtering*, Math. Comp., 68 (1999), pp. 1465–1496.

[4] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer-Verlag, Berlin, 2002.

[5] D. D. Holm and J. E. Marsden, *Momentum maps and measure-valued solutions (peakons, filaments and sheets) for the EPDiff equation*, in The Breadth of Symplectic and Poisson Geometry, Progr. Math. 232, Birkhäuser Boston, Boston, MA, 2005, pp. 203–235.

[6] D. D. Holm, V. Putkaradze, and S. N. Stechman, *Rotating concentric circular peakons*, Nonlinearity, 17 (2004), pp. 1–24.

[7] D. D. Holm, J. T. Ratnanather, A. Trouvé, and L. Younes, *Soliton dynamics in computational anatomy*, NeuroImage, 23 (2004), pp. S170–S178.

[8] D. D. Holm and M. F. Staley, *Wave structure and nonlinear balances in a family of evolutionary PDEs*, SIAM J. Appl. Dyn. Syst., 2 (2003), pp. 323–380.

[9] D. D. Holm and M. F. Staley, *Interaction Dynamics of Singular Wave Fronts*, manuscript.

[10] T. Y. Hou, J. S. Lowengrub, and M. J. Shelley, *The long-time motion of vortex sheets with surface tension*, Phys. Fluids, 9 (1997), pp. 1933–1954.

[11] D. D. Holm, M. Nitsche, and V. Putkaradze, *Euler-alpha and vortex blob regularization of vortex filament and vortex sheet motion*, J. Fluid Mech., 555 (2006), pp. 149–176.

[12] B. Khesin, *Topological fluid dynamics*, Notices Amer. Math. Soc., 52 (2005), pp. 9–19.

[13] R. Krasny, *A study of singularity formation in a vortex sheet by the point-vortex approximation*, J. Fluid Mech., 167 (1986), pp. 65–93.

[14] A. J. Majda and A. L. Bertozzi, *Vorticity and Incompressible Flow*, Cambridge University Press, Cambridge, UK, 2002.

[15] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, 2nd ed., Springer-Verlag, Berlin, 1999.

[16] S. Marsland and C. Twining, *Constructing diffeomorphic representations for the group-wise analysis of non-rigid registrations of medical images*, IEEE Trans. Medical Imaging, 23 (2004), pp. 1006–1020.

[17] R. I. McLachlan and S. R. Marsland, *N-particle dynamics of the Euler equations for planar diffeomorphisms*, Dynam. Systems, to appear.

[18] R. I. McLachlan and G. R. W. Quispel, *Geometric integrators for ODEs*, J. Phys. A, 39 (2006), pp. 5251–5286.

[19] M. I. Miller, A. Trouvé, and L. Younes, *On metrics and the Euler-Lagrange equations of computational anatomy*, Ann. Rev. Biomed. Engrg., 4 (2002), pp. 375–405.

[20] D. W. Moore, *The spontaneous appearance of a singularity in the shape of an evolving vortex sheet*, Proc. Roy. Soc. London Ser. A, 365 (1979), pp. 105–119.

[21] D. W. Moore and R. Griffiths-Jones, *Stability of an expanding circular vortex sheet*, Mathematika, 21 (1974), pp. 128–133.

[22] D. Mumford, *Pattern theory: The mathematics of perception*, in Proceedings of the International Congress of Mathematicians, Vol. III, World Scientific, Singapore, 2002, pp. 401–422.

[23] M. Nitsche, *Axisymmetric vortex sheet motion: Accurate evaluation of the principal value integral*, SIAM J. Sci. Comput., 21 (1999), pp. 1066–1084.

[24] M. Oliver and S. Shkoller, *The vortex blob method as a second-grade non-Newtonian fluid*, Comm. Partial Differential Equations, 26 (2001), pp. 295–314.
[25] M. A. Pinsky, *Introduction to Fourier Analysis and Wavelets*, Brooks/Cole, Pacific Grove, CA, 2002.
[26] A. Sidi and M. Israeli, *Quadrature methods for periodic singular and weakly singular Fredholm integral equations*, J. Sci. Comput., 3 (1988), pp. 201–231.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 759–782

Transition from Rotating Waves to Modulated Rotating Waves on the Sphere*

Adela N. Comanici†

Abstract. In this article, we consider parameter-dependent systems of reaction-diffusion equations on the sphere, which are equivariant under the group $SO(3)$ of all rigid rotations on the sphere. It is known that the transition from rotating waves to modulated rotating waves on the sphere can be explained via a supercritical Hopf bifurcation from a rotating wave, $SO(3)$-symmetry, and finite-dimensional equivariant center manifold reduction. Using Floquet theory, it is easy to get the decomposition of these modulated rotating waves into the primary frequency vector part and the periodic part. Going further, we use the Baker–Campbell–Hausdorff (BCH) formula in the Lie algebra $so(3)$ to get the closed form of the reduced differential equations on $so(3)$, and then closed formulas for the primary frequency vectors and for the periodic parts associated to the bifurcating modulated rotating waves. As a consequence, we get the explicit characterization of these modulated rotating waves and all possible tip motions on the sphere: quasi-periodically meandering and slowly drifting. This approach does not treat separately the resonant and nonresonant Hopf bifurcation of a rotating wave on the sphere; the change that appears in the resonant case in the Taylor expansions for the primary frequency vectors is implicitly captured by using the BCH formula in $so(3)$, and it is independent of the normal forms theory developed in [B. Fiedler and D. Turaev, *Arch. Ration. Mech. Anal.*, 145 (1998), pp. 129–159]. When systems with two parameters are involved and the norm of the frequency vector of the initial rotating wave undergoing Hopf bifurcation is a multiple integer of the critical eigenvalue leading to Hopf bifurcation, we give a shorter and more intuitive proof for the following fact stated in [C. Wulff, *Doc. Math.*, 5 (2000), pp. 227–274] for a general Lie group: the primary frequency vectors of a branch of these modulated rotating waves are generically orthogonal to the frequency vector of the initial rotating wave undergoing Hopf bifurcation, and their tip motions are slowly drifting along the equator of the sphere. Due to the computational nature of the BCH formula in $so(3)$, this approach can be translated in a computer-implemented method which will allow us to better control the tip motions of the modulated rotating waves on the sphere.

Key words. equivariant center manifold, rotating wave, modulated rotating wave, Hopf bifurcation, Baker–Campbell–Hausdorff formula in $so(3)$

AMS subject classifications. 34, 37, 35

DOI. 10.1137/050631872

1. Introduction.

The main motivation of this article is the presence of spiral waves in excitable media, especially in cardiac tissue. Spiral waves arise as stable spatio-temporal patterns in various chemical, physical, and biological systems, as well as numerical simulations of reaction-diffusion systems on excitable media with various geometries [19]. Excitable media are extended nonequilibrium systems having a uniform rest state that is linearly stable but susceptible to finite perturbations. Spiral waves have been observed experimentally, for instance, in catalysis of platinum surfaces [23], Belousov–Zhabotinsky chemical reactions [17], Rayleigh–Bernard convection [25], slime-mold cells [13], and most importantly, cardiac tis-

*Received by the editors May 18, 2005; accepted for publication (in revised form) by B. Sandstede September 16, 2006; published electronically December 5, 2006.
http://www.siam.org/journals/siads/5-4/63187.html
†Department of Mathematics, Rice University, Houston, TX 77005-1892 (anc1@rice.edu).

sue [8]. Numerical simulations of the spiral waves have been done in [2, 3] for the planar case, and in [1, 15, 22, 32, 33, 34] for the spherical case.

It is now believed that spiral and scroll waves that appear in the heart muscle can lead to cardiac arrhythmias (abnormal rhythms in the heart), giving rise to atrial fluttering or ventricular fibrillation. In normal hearts cardiac arrhythmias are rare, but in diseased hearts cardiac arrhythmias can become more common. For example, if chambers of the heart become abnormally large, they are susceptible to serious arrhythmias in which waves are believed to circulate in a fashion that is similar to the circulation of the Belousov–Zhabotinsky waves in a chemical medium. Real human hearts are enormously complex three-dimensional structures. In this article, we assume that the geometry of the excitable media is a sphere, which in the case of cardiac tissue is clearly an approximation.

In the planar case, a rigidly rotating spiral wave is an example of wave pattern rotating around a center and being well approximated by an Archimedean spiral wave far from the rotation center. Near the rotation center, there is a core region of the spiral wave, where the front of the wave has a tip, whose structure is considered to be the most important in understanding the behavior of the whole spiral wave. Barkley [2] was the first who performed a numerical linear stability analysis for the basic-time periodic spiral wave solution in a reaction-diffusion system on the unbounded plane and showed evidence of a Hopf bifurcation. Later, using an ad hoc model, Barkley [3] was the first to realize the key importance of the group $SE(2)$ of all planar translations and rotations in describing the dynamics and bifurcations of planar spiral waves. It is well known now that the tip of the planar spiral wave rotates steadily or meanders or linearly drifts in the plane [14].

From a mathematical point of view, for the planar case, rigidly rotating spiral waves are examples of rotating waves, meandering spiral waves are examples of modulated rotating waves, and linearly drifting spiral waves are examples of modulated travelling waves. The first rigorous mathematical theory of the planar spiral waves was done by Wulff [30]. The main difficulty comes from the fact that the group $SE(2)$ is noncompact and the action of $SE(2)$ on the usual spaces of functions is not smooth. The proofs in [30] are based on the basic assumption that the linearization at the rotating wave in the corotating frame does not exhibit continuous spectrum near the imaginary axis, and the proofs are done using Liapunov–Schmidt reduction on scales of Banach spaces.

Later, in [28], Sandstede, Scheel, and Wulff proved a finite-dimensional equivariant center manifold reduction theorem near a relative equilibrium Gu_0 of an infinite-dimensional vector field on a Banach space X on which acts a finite-dimensional Lie group (not necessarily compact). In [10], Fiedler et al. showed that the reduced differential equations on this center manifold have a skew-product form. All these were done by Field in [12] and Krupa in [18] for finite-dimensional spaces.

Using the results of [28], Hopf bifurcations from one-armed and multiarmed rotating planar spiral waves to meandering waves were studied in [10, 14].

In [26], Renardy considered bifurcations from rotating waves of semilinear equations that are equivariant under a general compact Lie group and applied his results to the Laser equations. His results do not cover the resonance case. The theorems were proved using a generalized implicit function theorem on scales of Banach spaces.

In [31], a G-equivariant semilinear system of parabolic equations (where G is a finite-

dimensional possibly noncompact Lie group) is studied. In particular, the periodic forcing of relative equilibria and resonant periodic forcing of relative equilibria to relative periodic orbits, as well as Hopf bifurcation from relative equilibria to relative periodic orbits, are treated using Liapunov–Schmidt reduction. Resonant drift phenomena are also studied using the same method in Proposition 3.4 of [31]. Then, these results are applied to the planar spiral waves for both Hopf bifurcation and periodic forcing, and to the case $G = SO(3)$ for the periodic forcing in Example 2.8 and Example 2.14 of [31].

However, the transition from rotating waves to modulated rotating waves on the sphere can be explained via a supercritical Hopf bifurcation from a rotating wave, $SO(3)$-symmetry, and finite-dimensional equivariant center manifold reduction developed in [28]. Using [10] and [18], the reduced differential equations on the center manifold have a skew-product form. Therefore, Hopf bifurcation from a rotating wave on a sphere can be described by a Hopf bifurcation of an equilibrium combined with a motion along the group directions. Moreover, on the group directions using Floquet theory, it is easy to get the decomposition of the modulated rotating waves into the primary frequency vector part and the periodic part is proved.

In 2004, Comanici used in her Ph.D. thesis [7] the Baker–Campbell–Hausdorff (BCH) formula in $so(3)$ and the reduction to a differential equation on $so(3)$ to further study the Hopf bifurcation from rotating waves on the sphere, i.e., to get a closed formula for the primary frequency vectors and for the periodic parts associated to the bifurcating modulated rotating waves and a closed formula for the angle between the axis of rotation of the initial rotating wave and the primary frequency vector of these modulated rotating waves. As a consequence, we get the explicit characterization of these bifurcating modulated rotating waves and all possible motions of their tips on the sphere: meandering or slow drifting.

When systems with two parameters are involved and the norm of the frequency vector of the initial rotating wave undergoing Hopf bifurcation is a multiple integer of the critical eigenvalue leading to Hopf bifurcation, we give a shorter and more intuitive proof for the following fact: the primary frequency vectors of a branch of bifurcating modulated rotating waves are generically orthogonal to the frequency vector of the initial rotating wave undergoing Hopf bifurcation. This result is stated in Proposition 3.4 in [31] for a general Lie group and is illustrated in Example 2.8 of [31] for the rotation group $SO(3)$ for periodic forcing. This proof is not only more intuitive, but also if it is combined with the angle formula mentioned before, it shows us how we can control the tip motions of these modulated rotating waves on the sphere if this approach is translated into a computer-implemented method.

Independently in 2005, Chan used the normal form theory developed for general Lie groups in [11], combined with an expansion in the bifurcation parameter and the BCH formula, to treat the same problem in [5] and obtained similar results. He made a mistake in the resonant case and he corrected it after he was aware of my independent work (see page 2 in [5]). Also, Chan obtained leading order approximations for the primary frequency vectors and for the characterization of the modulated rotating waves.

My approach does not treat separately the resonant and nonresonant Hopf bifurcation of the rotating wave. The change that appears in the resonant case in the Taylor expansions of the primary frequency vectors is implicitly captured by the BCH formula in $so(3)$, and it is independent of the normal form theory developed in [11]. Ultimately, we do consider the nonresonant and resonant cases separately just to emphasize the difference between them.

This method can be generalized for any compact Lie groups, as it will be briefly discussed at the end of sections 3 and 4.

The interest in considering spiral waves on nonplanar surfaces is motivated by the applicability to problems in physiology, biology, and chemistry. In the case of the spiral waves on the sphere, the dynamics are expected to be quite different, because any spiral wave starting from a rotating center cannot end at a point. The number of tips of a wave front cannot be odd, and therefore, the dynamics of spiral waves may acquire a new feature qualitatively different from the planar case. The dynamics of spiral waves in an excitable reaction-diffusion system on a sphere was numerically investigated by [1, 15, 33, 34], and [32] who employ a spectral method using spherical harmonics as basis functions. Maselko [20], as well as Maselko and Showalter [21], performed experiments with Belousov–Zhabotinsky chemical waves propagating on the surface of a sphere.

In [33], two different asymptotic regimes are observed. The first regime is a rigid rotation of an excitation wave around the symmetry axis of the domain. The second one is a compound rotation including a drift of the rotation center of the spiral wave along the equator of the sphere. In this case the shape of the wave and its rotation velocity are periodically changing in time.

In [15], it is shown that the tips of the double spiral wave can either perform a meandering motion or rigidly rotate around a fixed center, depending on the system control parameter. It is observed that the rotation of the spiral wave on a spherical surface is similar to the one obtained on the planar surface, except that in the absence of the boundary on a spherical surface some parts of the wave can undergo self-annihilation in contrast to the spiral wave behavior on bounded planar surfaces. In [1], meandering waves on a static sphere are obtained and it is shown that these waves are stable.

In this article we make the assumption that the relative equilibrium has trivial isotropy (no spatial symmetry), because as far as we are aware, there has been no observation of m-armed double spiral waves ($m > 1$) on spherical surfaces. Also, in [6] it was numerically checked that there is a critical size of the sphere below which self-sustained spiral waves cannot exist. Therefore, we consider the sphere of an arbitrary radius r.

In section 2, we recall that the functional-analytical framework can be found in [31] for the general Lie groups and we present only the definition of the relative periodic orbit, the frequency vector of a rotating wave, and the primary frequency vector (or average rotation frequency vector [31]) of a modulated rotating wave. Also, using the equivariant center-manifold reduction developed in [10, 18, 28], we recall the reduced skew-product differential equations for a Hopf bifurcation from a rotating wave on the sphere. It is known that generically we get modulated rotating waves. Decomposition of a modulated rotating wave obtained by Hopf bifurcation into a primary frequency part and the associated periodic part is easily obtained using Floquet theory and briefly described in this section.

Our main results are contained in sections 3, 4, and 5. In section 3, we get the closed form of the reduced differential equations on $so(3)$ in Theorem 3.4 using the BCH formula in the Lie algebra $so(3)$, the properties of the exponential map of $SO(3)$ and its differential, and the properties of the adjoint representations of $SO(3)$ and $so(3)$. Our approach follows the one presented in [14].

In section 4, we get closed formulas for the primary frequency vectors and the associated

periodic parts, as well for the angle between the axis of rotation for the initial rotating wave undergoing Hopf bifurcation and the primary frequency vectors of the bifurcating modulated rotating waves. As a consequence, the explicit characterization of these modulated waves is obtained in Theorem 4.4, and we show that there are two possible types of tip motions for the modulated rotating waves: quasi-periodically meandering of order $O(1)$ about the primary frequency vectors, with the angle between primary frequency vector and the axis of rotation for the initial rotating wave of order $O(\lambda^{\frac{1}{2}})$ for the nonresonant case, or slowly drifting of order $O(\lambda^{\frac{|k|}{2}})$ about the primary frequency vectors for the $k:1$ resonant case. Some pictures illustrating the theoretical results obtained in this section are shown. The details about these pictures can be found in [7].

In section 5, we discuss Hopf bifurcation with two parameters from a rotating wave on the sphere, assuming that the amplitude of the frequency vector of the initial rotating wave is k times the critical eigenvalue leading to Hopf bifurcation. We give a shorter proof for the following fact stated in [31] for a general Lie group: there exists generically a unique branch of modulated rotating waves on the sphere with primary frequency vectors orthogonal to the frequency vector of the initial rotating wave undergoing Hopf bifurcation, and their tip motions are slowly drifting of order $O(\lambda^{\frac{|k|}{2}})$ along the equator of the sphere. This proof makes use of the implicit function theorem and the BCH formula in $so(3)$.

Section 6 presents the proof of Theorem 3.4. An appendix is included. It contains some computations relevant for the results in section 4.

2. Hopf bifurcation from rotating waves on the sphere.

2.1. Rotating waves and modulated rotating waves on the sphere. Throughout this paper we denote by $\mathbf{S^2}$ the unit sphere in $\mathbb{R}^3$ and by $r\mathbf{S^2}$ the sphere of radius $r>0$. We consider the reaction-diffusion system

$$u_t = D\Delta_S u + F(u) \text{ on } r\mathbf{S^2}, \tag{2.1}$$

where $u=(u_1,u_2,\dots,u_N):\mathbb{R}\times r\mathbf{S^2}\to\mathbb{R}^N$ with $N\geq 1$,

$$D=\begin{pmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_N \end{pmatrix}$$

with $d_i\geq 0$ for $i=1,2,\dots,N$ the diffusion coefficients, Δ_S is the Laplace–Beltrami operator on $r\mathbf{S^2}$, and $F=(F_1,F_2,\dots,F_N):\mathbb{R}^N\to\mathbb{R}^N$ is a C^{k+2} function with $0\leq k\leq\infty$. The functional-analytical framework is well developed in [31] for general Lie groups G. Hence, the reaction-diffusion system (2.1) is considered on the function space

$$\mathbf{Y}=\begin{cases}\mathbf{L^2}(r\mathbf{S^2},\mathbb{R}^N) & \text{if } d_i>0 \text{ for } i=1,2,\dots,N;\\ \mathbf{H^2}(r\mathbf{S^2},\mathbb{R}^N) & \text{if there exists } i\in\{1,2\dots,N\} \text{ such that } d_i=0.\end{cases} \tag{2.2}$$

If $\mathbf{Y}=\mathbf{L^2}(r\mathbf{S^2},\mathbb{R}^N)$, then we recall that for any $\alpha\in(\frac{1}{2},1)$, $\mathbf{Y}^\alpha=D((I-D\Delta_S)^\alpha)$ are the fractional spaces of $\mathbf{Y}$ relative to the sectorial operator $-D\Delta_S$. Using [31], the reaction-diffusion system (2.1) defines a sufficiently smooth local semiflow $\mathbf{\Phi}$ on the function space $\mathbf{Y}^\alpha$ if

$\mathbf{Y} = \mathbf{L^2}(r\mathbf{S^2}, \mathbb{R}^N)$ or on $\mathbf{Y}$ if $\mathbf{Y} = \mathbf{H^2}(r\mathbf{S^2}, \mathbb{R}^N)$. Namely, the flow is C^{k+2} if $\mathbf{Y} = \mathbf{L^2}(r\mathbf{S^2}, \mathbb{R}^N)$ and C^k if $\mathbf{Y} = \mathbf{H^2}(r\mathbf{S^2}, \mathbb{R}^N)$. Also, in what follows we work with $\mathbf{Y}^\alpha$, but everything is valid for $\mathbf{Y} = \mathbf{H^2}(r\mathbf{S^2}, \mathbb{R}^N)$.

The reaction-diffusion system (2.1) is $SO(3)$-equivariant, where we recall that the C^1 unitary representation T of $SO(3)$ on $\mathbf{Y}^\alpha$ is defined by

$$T(A)u(x) = u(A^{-1}x), \text{ where } A \in SO(3),\ u \in \mathbf{Y}^\alpha, \text{ and } x \in r\mathbf{S^2},$$

as discussed, for example, in [31].

The notions of *tip position* function of $SO(3)$ on $\mathbf{Y}^\alpha$, *tip motion* of $u_0 \in \mathbf{Y}^\alpha$, *relative equilibrium* for (2.1), *rotating wave* for (2.1), *frequency vector* of a rotating wave, *relative periodic orbit* for (2.1), *modulated rotating wave* for (2.1), and *primary frequency vector* of a modulated rotating wave for (2.1) are well known in the literature and can be found in [31] for the general Lie groups. We present only a few remarks regarding these notions for $G = SO(3)$.

Throughout this paper we assume that the relative equilibrium and the relative periodic orbit have trivial isotropy (no spatial symmetry). We recall the definition of a *relative periodic orbit* for (2.1), because it is formulated slightly differently from the one given in section 1.7, page 9 in [31], in the sense that we also make use of the surjectivity of the exponential map of $SO(3)$, i.e., $g^* = e^{X_0 T}$.

Definition 2.1. *Let $u_0 \in \mathbf{Y}^\alpha$ be such that the stabilizer of u_0 is $\Sigma_{u_0} = I_3$, where I_3 is the 3×3 identity matrix. The set defined by $\{A\mathbf{\Phi}(t, u_0) \mid A \in SO(3),\ t \in [0, \infty)\}$ is called a* relative periodic orbit *for (2.1) if it is not a relative equilibrium and there exist a number $T > 0$ and a matrix $X_0 \in so(3)$ such that*

$$\mathbf{\Phi}(T, u_0) = e^{X_0 T} u_0 \text{ and } \mathbf{\Phi}(t, u_0) \notin SO(3)u_0 \text{ for any } t \in (0, T). \tag{2.3}$$

Given a relative periodic orbit, we say that T is the relative period *of the relative periodic orbit. If $|X_0|\, T \notin 2\pi\mathbb{Z}$, then any solution $A\mathbf{\Phi}(., u_0)$ with $A \in SO(3)$ of the reaction-diffusion system (2.1) is called a* modulated rotating wave *for (2.1).*

If $|X_0|\, T \in 2\pi\mathbb{Z}$, then $\mathbf{\Phi}(t, u_0)$ is a T-periodic solution of the reaction-diffusion (2.1).

From [4, 24, 27, 29], it is known that $SO(3)$ is diffeomorphic as a manifold to the real projective space $\mathbb{R}P^3$. If

$$Y = \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix} \in so(3), \text{ then we define } \overrightarrow{Y} = \begin{pmatrix} c \\ b \\ a \end{pmatrix}, \tag{2.4}$$

and also $|Y| = \left\| \overrightarrow{Y} \right\|$.

Let $\mathbf{\Phi}(., Au_0)$ be a rotating wave (respectively, a modulated rotating wave) for (2.1), where $u_0 \in \mathbf{Y}^\alpha$ and $A \in SO(3)$. By equivariance we have $\mathbf{\Phi}(t, Au_0) = Ae^{X_0 t}A^{-1}Au_0 = e^{AX_0A^{-1}t}Au_0$ (respectively, $\mathbf{\Phi}(T, Au_0) = Ae^{X_0 T}A^{-1}Au_0 = e^{AX_0A^{-1}T}Au_0$). Therefore, the vector $\overrightarrow{AX_0A^{-1}}$ is called the *frequency vector* of the rotating wave (respectively, a *primary frequency vector* of a modulated rotating wave).

Let the north hemisphere of $\mathbf{S}^2$ be $N = \{(x, y, z) \in \mathbf{S}^2 : z > 0 \text{ or } z = 0,\ x \in [-1, 1),\ y \in [0, 1]\}$, and the south hemisphere be the set $S = \mathbf{S}^2 \setminus N$. If the vector $\frac{1}{|X_0|}\overrightarrow{AX_0A^{-1}}$ is in

the north hemisphere, then $\omega_0 = |X_0|$ is called the *frequency* of the rotating wave (respectively, a *primary frequency* of a modulated rotating wave). Otherwise, $-\omega_0$ is called the *frequency* of the rotating wave (respectively, a *primary frequency* of a modulated rotating wave).

It is easy to see that a primary frequency vector $\overrightarrow{X_1}$ of a modulated rotating wave $\mathbf{\Phi}(t, Au_0)$ is unique up to an integer multiple of $\frac{2\pi}{|X_1|T}\overrightarrow{X_1}$ (see [7]).

As a geometrical interpretation, it is easy to see that for a rotating wave (respectively, a modulated rotating wave) $\mathbf{\Phi}(t, u_0)$, the tip motion $x_{tip}(\mathbf{\Phi}(t, u_0))$ is a circle (respectively, $x_{tip}(\mathbf{\Phi}(kT, u_0))$ for $k \in \mathbb{Z}$ are points of a circle) on the sphere $r\mathbf{S}^2$ with the center on the line having the direction of the frequency vector (respectively, the primary frequency vector) of $\mathbf{\Phi}(t, u_0)$; this is independent of the choice of the tip position function (see [7]).

2.2. Skew-product differential equations on the center manifold. When we talk about a supercritical Hopf bifurcation from an equilibrium, the bifurcating periodic solution has amplitude of order $\lambda^{\frac{1}{2}}$. For this reason we introduce the following definition.

Definition 2.2. *Let M be a smooth manifold, X be a normed space or the empty set, $p \geq 0$ be an integer, and $Y : X \times [0, \lambda_0) \times \mathbb{R}^p \to M$ for $\lambda_0 > 0$ small. The function Y is called CS on $X \times [0, \lambda_0) \times \mathbb{R}^p$ if the function $Z : X \times [0, \varepsilon_0) \times \mathbb{R}^p \to M$ defined by $Z(x, \varepsilon, \mu) = Y(x, \varepsilon^2, \mu)$ is smooth on $X \times [0, \varepsilon_0) \times \mathbb{R}^p$, where $\varepsilon_0 = \sqrt{\lambda_0}$. Moreover, Y is called* sufficiently CS *if Z is sufficiently smooth. We mention that CS stands for continuous with respect to λ and smooth with respect to $\varepsilon = \lambda^2$.*

We consider the one-parameter–dependant reaction-diffusion system

$$u_t = D\Delta_S u + F(u, \lambda) \text{ on } r\mathbf{S^2}, \tag{2.5}$$

where all the notation related to (2.5) is as in section 2, and $F = (F_1, F_2, \ldots, F_N) : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ is a C^{k+2} function with $3 \leq k \leq \infty$. We recall that the reaction-diffusion system (2.5) is considered on the function space $\mathbf{Y}$ defined by (2.2) as in section 2.

Let $u_0 \in \mathbf{Y}^\alpha$ be a relative equilibrium that is not an equilibrium for (2.5) at $\lambda = 0$ and such that the isotropy subgroup of u_0 is $\Sigma_{u_0} = I_3$. Moreover, let $\mathbf{\Phi}(t, u_0, 0) = e^{X_0 t}u_0$.

Consider the linearization L of the right-hand side of (2.5) with respect to the rotating wave $\mathbf{\Phi}(t, u_0, 0) = e^{X_0 t}u_0$ at $\lambda = 0$ in the corotating frame; that is,

$$L = D\Delta_S + D_u F(u_0, 0) - X_0.$$

We assume that L has five eigenvalues on the imaginary axis: three eigenvalues due to the rotation symmetry group $SO(3)$ (i.e., $0, \pm i\,|X_0|$) and another two pure eigenvalues, $\pm i\omega_H$, and all other eigenvalues are contained in the left-hand side plane.

We apply the equivariant center manifold reduction developed in [10, 28] by Fiedler et al. to reduce the study of the dynamics of (2.5) near a supercritical Hopf bifurcation, to the study of the following skew-product differential equations on $SO(3) \times \mathbb{C}$:

$$\begin{aligned} \dot{A} &= AX_G(q, \lambda), \\ \dot{q} &= X_N(q, \lambda), \end{aligned} \tag{2.6}$$

where $X_G : \mathbb{C} \times \mathbb{R} \to so(3)$ and $X_N : \mathbb{C} \times \mathbb{R} \to \mathbb{C}$ are sufficiently smooth functions, $X_G(0, 0) = X_0$, and $X_N(0, 0) = 0$. Moreover, $X_N(0, \lambda) = 0$, $D_q X_N(0, \lambda)$ has the eigenvalues $\alpha(\lambda) \pm i(\omega_H + \beta(\lambda))$ for $|\lambda|$ small, and $\alpha(0) = \beta(0) = 0$.

We also assume that $\alpha'(0) > 0$, such that a supercritical Hopf bifurcation takes place.

Then, there exists a unique sufficiently CS branch of periodic solutions $q(t,\lambda)$ bifurcating from $q = 0$ such that $q(t,\lambda) = O(\lambda^{\frac{1}{2}})$ with frequency $\omega_\lambda = \omega_H + O(\lambda)$ for $\lambda \geq 0$ small.

We recall that any solution $\mathbf{\Phi}(t, u_\lambda, \lambda)$ of the reaction-diffusion system (2.5) on $SO(3)\times\mathbb{C}$ is given by $\mathbf{\Phi}(t, u_\lambda, \lambda) = A_1(t,\lambda)\mathbf{\Psi}(q_1(t,\lambda))$ for $|\lambda|$ small and $A_1(t,\lambda)$ and $q_1(t,\lambda)$ are solutions of the differential equations (2.6), and $\mathbf{\Psi}$ is a local diffeomorphism from $\mathbf{C}$ into $\mathbf{Y}^\alpha$.

As a consequence, there exists a sufficiently CS branch $\mathbf{\Phi}(t, u_\lambda, \lambda)$ of solutions for the reaction-diffusion system (2.5) such that $\mathbf{\Phi}(t, u_0, 0) = e^{X_0 t}u_0$ and for $\lambda > 0$ small, $\mathbf{\Phi}(t, u_\lambda, \lambda)$ is either an orbitally stable modulated rotating wave with a primary frequency vector $\overrightarrow{X(\lambda)}$ and with the secondary frequency ω_λ or an orbitally stable periodic solution with the period $\frac{2\pi}{|\omega_\lambda|}$; $\overrightarrow{X(\lambda)}$ is discussed in Lemma 2.3 below.

If we define the sufficiently CS function

$$X^G(t,\lambda) = X_G(q(t,\lambda),\lambda) \text{ for } \lambda > 0 \text{ small and } t \in [0,\infty), \tag{2.7}$$

where $X_G(q,\lambda)$ appears in the differential equations (2.6), then $X^G(t,\lambda)$ is a $\frac{2\pi}{|\omega_\lambda|}$-periodic function in t for $\lambda > 0$ small. Let us consider the initial value problem (IVP)

$$\begin{aligned} \dot{A} &= AX^G(t,\lambda), \\ A(0) &= I_3, \end{aligned} \tag{2.8}$$

where $X^G(t,\lambda)$ is defined in (2.7) and I_3 is the 3×3 identity matrix. Since $SO(3)$ is a compact manifold and X^G is a sufficiently CS function, the IVP (2.8) has a unique sufficiently CS solution $A(t,\lambda)$ which is globally defined.

Lemma 2.3 (decomposition of $A(t,\lambda)$). *Suppose that the assumptions made in this section hold. Then, there exists a sufficiently CS solution to the IVP (2.8) given by*

$$A(t,\lambda) = e^{X(\lambda)t}B(t,\lambda), \tag{2.9}$$

where $X(\lambda) \in so(3)$ is the primary frequency vector of $\mathbf{\Phi}(t, u_\lambda, \lambda)$ and $B(t,\lambda)$ is a $\frac{2\pi}{|\omega_\lambda|}$-periodic function such that $B(0,\lambda) = I_3$.

The function $B(t,\lambda)$ is called the *periodic part* of $\mathbf{\Phi}(t, u_\lambda, \lambda)$ associated to $\overrightarrow{X(\lambda)}$. The proof of Lemma 2.3 uses Floquet theory and it is very briefly presented below.

Proof of Lemma 2.3. Since the exponential map $\exp : so(3) \to SO(3)$ is surjective, there exists a matrix $X(\lambda) \in so(3)$ such that $A(T(\lambda)) = e^{X(\lambda)T(\lambda)}$, where $T(\lambda) = \frac{2\pi}{|\omega_\lambda|}$.

Since $X_G(t,\lambda)$ is $T(\lambda)$-periodic, using Floquet theory we get that

$$A(t,\lambda) = e^{X(\lambda)t}B(t,\lambda) \text{ with } B(0,\lambda) = I_3 \text{ and } B(t,\lambda)\ \frac{2\pi}{\omega_\lambda}\text{-periodic.} \tag{2.10}$$

We have only to check that $\overrightarrow{X(\lambda)}$ is a primary frequency vector for $\mathbf{\Phi}(t, u_\lambda, \lambda)$ in the sense of Definition 2.1. We have

$$\mathbf{\Phi}(t, u_\lambda, \lambda) = A(t,\lambda)\mathbf{\Psi}(q(t,\lambda)) = e^{X(\lambda)t}B(t,\lambda)\mathbf{\Psi}(q(t,\lambda)),$$

where we recall that $\mathbf{\Psi}$ is a local diffeomorphism from $\mathbf{C}$ into $\mathbf{Y}^\alpha$ with $\mathbf{\Psi}(q(0,\lambda)) = u_\lambda$.

If we denote $q_1(t,\lambda) = B(t,\lambda)\mathbf{\Psi}(q(t,\lambda))$, then $\mathbf{\Phi}(t,u_\lambda,\lambda) = e^{X(\lambda)t}q_1(t,\lambda)$. Therefore,

$$\mathbf{\Phi}(T(\lambda),u_\lambda,\lambda) = e^{X(\lambda)T(\lambda)}u_\lambda. \tag{2.11}$$

Therefore, we have only to check that at least generically $\mathbf{\Phi}(t,u_\lambda,\lambda) \notin SO(3)u_\lambda$ for any $t \in (0,T(\lambda))$. Let $T_1(\lambda) \in (0,T(\lambda)]$ be the first positive number such that $\mathbf{\Phi}(T_1(\lambda),u_\lambda,\lambda) \in SO(3)u_\lambda$. We have to show that $T_1(\lambda) = T(\lambda)$.

It is easy to see that $T(\lambda) \in T_1(\lambda)\mathbb{Z}$. Namely, let $T(\lambda) = lT_1(\lambda) + q$ with $q \in [0,T_1(\lambda))$ and $l \in \mathbb{Z}$. Let

$$\mathbf{\Phi}(T_1(\lambda),u_\lambda,\lambda) = e^{Y(\lambda)T_1(\lambda)}u_\lambda \tag{2.12}$$

with $Y(\lambda) \in so(3)$. Using the $SO(3)$-equivariance of the semiflow $\mathbf{\Phi}$ and $T(\lambda) = lT_1(\lambda) + q$, we get

$$\mathbf{\Phi}(T(\lambda),u_\lambda,\lambda) = \mathbf{\Phi}(q,\mathbf{\Phi}(lT_1(\lambda),u_\lambda,\lambda),\lambda) = e^{lY(\lambda)T_1(\lambda)}\mathbf{\Phi}(q,u_\lambda,\lambda), \tag{2.13}$$

which implies

$$\mathbf{\Phi}(q,u_\lambda,\lambda) = e^{-lY(\lambda)T_1(\lambda)}e^{X(\lambda)T(\lambda)}u_\lambda \in SO(3)u_\lambda.$$

But $q \in [0,T_1(\lambda))$ and the definition of $T_1(\lambda)$ implies $q = 0$; i.e., $T(\lambda) = lT_1(\lambda)$. Further, we have to show that $l = 1$.

But $q = 0$, $T(\lambda) = lT_1(\lambda)$, and (2.13) implies

$$\mathbf{\Phi}(T(\lambda),u_\lambda,\lambda) = e^{Y(\lambda)T(\lambda)}u_\lambda,$$

which combined with (2.11) gives

$$e^{X(\lambda)T(\lambda)}u_\lambda = e^{Y(\lambda)T(\lambda)}u_\lambda.$$

Hence $\Sigma_{u_\lambda} = I_3$ gives $e^{X(\lambda)T(\lambda)} = e^{Y(\lambda)T(\lambda)}$.

Therefore,

$$\mathbf{\Phi}(t,u_\lambda,\lambda) = e^{Y(\lambda)t}B_1(t,\lambda)\mathbf{\Psi}(q(t,\lambda))$$

for some $T(\lambda)$-periodic function $B_1(t,\lambda)$. Hence,

$$\mathbf{\Phi}(T_1(\lambda),u_\lambda,\lambda) = e^{Y(\lambda)T_1(\lambda)}B_1(T_1(\lambda),\lambda)\mathbf{\Psi}(q(T_1(\lambda),\lambda)).$$

On the other hand

$$\mathbf{\Phi}(T_1(\lambda),u_\lambda,\lambda) = e^{Y(\lambda)T_1(\lambda)}u_\lambda$$

implies

$$\mathbf{\Psi}(q(T_1(\lambda),\lambda)) = (B(T_1(\lambda),\lambda))^{-1}u_\lambda. \tag{2.14}$$

Since u_λ is close to u_0 and $SO(3)u_0 \cap \mathbf{\Psi}(\mathbb{C}) = \{u_0\}$, we get generically $SO(3)u_\lambda \cap \mathbf{\Psi}(\mathbb{C}) = \{u_\lambda\}$.

Now $q(T_1(\lambda),\lambda) \in \mathbb{C}$ and $(B(T_1(\lambda),\lambda))^{-1}u_\lambda \in SO(3)u_\lambda$ combined with (2.14) and $\Sigma_{u_\lambda} = I_3$ give

$$B_1(T_1(\lambda),\lambda) = I_3.$$

Then, (2.14) implies

$$\mathbf{\Psi}(q(T_1(\lambda),\lambda)) = u_\lambda = \mathbf{\Psi}(q(0,\lambda)).$$

Since $\mathbf{\Psi}$ is a local diffeomorphism from $\mathbf{C}$ into $\mathbf{Y}^\alpha$, $q(t,\lambda)$ is a solution of $\dot{q} = X_N(q,\lambda)$ such that

$$q(T_1(\lambda),\lambda) = u_\lambda = q(0,\lambda).$$

Therefore, $q(t,\lambda)$ is $T_1(\lambda)$-periodic. On the other hand, it has the principal period $T(\lambda)$. Hence, $T(\lambda) = T_1(\lambda)$. ■

For the Euclidean group $SE(2)$, it is well known that the differential equations (2.6) can be solved explicitly due to the fact that $SE(2) \simeq \mathbb{C} \times SO(2)$ (see [14]). A similar treatment for the differential equations (2.6) can be done *theoretically* using the BCH formula in $so(3)$, the decomposition obtained in Lemma 2.3, and the closed form of the reduced differential equations on $so(3)$ obtained in Theorem 3.4 (see section 3). Our approach follows the one presented in [14].

3. Reduced differential equations on $so(3)$—closed form. Throughout this section we suppose that the assumptions made in section 2 hold. We recall the BCH formula in $so(3)$ below. A model for the real projective space $\mathbb{R}P^3$ diffeomorphic to $SO(3)$ is the set

$$D = \{\overrightarrow{y} \in \mathbb{R}^3 \mid \|\overrightarrow{y}\| \le \pi, \text{ with the antipodal points of the norm } |y| = \pi \text{ identified}\},$$

where $\vec{y}$ associated to $y \in so(3)$ is defined in (2.4). In fact, D is the quotient set $E/\!\sim$, where $\sim$ is the equivalence relation $\vec{y} \sim \vec{z}$ iff $z = -y$, $|y| = \pi$, and $E = \{\overrightarrow{y} \in \mathbb{R}^3 \mid \|\overrightarrow{y}\| \le \pi\}$. Sometimes, instead of $\overrightarrow{y} \in \mathbb{R}^3$ with $|\overrightarrow{y}| \le \pi$, we use $y \in so(3)$, in which case we denote the equivalence class $[y] = [\overrightarrow{y}]$.

From [7, 9] it is known that there exists a unique smooth function $\mathrm{Log} : SO(3) \to D$ such that $e^{\mathrm{Log}(A)} = A$.

Definition 3.1. *We define $BCH(X,Y) = \mathrm{Log}(e^X e^Y)$ for any $X, Y \in so(3)$. Clearly, we have $e^{BCH(X,Y)} = e^X e^Y$ and $BCH(X,Y) \in D$ for any $X, Y \in so(3)$.*

Theorem 3.2 (BCH formula in $so(3)$, [7, 9]). *The BCH formula in $so(3)$ has the form*

$$BCH(X,Y) = [\alpha(X,Y)X + \beta(X,Y)Y + \gamma(X,Y)[X,Y]] \text{ for } X, Y \in so(3), \tag{3.1}$$

where there are closed formulas for α, β, and γ (see [7, 9]). The functions $\alpha^2, \beta^2, \gamma^2$ are smooth on $so(3) \times so(3)$ and $|\alpha|, |\beta|, |\gamma|$ are continuous on $so(3) \times so(3)$. Also, the function BCH is smooth from $so(3) \times so(3)$ into D.

The BCH formula in $so(3)$ is computational (i.e., it can be translated into a computer-implemented program) and acts as a "logarithm."

Throughout this paper $u_0 \in \mathbf{Y}^\alpha$ is a relative equilibrium for (2.5) at $\lambda = 0$ with trivial isotropy subgroup $\Sigma_{u_0} = I_3$. Also, let $\mathbf{\Phi}(t, u_0, 0) = e^{X_0 t} u_0$.

The main result of this section is the closed form for the reduced differential equations on $so(3)$ in Theorem 3.4. Its proof is based on Lemma 3.3. We recall that the function $X^G(t, \lambda)$ is a $\frac{2\pi}{|\omega_\lambda|}$-periodic function defined in (2.7) and we denote by O_3 the 3×3 zero matrix.

Lemma 3.3. *We consider the IVP*

$$\begin{aligned} \overrightarrow{\dot{Z}} &= g(Z)\overrightarrow{X^G(t,\lambda)}, \\ Z(0) &= O_3, \end{aligned} \tag{3.2}$$

where $g : so(3) \to so(3)$ is a sufficiently smooth function such that there exist positive constants $M_g \geq 0$, $K_g > 0$, and $C > 0$ for which $|g(Z)| \leq M_g + K_g |Z|$ for $|Z| \leq C$.

Then, there exists a positive integer n independent of λ such that the IVP (3.2) has a unique sufficiently CS solution $Z(t, \lambda)$ such that $|Z(t, \lambda)| \leq C$ on $t \in [0, \frac{2\pi}{n|\omega_\lambda|}]$ for $\lambda \geq 0$ small.

The integer n in Lemma 3.3 can be found explicitly and it is not 1 in general. Therefore, the function $Z(t, \lambda)$ is not defined in general on the entire interval $[0, \frac{2\pi}{|\omega_\lambda|}]$ for $\lambda \geq 0$ small. The proof shows how we can get the maximal interval $[0, \frac{2\pi}{n|\omega_\lambda|}]$.

Proof of Lemma 3.3. The proof is straightforward and is based on the following result from [16].

Lemma. Let $g(t, u)$ be a continuous function on an open connected set $[a_1, b_1) \times [0, \infty) \subset \Omega \subset \mathbb{R}^2$ and such that the IVP for the scalar equation $\dot{u} = g(t, u)$ has a unique solution $u(t) \geq 0$ on $t \in [a_1, b_1)$. If $f : [a_1, b_1) \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous and $\|f(t, x)\| \leq g(t, \|x\|)$ for $t \in [a_1, b_1)$ and $x \in \mathbb{R}^n$, then the solutions of $\dot{x} = f(t, x)$, $\|x(a_1)\| \leq u(a_1)$ exists on $[a_1, b_1)$, and $\|x(t)\| \leq u(t)$ for $t \in [a_1, b_1)$.

Let $M > 0$ (independent of $\lambda \geq 0$ small) be such that $\left|X^G(t, \lambda)\right| < M$ for $t \in [0, \infty)$ and $\lambda \geq 0$ small. For any $|Z| < C$ and $\lambda \geq 0$ small, we have

$$\left\| g(Z)\overrightarrow{X^G(t,\lambda)} \right\| \leq |g(Z)|\, M \leq M\left(M_g + K_g\, |Z|\right). \tag{3.3}$$

The IVP

$$\begin{aligned} \dot{a} &= M(M_g + K_g a), \\ a(0) &= 0 \end{aligned} \tag{3.4}$$

has solution $a(t) \leq C$ on a maximal interval $[0, b_{max}]$ with b_{max} independent of λ.

Therefore, using the previous lemma from [16] with $a_1 = 0$, it follows that the sufficiently CS solution $Z(t, \lambda)$ of the IVP (3.2) is defined for any $t \in [0, b_{max}]$ and $\lambda \geq 0$ small, and that $|Z(t, \lambda)| \leq C$ for any $t \in [0, b_{max}]$ and $\lambda \geq 0$ small.

We recall that $T(\lambda) = \frac{2\pi}{|\omega_\lambda|}$. If we choose a positive integer $n > 0$ such that $n > \frac{T(\lambda)}{b_{max}}$, then the IVP (3.2) has a solution $Z(t, \lambda)$ defined and sufficiently CS on $t \in [0, \frac{T(\lambda)}{n}]$ and $\lambda \geq 0$ small. Since $T(\lambda) = T(0) + O(\lambda)$, it is clear that we can choose n independent of λ for $\lambda \geq 0$ small. ■

The main result of this section is stated in the following theorem.

Theorem 3.4 (differential equations on $so(3)$). *There exists a sufficiently CS function $Z(t,\lambda)$ such that $|Z(t,\lambda)| < 2\pi$ and $A(t,\lambda) = e^{Z(t,\lambda)}$ for $t \in [0,\infty)$ and $\lambda \geq 0$ small, where $A(t,\lambda)$ is the solution of the IVP (2.8).*

Moreover, $Z(t,\lambda)$ satisfies the following IVP for $t \in [0,\infty)$ and $\lambda \geq 0$ small:

$$\begin{aligned} \overrightarrow{\dot{Z}} &= \left(I_3 + \tfrac{1}{2}Z + f(|Z|)\tfrac{Z^2}{|Z|^2}\right)\overrightarrow{X^G(t,\lambda)} \pmod{2\pi}, \\ Z(0) &= O_3, \end{aligned} \tag{3.5}$$

where $f(a) = 1 - \frac{a}{2}\cot\frac{a}{2}$ if $a \neq 0$ and $f(0) = 0$.

The proof of Theorem 3.4 shows how we can construct the solution $Z(t,\lambda)$ inductively using the BCH formula in $so(3)$. It will be presented in section 6. As mentioned before, this construction can be translated into a computer-implemented method. At the end of this section a simplified version of the method is presented. It refers to both sections 3 and 4.

Because we do not have to decide if we are in the resonant or nonresonant case, the reduction to differential equations on $so(3)$ can be sometimes more suitable for applications to real life, where small variations can get things away from the resonant case. As mentioned before, it does matter if we are in the resonant or nonresonant case, but the change that appears in the resonant case in the Taylor expansions is implicitly taken care of.

In fact, we can generalize this approach to get the reduced differential equations on the Lie algebra g associated to any compact Lie group $G \subset O(N)$. The way this generalization works is briefly presented in section 6 after the proof of Theorem 3.4.

Simplified version of the method for sections 3 and 4. The first step is to get from the IVP (2.8) to the reduced differential equations (3.5) on $so(3)$. We use the surjectivity of the exponential map exp of $SO(3)$ and we make the change of variable $A = e^Z$ in the IVP (2.8). The expression for the differential of the exponential map combined with some geometrical properties of the adjoint representation of $so(3)$ [24, 27] are used to get the closed form of the reduced differential equations (3.5) on $so(3)$.

In the second step we construct the solution $|Z(t,\lambda)| < 2\pi$ inductively. We show that the IVP (2.8) has a solution $Z_1(t,\lambda)$ on a fixed small interval $[0,t_\lambda]$ with $t_\lambda = \frac{2\pi}{n|\omega_\lambda|}$ for some $n \in \mathbb{Z}$. In general $n > 1$. The main property used is the boundedness of the right-hand side of (3.5) with respect to $|Z|$. Then, we make the change of variable $B = e^{-Z_1(t,\lambda)}A$ and we end up with the IVP $\dot{B} = BX^G(t,\lambda)$ and $B(t_\lambda) = I_3$. We show that it has a solution $Z_2(t,\lambda)$ on $[t_\lambda, 2t_\lambda]$. We continue and in the end we use the BCH formula on $so(3)$ to put together all these solutions defined on "partial" intervals, such that we have constructed a solution $|Z(t,\lambda)| < 2\pi$ on $[0, \frac{2\pi}{|\omega_\lambda|}]$. It is very easy to extend the solution $Z(t,\lambda)$ to $[0,\infty)$. The fact that $|Z(t,\lambda)| < 2\pi$ is the same as taking mod 2π in the first differential equation in the IVP (2.8).

The third step is to get a closed formula for the primary frequency vectors $\overrightarrow{X(\lambda)}$ using the solution $Z(t,\lambda)$ constructed above. The scaling of the primary frequency vectors is different depending on the resonant or nonresonant case. That is why we introduce the primary frequency vectors branch $\overrightarrow{X^f(\lambda)}$ in the nonresonant case. We use the BCH formula in $so(3)$ to get the closed formula for the periodic parts and the characterization of the bifurcating modulated rotating waves.

4. Characterization of the bifurcating modulated rotating waves on the sphere. Throughout this section we suppose that the assumptions made in section 2 hold. From Lemma 2.3 we know that a primary frequency vector $\overrightarrow{X(\lambda)}$ of $\mathbf{\Phi}(t, u_\lambda, \lambda)$ is given by $A(\frac{2\pi}{|\omega_\lambda|}, \lambda) = e^{X(\lambda)\frac{2\pi}{|\omega_\lambda|}}$ for $\lambda > 0$ small.

Let $X_0^1 = \frac{1}{|X_0|}X_0$. Then, there exist unit vectors $X_1, X_2 \in so(3)$ such that the set $\{\overrightarrow{X}_0^1, \overrightarrow{X}_1, \overrightarrow{X}_2\}$ is an orthonormal basis in $\mathbb{R}^3$ satisfying $\overrightarrow{X}_0^1 \times \overrightarrow{X}_1 = \overrightarrow{X}_2$, $\overrightarrow{X}_1 \times \overrightarrow{X}_2 = \overrightarrow{X}_0^1$, and $\overrightarrow{X}_2 \times \overrightarrow{X}_0^1 = \overrightarrow{X}_1$. We recall that $[X, Y] = \mathrm{ad}(X)Y = XY - YX$ for any $X, Y \in so(3)$.

Since $so(3)$ is isomorphic to $\mathbb{R}^3$ and $\overrightarrow{[X,Y]} = \overrightarrow{X} \times \overrightarrow{Y}$, $\{X_0^1, X_1, X_2\}$ is a basis of the Lie algebra $so(3)$ such that $[X_0^1, X_1] = X_2$, $[X_1, X_2] = X_0^1$, and $[X_2, X_0^1] = X_1$.

Proposition 4.1 (primary frequency vectors). *There exists a unique sufficiently CS branch $X(\lambda)$ such that $|X(\lambda)| < |\omega_\lambda|$ and $\overrightarrow{X(\lambda)}$ is a primary frequency vector of $\mathbf{\Phi}(t, u_\lambda, \lambda)$ for $\lambda > 0$ small, and $e^{X(0)\frac{2\pi}{|\omega_\lambda|}} = e^{X_0\frac{2\pi}{|\omega_\lambda|}}$. Moreover,*

$$X(\lambda) = \frac{|\omega_\lambda|}{2\pi} Z\left(\frac{2\pi}{|\omega_\lambda|}, \lambda\right) \tag{4.1}$$

for $\lambda \geq 0$ small, where $Z(t, \lambda)$ is given in Theorem 3.4 in section 3. Moreover, if $|X_0| = k\,|\omega_H|$ for some $k \in \mathbb{Z}$, then $|X(\lambda)| = O(\lambda^{\frac{|k|}{2}})$.

Proof of Proposition 4.1. It is clear that $\overrightarrow{X(\lambda)}$ defined by

$$X(\lambda) = \frac{1}{T(\lambda)} Z(T(\lambda), \lambda) \tag{4.2}$$

for $\lambda \geq 0$ small is a primary frequency vector of the modulated rotating wave $\mathbf{\Phi}(t, u_\lambda, \lambda)$ (i.e., $A(T(\lambda), \lambda) = e^{X(\lambda)T(\lambda)}$), and the branch $X(\lambda)$ is sufficiently CS for $\lambda \geq 0$ small. From the proof of Theorem 3.4 it is clear that the branch $X(\lambda)$ defined by (4.2) is the unique one such that $|X(\lambda)| < 2\pi$ for $\lambda \geq 0$ small. Moreover, $A(t, 0) = e^{X_0 t}$ implies $e^{X(0)\frac{2\pi}{|\omega_\lambda|}} = e^{X_0\frac{2\pi}{|\omega_\lambda|}}$.

The scaling results for $X(\lambda)$ can be found from the reduced differential equations on $so(3)$ either using the approach of Wulff in [31, pp. 29–32] or using the normal forms theory of Fiedler and Turaev in [11], as they are described in [5, pp. 9–13]. ∎

We have a closed formula for the angle $\theta(\lambda)$ between the primary frequency vector $\overrightarrow{X(\lambda)}$ and the rotation frequency vector $\overrightarrow{X_0}$. Namely,

$$\cos(\theta(\lambda)) = \frac{\overrightarrow{X(\lambda)} \cdot \overrightarrow{X_0}}{|X(\lambda)|\,|X_0|}, \tag{4.3}$$

which combined with Proposition 4.1 can be used to control the tip motions of the bifurcating modulated rotating waves on the sphere.

The branch $X(\lambda)$ from Proposition 4.1 does not have the property that $X(0) = X_0$. The next corollary shows that we can construct a branch $\overrightarrow{X^f(\lambda)}$ of primary frequency vectors of $\mathbf{\Phi}(t, u_\lambda, \lambda)$ such that $X^f(0) = X_0$.

Corollary 4.2. *There is a branch $X^f(\lambda)$ of primary frequency vectors of $\mathbf{\Phi}(t, u_\lambda, \lambda)$ for $\lambda \geq 0$ such that $X^f(0) = X_0$. Moreover,*

1. *if* $|X_0| \notin \omega_H\mathbb{Z}$, *then the branch* $X^f(\lambda)$ *is sufficiently CS for* $\lambda \geq 0$ *small and*

$$X^f(\lambda) = [|X_0| + O(\lambda)]\, X_0^1 + O(\lambda^{\frac{1}{2}})X_1 + O(\lambda^{\frac{1}{2}})X_2; \tag{4.4}$$

2. *if* $|X_0| \in \omega_H\mathbb{Z}$, *then* $\left|X^f(\lambda)\right|$ *is continuous for* $\lambda \geq 0$ *small.*

Although Corollary 4.2 deals with the Taylor series of $X^f(\lambda)$, we emphasize that there is a closed formula for $X^f(\lambda)$, as can be seen from the proof of this corollary. Also, either Proposition 4.1 or Corollary 4.2 shows that there is no frequency-locking phenomenon for the modulated rotating waves $\mathbf{\Phi}(t, u_\lambda, \lambda)$.

Proof of Corollary 4.2. Let $|X_0| = \alpha_0 + k\omega_H$ with $\alpha_0 \in [0, |\omega_H|)$, $k \in \mathbb{Z}$ and $T(0) = \frac{2\pi}{|\omega_H|}$.

Since $A(T(0), 0) = e^{X_0 T(0)} = e^{X(0)T(0)} = e^{Z(T(0),0)}$, then by the definition of $X(0)$ in (4.2) we get $X(0) = \alpha_0 X_0^1$. Hence, $|X(0)| = \alpha_0$.

We define $X^f(0) = X_0$. If $|X(\lambda)| \neq 0$, then we define $X^f(\lambda) = \frac{|X(\lambda)|+k\omega_\lambda}{|X(\lambda)|}X(\lambda)$.

If $|X(\lambda)| = 0$, then we define $X^f(\lambda) = X(\lambda) + k\omega_\lambda Q(\lambda)$, where $Q(\lambda) \in so(3)$ such that $|Q(\lambda)| = 1$ and $\lambda > 0$.

Hence, we get that $e^{X^f(\lambda)T(\lambda)} = e^{X(\lambda)T(\lambda)}$; i.e., $\overrightarrow{X^f(\lambda)}$ is a primary frequency vector for the bifurcating modulated rotating waves.

It is clear that $\left|X^f(\lambda)\right| = |X(\lambda)| + k\omega_\lambda$ for $\lambda \geq 0$ small. Hence, $\left|X^f(\lambda)\right|$ is continuous for $\lambda \geq 0$ small. If $|X_0| \notin \omega_H\mathbb{Z}$, then it is clear that $X^f(\lambda)$ is well defined and sufficiently CS. In this case the scaling of $X(\lambda)$ yields either from the approach of Wulff in [31] or from the normal forms theory of Fiedler and Turaev in [11], and it is

$$X(\lambda) = (\alpha_0 + O(\lambda))X_0^1 + O(\lambda^{\frac{1}{2}})X_1 + O(\lambda^{\frac{1}{2}})X_2,$$

which combined with the definition of $X^f(\lambda)$ gives us the scaling in (4.4). ∎

Using Lemma 2.3 we have $A(t, \lambda) = e^{X(\lambda)t}B(t, \lambda) = e^{X^f(\lambda)t}B^f(t, \lambda)$ for each bifurcating modulated rotating wave, where $B(t, \lambda)$ and $B^f(t, \lambda)$ are $\frac{2\pi}{|\omega_\lambda|}$-periodic in t, and $B(0, \lambda) = B^f(0, \lambda) = I_3$ for $\lambda > 0$ small. Also, let $B(t, 0)$ and $B^f(t, 0)$ be such that $A(t, 0) = e^{X(0)t}B(t, 0) = e^{X^f(0)t}B^f(t, 0)$ (i.e., $B^f(t, 0) = I_3$).

Proposition 4.3 (periodic parts).

1. *If* $|X_0| \notin \omega_H\mathbb{Z}$, *then we have*

$$B^f(t, \lambda) = e^{\lambda^{\frac{1}{2}}Y(t,\lambda)} \tag{4.5}$$

for $t \in [0, \infty)$ *and* $\lambda \geq 0$ *small, where* $Y(t, \lambda)$ *is* $\frac{2\pi}{|\omega_\lambda|}$*-periodic in* t *and sufficiently CS.*

2. *If* $|X_0| \in \omega_H\mathbb{Z}$, *then we have*

$$B(t, \lambda) = e^{X_0 t + \lambda^{\frac{1}{2}}H(t,\lambda)} \tag{4.6}$$

for $t \in [0, \infty)$ *and* $\lambda \geq 0$ *small, where* $H(t, \lambda)$ *is sufficiently CS and* $e^{X_0 \frac{2\pi}{|\omega_\lambda|} + \lambda^{\frac{1}{2}}H(\frac{2\pi}{|\omega_\lambda|},\lambda)} = I_3$.

Although Proposition 4.3 deals with the Taylor series of the periodic parts $B(t,\lambda)$ and $B^f(t,\lambda)$, we emphasize that there are closed formula for them. Namely,

$$B(t,\lambda) = e^{BCH(-X(\lambda)t, Z(t,\lambda))} \text{ and } B^f(t,\lambda) = e^{BCH(-X^f(\lambda)t, Z(t,\lambda))}, \tag{4.7}$$

where $X(\lambda)$ is given by (4.1), $X^f(\lambda)$ is defined in the proof of Corollary 4.2, and $Z(t,\lambda)$ is given in Theorem 3.4.

Proof of Proposition 4.3. We recall that in the proof of Theorem 3.4 (see section 6) we defined the following smooth function:

$$q([\overrightarrow{Y}]) = \begin{cases} Y & \text{if } \frac{1}{|Y|}\overrightarrow{Y},\ \frac{1}{|X_0|}\overrightarrow{X_0} \text{ are in the same hemisphere or } Y = O_3, \\ (-\frac{2\pi}{|Y|}+1)Y & \text{if } \frac{1}{|Y|}\overrightarrow{Y},\ \frac{1}{|X_0|}\overrightarrow{X_0} \text{ are in different hemispheres and } Y \neq O_3, \end{cases} \tag{4.8}$$

where points 0 and 2π are identified, and $[\overrightarrow{Y}]$ is an equivalence class of D. It is easy to see that $e^{q([Y])} = e^{[Y]}$ for any $[Y] \in D$. We also recall that the function BCH from Definition 3.1 is sufficiently smooth. If $|X_0| \notin \omega_H\mathbb{Z}$, then we have

$$B^f(t,\lambda) = e^{-X^f(\lambda)t}A(t,\lambda) = e^{-X^f(\lambda)t}e^{Z(t,\lambda)} = e^{q(BCH(-X^f(\lambda)t, Z(t,\lambda)))}. \tag{4.9}$$

It is clear that $Per^f(t,\lambda) = q(BCH(-X^f(\lambda)t, Z(t,\lambda)))$ is sufficiently CS for $t \in [0,\infty)$ and $\lambda \geq 0$ small. Since

$$e^{-X^f(0)t}e^{Z(t,0)} = A(t)^{-1}A(t) = I_3 = e^{BCH(-X^f(0)t, Z(t,0))},$$

it follows that $Per^f(t,0) = q(BCH(-X^f(0)t, Z(t,0))) = O_3$. Then, $Per^f(t,\lambda) = \lambda^{\frac{1}{2}}Y(t,\lambda)$.

The $T(\lambda)$-periodicity of $Y(.,\lambda)$ results from the fact that $B(t,\lambda)$ is $T(\lambda)$-periodic and the exponential map is a local diffeomorphism about O_3.

If $|X_0| \in \omega_H\mathbb{Z}$, then we have

$$B(t,\lambda) = e^{-X(\lambda)t}A(t,\lambda) = e^{-X(\lambda)t}e^{Z(t,\lambda)} = e^{q(BCH(-X(\lambda)t, Z(t,\lambda)))}. \tag{4.10}$$

Since

$$A(t,0) = e^{X_0 t} = e^{Z(t,0)} \text{ and } Per(t,0) = q(BCH(O_3, Z(t,0))) = q(BCH(O_3, X_0 t)),$$

it yields that $Per(t,0) = X_0 t \pmod{2\pi}$.

Hence, if we denote $r(t) = t \pmod{\frac{2\pi}{|X_0|}}$, we get $Per(t,\lambda) = X_0 r(t) + \lambda^{\frac{1}{2}}H_1(t,\lambda)$. Therefore, it is clear that $B(t,\lambda) = e^{X_0 t + \lambda^{\frac{1}{2}}H(t,\lambda)}$. Moreover, $B(t,\lambda)$ is $\frac{2\pi}{|\omega_\lambda|}$-periodic with $B(0,\lambda) = I_3$, which implies $e^{X_0\frac{2\pi}{|\omega_\lambda|} + \lambda^{\frac{1}{2}}H(\frac{2\pi}{|\omega_\lambda|},\lambda)} = I_3$. ∎

Using Proposition 4.1, Corollary 4.2, and Proposition 4.3, we get the characterization of the modulated rotating waves $\mathbf{\Phi}(t, u_\lambda, \lambda) = \lambda^{\frac{1}{2}}A(t,\lambda)r(t,\lambda)$ in Theorem 4.4.

Theorem 4.4 (characterization of the bifurcating MRW).

1. *If* $|X_0| \notin \omega_H\mathbb{Z}$, *then we have*

$$A(t,\lambda) = e^{\left[(|X_0|+O(\lambda))X_0^1 + O(\lambda^{\frac{1}{2}})X_1 + O(\lambda^{\frac{1}{2}})X_2\right]t}e^{\lambda^{\frac{1}{2}}Y(t,\lambda)},$$

where $Y(t,\lambda)$ *is sufficiently* CS *and* $\frac{2\pi}{|\omega_\lambda|}$*-periodic for* $\lambda \geq 0$ *small.*

2. *If* $|X_0| = k\omega_H$ *for some* $k \in \mathbb{Z}$, *then we have*

$$A(t,\lambda) = e^{X(\lambda)t} e^{X_0 t + \lambda^{\frac{1}{2}} H(t,\lambda)},$$

where $|X(\lambda)| = O(|\lambda|^{\frac{k}{2}})$, $H(t,\lambda)$ *is sufficiently CS, and* $e^{X_0 \frac{2\pi}{|\omega_\lambda|} + \lambda^{\frac{1}{2}} H(\frac{2\pi}{|\omega_\lambda|},\lambda)} = I_3$ *for* $\lambda \geq 0$ *small.*

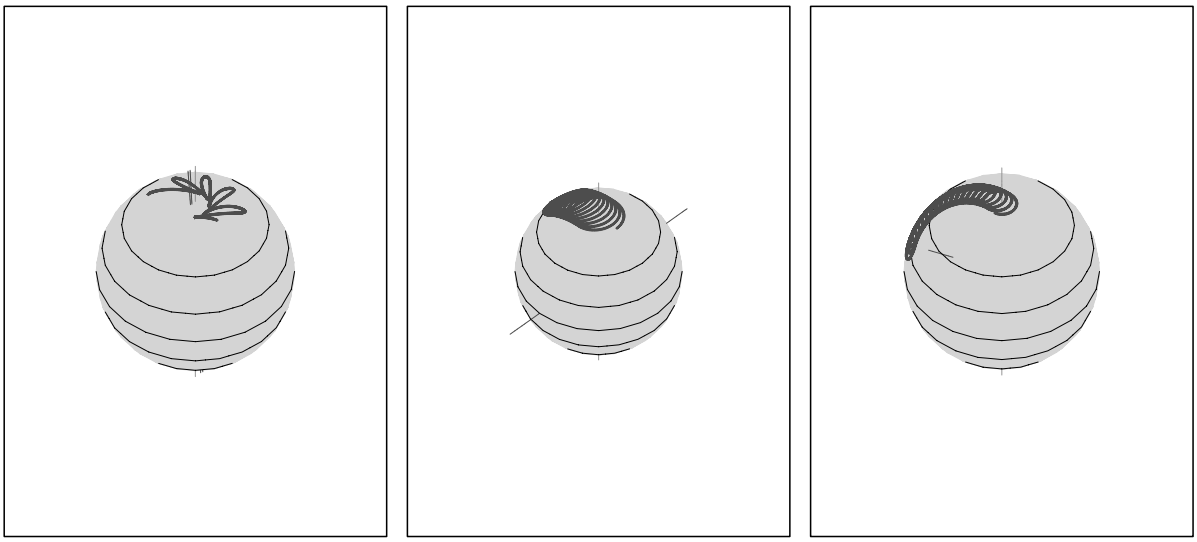
Although Theorem 4.4 deals with the Taylor series for $A(t,\lambda)$, we emphasize that there is an *explicit* characterization for $A(t,\lambda)$. Namely,

$$\begin{aligned} A(t,\lambda) &= e^{X(\lambda)t} e^{BCH(-X(\lambda)t, Z(t,\lambda))} \text{ and} \\ A(t,\lambda) &= e^{X^f(\lambda)t} e^{BCH(-X^f(\lambda)t, Z(t,\lambda))}, \end{aligned} \tag{4.11}$$

where $X(\lambda)$ is given by (4.1), $X^f(\lambda)$ is defined in the proof of Corollary 4.2, and $Z(t,\lambda)$ is given in Theorem 3.4.

Proof of Theorem 4.4. Using Lemma 2.3 we know that $A(t,\lambda) = e^{X(\lambda)} B(t,\lambda)$ and $A(t,\lambda) = e^{X^f(\lambda)} B^f(t,\lambda)$. The conclusion of Theorem 4.4 results by applying Proposition 4.1, Corollary 4.2, and Proposition 4.3. ■

From Theorem 4.4 it is easy to see that we obtain two possible types of tip motions for the modulated rotating waves $\mathbf{\Phi}(t, u_\lambda, \lambda)$ near a supercritical Hopf bifurcation with one parameter from a rotating wave on the sphere. They are illustrated in Figure 1 (see [7] for more details). These two types of tip motions for the bifurcating modulated rotating waves are quasi-periodically meandering of order $O(1)$, with the angle between primary frequency vector $\overrightarrow{X^f(\lambda)}$ and the axis of rotation $\overrightarrow{X_0}$ of the initial rotating wave of order $O(\lambda^{\frac{1}{2}})$, or slowly drifting of order $O(\lambda^{\frac{|k|}{2}})$ about the primary frequency vectors.



$|X_0| \notin \omega_H \mathbb{Z}$ $\quad$ $|X_0| = \omega_H$, $\overrightarrow{X(\lambda)} \perp \overrightarrow{X_0}$ $\quad$ $|X_0| = \omega_H$, $\overrightarrow{X(\lambda)}$ **is not** $\perp \overrightarrow{X_0}$

Figure 1. *Hopf bifurcation with one parameter* λ.

All the results presented in this section can be generalized for any compact Lie group $G \subset O(N)$. The main property used for $G = SO(3)$ is the existence of a sufficiently smooth function $BCH : g \times g \to \{U \in g : |U| < 2\pi\}$ such that $e^X e^Y = e^{BCH(X,Y)}$, where g is the Lie algebra of G. Moreover, for any Lie compact group G there is a series expansion for $BCH(X, Y)$ for X and Y close enough to O_3.

5. Drift phenomena and orthogonality for modulated rotating waves on the sphere. We consider the following reaction-diffusion system with parameters λ and μ:

$$u_t = D\Delta_S u + F(u, \lambda, \mu) \text{ on } r\mathbf{S^2}, \tag{5.1}$$

where the notation is the same as in section 2. Under the same framework and assumptions as the ones in section 2, the reduced differential equations on $SO(3) \times \mathbb{C}$ near a Hopf bifurcation are

$$\begin{aligned} \dot{A} &= AX_G(q, \lambda, \mu), \\ \dot{q} &= X_N(q, \lambda, \mu), \end{aligned} \tag{5.2}$$

where $X_G(0, 0, 0) = X_0$ and $X_N(0, 0, 0) = 0$. We suppose that $X_N(0, \lambda, \mu) = 0$ for $|\lambda|$ and $|\mu|$ small, and $D_q X_N(0, \lambda, \mu)$ has the eigenvalues $\alpha(\lambda, \mu) \pm i(\omega_H + \beta(\lambda, \mu))$ with $\alpha(0, 0) = \beta(0, 0) = 0$ and $\alpha_\lambda(0, 0) > 0$.

We have $\alpha(\lambda_H(\mu), \mu) = 0$ for some sufficiently smooth curve $\lambda = \lambda_H(\mu)$ with $\lambda_H(0) = 0$. This curve represents the Hopf points and we suppose that $\lambda_H(\mu) = 0$ for $|\mu|$ small.

Then, there exists a unique sufficiently CS branch of periodic solutions $q(t, \lambda, \mu)$ with period $\frac{2\pi}{|\omega_{\lambda,\mu}|}$ bifurcating from $q = 0$ that generically satisfies $q(t, \lambda, \mu) = \lambda^{\frac{1}{2}} r(t, \lambda, \mu)$. The secondary frequency is given by $\omega_{\lambda,\mu} = \omega_H + O(\mu) + \lambda s(\lambda, \mu)$.

Using [10, 28] and Proposition 4.1, there exist sufficiently CS branches $\mathbf{\Phi}(t, u_{\lambda,\mu}, \lambda, \mu)$ and $X(\lambda, \mu)$ such that for $|\mu|$ small

$$\mathbf{\Phi}(t, u_{0,\mu}, 0, \mu) = e^{X_G(0,0,\mu)t} u_{0,\mu} \text{ and } e^{X(0,\mu)\frac{2\pi}{|\omega_{0,\mu}|}} = e^{X_G(0,0,\mu)\frac{2\pi}{|\omega_{0,\mu}|}}, \tag{5.3}$$

and such that for $\lambda > 0$ small and $|\mu|$ small $\mathbf{\Phi}(t, u_{\lambda,\mu}, \lambda, \mu)$ has primary frequency vector $\overrightarrow{X(\lambda, \mu)}$.

Next, we give a more intuitive (from the computational point of view) proof for Proposition 3.4 in [31] for $G = SO(3)$, using the BCH formula in $so(3)$.

Theorem 5.1 (drift phenomena and orthogonality). *Suppose that all the assumptions made in this section hold and that $|X_0| = k\omega_H$ for some $k \in \mathbb{Z}$ with $k \neq 0$. Let*

$$X_G(q, \lambda, \mu) = x_0(q, \lambda, \mu)X_0^1 + x_1(q, \lambda, \mu)X_1 + x_2(q, \lambda, \mu)X_2. \tag{5.4}$$

If $(x_0)_\mu(0, 0, 0) \neq k(\omega_{0,\mu})_\mu|_{\mu=0}$, then there exists a sufficiently CS curve $\mu = \mu(\lambda)$ such that $\mu(0) = 0$ which gives a sufficiently CS branch of orbitally stable modulated rotating waves $\mathbf{\Phi}(t, u_{\lambda,\mu(\lambda)}, \lambda, \mu(\lambda))$ for $\lambda > 0$ small having the primary frequency vector $\overrightarrow{X(\lambda, \mu(\lambda))}$ orthogonal to $\overrightarrow{X_0}$.

Moreover, $X(\lambda, \mu(\lambda)) = O(\lambda^{\frac{|k|}{2}})X_1 + O(\lambda^{\frac{|k|}{2}})X_2$ for $\lambda \geq 0$ small.

Proof of Theorem 5.1. We recall that $T(\lambda,\mu)=\frac{2\pi}{|\omega_{\lambda,\mu}|}$. From (5.4) we have that

$$X_G(0,0,\mu)=a(\mu)X_0^1+b(\mu)X_1+c(\mu)X_2, \tag{5.5}$$

where we denoted $a(\mu)=x_0(0,0,\mu)=|X_0|+O(\mu)$, $b(\mu)=x_1(0,0,\mu)=O(\mu)$ and $c(\mu)=x_2(0,0,\mu)=O(\mu)$. From Lemma 2.3 we have $A(T(\lambda,\mu),\lambda,\mu)=e^{X(\lambda,\mu)T(\lambda,\mu)}$, where $X(\lambda,\mu)$ is given in Proposition 4.1.

Since $X(\lambda,\mu)$ is sufficiently CS and $X(0,0)=O_3$, we get

$$X(\lambda,\mu)=a_1(\lambda,\mu)X_0^1+b_1(\lambda,\mu)X_1+c_1(\lambda,\mu)X_2, \tag{5.6}$$

with a_1,b_1,c_1 such that $a_1(0,0)=0$ and $b_1(0,\mu)=O(\mu)$, $c_1(0,\mu)=O(\mu)$.

We have to find a unique sufficiently CS branch $\mu=\mu(\lambda)$ such that $\mu(0)=0$ and $a_1(\lambda,\mu(\lambda))=0$ for $\lambda\geq 0$ small. If $(a_1)_\mu(0,0)\neq 0$, then the implicit function theorem gives the existence of the required sufficiently CS branch. We show that $(a_1)_\mu(0,0)=a'(0)-k(\omega_{0,\mu})'_\mu|_{\mu=0}$.

From (5.3) we get $A(T(0,\mu),0,\mu)=e^{X(0,\mu)T(0,\mu)}=e^{X_G(0,0,\mu)T(0,\mu)}$. This implies

$$X(0,\mu)T(\mu,0)=X_G(0,0,\mu)T(\mu,0)+\frac{2l(\mu)\pi}{|X_G(0,0,\mu)|}X_G(0,0,\mu) \text{ for some } l(\mu)\in\mathbb{Z}.$$

Hence, (5.5) and (5.6) yield

$$\begin{aligned}&\left[a_1(0,\mu)X_0^1+b_1(0,\mu)X_1+c_1(0,\mu)X_2\right]T(0,\mu)\\&=\left[a(\mu)X_0^1+b(\mu)X_1+c(\mu)X_2\right]T(0,\mu)+\frac{2l(\mu)\pi}{\sqrt{a(\mu)^2+O(\mu^2)}}\left[a(\mu)X_0^1+b(\mu)X_1+c(\mu)X_2\right]\end{aligned}$$

for $|\mu|$ small. This implies

$$a_1(0,\mu)=a(\mu)+\frac{a(\mu)}{\sqrt{a(\mu)^2+O(\mu^2)}}l(\mu)\omega_{0,\mu}\cdot\operatorname{sgn}(\omega_H), \tag{5.7}$$

where $\operatorname{sgn}(x)$ stands for the sign of x.

For $\mu=0$ (5.7) gives $l(0)=-|k|$, since $|X_0|=k\omega_H>0$ implies $\operatorname{sgn}(k)=\operatorname{sgn}(\omega_H)$. Thus $k\cdot\operatorname{sgn}(\omega_H)=|k|$. Since $l(\mu)\in\mathbb{Z}$ is continuous, we get $l(\mu)=-|k|$ for $|\mu|$ small. Taking into account that $\operatorname{sgn}(\omega_H)=\operatorname{sgn}(k)$, (5.7) becomes

$$a_1(0,\mu)=a(\mu)-k\frac{a(\mu)}{\sqrt{a(\mu)^2+O(\mu^2)}}\omega_{0,\mu}. \tag{5.8}$$

We differentiate (5.8) to get $(a_1)_\mu(0,0)=a'(0)-k(\omega_{0,\mu})_\mu|_{\mu=0}$.

The scaling results from Proposition 4.1 and [31]. ∎

As it follows from the above proof, the branch $\mu=\mu(\lambda)$ for $\lambda\geq 0$ small can be found using the BCH formula in $so(3)$. The results presented in sections 4 and 5 allow us to explicitly control the direction of the primary frequency vectors and, therefore, the tip motions of the bifurcating modulated rotating waves $\mathbf{\Phi}(t,u_\lambda,\lambda)$.

6. Proof of Theorem 3.4 from section 3.

Proof of Theorem 3.4. We divide the proof into three parts. In the first two parts we construct the solution $Z(t,\lambda)$ on the interval $[0,T(\lambda)]$. In the third part, we extend it to $[0,\infty)$. We recall that $T(\lambda)=\frac{2\pi}{|\omega_\lambda|}$.

(1) *The reduced differential equations on so*(3). Consider the IVP

$$\begin{aligned}\dot{A} &= AX^G(t,\lambda),\\ A(0) &= I_3.\end{aligned}\tag{6.1}$$

We make the change of variable $A=e^Z$ near I_3 in the IVP (6.1). The main property used to get the reduced differential equations on $so(3)$ is from [24, 27]. It states that the exponential map exp of $SO(3)$ is a smooth function on $so(3)$ with differential given by

$$(d(\exp))_X(Y)=e^X\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}(\mathrm{ad}(X))^n(Y)\tag{6.2}$$

for any $X,Y\in so(3)$. Moreover, it is a local diffeomorphism near any $X\in so(3)$ iff the operator $\mathrm{ad}(X)$ has no eigenvalues of the form $2i\pi k$ with $k\neq 0$; i.e., $|X|\neq 2k\pi$ for $k\in\mathbb{Z}$ and $k\neq 0$. Using (6.2) the IVP (6.1) becomes successively

$$\begin{aligned}e^Z\textstyle\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}(\mathrm{ad}\,Z)^n\dot{Z} &= e^ZX^G(t,\lambda),\\ Z(0) &= O_3,\end{aligned}$$

or

$$\begin{aligned}\textstyle\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}(\mathrm{ad}\,Z)^n\dot{Z} &= X^G(t,\lambda),\\ Z(0) &= O_3,\end{aligned}$$

or

$$\begin{aligned}\overrightarrow{\textstyle\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}(\mathrm{ad}\,Z)^n\dot{Z}} &= \overrightarrow{X^G(t,\lambda)},\\ Z(0) &= O_3.\end{aligned}\tag{6.3}$$

The next important property used to get the reduced differential equations on $so(3)$ is from [27]. It states that for any $X,Y\in so(3)$ we have

$$\overrightarrow{(\mathrm{ad}(X))^nY}=X^n\overrightarrow{Y}\ \text{ for any integer } n\geq 1.\tag{6.4}$$

If we use (6.4) in (6.3), then we get

$$\begin{aligned}\textstyle\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}Z^n\overrightarrow{\dot{Z}} &= \overrightarrow{X^G(t,\lambda)},\\ Z(0) &= O_3.\end{aligned}\tag{6.5}$$

Moreover, in Appendix A we check that

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}Z^n=I_3+\frac{\cos|Z|-1}{|Z|^2}Z+\frac{|Z|-\sin|Z|}{|Z|^3}Z^2.\tag{6.6}$$

If we substitute (6.6) into (6.5), then we get

$$\begin{aligned}\left(I_3+\tfrac{\cos|Z|-1}{|Z|^2}Z+\tfrac{|Z|-\sin|Z|}{|Z|^3}Z^2\right)\overrightarrow{\dot{Z}} &= \overrightarrow{X^G(t,\lambda)},\\ Z(0) &= O_3.\end{aligned}\tag{6.7}$$

In Appendix A we show that for any $|Z|<2\pi$

$$\left(\frac{\cos|Z|-1}{|Z|^2}Z+\frac{|Z|-\sin|Z|}{|Z|^3}Z^2\right)^{-1}=I_3+\frac{1}{2}Z+\left(1-\frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{Z^2}{|Z|^2}.\tag{6.8}$$

Since we are looking for $|Z|\le\pi<2\pi$, the IVP (6.7) becomes

$$\begin{aligned}\overrightarrow{\dot{Z}} &= \left[I_3+\tfrac{1}{2}Z+\left(1-\tfrac{|Z|}{2}\cot\tfrac{|Z|}{2}\right)\tfrac{Z^2}{|Z|^2}\right]\overrightarrow{X^G(t,\lambda)},\\ Z(0) &= O_3,\end{aligned}\tag{6.9}$$

where we are looking for a solution Z such that $|Z|\le\pi$. If we denote $f(a)=1-\frac{a}{2}\cot\frac{a}{2}$ if $a\neq 0$ and $f(0)=0$, then we get the IVP (3.5) from the statement of Theorem 3.4. It remains to construct a sufficiently CS solution $Z(t,\lambda)$ of the IVP (3.5) such that $|Z(t,\lambda)|<2\pi$; i.e., we have to consider the differential equations in IVP (3.5) mod 2π.

(2) *The solution $Z(t,\lambda)$ on $[0,T(\lambda)]$.* The main property used for now is the boundedness of the right-hand side of the IVP (6.9) with respect to $|Z|$.

We recall that the sufficiently smooth function BCH is given in Definition 3.1. We check that we can apply Lemma 3.3 for the IVP (6.9). Let $M>0$ be such that $\left|X^G(t,\lambda)\right|<M$ for $t\in[0,\infty)$ and $\lambda\ge 0$ small. For $\lambda\ge 0$ small and any $|Z|\le\pi$, we have successively

$$\begin{aligned}&\left\|\left[I_3+\frac{1}{2}Z+\left(1-\frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{Z^2}{|Z|^2}\right]\overrightarrow{X^G(t,\lambda)}\right\|\\ &\le\left[\|I_3\|+\frac{1}{2}\|Z\|+\left(1+\frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{\|Z\|^2}{|Z|^2}\right]M,\\ \text{or }&\left\|\left[I_3+\frac{1}{2}Z+\left(1-\frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{Z^2}{|Z|^2}\right]\overrightarrow{X^G(t,\lambda)}\right\|\\ &\le M\left[\sqrt{3}+\frac{1}{2}\sqrt{2}\,|Z|+\left(1+\frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{2\,|Z|^2}{|Z|^2}\right],\end{aligned}$$

or, because the function $x\to x\cot x$ is decreasing on $[0,\frac{\pi}{2}]$,

$$\begin{aligned}&\left\|\left[I_3+\frac{1}{2}Z+\left(1-\frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{Z^2}{|Z|^2}\right]\overrightarrow{X^G(t,\lambda)}\right\|\\ &\le M\left[\sqrt{3}+\frac{1}{2}\sqrt{2}\,|Z|+2(1+1)\right]\le M(6+|Z|).\end{aligned}$$

We apply Lemma 3.3 with $M_g=6$, $K_g=1$, and $C=\pi$ and we get $A(t,\lambda)=e^{Z_1(t,\lambda)}$ for any $t\in[0,\frac{T(\lambda)}{n}]$ and $\lambda\ge 0$ small, where $Z_1(t,\lambda)$ is the unique sufficiently CS solution of (6.9). We denote $t_\lambda=\frac{T(\lambda)}{n}$ from now on.

We make the change of variable $B = (A(t_\lambda, \lambda))^{-1}A$. We get that the solution of the IVP (6.1) is given by

$$A(t, \lambda) = e^{Z_1(t_\lambda, \lambda)}B(t, \lambda),$$

where $B(t, \lambda)$ is the solution of the IVP

$$\begin{aligned} \dot{B} &= BX^G(t, \lambda), \\ B(t_\lambda) &= I_3. \end{aligned} \tag{6.10}$$

We look for a solution $B = e^Z$. Therefore, (6.10) becomes

$$\begin{aligned} \overrightarrow{\dot{Z}} &= \left[I_3 + \tfrac{1}{2}Z + \left(1 - \tfrac{|Z|}{2}\cot\tfrac{|Z|}{2}\right)\tfrac{Z^2}{|Z|^2}\right]\overrightarrow{X^G(t, \lambda)}, \\ Z(t_\lambda) &= O_3. \end{aligned} \tag{6.11}$$

The IVP

$$\begin{aligned} \dot{a} &= M(6 + a), \\ a(t_\lambda) &= 0 \end{aligned} \tag{6.12}$$

has the solution satisfying $a(t) \leq \pi$ for any $t \in [t_\lambda, 2t_\lambda]$. Using the same argument as the one in the proof of Lemma 3.3, we get $B(t, \lambda) = e^{Z_2(t,\lambda)}$ for any $t \in [t_\lambda, 2t_\lambda]$ and $\lambda \geq 0$ small, where $Z_2(t, \lambda)$ is the unique sufficiently CS solution of (6.11).

If we continue this, then we get

$$A(t, \lambda) = e^{Z(t,\lambda)},$$

where $Z(t, \lambda)$ is given by

$$e^{Z(t,\lambda)} = e^{Z_1(t_\lambda,\lambda)}e^{Z_2(2t_\lambda,\lambda)} \dots e^{Z_i(t,\lambda)}, \tag{6.13}$$

where Z_i is the solution of the following IVP on the interval $[(i-1)t_\lambda, it_\lambda]$ for any $\lambda \geq 0$ small and for $i = 1, 2, \dots, n$:

$$\begin{aligned} \overrightarrow{\dot{Z}} &= \left[I_3 + \tfrac{1}{2}Z + \left(1 - \tfrac{|Z|}{2}\cot\tfrac{|Z|}{2}\right)\tfrac{Z^2}{|Z|^2}\right]\overrightarrow{X^G(t, \lambda)}, \\ Z((i-1)t_\lambda) &= O_3. \end{aligned} \tag{6.14}$$

From (6.13) we get

$$[Z(t, \lambda)] = BCH(Z_1(t_\lambda, \lambda), Z_2(2t_\lambda, \lambda), \dots, Z_i(it_\lambda, \lambda)) \tag{6.15}$$

for any $t \in [(i-1)t_\lambda, it_\lambda]$ and $\lambda \geq 0$ small, for $i = 1, 2, \dots, n$, where Z_i is the solution of the IVP (6.14). From (6.15) it is clear that $[Z(t, \lambda)]$ is sufficiently CS on $[0, T(\lambda)]$ and $\lambda \geq 0$ small.

(3) *The solution $Z(t, \lambda)$ on $[0, \infty)$.* Since

$$A(t + T(\lambda), \lambda) = e^{X(\lambda)T(\lambda)}A(t, \lambda) = A(T(\lambda), \lambda)A(t, \lambda) = e^{Z(T(\lambda),\lambda)}e^{Z(t,\lambda)},$$

it is easy to see that we can define

$$[Z(t,\lambda)] = BCH(Z(T(\lambda),\lambda), Z(t-T(\lambda),\lambda)) \text{ for } t \in [T(\lambda), 2T(\lambda)]. \tag{6.16}$$

It is clear that $[Z(t,\lambda)]$ is sufficiently CS for any $t \in [T(\lambda), 2T(\lambda)]$ and for $\lambda \geq 0$ small. We repeat this argument to get a sufficiently CS function $Z(t,\lambda)$ defined on $[0,\infty)$ such that $A(t,\lambda) = e^{Z(t,\lambda)}$, where $A(t,\lambda)$ is the solution to IVP (6.1).

We define the smooth function

$$q([\overrightarrow{Y}]) = \begin{cases} Y & \text{if } \frac{1}{|Y|}\overrightarrow{Y},\ \frac{1}{|X_0|}\overrightarrow{X_0} \text{ are in the same hemisphere or } Y = O_3, \\ (-\frac{2\pi}{|Y|}+1)Y & \text{if } \frac{1}{|Y|}\overrightarrow{Y},\ \frac{1}{|X_0|}\overrightarrow{X_0} \text{ are in different hemispheres and } Y \neq O_3, \end{cases} \tag{6.17}$$

where points 0 and 2π are identified, and $[\overrightarrow{Y}]$ is an equivalence class of D. Then, $e^{q([Y])} = e^{[Y]}$ for any $[Y] \in D$. If we define $Z^*(t,\lambda) = q([Z(t,\lambda)])$, then $Z^*(t,\lambda)$ is sufficiently CS and $A(t,\lambda) = e^{Z(t,\lambda)} = e^{[Z(t,\lambda)]} = e^{Z^*(t,\lambda)}$. Thus, $Z^*(t,\lambda)$ is the solution of the IVP (3.5). We relabel $Z^*(t,\lambda)$ to $Z(t,\lambda)$.

We discuss the change that needs to be done in order to get the generalization to any compact Lie group $G \subset O(N)$. We make the change of variable $C = e^{-X_G(0,\lambda)t}A$. Then, the IVP (6.1) becomes $\dot{C} = -X_G(0,\lambda)C + CX^G(t,\lambda)$ and $C(0) = I_N$, where I_N is the identity matrix. Next, we use the fact that the exponential map exp of G is a smooth function on g with differential given by (6.2) (see [24, 27]). We make the change of variable $C = e^Z$ in the new IVP and we end by looking for solutions $|Z(t,\lambda)| = O(\lambda^{\frac{1}{2}})$ to $\dot{Z} = h(Z,t,\lambda)$ and $Z(0) = O_N$, where $h(t,Z,\lambda)$ is bounded and O_N denotes the zero matrix. ■

Appendix A. We now show that (6.8) from section 6 holds. It is enough to check that

$$\left(I_3 + \frac{\cos|Z|-1}{|Z|^2}Z + \frac{|Z|-\sin|Z|}{|Z|^3}Z^2\right)\left[I_3 + \frac{1}{2}Z + \left(1 - \frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{Z^2}{|Z|^2}\right] = I_3.$$

From [27] we know that for any $X \in so(3)$,

$$\begin{aligned} X^{2n} &= (-1)^{n-1}|X|^{2(n-1)}X^2 \text{ for any } n \geq 1, \\ X^{2n+1} &= (-1)^n|X|^{2n}X \text{ for any } n \geq 0. \end{aligned} \tag{A.1}$$

Using $Z^3 = -|Z|^2 Z$ and $Z^4 = -|Z|^2 Z^2$, it yields

$$\begin{aligned} &\left(I_3 + \frac{\cos|Z|-1}{|Z|^2}Z + \frac{|Z|-\sin|Z|}{|Z|^3}Z^2\right)\left[I_3 + \frac{1}{2}Z + \left(1 - \frac{|Z|}{2}\cot\frac{|Z|}{2}\right)\frac{Z^2}{|Z|^2}\right] \\ &= \left(I_3 + \frac{\cos|Z|-1}{|Z|^2}Z + \frac{|Z|-\sin|Z|}{|Z|^3}Z^2\right)\left[I_3 + \frac{1}{2}Z + \left(\frac{1}{|Z|^2} - \frac{\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}|Z|}\right)Z^2\right] \\ &= I_3 + \frac{1}{2}Z + \left(\frac{1}{|Z|^2} - \frac{\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}|Z|}\right)Z^2 + \frac{\cos|Z|-1}{2|Z|^2}Z^2 \\ &+ \frac{\cos|Z|-1}{|Z|^2}Z + \frac{\cos|Z|-1}{|Z|^2}\left(\frac{1}{|Z|^2} - \frac{\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}|Z|}\right)Z^3 + \frac{|Z|-\sin|Z|}{|Z|^3}Z^2 + \frac{|Z|-\sin|Z|}{2|Z|^3}Z^3 \end{aligned}$$

$$
\begin{aligned}
&+\frac{|Z|-\sin|Z|}{|Z|^3}\left(\frac{1}{|Z|^2}-\frac{\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}|Z|}\right)Z^4\\
&=I_3+\left[\frac{1}{2}+\frac{\cos|Z|-1}{|Z|^2}-\frac{\cos|Z|-1}{|Z|^2}+\frac{\cos|Z|-1}{|Z|^2}\frac{|Z|\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}}-\frac{1}{2}+\frac{\sin|Z|}{2|Z|}\right]Z\\
&\quad+\left[\frac{1}{|Z|^2}-\frac{\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}|Z|}+\frac{\cos|Z|-1}{2|Z|^2}+\frac{|Z|-\sin|Z|}{|Z|^3}-\frac{|Z|-\sin|Z|}{|Z|}\frac{1}{|Z|^2}\right.\\
&\qquad\left.+\frac{|Z|-\sin|Z|}{|Z|}\frac{\cos\frac{|Z|}{2}}{2\sin\frac{|Z|}{2}|Z|}\right]Z^2\\
&=I_3+\left[\frac{-\sin^2\frac{|Z|}{2}}{|Z|}\frac{\cos\frac{|Z|}{2}}{\sin\frac{|Z|}{2}}+\frac{\sin|Z|}{2|Z|}\right]Z+\left[\frac{\cos^2\frac{|Z|}{2}}{|Z|^2}-\frac{2\sin\frac{|Z|}{2}\cos\frac{|Z|}{2}}{2|Z|^2}\frac{\cos\frac{|Z|}{2}}{\sin\frac{|Z|}{2}}\right]Z^2=I_3.
\end{aligned}
$$

The proof of (6.6) follows using (A.1):

$$
\begin{aligned}
\sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}Z^n&=I_3+\sum_{k=1}^{\infty}\frac{(-1)^{2k}}{(2k+1)!}Z^{2k}+\sum_{k=0}^{\infty}\frac{(-1)^{2k+1}}{(2k+2)!}Z^{2k+1}\\
&=I_3+\left(-\frac{1}{2!}+\frac{1}{4!}|Z|^2-\frac{1}{6!}|Z|^4+\dots\right)Z\\
&\quad+\left(\frac{1}{3!}-\frac{1}{5!}|Z|^2+\frac{1}{7!}|Z|^4-\dots\right)Z^2\\
&=I_3+\frac{\cos|Z|-1}{|Z|^2}Z+\frac{|Z|-\sin|Z|}{|Z|^3}Z^2,
\end{aligned}
$$

where we used the Taylor expansions for sin and cos.

Acknowledgments. I wish to thank to my Ph.D. supervisor, Victor G. LeBlanc, for his help in writing this article. I am also grateful to the Natural Sciences and Engineering Research Council of Canada for a PGS-B Scholarship. I would like to thank the referees for their helpful comments.

REFERENCES

[1] F. Amjadi, *Spiral waves on static and moving spherical domains*, J. Comput. Appl. Math., 182 (2005), pp. 472–486.
[2] D. Barkley, *Linear stability analysis of rotating spiral waves in excitable media*, Phys. Rev. Lett., 68 (1992), pp. 2090–2093.
[3] D. Barkley, *Euclidean symmetry and the dynamics of rotating waves*, Phys. Rev. Lett., 72 (1994), pp. 164–167.
[4] T. Bröcker and T. tom Dieck, *Representations of Compact Lie Groups*, Springer-Verlag, New York, 1985.
[5] D. Chan, *Hopf bifurcations from relative equilibria with spherical geometry*, J. Differential Equations, 226 (2006), pp. 118–134.
[6] F. Chavez, R. Kapral, G. Rousseau, and L. Glass, *Scroll waves in spherical shell geometries*, Chaos, 11 (2001), pp. 757–765.

[7] A. N. Comanici, *Spiral Waves on Spherical Domains: A Dynamical Systems Approach*, Ph.D. thesis, University of Ottawa, Ottawa, Canada, 2004.
[8] J. M. Davidenko, A. V. Pertsov, R. Salomonsz, W. Baxter, and J. Jalife, *Stationary and drifting spiral waves of excitation in isolated cardiac muscle*, Nature, 355 (1992), pp. 349–351.
[9] K. Engo, *On the BCH-formula in so*(3), BIT, 41 (2001), pp. 629–632.
[10] B. Fiedler, B. Sandstede, A. Scheel, and C. Wulff, *Bifurcation from relative equilibria of non-compact group actions: Skew products, meanders, and drifts*, Doc. Math., 1 (1996), pp. 479–505.
[11] B. Fiedler and D. Turaev, *Normal forms, resonances, and meandering tip positions near relative equilibria of Euclidean group actions*, Arch. Ration. Mech. Anal., 145 (1998), pp. 129–159.
[12] M. Field, *Local structure of equivariant dynamics*, in Singularity Theory and Its Applications Warwick Part II, Lecture Notes in Math. 1463, Springer-Verlag, Berlin, 1991, pp. 142–166.
[13] G. Gerisch, *Periodic signals control the formation of spatial patterns in cell aggregates*, Naturwissenschaften, 58 (1971), p. 420.
[14] M. Golubitsky, V. G. LeBlanc, and I. Melbourne, *Meandering of the spiral tip: An alternative approach*, J. Nonlinear Sci., 7 (1997), pp. 557–586.
[15] J. Gomatam and F. Amdjadi, *Reaction-diffusion on a sphere: Meandering of spiral waves*, Phys. Rev. E (3), 56 (1997), pp. 3913–3919.
[16] J. K. Hale, *Ordinary Differential Equations*, Wiley, New York, 1969.
[17] W. Jahnke, W. E. Skaggs, and A. T. Winfree, *Chemical vortex dynamics in the Belousov-Zhabotinsky reaction and in the two-variable Oregonator model*, J. Chem. Phys., 93 (1989), pp. 740–749.
[18] M. Krupa, *Bifurcations of relative equilibria*, SIAM J. Math. Anal., 21 (1990), pp. 1453–1486.
[19] R. Kapral and K. Showalter, *Chemical Waves and Patterns*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.
[20] J. Maselko, *Symmetrical double rotor spiral waves on spherical surfaces*, J. Chem. Soc., Faraday Trans., 94 (1998), pp. 2343–2345.
[21] J. Maselko and K. Showalter, *Chemical waves on spherical surfaces*, Letters to Nature, 339 (1989), pp. 609–611.
[22] P. McQuillan and J. Gomatam, *Rotating chemical waves on the sphere*, J. Phys. Chem., 100 (1996), pp. 5157–5159.
[23] S. Nettesheim, A. von Oertzen, H. H. Rotermund, and G. Ertl, *Reaction diffusion patterns in the catalytic CO-oxidation on Pt*(110)*-front propagation and spiral waves*, J. Chem. Phys., 98 (1993), pp. 9977–9985.
[24] A. L. Onischchik, *Lie Groups and Lie Algebras, Vol.* I, Springer-Verlag, Berlin, 1993.
[25] B. B. Plapp and E. Bodenshatz, *Core dynamics of multi-armed spirals in Rayleigh-Benard convection*, Phys. Scripta, T67 (1996), pp. 111–116.
[26] M. Renardy, *Bifurcation from rotating waves*, Arch. Ration. Mech. Anal., 79 (1982), pp. 49–84.
[27] W. Rossmann, *Lie Groups: An Introduction through Linear Groups*, Oxford University Press, Oxford, UK, 2002.
[28] B. Sandstede, A. Scheel, and C. Wulff, *Dynamics of spiral waves on unbounded domains using center-manifold reductions*, J. Differential Equations, 141 (1997), pp. 122–149.
[29] V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations*, Prentice–Hall, Englewood Cliffs, NJ, 1974.
[30] C. Wulff, *Theory of Meandering and Drifting Spiral Waves in Reaction-Diffusion Systems*, Ph.D. thesis, Freie Universität, Berlin, Germany, 1996.
[31] C. Wulff, *Transitions from relative equilibria to relative periodic orbits*, Doc. Math., 5 (2000), pp. 227–274.
[32] H. Yagisita, M. Mimura, and M. Yamada, *Spiral wave behaviors in an excitable reaction-diffusion system on a sphere*, Phys. D, 124 (1998), pp. 126–136.
[33] V. S. Zykov and S. C. Müller, *Spiral waves on circular and spherical domains of excitable medium*, Phys. D, 97 (1996), pp. 322–332.
[34] V. S. Zykov, A. S. Mikailov, and S. C. Müller, *Controlling spiral waves in confined geometries by global feedback*, Phys. Rev. Lett., 78 (1997), pp. 3398–3401.

SIAM J. Applied Dynamical Systems
Vol. 5, No. 4, pp. 783–807

Wave Radiation by Balanced Motion in a Simple Model*

J. Vanneste†

Abstract. We introduce and study a toy model which captures some essential features of wave radiation by slow (or balanced) motion in the atmosphere and the ocean. Inspired by the widely studied five-component model due to Lorenz, the model describes the coupling of a nonlinear pendulum with linear waves. The waves obey a one-dimensional linear Klein–Gordon equation, so their dispersion relation is identical to that of inertia-gravity waves in a rotating shallow-water fluid. The model is Hamiltonian. We examine two physically relevant asymptotic regimes in which there is some time-scale separation between the slow pendulum motion and the fast waves: in regime (i), the time-scale separation breaks down for waves with asymptotically large wavelengths; in regime (ii), the time-scale separation holds for all wavelengths. We study the generation of waves in each regime using distinct asymptotic methods. In regime (i), long waves are excited resonantly in a manner that is analogous to the Lighthill radiation of sound waves in weakly compressible flows, and to the radiation of gravitational waves by slow mass motion in general relativity. Matched asymptotics provides the functional form of the waves radiated, and leads, at higher order, to a closed model describing the pendulum dynamics while accounting for the dissipative effect of wave radiation. In regime (ii), an exponentially accurate slow manifold can be defined, and the waves radiated are exponentially small. They are captured using an exponential-asymptotic technique combining complex-time matching with Borel summation. The asymptotic results obtained in each regime are tested against numerical simulations of the model.

Key words. slow manifold, wave radiation, inertia-gravity wave, exponential asymptotics

AMS subject classifications. 37N10, 76B15, 76U05, 37K05

DOI. 10/1137/050644185

1. Introduction.

Systems with widely separated time scales abound, and numerous mathematical techniques have been devised to take advantage of their time-scale separation. In many such systems, the fast degrees of freedom are only weakly excited; it is then natural to attempt to eliminate them [33] by reducing the dynamics to a slow manifold, that is, to a submanifold of the state space which is nearly invariant and on which the dynamics is slow. We refer the reader to the recent paper by MacKay [24] for a comprehensive discussion of the concept of slow manifold and for several examples of applications.

The particular application which motivates the present paper is provided by geophysical fluid dynamics. The dynamics of the atmosphere and the ocean at midlatitudes is dominated by the large-scale, slow motion usually referred to as "balanced motion," but much faster motion in the form of inertia-gravity waves is also possible. (The even faster sound waves are generally filtered out at the outset by using incompressible, hydrostatic, or anelastic fluid

*Received by the editors November 3, 2005; accepted for publication (in revised form) by T. Kaper September 11, 2006; published electronically December 26, 2006. This research was supported by a NERC Advanced Research Fellowship.

http://www.siam.org/journals/siads/5-4/64418.html

†School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JZ, UK (J.Vanneste@ed. ac.uk).

models.) The time-scale separation between the two types of motion is large, with typical time scales of the order of a few days or weeks for the balanced motion in the atmosphere or the ocean, respectively, and inertia-gravity wave periods of the order of a few minutes. This has led to the development of a variety of "balanced models" describing the reduced dynamics on a slow manifold (see, e.g., [36, 5, 25] and references therein), the simplest of which is the well-known quasi-geostrophic model. Although balanced models are today mostly theoreticians' tools, the concept of slow manifold is used in weather forecasting in the process of initialization [9]: initial data are prepared by projection onto a slow manifold to reduce the level of (mostly spurious) inertia-gravity wave activity.

The reliance on balanced models has led many researchers to investigate the fundamental limitations of the concepts of slow manifold and balance. This has been largely carried out using low-order models consisting of a few ordinary differential equations (ODEs), typically derived from the fluid equations by spectral expansion and severe truncation. The most widely studied among these is the five-component model due to Lorenz [21], also referred to as the Lorenz–Krishnamurthy (LK) model [22]. It can be reduced to four ODEs which describe the dynamics of a nonlinear pendulum, representing slow balanced motion, coupled to a stiff spring, representing the fast waves [8, 6]. Another, essentially equivalent, model is the swinging spring [23], or spring pendulum [24].

For these ODE models, the status of the slow manifold is now well understood. In the absence of dissipation, thought to be negligible in the geophysical context, the slow manifolds are elliptic and in general not invariant [14, 24]. However, a systematic improvement procedure provides slow manifolds that are invariant up to an $O(\epsilon^N)$ error, where $\epsilon \ll 1$ is the ratio between the slow and fast time scales, for any $N \geq 1$. An optimal choice of N then leads to an exponentially small error, under an assumption of analyticity [14, 24, 38]. The physical implications are clear: regardless of how well prepared the initial data are, the generation of fast oscillations is unavoidable. For initial data lying on an optimal slow manifold, these oscillations are very weak and are exponentially small in ϵ. In the atmospheric context, this provides a mechanism for the generation of inertia-gravity waves which is often referred to as "spontaneous" generation, to emphasize the difference from the generation that results from the adjustment of poorly prepared initial data (see, e.g., [29] for a recent analysis).

In spite of the fact that their amplitude is beyond all orders in ϵ, the fast oscillations generated spontaneously can be studied perturbatively, using the techniques of exponential asymptotics [32]. Such a study reveals the mechanism of generation to be an instance of the Stokes phenomenon (see, e.g., [3, 28]) and provides explicit estimates for the wave amplitudes. Results of this type have been obtained in [34] for the LK model and in [35, 27] for particular solutions of the fluid equations that are also governed by ODEs.

Low-order models such as the LK model have proved very useful for understanding the rather subtle questions raised by the concepts of slow manifold and balance. However, the drastic simplification entailed by the reduction from partial differential equations (PDEs) to ODEs means that a number of issues cannot be addressed using these models. To examine some of these issues, it is therefore useful to introduce a new simplified model, in the spirit of the LK and swinging-spring models, but retaining a PDE component. This is the purpose of this paper.

In fluids, the fast oscillations are propagating waves, with frequencies that depend on

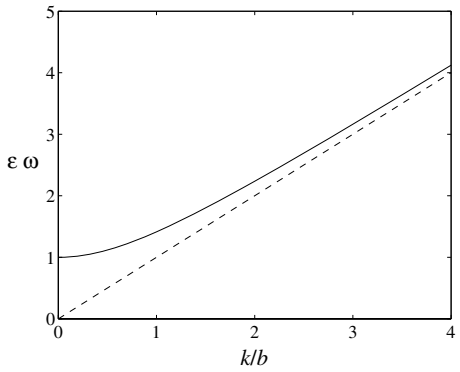


Figure 1. *Dispersion relation $\epsilon\omega = (1 + k^2/b^2)^{1/2}$ for shallow-water inertia-gravity waves (solid line). The nondispersive limit, valid for $k \gg b$, is also shown (dashed line).*

wavenumber. By fixing the wavenumbers involved, the derivation of LK-type models suppresses the possibility of interactions between very different wavenumbers. This possibility, however, is at the heart of one mechanism of wave generation which appears in several physical systems: gravitational waves in general relativity [10], [18, section 110], sound waves in weakly compressible fluids [20], [19, section 75], and, in the geophysical context, inertia-gravity waves in rotating shallow water [11, 12]. In these systems, the wave frequencies decrease with wavenumbers in such a manner that there always are some resonant interactions between the slow motion and waves of sufficiently large scales. The wave generation is then relatively inefficient—because of the mismatch between the spatial scales of the slow motion and waves—but nevertheless scales like some power of the relevant small parameter rather than exponentially.

For the rotating shallow-water model and for more realistic models of geophysical flows, this resonant mechanism of wave generation (which we will refer to as Lighthill radiation following [11, 12] and the sound-wave analogy) or the nonresonant mechanism captured in the LK and other ODE models may be relevant, depending on the flow regime. To see why, consider the dispersion relation

$$\omega^2 = \epsilon^{-2}(1 + k^2/b^2) \tag{1.1}$$

of shallow-water inertia-gravity waves, displayed in Figure 1. Here, ω is the frequency and k the wavenumber, and both are nondimensionalized using the characteristic frequency U/L and scale L of the balanced motion. There are two independent parameters: the Rossby number, $\epsilon = U/(fL)$, where f is the Coriolis parameter (measuring the earth's rotation rate), and the rotational Froude number $b = fL/(gh)^{1/2}$, where g is the earth's gravity and h the fluid depth. Two main asymptotic regimes are thought to be relevant: (i) the small-Froude-number regime, with $b \ll 1$ and $\epsilon = O(1)$, and (ii) the small-Rossby-number (or quasi-geostrophic) regime, with $\epsilon \ll 1$ and $b = O(1)$. In regime (i), there is no time-scale separation between (slow) balanced motion and long waves (with $O(b)$ wavenumbers), and Lighthill radiation occurs. In regime (ii), on the other hand, there is a time-scale separation between balanced motion and waves for all wavenumbers, since the wave frequency is bounded from below by

$\epsilon^{-1} \gg 1$. Thus Lighthill radiation cannot occur, and one can expect exponentially small wave radiation of the type studied in the LK model [12, 31, 13].

The main advantage of the model that we introduce in this paper is that it makes it possible to analyze both regimes (i) and (ii) and, correspondingly, both types of wave generation in as simple a set-up as possible. The rotating shallow water may seem to be suitable for such an analysis: indeed, Ford, McIntyre, and Norton [12] succeeded in capturing the Lighthill radiation and their feedback in regime (i) using matched asymptotics. This, however, requires a large amount of algebra which might deter many readers. Worse still, the asymptotic treatment of regime (ii) seems hopeless in the absence of a well-developed theory of exponential asymptotics for PDEs. By contrast, our model can be analyzed in regimes (i) and (ii) by relatively simple means.

Another advantage of our model compared to low-order models is that, by keeping a PDE component, it introduces the possibility of wave radiation to infinity. Thus, the waves move away from their region of generation, thereby providing a source of dissipation for the balanced motion. This is probably a good approximation for what is happening in the atmosphere and ocean where the waves can escape before being ultimately damped by breaking or viscous dissipation. Our model can thus be used to examine how efficient wave generation and radiation can be as mechanisms for the dissipation of the energy of the balanced motion. This is an issue of current interest in oceanography (see, e.g., [26]). It would be best studied by adding some forcing (perhaps random) to the model so that the properties of the statistical equilibrium arising from the balance between forcing and wave radiation can be established. In this paper, however, we limit our considerations to the unforced version of the model.

The new model is introduced in section 2. It is a simple modification of the LK model in which the linear oscillator described by the fast variables is replaced by a linear, one-dimensional Klein–Gordon equation [37, Chapter 11] with dispersion relation (1.1). Thus, the slow component of the model remains governed by ODEs, but the fast component is governed by PDEs. The coupling between the spatially dependent (fast) variables and the spatially independent (slow) variables is through an arbitrary localized shape function, which we take to be the derivative of a Gaussian. This has zero average and is odd so that, by symmetry, the spatial dynamics can be reduced to the half line $\mathbb{R}_+$. The model is Hamiltonian. It is defined by three parameters: ϵ and b, which appear in the dispersion relation (1.1), and the amplitude a of the shape function. The model is not derived in any way from the fluid equations, nor does it obviously represent any simple mechanical device. This is not a significant drawback, however. What is important for our purpose is that the parameters ϵ and b play the same role as they do for the rotating shallow-water model. Because of this, we refer to them as the Rossby number and Froude number, respectively.

After introducing the model, we discuss the asymptotic behavior of its solutions. Section 3 is devoted to the small-Froude-number regime $b \ll 1$, $\epsilon = O(1)$. As mentioned above, this is the regime where Lighthill radiation occurs. Using matched asymptotics, we obtain an approximation for the waves generated spontaneously by the balanced motion. We further derive a reduced model, which describes the evolution of the slow variables while accounting for the energy loss due to wave radiation. We term this model "postbalanced" by analogy with the post-Newtonian models used in general relativity to describe gravitational-wave radiation and its feedback on compact sources (see, e.g., [4] for a review). Our postbalanced model can

then be seen as a toy version of the one derived by Ford, McIntyre, and Norton [12] for the rotating shallow-water equations.

Section 4 is devoted to the small-Rossby-number regime $\epsilon \ll 1$ and $b = O(1)$. This regime is similar to that studied in low-order models in that approximately invariant slow manifolds can be defined to arbitrary order $O(\epsilon^N)$, and wave radiation is exponentially weak. We estimate the amplitude of the waves radiated using exponential asymptotics. This shows, in particular, that the waves are near inertial, that is, have frequencies close to ϵ^{-1}, with large spatial scales of the order of $\epsilon^{-1/2}$. The asymptotic results of both sections 3 and 4 are compared with numerical simulations. The numerical formulation, which implements nonreflecting boundary conditions, is described in Appendix A. The paper concludes with a discussion in section 5.

2. Model.

The LK model [21, 22], obtained by truncation of a spectral expansion of the rotating shallow-water equations on the plane, can be written as the system of five ODEs

$$\dot{u}_{\mathrm{L}} = -v_{\mathrm{L}} w_{\mathrm{L}} + b v_{\mathrm{L}} y_{\mathrm{L}}, \tag{2.1}$$

$$\dot{v}_{\mathrm{L}} = w_{\mathrm{L}} u_{\mathrm{L}} - b u_{\mathrm{L}} y_{\mathrm{L}}, \tag{2.2}$$

$$\dot{w}_{\mathrm{L}} = -u_{\mathrm{L}} v_{\mathrm{L}}, \tag{2.3}$$

$$\delta \dot{x}_{\mathrm{L}} = -y_{\mathrm{L}}, \tag{2.4}$$

$$\delta \dot{y}_{\mathrm{L}} = x_{\mathrm{L}} + b\delta u_{\mathrm{L}} v_{\mathrm{L}}. \tag{2.5}$$

Here, the small parameter is δ; it is related to the Rossby and Froude numbers by $\delta = \epsilon b/(1+b^2)^{1/2}$ so that both the small-Froude-number and small-Rossby-number regimes lead to $\delta \ll 1$. The slow variables $(u_{\mathrm{L}}, v_{\mathrm{L}}, w_{\mathrm{L}})$ describe the evolution of a rigid body or, after reduction using the constancy of $u_{\mathrm{L}}^2 + v_{\mathrm{L}}^2$, of a pendulum with $O(1)$ frequency. The fast variables $(x_{\mathrm{L}}, y_{\mathrm{L}})$ describe a linear oscillator with frequency ϵ^{-1} [8, 6].

We propose the following modification of the LK model. The three slow variables, which we denote by (u, v, w), remain functions of t only, but the fast variables, denoted by (x, y), are functions of t and of a spatial coordinate $s \in \mathbb{R}$. Choosing some localized function $f(s)$ (e.g., Gaussian or compactly supported), we write the new model as the mixed ODE–PDE system

$$\dot{u} = -vw + v \int f(s) y(s,t)\, \mathrm{d}s, \tag{2.6}$$

$$\dot{v} = wu - u \int f(s) y(s,t)\, \mathrm{d}s, \tag{2.7}$$

$$\dot{w} = -uv, \tag{2.8}$$

$$\epsilon x_t = -y, \tag{2.9}$$

$$\epsilon y_t = x - x_{ss}/b^2 + \epsilon f(s) uv, \tag{2.10}$$

where either ϵ or b are now the small parameters. Note that we use both overdots and subscripts t to denote time derivatives in order to distinguish between ODEs and PDEs. In (2.6)–(2.7) and in what follows, unspecified limits of integrations are $(-\infty, \infty)$. As announced, in the linear approximation, the fast variables (x, y) satisfy a Klein–Gordon equation [37,

Chapter 11]. As boundary conditions for these variables, we impose outward radiation as $s \to \pm\infty$.

In what follows, we make the choice

$$f(s) = \frac{a}{(2\pi)^{1/2}} \frac{\mathrm{d}}{\mathrm{d}s} \mathrm{e}^{-s^2/2} = \frac{-as\,\mathrm{e}^{-s^2/2}}{(2\pi)^{1/2}},$$

where a is a fixed amplitude. Three qualitative properties of this function matter: the rapid decay as $|s| \to \infty$, the vanishing of its zeroth moment, and the nonvanishing of its first moment. Our results would be qualitatively the same for other choices of $f(s)$ satisfying these properties. The oddness of $f(s)$ is inessential but convenient since it is inherited by $x(s,t)$ and $y(s,t)$ and allows computations to be limited to the half line $s \in \mathbb{R}_+$.

The three properties satisfied by $f(s)$ are motivated by the geophysical context: $f(s)$ can be thought of as the analogue of the spatial distribution of potential vorticity in the shallow-water model. This distribution is localized in space and has a zeroth moment that is time-independent, although not necessarily zero. (The first moment of the potential vorticity is also time-independent, and the analogy between our model and the shallow-water model would arguably have been closer had we taken the first moment of $f(s)$ to also vanish; however, this would only make the analytical treatment of sections 3 and 4 more cumbersome without changing our conclusions significantly.)

Before examining solutions of (2.6)–(2.10), we reduce this system to four equations and mention some of its properties. Note first that the model conserves

$$C = \frac{u^2 + v^2}{2} \quad \text{and} \quad H = \frac{1}{2}\left(-u^2 + w^2\right) + \frac{1}{2}\int \left(x_s^2/b^2 + x^2 + y^2\right) \mathrm{d}s.$$

In fact, it is Hamiltonian and, when reduced to four equations, canonical (cf. [8, 6]). To see this, we first take $C = 1$ without loss of generality. We then introduce the new coordinate ϕ, with

$$u = \sqrt{2}\cos\phi \quad \text{and} \quad v = \sqrt{2}\sin\phi,$$

and obtain the equations

$$\dot{\phi} = w - \int f(s) y(s,t)\,\mathrm{d}s, \tag{2.11}$$

$$\dot{w} = -\sin(2\phi), \tag{2.12}$$

$$\epsilon x_t = -y, \tag{2.13}$$

$$\epsilon y_t = x - x_{ss}/b^2 + \epsilon f(s)\sin(2\phi). \tag{2.14}$$

The further change of variable

$$\theta = \phi - \epsilon \int f(s) x(s,t)\,\mathrm{d}s$$

transforms the system into

$$\dot{\theta} = w, \tag{2.15}$$

$$\dot{w} = -\sin\left[2\theta + 2\epsilon \int f(s)x(s,t)\,\mathrm{d}s\right], \tag{2.16}$$

$$\epsilon x_t = -y, \tag{2.17}$$

$$\epsilon y_t = x - x_{ss}/b^2 + \epsilon f(s)\sin\left[2\theta + 2\epsilon \int f(s)x(s,t)\,\mathrm{d}s\right]. \tag{2.18}$$

This is a Hamiltonian system, with Hamiltonian

$$H = \frac{w^2}{2} - \frac{1}{2}\cos\left[2\theta + 2\epsilon \int f(s)x(s,t)\,\mathrm{d}s\right] + \frac{1}{2}\int \left(x_s^2/b^2 + x^2 + y^2\right)\mathrm{d}s \tag{2.19}$$

and symplectic form

$$\Omega = \mathrm{d}\theta \wedge \mathrm{d}w + \epsilon \mathrm{d}y \wedge \mathrm{d}x.$$

Below we use the energy flux F to diagnose the wave radiation. The flux emerges in the derivation of the conservation law for H written in the form

$$\frac{\mathrm{d}H}{\mathrm{d}t} = -\int \partial_s F\,\mathrm{d}s = 0, \quad \text{with} \quad F = -x_t x_s/b^2. \tag{2.20}$$

In the form (2.15)–(2.18) the model can be recognized as describing the dynamics of a pendulum coupled nonlinearly with a Klein–Gordon wave equation, with dispersion relation (1.1). For waves with $O(1)$ wavenumbers, there is a time-scale separation between the slow pendulum and the wave motion if either $b \ll 1$ or $\epsilon \ll 1$. At leading order, a slow manifold is simply given $x = y = 0$, and the corresponding balanced model corresponds to (2.15)–(2.16) with $x = 0$. The problem is then to derive reduced models, governing the slow evolution of θ and w only, which are more accurate than this first-order balanced model. In the next two sections, we examine this problem for each of the two regimes $b \ll 1$ and $\epsilon \ll 1$. We use either of the two formulations (2.6)–(2.10) or (2.15)–(2.18)—whichever is more convenient.

3. Small-b behavior.

We start our asymptotic study of the new model by the small-Froude-number regime $b \ll 1$ and $\epsilon = O(1)$. As mentioned, the time-scale separation is not complete in this regime: long-wave solutions of the Klein–Gordon equation with $k = O(b)$ have an $O(1)$ frequency (see (1.1)) which can match the pendulum frequency. This is the origin of the Lighthill radiation which we now examine. Because this radiation appears at a small power of the small parameter ($O(b^2)$ in the present case), it is easy to derive a postbalanced model, which reduces the dynamics to the two dependent variables θ and w but nevertheless describes the effect of wave radiation. This asymptotic model, which we now derive, is analogous to the post-Newtonian models developed for general relativity [4], and to the Ford, McIntyre, and Norton model for rotating shallow water [12].

3.1. Lighthill radiation and postbalanced model.

Following [36], we expand only x and y in powers of the small parameter b, leaving (2.15)–(2.16) for θ and w unexpanded. Introducing the expansion

$$x = b^2 x^{(0)} + b^3 x^{(1)} + \cdots \tag{3.1}$$

into (2.17)–(2.18) leads to

$$x^{(0)}_{ss} = \epsilon f(s)\sin(2\theta) \quad \text{and} \quad x^{(1)}_{ss} = 0.$$

Solving and imposing oddness in s give

$$x^{(0)} = \epsilon G(s)\sin(2\theta) + A(t)s \quad \text{and} \quad x^{(1)} = B(t)s, \tag{3.2}$$

where

$$G(s) = \frac{a}{(2\pi)^{1/2}}\int_0^s \mathrm{e}^{-s'^2/2}\,\mathrm{d}s',$$

and $A(t)$ and $B(t)$ are functions to be determined.

The expansion (3.1) breaks down for $s = O(b^{-1})$. This reflects the breakdown of the time-scale separation for long waves with $k = O(b)$. In the outer region $s = O(b^{-1})$, we use the rescaled variable

$$S = bs > 0$$

and expand

$$x = b^2 X^{(0)}(S,t) + b^3 X^{(1)}(S,t) + \cdots.$$

Because of the rapid decay of $f(s)$ as $|s| \to \infty$, each $X^{(i)}$, $i = 0, 1, \ldots$, satisfies the free Klein–Gordon equation

$$\epsilon^2 X^{(i)}_{tt} - X^{(i)}_{SS} + X^{(i)} = 0. \tag{3.3}$$

The solutions decaying as $S \to \infty$ are written in terms of their Laplace transform in time. Denoting the Laplace transform by a tilde and the Laplace variable by σ, we find from (3.3) that

$$\tilde{X}^{(i)}(S,\sigma) = \mathrm{e}^{-(1+\epsilon^2\sigma^2)^{1/2}S}\xi^{(i)}(\sigma), \tag{3.4}$$

where the $\xi^{(i)}$ remain to be determined. Note that we have assumed vanishing initial conditions for X. Matching the Laplace transform of (3.2) with (3.4) and noting that $G(\infty) = a/2$ give

$$A(t) = 0, \quad \xi^{(0)}(\sigma) = \frac{a\epsilon}{2}\widetilde{\sin(2\theta)}, \quad \tilde{B}(\sigma) = -(1+\epsilon^2\sigma^2)^{1/2}\xi^{(0)}(\sigma), \quad \text{and} \quad \xi^{(1)}(\sigma) = 0.$$

Thus, to leading order, the waves generated by the balanced motion are given by

$$\tilde{x}(s,\sigma) \approx b^2\tilde{X}^{(0)}(S,\sigma) = \frac{b^2 a\epsilon}{2}\mathrm{e}^{-(1+\epsilon^2\sigma^2)^{1/2}S}\widetilde{\sin(2\theta)} \quad \text{for} \quad S = bs = O(1).$$

This is also the result of a Lighthill-like approximation [20]: this takes advantage of the spatial scale separation between the waves and $f(s)$ to regard (2.17)–(2.18) as a forced Klein–Gordon equation with the localized forcing $\epsilon f(s)\sin(2\theta)$ approximated by $a\epsilon\delta'(s)\sin(2\theta) = b^2 a\epsilon\delta'(S)\sin(2\theta)$. The feedback of the waves on θ and w arises at $O(b^3)$, through the nonzero

$$\tilde{B}(\sigma) = -\frac{a\epsilon(1+\epsilon^2\sigma^2)^{1/2}}{2}\widetilde{\sin(2\theta)}.$$

Inverting the Laplace transform using the convolution theorem gives

$$B(t) = -a\epsilon^2 w\cos(2\theta) - \frac{a\epsilon J_1(t/\epsilon)}{2t} \star \sin(2\theta), \tag{3.5}$$

where $J_1(\cdot)$ is a Bessel function and $\star$ denotes convolution in time defined as

$$(h_1 \star h_2)(t) = \int_0^t h_1(t-\tau)h_2(\tau)\,\mathrm{d}\tau \quad \text{for any two functions } h_1(t) \text{ and } h_2(t).$$

Using (3.2), we compute

$$\int f(s)x(s,t)\,\mathrm{d}s = -\frac{b^2a^2\epsilon}{2\pi^{1/2}}\sin(2\theta) - b^3aB(t) + O(b^4) \tag{3.6}$$

and reduce (2.15)–(2.16) to

$$\dot\theta = w, \tag{3.7}$$

$$\dot w = -\sin\left[2\theta - b^2a^2\epsilon^2\sin(2\theta)/\pi^{1/2} - 2b^3a\epsilon B(t)\right]. \tag{3.8}$$

This is the sought postbalanced model. With $B(t)$ given in (3.5), it is a closed system of ordinary integro-differential equations for θ and w which accounts for the effect of wave radiation. The model is not Hamiltonian because of the $O(b^3)$ term in (3.8). This term describes the loss of pendulum energy caused by the waves; its $O(b^3)$ scaling is consistent with the scaling of the wave-energy flux F. Estimated from the inner solution (3.2) as $s \to \infty$, the flux is found as

$$F = -x_t x_s/b^2 \approx -b^3\, x_t^{(0)} x_s^{(1)}\Big|_{s\to\infty} = -b^3a\epsilon w\cos(2\theta)B(t). \tag{3.9}$$

Note that since the system (3.7)–(3.8) is integrable when $b = 0$, closed-form solutions for small b could be derived by averaging. We do not pursue this here, since this possibility is a fragile particularity of our model.

A number of conclusions can be drawn from the above analysis. First, the wave radiation scales like a power, here the square, of the small parameter b. Next, the waves radiated are long, with $O(b^{-1})$ wavelength, so that they can resonate with the slow pendulum motion. Because of this large scale, a Lighthill-like theory can be applied; this describes the waves as generated by a Dirac-type source which can be estimated by the first nontrivial moment of the balanced part of the flow (here represented by the first moment of $f(s)$). Finally, the feedback of the waves on the flow can be captured asymptotically. This leads to a postbalanced model, which evolves on the slow time scale only and describes to leading order the impact of wave radiation on balanced motion. This model is dissipative and involves time integrals as well as derivatives, as a result of dispersion. These conclusions are identical (except for the specific scalings) to those that can be drawn from the analysis of the much more complex shallow-water equations in [12].

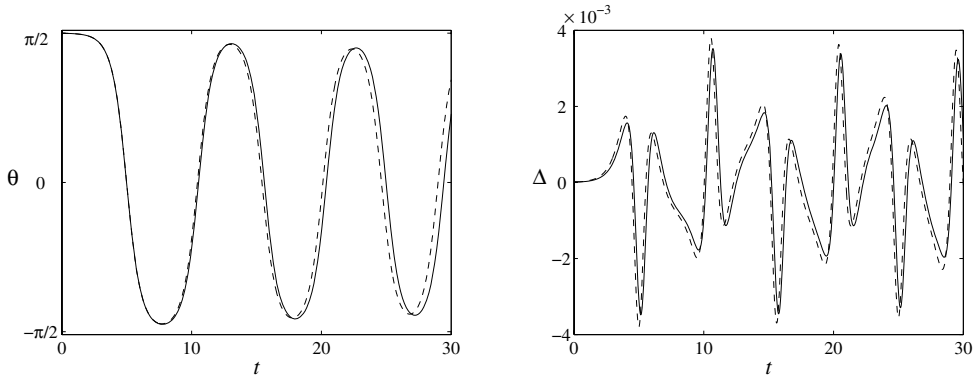


Figure 2. *Comparison of a numerical solution of the model* (2.15)–(2.18) *(solid curves) with the solution of the postbalanced approximation* (3.7)–(3.8) *derived in the limit* $b \ll 1$ *(dashed curves). The left panel shows the evolution of the angle* θ*; the right panel compares the* $O(b^3)$ *quantity* Δ *defined in* (3.11) *with its postbalanced approximation* $-b^3 aB(t)$*. The parameters chosen are* $b = 0.15$*,* $\epsilon = a = 1$*, and the initial conditions have been taken on the unperturbed heteroclinic trajectory* (3.10)*.*

3.2. Numerical results. We confirm the validity of the postbalanced approximation (3.7)–(3.8) by presenting the results of a numerical experiment. We compare the numerical solution of the full ODE–PDE model (2.15)–(2.18), implemented as described in Appendix A, with the numerical solution of (3.7)–(3.8) for particular initial conditions. The fast variables are taken as $(x, y) = (0, 0)$, and the slow ones are chosen so that, for $b = 0$, the angle $\theta(t)$ follows the separatrix joining $\theta = \pi/2$ to $\theta = -\pi/2$. (We refer to this trajectory as heteroclinic, although it can also be viewed as homoclinic, if one identifies $\pi/2$ with $-\pi/2$ as is done in [8].) For our system, this unperturbed solution satisfies

$$\cos\theta = \operatorname{sech}\left[\sqrt{2}(t - t_0)\right] \quad \text{and} \quad \sin\theta = -\tanh\left[\sqrt{2}(t - t_0)\right] \tag{3.10}$$

for some t_0, which we have taken to be $t_0 = 5$. When $b \neq 0$, the dissipation introduced by the wave radiation means that the heteroclinic trajectory is replaced by damped oscillations; with $b \ll 1$, these have long, slowly decreasing periods.

This is illustrated by Figure 2, which shows the evolution of θ for $b = 0.15$, with $\epsilon = a = 1$. (The results are typical of a wide range of values satisfying $b \ll 1$ and $\epsilon, a = O(1)$ with which we have experimented.) The period of the oscillations introduced by wave radiation is of the order of 10. This is consistent with the order of magnitude $\log(b^{-3})$ obtained by noting that the pendulum energy is decreased by an $O(b^3)$ amount from the separatrix energy; this scaling has been confirmed by computations with several values of b. The figure compares the solution of the full model (solid curve) with that of the postbalanced model (dashed curve) and confirms the validity of the latter: the postbalanced approximation captures well the transition to damped oscillations, even though the period differs slightly from that obtained with the full model. This difference in periods can be attributed to the $O(b^4)$ terms neglected in our asymptotics which lead to an $O(b)$ error in the period of the postbalanced model.

The effectiveness of the postbalanced model is better judged from the right panel of Figure 2. This figure assesses the accuracy of the crucial approximation (3.6) by comparing

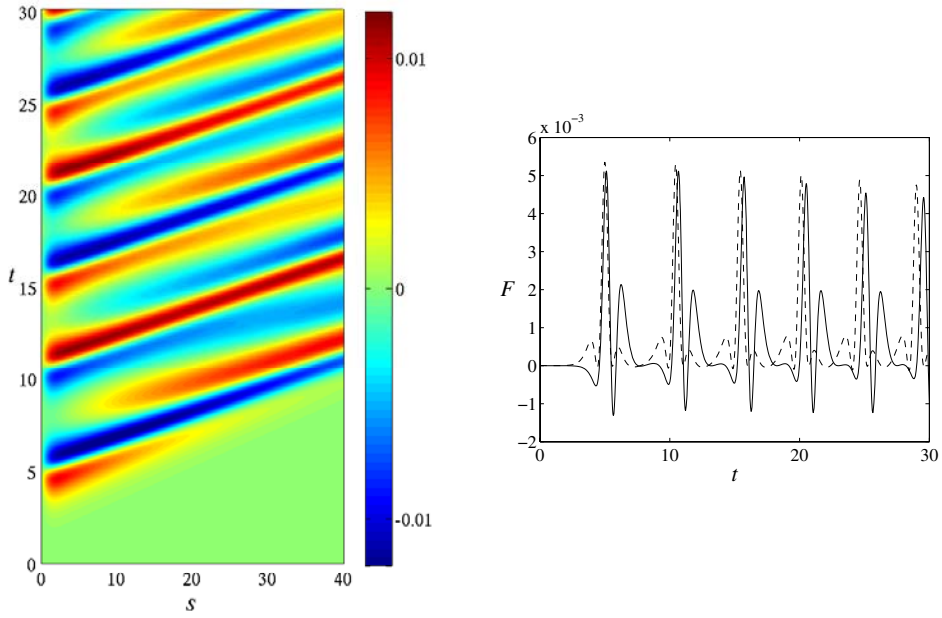


Figure 3. *Wave generation in the small-b limit. The left panel shows the evolution of $x(s,t)$ for the same parameters and initial conditions as in Figure 2. The right panel compares the energy flux $F = x_s x_t/b^2$ evaluated at $s = 2$ (solid curve), with the postbalanced approximation (3.9) (dashed curve).*

$$\Delta = \int f(s)x(s,t)\,\mathrm{d}s + \frac{b^2a^2\epsilon}{2\pi^{1/2}}\sin(2\theta) \tag{3.11}$$

with its approximation $-b^3aB(t)$. Because both functions are computed from the full-model solution, the phase shift present in the postbalanced solution disappears. The figure confirms the validity of (3.6).

The left panel of Figure 3 shows the structure of the waves that are radiated from the balanced motion. As expected, these are long waves, with $O(b^{-1})$ wavelengths. The right panel compares the flux (2.20) evaluated at $s = 2$, with the flux (3.9) obtained in the asymptotic treatment. The chosen value $s = 2$ is roughly consistent with the idea that (3.9) holds in an intermediate region between $s = O(1)$ and $s = O(b^{-1})$. If we factor out the phase shift mentioned above, there is a good agreement between the asymptotic and full numerical results. It is interesting to note that the actual flux changes sign while, by construction, the asymptotic one does not. There is a fair amount of cancellation in the actual flux which makes its time-integrated effect well described by the asymptotic one.

4. Small-ϵ behavior. We now turn to the regime $\epsilon \ll 1$, with $b = O(1)$, which in the shallow-water context corresponds to the quasi-geostrophic regime. In this regime, there is a frequency separation between balanced motion and waves at all scales, and the wave generation can be expected to be exponentially weak. In this section, we use exponential asymptotics to estimate the amplitude of the waves generated.

4.1. Exponential asymptotics. A balanced solution of (2.15)–(2.18) can be sought by expansion of θ, w, x, and y in powers of ϵ or, alternatively, by first deriving an expression for a slow manifold in the form of slaving relations $x = x_{\rm bal}(s,\theta,w)$ and $y = y_{\rm bal}(s,\theta,w)$, and then reducing the dynamics onto it. The slaving relations can be derived order-by-order in ϵ, through either an iteration or an expansion procedure. Choosing the latter, we write

$$x_{\rm bal}(s,\theta,w) = \sum_{n=0}^{\infty} \epsilon^{2n+1} x_{\rm bal}^{(n)}(s,\theta,w) \quad \text{and} \tag{4.1}$$

$$y_{\rm bal}(s,\theta,w) = \sum_{n=0}^{\infty} \epsilon^{2n+2} y_{\rm bal}^{(n)}(s,\theta,w) \tag{4.2}$$

and introduce these expressions into (2.15)–(2.18). This leads to a sequence of ODEs in s for the $x_{\rm bal}^{(n)}(s,\theta,w)$ which are best solved in the Fourier domain. Denoting the Fourier transform in s by a hat, with k as the Fourier variable, we find

$$\hat{x}_{\rm bal}^{(0)} = -\frac{\hat{f}(k)}{\Omega^2}\sin(2\theta), \quad \hat{y}_{\rm bal}^{(0)} = \frac{2\hat{f}(k)}{\Omega^2} w\cos(2\theta), \quad \text{etc.},$$

where

$$\hat{f}(k) = \frac{{\rm i}ak}{2\pi}{\rm e}^{-k^2/2} \tag{4.3}$$

and Ω is the scaled frequency satisfying

$$\Omega^2 = \epsilon^2\omega^2 = 1 + k^2/b^2. \tag{4.4}$$

The Fourier transform can be inverted. For instance, we find that

$$x_{\rm bal}^{(0)} = -\frac{ab^2{\rm e}^{b^2/2}}{4}\left({\rm e}^{bs}\,{\rm erfc}\,\frac{b+s}{\sqrt{2}} - {\rm e}^{-bs}\,{\rm erfc}\,\frac{b-s}{\sqrt{2}}\right)\sin(2\theta). \tag{4.5}$$

By contrast with the small-b case, there are no obstacles to carrying out this calculation to obtain, in principle, $x_{\rm bal}^{(N)}$ and $y_{\rm bal}^{(N)}$ for arbitrary order N. The decay of $x_{\rm bal}^{(N)}$ and $y_{\rm bal}^{(N)}$ as $|s| \to \infty$, in particular, is satisfied. Indeed a rough estimate indicates that $\hat{x}^{(N)} \propto \hat{f}(k)/\Omega^{2(N+1)}$, leading to

$$x^{(N)} \propto s^N{\rm e}^{-b|s|} \quad \text{as} \quad |s| \to \infty. \tag{4.6}$$

Of course, the series (4.1) diverges, and only finite values of N, typically up to $O(\epsilon^{-1})$, can be considered. This divergence reflects the existence of a subdominant solution which is switched on through a Stokes phenomenon [1, 3]. Physically, this subdominant solution represents waves, and the switching on corresponds to their spontaneous generation by the balanced motion. Indeed, introducing the decomposition $x = x_{\rm bal} + x_{\rm w}$, with $x_{\rm bal}$ defined by an optimally truncated series of the form (4.1) and $x_{\rm w} \ll x_{\rm bal}$, we find that $x_{\rm w}$ satisfies the

free Klein–Gordon equations at leading order in ϵ. Thus, the time dependence of the Fourier transform of $\hat{x}_{\rm w}$ is

$$\hat{x}_{\rm w}(k,t) \propto \exp(\mathrm{i}\omega t), \tag{4.7}$$

where the branch of ω defined by (1.1) is chosen for radiation at $|s| \to \infty$. If the solution is properly balanced, then $\hat{x}_{\rm w} = 0$. However, in the course of the evolution, a Stokes phenomenon can occur which switches $\hat{x}_{\rm w}$ to an exponentially small, nonzero value which we now estimate.

The Stokes phenomenon is associated with singularities of the balanced motion for complex values of t. By analyzing the equations near the relevant singularities, and matching with expressions (4.1) and (4.7) which are valid some distance away from these singularities, one can estimate $x_{\rm w}$. This is the essence of the so-called Kruskal–Segur method, which we apply in Appendix B. There, we obtain the amplitude of the waves switched on when the (real) time t crosses the Stokes line joining a pair of complex-conjugate singularities t_* and $\bar{t}_*$. In terms of Fourier transform, we find

$$\hat{x}_{\rm w}(k,t) \sim \frac{-\mathrm{i}k\lambda(k)}{\epsilon}\,\mathrm{e}^{-\alpha\Omega/\epsilon - k^2/2}\cos[\Omega(t-\beta)/\epsilon], \tag{4.8}$$

where

$$\alpha = \operatorname{Im} t_* > 0 \quad \text{and} \quad \beta = \operatorname{Re} t_*.$$

In this expression, $\lambda(k)$ is an $O(1)$ function defined for small k that can be obtained numerically as described in Appendix B by solving the discretized version of an infinite set of recurrence relations. The switching on takes place for $t = \beta$ so that (4.8) needs to be multiplied by the Heaviside function $H(t-\beta)$ (which could be smoothed out as an error function of width $\epsilon^{1/2}$ using Berry's result [3]), and only the singularities $(t_*, \bar{t}_*)$ nearest to the real axis need to be taken into account.

The Fourier transform can be inverted to derive an approximation for $x_{\rm w}(s,t)$ from (4.8); using the steepest-descent method, we find the final estimate

$$x_{\rm w}(s,t) \sim (2\pi)^{1/2} b^3 \lambda(0)\,\mathrm{e}^{-\alpha/\epsilon} S \operatorname{Re}\frac{\mathrm{e}^{\mathrm{i}(t-\beta)/\epsilon - b^2S^2/[2(\alpha - \mathrm{i}(t-\beta))]}}{[\alpha - \mathrm{i}(t-\beta)]^{3/2}}, \tag{4.9}$$

valid for $S = \epsilon^{1/2}s = O(1)$. This provides a useful closed-form approximation which requires only the computation of the value of $\lambda(0)$. Figure 4 shows $\lambda(0)$ as a function of a and b; it also shows $\lambda(0)/a$ to demonstrate that $\lambda(0) \sim a$ for $a \to 0$ or $b \to 0$. In fact, it can be established that these limits correspond to the decoupling of θ and w from x and y, and that

$$\lambda(k) \sim a \quad \text{as} \quad a \to 0 \quad \text{or} \quad b \to 0. \tag{4.10}$$

Formula (4.9) shows that the waves emitted are "near-inertial" waves, that is, have a frequency close to the minimum frequency ϵ^{-1} and have large, $O(\epsilon^{-1/2})$ spatial scales.

It is interesting to examine what a Lighthill-like approach would predict for the wave amplitude in the regime $\epsilon \ll 1$. This approach amounts to solving (2.17)–(2.18), with the last term of (2.18) approximated by $\epsilon \sin(\theta^{(0)})$, where $\theta^{(0)}$ is the leading-order approximation to θ.

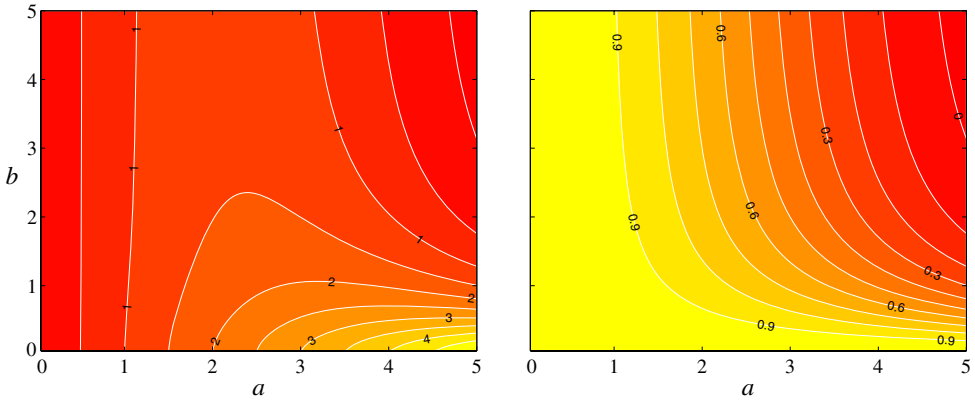


Figure 4. *Value of $\lambda(0)$, governing the amplitude of the waves generated spontaneously in the limit of small ϵ, as a function of a and b (left panel). The right panel shows $\lambda(0)/a$ and demonstrates that $\lambda(0) \sim a$ for $a \to 0$ or $b \to 0$.*

The computation of the wave radiated in this case is straightforward in Fourier space, where it follows the computation carried out originally for the LK model in [22]. The result has the form (4.8), with $\lambda(k)$ replaced by a, consistent with (4.10). For finite a and b, $\lambda(k)$ differs significantly from a (see Figure 4). Thus, a Lighthill-like approach does not give the correct asymptotics for amplitude of the Fourier modes even to leading order (although the controlling behavior, that is, the asymptotics of the log of the amplitude, is correct). This is because the wave amplitude at leading order in ϵ depends on the structure of the balanced motion to all orders. Thus, in contrast with the small-b situation, an essentially complete description of the balanced motion is necessary to estimate even the leading-order wave amplitude. We say "essentially" here because this complete description is in fact needed only in the vicinity of the singular points t_* nearest to the real axis rather than for all t. This is the simplification that is exploited in the asymptotic approach of Appendix B.

4.2. Numerical results. We compare the theoretical predictions of exponential asymptotics with numerical results for two sets of parameters. For the first, we choose the same initial condition as in section 3.2, that is, on the unperturbed ($\epsilon = 0$) heteroclinic trajectory (3.10). For this trajectory, there is only one pair of complex-conjugate singularities of the balanced motion, t_* and $\bar{t}_*$ with

$$t_* = t_0 + \mathrm{i}\frac{\pi}{2\sqrt{2}},$$

so that $\alpha = \pi/(2\sqrt{2})$ and $\beta = t_0$. An initial condition on the heteroclinic trajectory makes the wave generation particularly easy to identify and avoids the difficulties associated with the initialization of the fast variables (x, y). By taking t_0 sufficiently large, all the time derivatives of w and θ at $t = 0$ can be made arbitrarily small, leading to arbitrarily small values for the coefficients $x_{\mathrm{bal}}^{(i)}$, $i > 0$, and $y_{\mathrm{bal}}^{(i)}$, $i \geq 0$, of the balanced part of x and y. Thus, taking

$$x(0) = \epsilon x_{\mathrm{bal}}^{(0)}, \quad y(0) = 0,$$

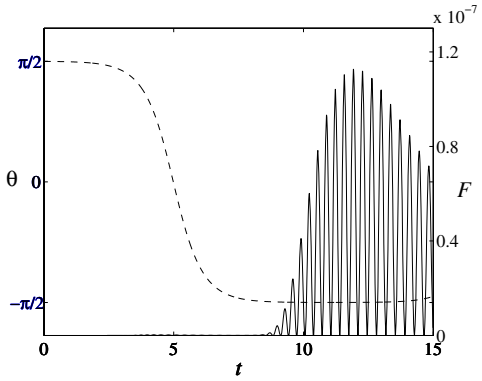


Figure 5. *Wave generation for $\epsilon = 0.125$, with $a = 2$ and $b = 1$ for initial conditions on the unperturbed heteroclinic trajectory. The angle θ (dashed curve, left axis) and wave energy flux F at $s = 25$ (solid curve, right axis) are shown as a function of time.*

where $x_{\rm bal}^{(0)}$ is given by (4.5) with $\theta = \theta(0)$, provides a solution that can be balanced to an arbitrary accuracy simply by taking t_0 large enough. We have taken $t_0 = 5$ which turns out to be sufficient to make the wave radiation by initial adjustment negligible.

Figure 5 shows the evolution of θ obtained for the parameter choice $\epsilon = 0.125$, $a = 2$, and $b = 1$. As expected, θ follows closely the heteroclinic trajectory; it does not tend to $\pi/2$ for large t, however, and a longer simulation would reveal that θ undergoes a series of weakly damped oscillations. The damping is of course associated with the wave radiation. This is first diagnosed by showing the wave-energy flux F computed at the boundary $s = 25$ of our integration domain. The flux is always positive, confirming the proper implementation of the radiation boundary condition, and it oscillates rapidly, with a frequency $2/\epsilon$; its maximum value, of the order of 10^{-7}, is consistent with the crude estimate $\exp(-2\alpha/\epsilon) \approx 2\times 10^{-8}$ which follows from (4.9).

A more complete picture of the wave radiation emerges from Figure 6. This shows $x - x_{\rm bal}^{(0)}$ as a function of s and t. The variable $x - x_{\rm bal}^{(0)}$ includes an $O(\epsilon^3)$-contribution of the balanced part of x as well as the exponentially small wave part $x_{\rm w}$. However, the balanced contribution decreases rapidly for $t > t_0$ and for $s > 1$, so that the wave contribution is clearly isolated. This consists of rapidly propagating waves, with frequency approximately equal to ϵ^{-1}, that are emitted for $t \approx t_0 = 5$ from the region $s = O(1)$. The right panel of Figure 6, showing the estimate (4.9) for $x_{\rm w}$, confirms the validity of our asymptotic treatment. Note that for the chosen parameters $a = 2$ and $b = 1$, we obtain $\lambda(0) = 1.69$ using the method described in Appendix B; a Lighthill-like approach would therefore overestimate the amplitude of the wave generation by the factor $a/\lambda(0) \approx 1.2$.

For the second numerical experiment, we have chosen an initial condition on a periodic trajectory of the undisturbed system. The periodic trajectories can be written explicitly as

$$\theta = -\,\mathrm{am}\,(\sqrt{2}(t - t_0)/k, k), \tag{4.11}$$

where am is the amplitude of the Jacobian elliptic functions [2, Chapter 16], and $k \geq 1$ and t_0 are fixed by the initial conditions. The period of (4.11) is $2\sqrt{2}kK(k) = 2\sqrt{2}K(1/k)$, where

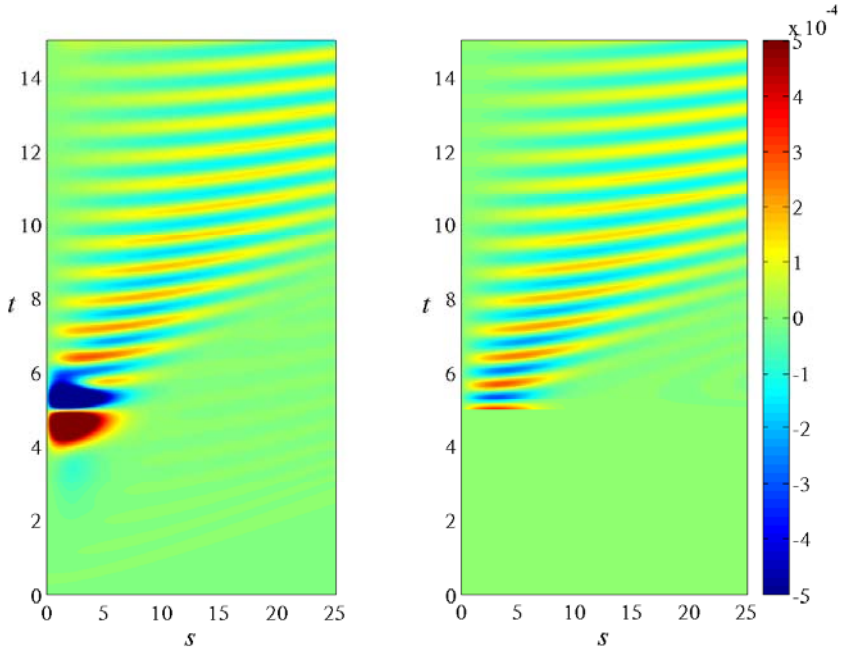


Figure 6. *Wave generation for the same parameters as in Figure 5. The $O(\epsilon^3)$ quantity $x - x^{(0)}_{\rm bal}$ (left panel) is compared with the asymptotic result (4.9) for the wave part $x_{\rm w}$ (right panel). The color code, used for both panels, does not cover the large values of $|x - x^{(0)}_{\rm bal}|$ for $4 < t < 6$ which reach 1.8×10^{-3}.*

K is the elliptic integral of the first kind. The solutions (4.11) have complex-time poles, with the nearest to the real axis located at

$$t_{*n} = t_0 + n\sqrt{2}K(1/k) + \mathrm{i}K'(1/k)/\sqrt{2} \quad \text{and} \quad \bar{t}_{*n},$$

where $n \in \mathbb{Z}$ and K' is the complementary elliptic integral. We have taken the parameters $\epsilon = 0.15$, $a = 2.5$, and $b = 0.5$. For initial conditions we have chosen $\theta = \pi/4$ and $w = 0$, giving

$$k = \sqrt{2}, \quad t_0 = K(1/\sqrt{2})/\sqrt{2} = 1.311\ldots,$$

and poles $t_{*n} = (2n + 1 + \mathrm{i})t_0$. Correspondingly, for real t, Stokes lines are crossed for $t = t_n = (2n + 1)t_0$, when the unperturbed trajectory has vanishing angle θ.

For periodic unperturbed solutions, it is difficult to compute balanced initial conditions accurately: in principle, there is no alternative to the computation of $x_{\rm bal}$ and $y_{\rm bal}$ by optimally truncated series expansion. In practice, however, we found that the lack of balance that results from truncating these series to $O(\epsilon)$, that is, from taking $x = \epsilon x^{(0)}_{\rm bal}$ given in (4.5) with $\theta = \pi/4$, and $y = 0$ is not problematic. A small-amplitude wavepacket is emitted at the initial time which quickly leaves the computational domain while adjusting the solution to a well-balanced state.

Figure 7 shows the evolution of θ and of the wave-energy flux F at the boundary of our computational domain, taken to be $s = 50$. The angle θ oscillates according to (4.11), with

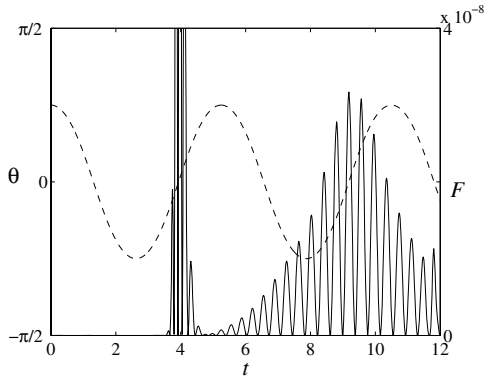


Figure 7. *Wave generation for $\epsilon = 0.15$, with $a = 2.5$ and $b = 0.5$ for initial conditions on a periodic unperturbed trajectory. Angle θ (dashed curve, left axis) and wave energy flux F at $s = 50$ (solid curve, right axis) as a function of time.*

small corrections. The flux F has a large peak for $t \approx 4$ that corresponds to the passage of the wavepacket emitted by the initial adjustment. Thereafter, the flux is associated with the exponentially small waves radiated spontaneously when t crosses Stokes lines. The estimate $\exp(-2\alpha/\epsilon)$, with $\alpha = \operatorname{Im} t_{*n} = t_0 = 1.311\ldots$, gives 2.5×10^{-8}, in rough agreement with the observed flux.

The left panel of Figure 8 shows $x - x_{\mathrm{bal}}^{(0)}$ as a function of s and t. Although this field is dominated by the balanced component for $s = O(1)$, the exponential decay of this component with s (see (4.6)) ensures that the wave part emerges for large s, say, for $s > 10$. After the wavepacket created by the initial adjustment leaves the domain, the wave field consists of a superposition of large-scale wavepackets emitted at $t = t_n$, creating an interference pattern. The right panel of Figure 8 shows the asymptotic prediction for x_{w}. To construct the figure, we have used (4.9), with $\alpha = \operatorname{Im} t_{*0} = 1.311$, $\beta = \operatorname{Re} t_{*0} = 1.311$, and $\lambda(0) = 2.235$, to estimate the wave radiated when the first Stokes line is crossed, at $t = t_0$. The full wave field is then the superposition of similar contributions generated for each t_n. If we denote the first contribution by $x_{\mathrm{w}0}(s, t - t_0)$, the full wave field is given by

$$x_{\mathrm{w}}(s,t) = \sum_{n=0}^{\infty} (-1)^n x_{\mathrm{w}0}(s, t - t_n) H(t - t_n),$$

where $H(\cdot)$ denotes the Heaviside function, and the factors $(-1)^n$ appear because the signs of the balanced variables near the poles t_{*n} alternate with n. On the whole, this asymptotic expression gives a good picture of the wave field, even though the amplitudes are somewhat overestimated, and the interference pattern appears slightly distorted. However, keeping in mind that (4.9) has an $O(\epsilon)$ error and is only valid for $s = O(\epsilon^{1/2})$, the agreement between the asymptotic prediction and the numerical result is satisfactory.

5. Discussion. In this paper, we introduce the simple mixed ODE–PDE system (2.15)–(2.18) as a toy model for wave radiation by slow (or balanced) motion in the atmosphere and ocean. Though no more than a caricature of the real atmosphere and oceans, the model has a number of appealing features: wave dispersion relation identical to that of shallow-water

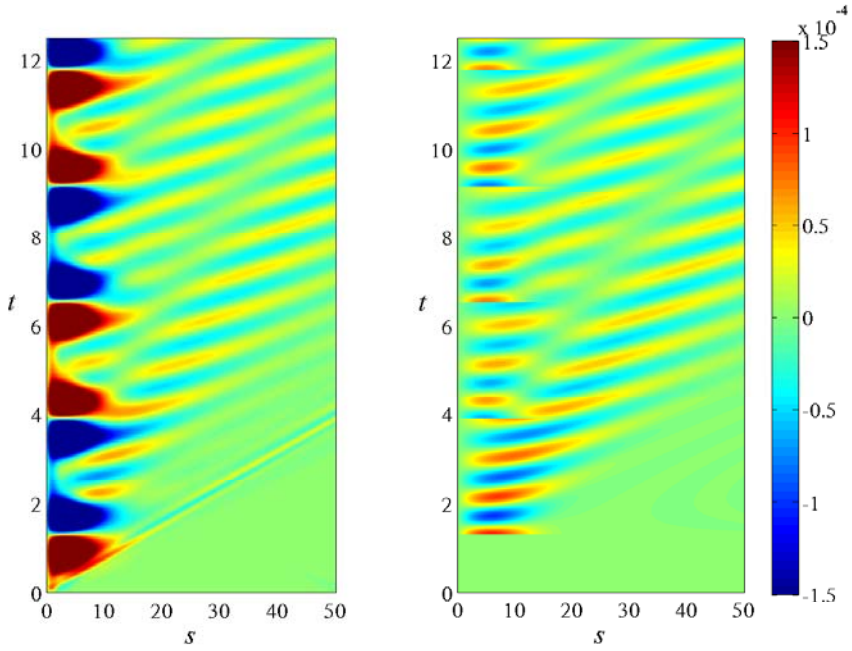


Figure 8. *Wave generation for the same parameters as in Figure 7. The $O(\epsilon^3)$ quantity $x - x_{\rm bal}^{(0)}$ (left panel) is compared with the asymptotic result (4.9) for the wave part $x_{\rm w}$ (right panel). The color code, used for both panels, does not cover the large values of $|x - x_{\rm bal}^{(0)}|$ for $0 \le s < 10$ which reach 5.8×10^{-4}.*

inertia-gravity waves, Hamiltonian structure, possibility of wave radiation to infinity, and, of course, great simplicity. Compared to low-order models such as the LK and swinging-spring models, the inclusion of a PDE component is a major advance toward realism. It makes it possible to examine the interaction between waves and balanced motion of vastly different spatial scales that is at the core of the Lighthill-like radiation (or gravitational-wave-like) radiation in the small-Froude-number regime. The simplicity of the model also allows the estimation of wave radiation in the small-Rossby-number regime using exponential asymptotics. It emerges from the analysis that the wave amplitude in the latter regime cannot be estimated as it is in the former regime, by simply regarding the wave equation as forced by terms computed from the leading-order balanced motion. Indeed, estimating the wave amplitude in the small-Rossby-number regime requires knowledge of the balanced motion to all algebraic orders in ϵ, at least near complex-time singularities. This makes the study of spontaneous wave generation in the more realistic context of, say, the shallow-water model particularly challenging.

Another phenomenon that the PDE component makes it possible to represent is the loss of energy through wave radiation. This is of great interest since some form of dissipation is key to the maintenance of balance in the atmosphere and ocean. This dissipation is poorly modeled by simply adding a small damping to the fast equations, however, because the fast variables are in fact dominated by their balanced, slow component, which thus becomes significantly affected by dissipation. In real fluids, however, dissipation is thought to affect mostly the

truly fast component of the dynamics, through wave radiation and possibly nonlinear forward energy cascades. The new model realistically reproduces at least the first of these processes, and it could be used to investigate how close to slow manifolds the evolution of the system remains owing to wave radiation. Discussing these aspects will of course involve typical PDE issues, such as the choice of a norm suitable to measure wave activity in the vicinity of the origin. It is worth noting that the energy dissipation caused by wave radiation does not necessarily imply that the balanced system relaxes to a state of rest. In fact, it is not even the case for our model, when formulated in terms of the original slow variables (u, v, w), since the ultimate state $w = 0$ and $\theta = 0$ corresponds to $u = \sqrt{2}$, $v = 0$, $w = 0$. One can interpret this by recognizing that the conservation of $u^2 + v^2$, left unaffected by wave radiation, prevents the system from relaxing to a resting state. This observation is trivial for our model, but less so for the fluid systems: for these the relaxation will be toward a state of minimum energy for a fixed distribution of potential vorticity. Since this state is typically not smooth, this means that wave radiation leads to the formation of complicated, rough structures.

There are several generalizations of the model introduced in this paper which may be worthwhile to consider. A first, mentioned in the introduction, is the addition of a slow forcing. The aim of this addition is to study the statistical equilibrium that can emerge from the balance between forcing and radiation. A second modification is to alter the system so as to make the balanced dynamics chaotic. This is useful to remove some of the peculiarities that arise because of the integrability of the one-degree-of-freedom balanced system. Wirosoetisno and Shepherd [39] introduced a modification of the LK model which amounts to replacing the constant value of $u_{\mathrm{L}}^2 + v_{\mathrm{L}}^2$ by an arbitrarily chosen function of time. Our model could be modified in exactly the same way.

We conclude by noting that the study of wave radiation by balanced motion need not be limited to the regimes of small Froude number b or small Rossby number ϵ as is the case here. Although balance is difficult to define in the absence of a small parameter, it is often remarked that the qualitative features observed for small b or ϵ, in particular, the clear separation between wave- and flow-type motion, persist when these parameters are of order unity. The model introduced in this paper could very well be used as a starting point to examine this issue and to give a precise meaning to, as well as an explanation for, this observation.

Appendix A. Numerical implementation. We need to solve numerically the model equations (2.15)–(2.16) in a finite domain $[-L, L]$ in manner that allows outward radiation of wave energy. To this end, we implement exact nonreflecting boundary conditions. This takes advantage of the fact that $f(s)$ is localized, so that, for $|s| \gg 1$, (2.17)–(2.18) reduce to the linear Klein–Gordon equation and can be solved in closed form.

Consider the solution of the Klein–Gordon equation in $[L, \infty)$ with radiation boundary condition as $s \to \infty$. In terms of the Laplace transform $\tilde{x}(s, \sigma)$ of $x(s, t)$, where σ is the Laplace variable, this solution reads

$$\tilde{x}(s, \sigma) = \exp\left[-b(1 + \epsilon^2\sigma^2)^{1/2}s\right] \xi(\sigma)$$

for some function $\xi(\sigma)$. This gives the simple Dirichlet-to-Neumann map [15]

$$\tilde{x}_s(s, \sigma) = -b(1 + \epsilon^2\sigma^2)^{1/2}\tilde{x}(s, \sigma).$$

Upon transforming back to the time domain and evaluating at $s = L$, this reads

$$x_s(L,t) = -b\epsilon x_t(L,t) - b\frac{J_1(t/\epsilon)}{t} \star x(L,t), \tag{A.1}$$

where $J_1(\cdot)$ is a Bessel function, and $\star$ denotes convolution in time. This provides an exact nonreflecting condition at $s = L$ for the Klein–Gordon equation; we employ it for our system with L large enough that $f(L)$ is negligible.

We solve the model equations (2.15)–(2.18) by finite differences, using a uniform grid for $s \in [0, L]$ and a Störmer–Verlet scheme both in t and s. This discretization is symplectic for the ODE part of the model, and multisymplectic for the PDE part [7]. The boundary condition (A.1) is discretized in a straightforward manner, using a backward difference for the s- and t-derivatives, and the trapezoid rule for the convolution. This discretization of the exact boundary condition can lead to spurious wave reflection at $s = L$ [30], but we have found this to be negligible.

Appendix B. Exponential asymptotics for $\epsilon \ll 1$. In this appendix, we use exponential asymptotics to obtain the estimates (4.8)–(4.9) for the amplitude of the waves generated by the balanced motion in the regime $\epsilon \ll 1$, $b = O(1)$. The method is similar to that used in [34] for the LK model, and we refer the reader to that paper for further details.

The balanced solutions of (2.15)–(2.16) may be obtained by expansion in powers of ϵ of all the variables. Because it is convenient to use the formulation (2.6)–(2.10), we expand

$$u(t) = \sum_{n=0}^{\infty} \epsilon^n u^{(n)}(t) \quad \text{and similarly for } v \text{ and } w, \tag{B.1}$$

and

$$x(s,t) = \sum_{n=0}^{\infty} \epsilon^{n+1} x^{(n)}(s,t) \quad \text{and} \quad y(s,t) = \sum_{n=0}^{\infty} \epsilon^{n+2} y^{(n)}(s,t). \tag{B.2}$$

These series, whose coefficients can be computed recursively, describe the balanced part of the solution. The leading-order coefficients $u^{(0)}(t)$, $v^{(0)}(t)$, and $w^{(0)}(t)$, in particular, solve

$$\dot{u}^{(0)} = -v^{(0)}w^{(0)}, \quad \dot{v}^{(0)} = w^{(0)}u^{(0)}, \quad \dot{w}^{(0)} = -u^{(0)}v^{(0)}$$

and are given in terms of Jacobi elliptic functions. The wave part of the solution is exponentially smaller, and to leading order given by the solution $x_{\mathrm{w}}(s,t)$ of a Klein–Gordon equation. In terms of Fourier transform, this reads

$$\hat{x}_{\mathrm{w}}(k,t) = C(k)\mathrm{e}^{\mathrm{i}\omega t} + \text{c.c.}, \tag{B.3}$$

where $C(k)$ is a Stokes multiplier switching from zero to an exponentially small value as a Stokes line is crossed by t.

The coefficients in (B.1)–(B.2) have singularities in the complex plane near which the expansions (B.1)–(B.2) cease to be valid. In the vicinity of these singularities, the balanced and wave parts of the solution are not well separated, and (B.1)–(B.2), which we can regard

as an outer solution, must be replaced by a different, inner, solution. We now consider this inner solution near a singularity t_* with $\operatorname{Im} t_* > 0$.

The behavior of the outer solution near t_* is readily obtained as

$$u \sim \frac{-\mathrm{i}}{t - t_*}, \quad v \sim \frac{-1}{t - t_*}, \quad w \sim \frac{\mathrm{i}}{t - t_*}$$

(up to pairwise changes of signs). Similarly, x and y behave like $\epsilon(t - t_*)^{-2}$ and $\epsilon^2(t - t_*)^{-3}$, respectively. This behavior is valid for $\epsilon \ll |t - t_*| \ll 1$ and should match the inner solution. This suggests the inner scaling

$$u = \epsilon^{-1}U(\tau), \quad v = \epsilon^{-1}V(\tau), \quad w = \epsilon^{-1}W(\tau), \quad x = \epsilon^{-1}X(s,\tau), \quad \text{and} \quad y = \epsilon^{-1}Y(s,\tau),$$

where τ is a rescaled time defined by

$$t = t_* + \epsilon\tau.$$

Introducing this into (2.6)–(2.10) leads to essentially identical equations with lowercase variables replaced by their uppercase counterparts, time derivatives replaced by τ-derivatives, and, crucially, ϵ set to 1. This is best solved using the Fourier transform of X and Y or, more conveniently, a small modification thereof. Noting the form (4.3) of $\hat{f}(k)$, we define

$$\bar{X} = k^{-1}\mathrm{e}^{k^2/2}\hat{X} \quad \text{and} \quad \bar{Y} = k^{-1}\mathrm{e}^{k^2/2}\hat{Y},$$

where, as before, the hat denotes the Fourier transform in s. This leads to the system

$$U' = -VW - \mathrm{i}aV \int k^2\mathrm{e}^{-k^2}\bar{Y}(k)\,\mathrm{d}k, \tag{B.4}$$

$$V' = WU + \mathrm{i}aU \int k^2\mathrm{e}^{-k^2}\bar{Y}(k)\,\mathrm{d}k, \tag{B.5}$$

$$W' = -UV, \tag{B.6}$$

$$\bar{X}' = -\bar{Y}, \tag{B.7}$$

$$\bar{Y}' = \Omega^2\bar{X} + \frac{\mathrm{i}a}{2\pi}UV, \tag{B.8}$$

where $' = \mathrm{d}/\mathrm{d}\tau$ and Ω is the scaled frequency defined in (4.4). The large-τ behavior of the solutions, needed for matching, can be obtained by expanding in inverse powers of τ. Specifically, we let

$$U = \mathrm{i}\sum_{n=1}^{\infty}\frac{U_n}{\tau^{2n-1}}, \quad V = \sum_{n=1}^{\infty}\frac{V_n}{\tau^{2n-1}}, \quad W = \mathrm{i}\sum_{n=1}^{\infty}\frac{W_n}{\tau^{2n-1}},$$
$$\bar{X} = \sum_{n=1}^{\infty}\frac{\bar{X}_n(k)}{\tau^{2n}}, \quad \text{and} \quad \bar{Y} = \sum_{n=1}^{\infty}\frac{\bar{Y}_n(k)}{\tau^{2n+1}},$$

and obtain a set of recurrence relations for the (real) coefficients U_n, V_n, etc. With the initial conditions

$$U_1 = -1, \quad V_1 = -1, \quad W_1 = 1, \quad \bar{X}_1 = \frac{a}{2\pi\Omega^2}, \quad \text{and} \quad \bar{Y}_1 = 2\bar{X}_1,$$

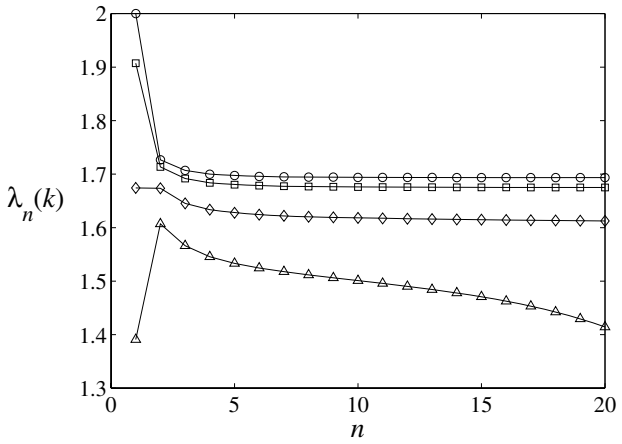


Figure 9. *Coefficients $\lambda_n(k)$ defined in (B.10) as a function of iteration number n for $a = 2$ and $b = 1$. Four values of k are shown, corresponding to the first four zeros of the Hermite polynomial H_{101}, namely, $k_0 = 0$ (○), $k_1 = 0.22$ (□), $k_2 = 0.44$ (◇), and $k_3 = 0.66$ (△). For $j = 0, 1, 2$, the $\lambda_n(k_j)$ converge to finite values $\lambda(k_j)$ as $n \to \infty$, but $\lambda(k_3)$ diverges (as do the $\lambda_n(k_j)$ for $j \geq 3$).*

found by matching with the leading-order outer solution, these recurrence relations can be solved numerically by discretizing the k-dependence of $\bar{X}_n$ and $\bar{Y}_n$ in some way. The integrals in (B.4)–(B.5) can be computed efficiently using Gauss–Hermite quadrature, so we take the discrete values k_j, $j = 0, 1, 2, \ldots, M$, of the wavenumber to be the nonnegative zeros of a Hermite polynomial of sufficiently high order $2M + 1$. We thus obtain $2M + 5$ coupled, nonlinear recurrence relations for the coefficients U_n, V_n, W_n, $\bar{X}_n(k_j)$, and $\bar{Y}_n(k_j)$ which we solve numerically.

What matters for the estimation of the Stokes multiplier $C(k)$ is the behavior of these coefficients for $n \gg 1$. We can concentrate on $\bar{X}_n$. By analogy with the treatment of the LK model [34], we expect the linear terms in the recurrence to dominate for $n \gg 1$, leading to the asymptotics

$$\bar{X}_n \sim \frac{(-1)^{n+1}(2n-1)!\lambda(k)}{2\pi\Omega^{2n}} \tag{B.9}$$

for some function $\lambda(k)$ which depends on the early-term behavior and needs to be determined numerically. (The factor 2π in the denominator is introduced for convenience.) The numerical determination of $\lambda(k)$ is achieved by computing

$$\lambda_n(k) = \frac{(-1)^{n+1}2\pi\Omega^{2n}}{(2n-1)!}\bar{X}_n \tag{B.10}$$

for n large enough. Our computations indicate that the convergence $\lambda_n(k) \to \lambda(k)$ occurs only for k small enough; in other words, the asymptotic behavior (B.9) holds true only for small k. This is illustrated by Figure 9, obtained for $a = 2$, $b = 1$, and $M = 50$. The divergence for k of order unity occurs because for such k, the recurrence relations are not dominated by their linear parts for $n \gg 1$ as is assumed to obtain (B.9). This divergence is

inconsequential, however: as shown below, the asymptotics of $x_{\rm w}$ for $\epsilon \to 0$ is controlled by the small-k behavior of $\lambda(k)$, indeed by $\lambda(0)$.

Assuming this, we rely on (B.9) and use a Borel-summation technique [16, 17] to relate the behavior of $\bar{X}(k,\tau)$ for $\mathrm{Re}\,\tau < 0$ and $\mathrm{Re}\,\tau > 0$. We write (formally)

$$\bar{X} = \int_0^\infty \frac{\mathrm{e}^{-s}}{s} B(s/\tau)\,\mathrm{d}s, \quad \text{where} \quad B(\cdot) = \sum_{n=1}^{\infty} \frac{\bar{X}_n}{(2n-1)!}(\cdot)^{2n}, \tag{B.11}$$

and note that the late behavior (B.9) is associated with the existence of a pole of $B(s/\tau)$, in the neighborhood of which

$$B(s/\tau) \sim \frac{\lambda(k)s^2}{2\pi(s^2+\Omega^2\tau^2)}.$$

This pole yields an oscillatory (or wave) contribution to (B.11) for $\mathrm{Re}\,\tau > 0$, i.e., after the Stokes line is crossed. Evaluating this leads to the wave part of $\bar{X}$ in the form

$$\bar{X}_{\rm w} = -\frac{\mathrm{i}\lambda(k)}{2}\mathrm{e}^{-\mathrm{i}\Omega\tau} = -\frac{\mathrm{i}\lambda(k)}{2}\mathrm{e}^{-\mathrm{i}\omega(t-t_*)},$$

with $\omega > 0$. Taking into account the contribution of the complex-conjugate pole $\bar{t}_*$ and matching with (B.3) give the wave amplitude (4.8).

Multiplying (4.8) by $\exp(\mathrm{i}ks)$ and integrating over k then give an estimate for $x_{\rm w}$. The smallness of ϵ can be used to approximate the resulting integral using the steepest-descent method. The saddle is located at $k = 0$, indicating that wavenumbers $k = O(\epsilon^{1/2})$ dominate the integral. This suggests the natural rescaling of the spatial variable as $S = \epsilon^{1/2}s$. Assuming this to be $O(1)$, we obtain (4.9). Note that an estimate valid uniformly in s could be obtained by including the term $\mathrm{i}ks$ in the saddle-point calculation.

Acknowledgments. The author thanks O. Bokhove and D. Wirosoetisno for useful comments on the manuscript.

REFERENCES

[1] M. J. Ablowitz and A. S. Fokas, *Complex Variables: Introduction and Applications*, Cambridge University Press, Cambridge, UK, 1997.
[2] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965.
[3] M. V. Berry, *Uniform asymptotic smoothing of Stokes's discontinuities*, Proc. Roy. Soc. London Ser. A, 422 (1989), pp. 7–21.
[4] L. Blanchet, *Gravitational radiation from post-Newtonian sources and inspiralling compact binaries*, Living Rev. Relativ., 5 (2002), p. 3. Also available online from http://www.livingreviews.org/lrr-2002-3.
[5] O. Bokhove, *Balanced models in geophysical fluid dynamics: Hamiltonian formulation, constraints and formal stability*, in Large-Scale Atmosphere-Ocean Dynamics, Vol. II, Geometric Methods and Models, I. Roulstone and J. Norbury, eds., Cambridge University Press, Cambridge, UK, 2002, pp. 1–63.
[6] O. Bokhove and T. G. Shepherd, *On Hamiltonian balanced dynamics and the slowest invariant manifold*, J. Atmospheric Sci., 53 (1996), pp. 276–297.
[7] T. J. Bridges and S. Reich, *Multi-symplectic integrators: Numerical schemes for Hamiltonian PDEs that conserve symplecticity*, Phys. Lett. A, 284 (2001), pp. 183–193.
[8] R. Camassa, *On the geometry of an atmospheric slow manifold*, Phys. D, 84 (1995), pp. 357–397.

[9] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[10] A. Einstein, *Über Gravitationswellen*, Sitzunberg. K. Preuss. Akad. Wiss., 1918.
[11] R. Ford, *Gravity wave radiation from vortex trains in rotating shallow water*, J. Fluid Mech., 281 (1994), pp. 81–118.
[12] R. Ford, M. E. McIntyre, and W. A. Norton, *Balance and the slow quasi-manifold: Some explicit results*, J. Atmospheric Sci., 57 (2000), pp. 1236–1254.
[13] R. Ford, M. E. McIntyre, and W. A. Norton, *Reply to comments by S. Saujani and T. G. Shepherd on "Balance and the slow quasi-manifold: Some explicit results,"* J. Atmospheric Sci., 59 (2002), pp. 2878–2882.
[14] V. Gelfreich and L. Lerman, *Almost invariant elliptic manifold in a singularly perturbed Hamiltonian system*, Nonlinearity, 15 (2002), pp. 447–457.
[15] T. Hagstrom, *Radiation boundary conditions for the numerical simulation of waves*, in Acta Numerica 1999, Acta Numer. 8, Cambridge University Press, Cambridge, UK, pp. 47–106.
[16] V. Hakim, *Computation of transcendental effects in growth problems: Linear solvability conditions and nonlinear methods—the example of the geometric model*, in Asymptotics beyond All Orders, H. Segur, S. Tanveer, and H. Levine, eds., Plenum Press, New York, 1991, pp. 15–28.
[17] V. Hakim, *Asymptotic techniques in nonlinear problems: Some illustrative examples*, in Hydrodynamics and Nonlinear Instabilities, C. Godrèche and P. Manneville, eds., Cambridge University Press, Cambridge, UK, 1998, pp. 295–386.
[18] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields*, 4th ed., Pergamon, Oxford, UK, 1975.
[19] L. D. Landau and E. M. Lifschitz, *Fluid Mechanics*, 2nd ed., Pergamon, Oxford, UK, 1987.
[20] M. J. Lighthill, *On sound generated aerodynamically,* I. *General theory*, Proc. Roy. Soc. London Ser. A, 211 (1952), pp. 564–587.
[21] E. N. Lorenz, *On the existence of a slow manifold*, J. Atmospheric Sci., 43 (1986), pp. 1547–1557.
[22] E. N. Lorenz and V. Krishnamurthy, *On the nonexistence of a slow manifold*, J. Atmospheric Sci., 44 (1987), pp. 2940–2950.
[23] P. Lynch, *The swinging spring: A simple model for atmospheric balance*, in Large-Scale Atmosphere-Ocean Dynamics, Vol. II, Geometric Methods and Models, I. Roulstone and J. Norbury, eds., Cambridge University Press, Cambridge, UK, 2002, pp. 64–108.
[24] R. S. MacKay, *Slow manifolds*, in Energy Localization and Transfer, T. Dauxois, A. Litvak-Hinenzon, R. S. MacKay, and A. Spanoudaki, eds., World Scientific, River Edge, NJ, 2004, pp. 149–192.
[25] M. E. McIntyre and I. Roulstone, *Are there higher-accuracy analogues of semi-geostrophic theory?*, in Large-Scale Atmosphere-Ocean Dynamics, Vol. II, Geometric Methods and Models, I. Roulstone and J. Norbury, eds., Cambridge University Press, Cambridge, UK, 2002, pp. 301–364.
[26] J. C. McWilliams, J. M. Molemaker, and I. Yavneh, *From stirring to mixing of momentum: Cascades from balanced flows to dissipation in the oceanic interior*, in Aha Huliko'a Proceedings 2001, University of Hawaii, Honolulu, HI, 2001, pp. 59–66.
[27] E. I. Ólafsdóttir, A. B. Olde Daalhuis, and J. Vanneste, *Multiple Stokes multipliers in an inhomogeneous differential equation with a small parameter*, Proc. Roy. Soc. London Ser. A, 461 (2005), pp. 2243–2256.
[28] R. B. Paris and A. D. Wood, *Stokes phenomenon demystified*, Bull. Inst. Math. Appl., 31 (1995), pp. 21–28.
[29] G. M. Reznik, V. Zeitlin, and M. B. Jelloul, *Nonlinear theory of geostrophic adjustment. Part* 1. *Rotating shallow-water model*, J. Fluid Mech., 445 (2001), pp. 93–120.
[30] C. W. Rowley and T. Colonius, *Discretely nonreflecting boundary conditions for linear hyperbolic systems*, J. Comput. Phys., 157 (2000), pp. 500–538.
[31] S. Saujani and T. G. Shepherd, *Comments on "Balance and the slow quasi-manifold: Some explicit results,"* J. Atmospheric Sci., 59 (2002), pp. 2874–2877.
[32] H. Segur, S. Tanveer, and H. Levine, eds., *Asymptotics beyond All Orders*, Plenum Press, New York, 1991.
[33] N. G. van Kampen, *Elimination of fast variables*, Phys. Rep., 124 (1985), pp. 69–160.
[34] J. Vanneste, *Inertia-gravity-wave generation by balanced motion: Revisiting the Lorenz–Krishnamurthy model*, J. Atmospheric Sci., 61 (2004), pp. 224–234.

[35] J. Vanneste and I. Yavneh, *Exponentially small inertia-gravity waves and the breakdown of quasi-geostrophic balance*, J. Atmospheric Sci., 61 (2004), pp. 211–223.
[36] T. Warn, O. Bokhove, T. G. Shepherd, and G. K. Vallis, *Rossby number expansions, slaving principles, and balance dynamics*, Quart. J. R. Met. Soc., 121 (1995), pp. 723–739.
[37] G. B. Whitham, *Linear and Nonlinear Waves*, Wiley, New York, 1974.
[38] D. Wirosoetisno, *Exponentially accurate balance dynamics*, Adv. Differential Equations, 9 (2004), pp. 177–196.
[39] D. Wirosoetisno and T. G. Shepherd, *Averaging, slaving, and balance dynamics in a simple atmospheric model*, Phys. D, 141 (2000), pp. 37–53.