SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 1, No. 1, pp. 1–43

Hamiltonian Systems Near Relative Periodic Orbits[*]

Claudia Wulff[†] and Mark Roberts[‡]

Abstract. We give explicit differential equations for a symmetric Hamiltonian vector field near a relative periodic orbit. These decompose the dynamics into periodically forced motion in a Poincaré section transversal to the relative periodic orbit, which in turn forces motion along the group orbit. The structure of the differential equations inherited from the symplectic structure and symmetry properties of the Hamiltonian system is described, and the effects of time reversing symmetries are included. Our analysis yields new results on the stability and persistence of Hamiltonian relative periodic orbits and provides the foundations for a bifurcation theory. The results are applied to a finite dimensional model for the dynamics of a deformable body in an ideal irrotational fluid.

Key words. relative periodic orbits, equivariant Hamiltonian systems, noncompact groups

AMS subject classifications. 37J15, 37J20, 53D20, 70H33

PII. S1111111101387760

1. Introduction.

Relative periodic orbits are periodic solutions of a flow induced by an equivariant vector field on a space of group orbits. In applications they typically appear as oscillations of a system which are periodic when viewed in some rotating or translating frame. They therefore generalize relative equilibria, for which the "shape" of the system remains constant in an appropriate frame. Relative periodic orbits are ubiquitous in Hamiltonian systems with symmetry. For example, generalizations of the Weinstein–Moser theorem show that they are typically present near stable relative equilibria [25, 39, 43] and can therefore be found in virtually any physical application with a continuous symmetry group. Specific examples for which relative periodic orbits have been discussed or could be found by applying the Weinstein–Moser theorem to stable relative equilibria include rigid bodies [1, 31, 28, 24], deformable bodies [8, 27, 13], gravitational N-body problems [32, 47], molecules [17, 19, 20, 34, 48], and point vortices [26, 46, 38].

Existing theoretical work on Hamiltonian relative periodic orbits includes results on their stability [41, 42] and on their persistence to nearby energy-momentum levels in the case of compact symmetry groups [33]. However, stability, persistence, and bifurcations are still a long way from being well understood, especially in the presence of actions of noncompact symmetry groups with nontrivial isotropy subgroups. Our main aim with this paper is to

[*] Received by the editors April 12, 2001; accepted for publication (in revised form) by J. Marsden October 3, 2001; published electronically April 8, 2002. This work was partially supported by UK EPSRC Research grant GR/L60029 and by European Commission funding for the Research Training Network "Mechanics and Symmetry in Europe (MASIE)."

http://www.siam.org/journals/siads/1-1/38776.html

[†] Institut für Mathematik I, Freie Universität Berlin, Arnimallee 2-6, 14195 Berlin, Germany (wulff@math.fu-berlin.de).

[‡] Mathematics and Statistics, University of Surrey, Guildford GU2 7XH, UK (m.roberts@surrey.ac.uk).

provide a local description of Hamiltonian vector fields near relative periodic orbits that can be used to develop stability and bifurcation theories.

In [55], the "bundle" structure of a general vector field near a relative periodic solution is analyzed for proper actions of arbitrary Lie groups. The dynamics near a relative periodic orbit is decomposed as periodically forced motion in a Poincaré section, which in turn forces motion along the group orbit. In this way, the study of bifurcations from relative periodic solutions is reduced to the study of bifurcations from discrete rotating waves in systems with compact symmetry. These are treated in [23]. The aim of this paper is to extend the results of [55] from general systems to Hamiltonian systems by taking into account the symplectic structure and conserved quantities of the problem.

In addition to [55], we draw on several other sources for inspiration. In the absence of symmetry, it is well known that the dynamics near a Hamiltonian periodic orbit can be described as periodically forced motion in a Poincaré section within an energy level set [2]. Finite symmetries are treated in [7]. We combine these ideas with the local description of Hamiltonian vector fields near a relative equilibrium given in [50], and, since these are present in most Hamiltonian systems, we include the effects of time reversing symmetries by extending the paper [22] to Hamiltonian vector fields.

In section 3.1, we show that a Poincaré section transverse to a relative periodic orbit of a Hamiltonian system decomposes into a part tangent to the energy level set describing rigid body motion, another part tangent to the energy level set describing vibrational motion, and a part parametrizing energy. Then, in section 3.2, we present our central result, the differential equations in these bundle coordinates. In section 4, we use them to deduce a number of new results on stability and persistence. These include Proposition 4.3, describing the block structure of the linearization of a Hamiltonian vector field at a relative periodic orbit, the stability result Corollary 4.5, Corollary 4.8 on the persistence of relative periodic orbits with generic momenta to nearby energy-momentum level sets, and Theorem 4.9 on persistence in the case of nongeneric momenta and finite isotropy subgroups, in the spirit of [33, 40]. Whereas the results of [33, 40] build on topological methods which require compact symmetry groups, our persistence results apply to noncompact symmetry groups as well. (For a more detailed comparison, see section 4.2.2.) Moreover, we will see that in the case of generic momenta, bifurcations from relative periodic orbits reduce to fixed point bifurcations of symplectic maps which are twisted semiequivariant with respect to compact symmetry groups. The latter bifurcations are studied, for example, in [6, 9] for equivariant symplectic maps and in [10] for reversible symplectic maps. All of these results are simple corollaries of the bundle equations. We will present a more general and detailed study of bifurcations and persistence in future work.

In this paper, we restrict our attention to algebraic symmetry groups. These are groups defined by polynomial equations and include compact, Euclidean, and the classical Lie groups, so this assumption is usually satisfied in applications. If this assumption is not satisfied, then there might be no comoving frame in which the relative periodic orbit becomes periodic (i.e., Lemma 2.1 would not apply). Since in this case the bundle structure near relative periodic orbits already becomes more complicated for general systems [55], we deal only with algebraic symmetry groups in this paper.

The paper is organized as follows. In section 2, we recall the bundle structure theorem of

[55, 22] on relative periodic orbits of general systems. In section 3, we study this structure for Hamiltonian systems and present the differential equations in bundle coordinates. In section 4, we discuss linear stability, persistence, and bifurcation from relative periodic solutions. In section 5, we illustrate these results with an application to the dynamics of a finite dimensional model of a deformable body in an ideal irrotational fluid. Section 6 is devoted to the proofs of the bundle structure theorems of section 3.

2. Relative periodic orbits in general systems.

In this section, we recall the results of [52], [55], and [22], giving a parametrization of a manifold in the neighborhood of a relative periodic orbit of a reversible equivariant vector field and the form that the vector field takes in these coordinates.

2.1. Reversible equivariant vector fields.

We consider an ordinary differential equation on a manifold $\mathcal{M}$,

$$\frac{dx}{dt} = f(x), \quad x \in \mathcal{M}, \tag{2.1}$$

that is equivariant with respect to a smooth, proper action of a finite dimensional algebraic Lie group Γ on $\mathcal{M}$:

$$\gamma f(x) = f(\gamma x) \quad \text{for all} \quad \gamma \in \Gamma. \tag{2.2}$$

If $x(t)$ is a solution of the equation and $\gamma \in \Gamma$, then $\gamma x(t)$ is also a solution. We call a diffeomorphism γ satisfying (2.2) a *symmetry* of the vector field $f(x)$.

We also include the possibility that the vector field is *reversible*, i.e., there exists a *reversing symmetry* ρ such that

$$\rho f(x) = -f(\rho x). \tag{2.3}$$

This implies that if $x(t)$ is a solution of (2.1), then so is $\rho x(-t)$. Note that if ρ is a reversing symmetry, then $\rho\gamma$ is also a reversing symmetry for every $\gamma \in \Gamma$.

In the reversible case, the symmetries and reversing symmetries together form the *reversing symmetry group* G of the vector field. The group of symmetries, Γ, is a normal subgroup of G of index two, i.e., the quotient G/Γ is isomorphic to $\mathbb{Z}_2$. It is useful to describe this structure by introducing a character (group homomorphism) $\chi : G \mapsto \{\pm 1\}$, such that $\chi(\gamma) = 1$ for all $\gamma \in \Gamma$, and $\chi(\rho) = -1$ for all $\rho \in G \setminus \Gamma$. This map is called a *reversible sign* or *temporal character* [50, 35]. Using this notation, (2.2) and (2.3) are equivalent to the single equation

$$gf(x) = \chi(g) f(gx) \quad \text{for all} \quad g \in G. \tag{2.4}$$

We say that a vector field f satisfying (2.4) is (infinitesimally) (G, χ)-*reversible-equivariant* or (G, χ)-*semiequivariant*. The corresponding flow $\Phi_t(\cdot)$ is (G, χ)-semiequivariant in the sense of diffeomorphisms:

$$g\Phi_t = (\Phi_t)^{\chi(g)} g \qquad \text{for all } g \in G \text{ and } t \in \mathbb{R}. \tag{2.5}$$

We usually omit the (G, χ) prefix when it is obvious from the context.

Throughout this paper, we assume that the symmetry group Γ is algebraic. Algebraic groups include all compact and Euclidean groups, and so the assumption is usually satisfied in applications.

2.2. Symmetry groups of relative periodic orbits. In this subsection, we recall the notion of a relative periodic orbit and its symmetry groups. Denote the isotropy subgroup of a point p in $\mathcal{M}$ by G_p:

$$G_p \;=\; \{g \in G \mid gp = p\},$$

and let $\Gamma_p = G_p \cap \Gamma$. Either $\Gamma_p = G_p$ or Γ_p is a normal subgroup of G_p of index 2. The groups G_p and Γ_p are compact because the action of G on $\mathcal{M}$ is assumed to be proper.

A solution $x(t)$ of (2.1) is said to lie on a *relative periodic orbit* if there exists $T > 0$ such that $x(T)$ lies in the group orbit $\Gamma x(0)$ of $x(0) = p$, i.e., there exists $\sigma \in \Gamma$ with

$$\Phi_T(p) \;=\; \sigma p.$$

The infimum of the numbers T with this property is called the *relative period* of the relative periodic orbit, and the corresponding σ is called a *spatio-temporal symmetry*, *phase-shift symmetry*, or *reconstruction phase* [5, 30, 31] of the relative periodic orbit with respect to p. Note that σ determines the drift direction of the relative periodic orbit. A simple calculation shows that σ must lie in $N_\Gamma(\Gamma_p)$, the normalizer of Γ_p in Γ. We will always assume that time has been parametrized so that the relative period is 1 and so $\Phi_1(p) = \sigma p$.

The relative periodic orbit itself is defined to be the submanifold of $\mathcal{M}$ given by

$$\mathcal{P} \;=\; \{\gamma\Phi_t(p) \mid \gamma \in \Gamma, \quad t \in \mathbb{R}\}.$$

Thus relative periodic orbits are periodic orbits for the induced flow on the space of orbits of the action of Γ on $\mathcal{M}$, just as relative equilibria are equilibria in the space of group orbits.

Note that the (G, χ)-semiequivariance of the flow on $\mathcal{M}$ does *not* imply that the flow descends to a flow on the space of orbits for the full action of G, and so it does not make sense to replace Γ by G in the definition of a relative periodic orbit. However, we can consider the action of $G/\Gamma \cong \mathbb{Z}_2$ on the space of Γ orbits and define a relative periodic orbit to be *reversible* if it is invariant under this action and to be *nonreversible* otherwise. If $\mathcal{P}$ is nonreversible, then $G_p = \Gamma_p$ for all $p \in \mathcal{P}$. It is shown in [22] that $\mathcal{P}$ is a reversible relative periodic orbit if and only if there exists a point $p \in \mathcal{P}$ such that G_p contains a reversing symmetry, and so Γ_p is a normal subgroup of G_p of index two. We call such a point a *brake point* of the relative periodic orbit and will always choose p in such a way. Moreover, it is easily shown that the spatio-temporal symmetry σ of a reversible relative periodic orbit satisfies $\rho\sigma\rho^{-1} \in \sigma^{-1}\Gamma_p$ for each $\rho \in G_p \setminus \Gamma_p$.

Examples of both reversible and nonreversible relative periodic orbits are provided by the (relative) nonlinear normal modes of relative equilibria. If the relative equilibrium is not reversible, then none of its normal modes will be reversible. If the relative equilibrium is elliptic, nonresonant, and reversible for some involutory reversing symmetry, then its normal modes are also reversible. Consider, for example, the relative equilibria of an ellipsoidal rigid body in an irrotational, ideal fluid modelled by Kirchhoff's equations [24]. Assume that the body is neutrally buoyant but has noncoincident centers of gravity and buoyancy so that it "feels" gravity. Then relative equilibria for which the body is translating vertically are not reversible since the time reversed motion cannot be obtained by a symmetry transformation which preserves the direction of gravity [56]. However, horizontally translating relative equilibria and their normal modes are reversible. Other examples of reversible relative periodic orbits of neutrally buoyant ellipsoidal deformable bodies in irrotational, ideal fluids can be found in section 5.1.

Let $\mathbf{g}$ denote the Lie algebra of Γ and hence also of G. The adjoint action of G on $\mathbf{g}$ is defined by

$$\mathrm{Ad}_g(\xi) \;=\; g\xi g^{-1}.$$

We define the χ-*dual* of any representation of G to be the new representation obtained by composing the map representing $g \in G$ with $\chi(g)$. Let $Z(g)$ denote the centralizer of $g \in G$, and let $\mathbf{z}(g)$ denote its Lie algebra. For a subgroup K of G, let $\mathbf{z}^\chi(K)$ denote the centralizer, or fixed point subspace, of K in $\mathbf{g}$ with respect to the χ-dual action of K on $\mathbf{g}$:

$$\mathbf{z}^\chi(K) = \{\ \xi \in \mathbf{g}\ :\ \ \chi(g)\mathrm{Ad}_g\xi = \xi \ \text{ for all } g \in K\ \}.$$

The following lemma states that every relative periodic orbit in a system with an algebraic symmetry group becomes periodic in a comoving frame which respects the isotropy of the relative periodic orbit.

Lemma 2.1 (see [22]). *Assume that Γ is an algebraic Lie group, and let $\tilde\sigma \in \Gamma$ be a spatio-temporal symmetry of a relative periodic orbit $\mathcal{P}$ with respect to $p \in \mathcal{P}$. Then there exists a choice of σ in $\tilde\sigma\Gamma_p$ and $\alpha \in \Gamma$, $\xi \in \mathbf{z}(\sigma)$, and $n \in \mathbb{N}$ such that*

$$\sigma = \alpha\exp(\xi), \quad \alpha^n = 1, \quad \textit{and} \quad \xi \in \mathbf{z}^\chi(G_p).$$

If Γ is not algebraic, then the conclusions of Lemma 2.1 are in general not satisfied, the bundle structure near relative periodic orbits becomes more complicated [55, 22], and Theorem 2.1 on the bundle structure near relative periodic orbits does not apply. Since most groups in applications are algebraic, we restrict our attention to such symmetry groups.

Following [22], we define the *twist diffeomorphism* $\phi : G_p \to G_p$ determined by $\sigma \in \Gamma$ to be

$$\phi(g_p) \;=\; \sigma^{-1}g_p\sigma^{\chi(g_p)}. \tag{2.6}$$

Lemma 2.1 implies that there is a choice of σ in $\tilde\sigma\Gamma_p$ such that the order of ϕ is finite and that we may replace σ by α in the definition of ϕ. If we denote the order of ϕ by k, then k divides n. In general, ϕ is not a group automorphism. However, its restriction $\phi|_{\Gamma_p}$ is the automorphism of Γ_p given by

$$\phi(\gamma_p) \;=\; \sigma^{-1}\gamma_p\sigma \qquad \text{for all } \gamma_p \in \Gamma_p.$$

For any multiple r of k, we define the group L_r to be the index r extension of G_p by an abstract element Q of order r such that

$$Q^{-1}g_pQ^{\chi(g_p)} \;=\; \phi(g_p) \quad \text{for all } g_p \in G_p. \tag{2.7}$$

If an operator Q satisfies this equation, we say that its inverse Q^{-1} is *twisted semiequivariant* or twisted reversible equivariant [22]. Replacing Q by α identifies L_n with the subgroup of G generated by G_p and α. For orientability reasons the index two extension $L_{2n} = L_n \times \mathbb{Z}_2$

of this group is needed in the results below. We call the groups L_r *reduced spatio-temporal symmetry groups* of the relative periodic orbit because the group L_{2n} or L_n (if the bundle is orientable) is the spatio-temporal symmetry group of the periodic orbit for the symmetry reduced dynamics; cf. section 2.3. We will label elements of L_r by pairs (g_p, i), where $g_p \in G_p$ and $i \in \mathbb{Z}_r$.

If the relative periodic orbit $\mathcal{P}$ is nonreversible and so $\Gamma_p = G_p$ for all $p \in \mathcal{P}$, then $L_r = \Lambda_r := \Gamma_p \rtimes \mathbb{Z}_r$. If the relative periodic orbit is reversible, we have $L_r = (\Lambda_r)_\rho$, where $\rho \in G_p \setminus \Gamma_p$ and $(\Lambda_r)_\rho$ is the index two extension of Λ_r generated by $\rho \in G_p \setminus \Gamma_p$ using (2.7). For a reversible relative periodic orbit, the group L_r/Γ_p is isomorphic to the dihedral group of order $2r$, $\mathbb{D}_{2r}$, while for a nonreversible relative periodic orbit $L_r/\Gamma_p \cong \mathbb{Z}_r$.

2.3. Differential equations near a relative periodic orbit.

The following theorems describe the bundle structure near a relative periodic orbit and the form that the differential equations (2.1) take in coordinates adapted to this structure. As mentioned in the introduction, these coordinates decompose the dynamics into a periodically forced motion inside a Poincaré section N which drives drift dynamics on the group.

Theorem 2.1 (see [55, 22]). *Let p lie on a relative periodic orbit $\mathcal{P}$ with relative period 1 so that $\Phi_1(p) = \sigma p$ for some $\sigma \in \Gamma$. If $\mathcal{P}$ is reversible, assume p is a brake point. Let $\sigma = \alpha \exp(\xi)$ as in Lemma 2.1. Then in a frame moving uniformly with velocity ξ, a G-invariant neighborhood $\mathcal{U}$ of $\mathcal{P}$ in $\mathcal{M}$ can be parametrized by*

$$\mathcal{U} \;\equiv\; (G \times \mathbb{R}/2n\mathbb{Z} \times N)/L_{2n}, \tag{2.8}$$

where N is a G_p-invariant complement to $T_p\mathcal{P}$ in $T_p\mathcal{M}$ at $p = (\mathrm{id}, 0, 0)$ and the quotient by L_{2n} is with respect to the following action of L_{2n} on $G \times \mathbb{R}/2n\mathbb{Z} \times N$:

$$(g_p, i)(g, \theta, v) \;=\; (\; g\alpha^{-i} g_p^{-1},\; \chi(g_p)(\theta + i),\; g_p Q_N^i v \;) \quad \textit{for all} \quad g_p \in G_p,\; i \in \mathbb{Z}_{2n}. \tag{2.9}$$

Here Q_N is a linear transformation of N of order $2n$ which is orthogonal with respect to a G_p-invariant inner product on N and such that Q_N^{-1} is G_p twisted semiequivariant.

Note that N is a Poincaré section transverse to the relative periodic orbit $\mathcal{P}$ at p. The transformation Q_N is determined by the linear map $\sigma^{-1}\mathrm{D}\Phi_1(p)$ at the relative periodic orbit; for details see [55, 22] and section 6.1 of this paper. In some cases, the action of L_{2n} can be replaced by an action of L_n, and the transformation Q can be chosen to have order n; see [55]. Whether or not this is possible depends on orientability properties of the bundle. For Hamiltonian systems it is always possible, as we will see in section 3.1.

The following theorem describes how the differential equation (2.1) lifts to a differential equation on $G \times \mathbb{R}/2n\mathbb{Z} \times N$ under the isomorphism given by Theorem 2.1.

Theorem 2.2 (see [55, 22]). *The differential equations in coordinates adapted to the bundle structure given by (2.8) have the form*

$$\begin{array}{rcl} \dot{g} & = & \chi(g) g f_G(\theta, v), \\ \dot{\theta} & = & \chi(g) f_\Theta(\theta, v), \\ \dot{v} & = & \chi(g) f_N(\theta, v), \end{array} \tag{2.10}$$

where f_G, f_Θ, and f_N are functions on $\mathbb{R}/2n\mathbb{Z}\times N$ taking values in $\mathbf{g}$, $\mathbb{R}$, and N, respectively, and are L_{2n}-semiequivariant:

$$\begin{aligned} f_G(\chi(g_p)\theta, g_pv) &= \chi(g_p)\mathrm{Ad}_{g_p}f_G(\theta,v), & f_G(\theta+1,Q_Nv) &= \mathrm{Ad}_\alpha f_G(\theta,v),\\ f_\Theta(\chi(g_p)\theta, g_pv) &= f_\Theta(\theta,v), & f_\Theta(\theta+1,Q_Nv) &= f_\Theta(\theta,v),\\ f_N(\chi(g_p)\theta, g_pv) &= \chi(g_p)g_pf_N(\theta,v), & f_N(\theta+1,Q_Nv) &= Q_Nf_N(\theta,v) \end{aligned} \tag{2.11}$$

for all $g_p \in G_p$.

Note that the vector field is in fact determined by its restriction to a Γ-invariant neighborhood of $\mathcal{P}$ in $\mathcal{M}$, and so the equations in the theorem can be restricted to $g \in \Gamma$. The coefficients $\chi(g)$ then "disappear" from the equations.

The (θ, v) equations form a closed subsystem that is semiequivariant with respect to the action of L_{2n} on $\mathbb{R}/2n\mathbb{Z}\times N$. In particular, f_N and f_Θ are $2n$-periodic in θ, and by a time reparametrization we can assume that $f_\Theta \equiv 1$ so that we obtain a periodically forced equation $\dot v = f_N(t, v)$ on the Poincaré section N. Furthermore, the relative periodic orbit $\mathcal{P}$ of (2.1) reduces to a periodic orbit of the (θ, v) subsystem with a finite order phase shift symmetry, a *discrete rotating wave* [23]. Thus the study of bifurcations from (reversible) relative periodic orbits reduces to that of bifurcations from (reversible) discrete rotating waves. For general nonreversible non-Hamiltonian systems, these are studied in [23].

3. Relative periodic orbits of Hamiltonian systems.

In this section, we combine the local bundle structure near relative periodic orbits of general systems described in section 2 with the methods used in [50] to obtain equations near Hamiltonian relative equilibria and thereby obtain local descriptions of Hamiltonian systems of equations near relative periodic orbits.

We consider a Hamiltonian ordinary differential equation on a smooth finite dimensional symplectic manifold $\mathcal{M}$ with symplectic two-form ω. For each $x \in \mathcal{M}$, the restriction of ω to the tangent space $T_x\mathcal{M}$ is denoted by ω_x. Let G be a finite dimensional Lie group, let $\chi : G \to \mathbb{Z}_2$ be a group homomorphism, and let $\Gamma = \ker\chi$. We say that G acts χ-*semisymplectically* on $\mathcal{M}$ if [35, 50]

$$\omega_{gx}(gu, gv) \;=\; \chi(g)\,\omega_x(u,v) \qquad \text{for all} \quad x\in\mathcal{M},\ g\in G,\ u,v\in T_x\mathcal{M}.$$

A Hamiltonian vector field

$$\dot x = f_H(x) \tag{3.1}$$

is generated by a smooth function, the *Hamiltonian* $H : \mathcal{M} \to \mathbb{R}$, via the relationship

$$\omega_x(f_H(x), v) \;=\; \mathrm{D}H(x)v, \qquad x\in\mathcal{M},\ v\in T_x\mathcal{M}. \tag{3.2}$$

If H is invariant under the action of G, then the vector field f_H is (G, χ)-semiequivariant. As before, we denote the flow of (3.1) by $\Phi_t(\cdot)$.

By Noether's theorem, locally there is a conserved quantity $\mathbf{J}_\xi$ for each continuous symmetry $\xi \in \mathbf{g}$ of the system; see, e.g., [1]. The map $\mathbf{J}_\xi(x) = \mathbf{J}(x)(\xi)$ is linear in ξ so that $\mathbf{J}$ is a map from a neighborhood of $x \in \mathcal{M}$ to $\mathbf{g}^*$, called a momentum map. Here $\mathbf{g}^*$ is the dual of the Lie algebra $\mathbf{g}$ of G. We assume that the momentum map $\mathbf{J} : \mathcal{M} \to \mathbf{g}^*$ exists globally and is G-equivariant with respect to the action of G on $\mathcal{M}$ and the χ-dual of the coadjoint action, or χ-coadjoint action, of G on $\mathbf{g}^*$ [35, 50]:

$$\mathbf{J}(gx) \;=\; \chi(g)(\mathrm{Ad}_g^*)^{-1}\mathbf{J}(x), \qquad x \in \mathcal{M},\; g \in G.$$

Here Ad_g^* is the dual operator to Ad_g, i.e., $\mathrm{Ad}_g^*\mu(\xi) = \mu(\mathrm{Ad}_g\xi)$ for all $\mu \in \mathbf{g}^*, \xi \in \mathbf{g}$. Note that, since we are interested only in the dynamics inside a G-invariant neighborhood $\mathcal{U}$ of the relative periodic orbit $\mathcal{P}$, it suffices to make the above assumptions about the momentum map on $\mathcal{U}$ or, alternatively, to set $\mathcal{M} := \mathcal{U}$.

Let $\mathcal{P}$ be a relative periodic orbit of the Hamiltonian system (3.1) of relative period 1, and assume that $p \in \mathcal{P}$ satisfies $\Phi_1(p) = \sigma p$ for $\sigma \in \Gamma$. If the relative periodic orbit is reversible, assume that p is a brake point. We will assume, without loss of generality, that $H(p) = 0$. As before, let G_p denote the isotropy subgroup of p. Let $\mu = \mathbf{J}(p)$ be the momentum of the point p, and let

$$G_\mu \;=\; \{g \in G \;:\; \chi(g)\mathrm{Ad}_g^{*-1}\mu = \mu\}$$

be the momentum isotropy subgroup for the χ-dual of the coadjoint action of G on $\mathbf{g}^*$.

3.1. Bundle structure near Hamiltonian relative periodic orbits.

As before, we assume that the symmetry group Γ is algebraic. Theorem 2.1 describes the bundle structure near relative periodic orbits of general (G, χ)-semiequivariant vector fields. In this subsection, we will describe the additional structure that is present for Hamiltonian systems.

Let $\mathcal{P}$ be a relative periodic orbit of (3.1), and let $p = \sigma^{-1}\Phi_1(p) \in \mathcal{P}$, $\sigma \in \Gamma$. Note that $G_p \subset G_\mu$, $\Gamma_p \subset \Gamma_\mu$ and that $\sigma \in \Gamma_\mu$ since

$$\sigma\mu \;=\; \sigma\mathbf{J}(p) \;=\; \mathbf{J}(\sigma p) \;=\; \mathbf{J}(\Phi_1(p)) \;=\; \mathbf{J}(p) \;=\; \mu.$$

As $\sigma \in N(\Gamma_p)$, we conclude that $\sigma \in N_{\Gamma_\mu}(\Gamma_p) = N(\Gamma_p) \cap \Gamma_\mu$, which gives a restriction on possible drift directions of Hamiltonian relative periodic orbits, as we will see in the examples in section 5.4. If Γ is algebraic, then so is $\Gamma_\mu = \{\gamma \in \Gamma, \gamma\mu = \mu\}$, and it follows immediately from Lemma 2.1 that we can choose σ such that it decomposes as $\sigma = \alpha\exp(\xi)$ with

$$\alpha \in \Gamma_\mu, \quad \alpha^n = 1, \quad \xi \in \mathbf{g}_\mu \cap \mathbf{z}(\sigma) \cap \mathbf{z}^\chi(G_p). \tag{3.3}$$

As before, identify $L_n \subset G_\mu$ with the compact group generated by α and G_p. Choose L_n-invariant complements $\mathbf{m}_\mu$ to $\mathbf{g}_p$ in $\mathbf{g}_\mu$ and $\mathbf{n}_\mu$ to $\mathbf{g}_\mu$ in $\mathbf{g}$. Then $\mathbf{g} = \mathbf{g}_p \oplus \mathbf{m}_\mu \oplus \mathbf{n}_\mu$, and $\mathbf{g}^* = \mathrm{ann}(\mathbf{m}_\mu \oplus \mathbf{n}_\mu) \oplus \mathrm{ann}(\mathbf{g}_p \oplus \mathbf{n}_\mu) \oplus \mathrm{ann}(\mathbf{g}_p \oplus \mathbf{m}_\mu)$. These choices of complements define L_n-equivariant linear isomorphisms [50]

$$\begin{array}{lclcl} \mathrm{ann}(\mathbf{n}_\mu) & \cong & \mathbf{g}_\mu^*, & & \\ \mathrm{ann}(\mathbf{m}_\mu \oplus \mathbf{n}_\mu) & \cong & \mathrm{ann}_{\mathbf{g}_\mu^*}(\mathbf{m}_\mu) & \cong & \mathbf{g}_p^*, \\ \mathrm{ann}(\mathbf{g}_p \oplus \mathbf{n}_\mu) & \cong & \mathrm{ann}_{\mathbf{g}_\mu^*}(\mathbf{g}_p) & \cong & (\mathbf{g}_\mu/\mathbf{g}_p)^*, \end{array} \tag{3.4}$$

where $\mathrm{ann}(\cdot)$ denotes an annihilator in $\mathbf{g}^*$ and $\mathrm{ann}_{\mathbf{g}_\mu^*}(\cdot)$ an annihilator in $\mathbf{g}_\mu^*$.

Theorem 2.1 states that in a frame moving with velocity $\xi \in \mathbf{g}_\mu$ the bundle near the relative periodic orbit $\mathcal{P}$ is periodic with period $2n$. The following result shows that in the

Hamiltonian case the period can be reduced to n and the Poincaré section N can be further decomposed into three subspaces.

Theorem 3.1. *Let $\mathcal{P}$ be a relative periodic orbit, and let $p = \sigma^{-1}\Phi_1(p) \in \mathcal{P}$. Then the G_p-invariant Poincaré section N at p of Theorem 2.1 can be chosen to decompose as*

$$N = N_0 \oplus N_1 \oplus N_2, \tag{3.5}$$

where

$$\begin{array}{rcll} N_0 & = & \ker \mathrm{D}H(p) \cap (\ker \mathrm{D}\mathbf{J}(p))^{\perp} \cap N & \simeq \; (\mathbf{g}_\mu/\mathbf{g}_p)^*, \\ N_1 & = & \ker \mathrm{D}H(p) \cap \ker \mathrm{D}\mathbf{J}(p) \cap N, & \\ N_2 & = & (\ker \mathrm{D}H(p))^{\perp} \cap \ker \mathrm{D}\mathbf{J}(p) \cap N & \simeq \; \mathbb{R}. \end{array} \tag{3.6}$$

Here $\perp$ denotes orthogonal complements with respect to an appropriate G_p-invariant inner product on $T_p\mathcal{M}$. The spaces N_0, N_1, and N_2 are all G_p-invariant, and N_1 is a symplectic subspace of $T_p\mathcal{M}$.

The operator Q_N in Theorem 2.1 can be chosen to have order n, and so the action of the group L_{2n} on N factors through an action of L_n. The actions of G_p and Q_N on N now have the forms

$$g_p(\nu, w, E) \; = (\chi(g_p)(\mathrm{Ad}^*_{g_p})^{-1}\nu, \; g_p w, \; E) \quad \textit{for all} \quad g_p \in G_p \tag{3.7}$$

and

$$Q_N(\nu, w, E) \; = (Q_0\nu, \; Q_1 w, \; E) \quad \textit{with} \quad Q_0 = (\mathrm{Ad}^*_\alpha)^{-1}. \tag{3.8}$$

The linear map $Q_1 : N_1 \to N_1$ is orthogonal with respect to the restricted G_p-invariant inner product on N_1 and symplectic with respect to the restricted G_p-semi-invariant symplectic form $\omega_{N_1} := \omega|_{N_1}$. Its inverse Q_1^{-1} is twisted semiequivariant with respect to the action of G_p on N_1. Moreover, the identification (2.8) of a G-invariant neighborhood $\mathcal{U}$ with $(G \times \mathbb{R}/n\mathbb{Z} \times N)/L_n$ is a symplectomorphism, and the Γ-reduced phase space $\mathcal{U}/\Gamma \equiv (\mathbb{R}/n\mathbb{Z} \times N)/(\Gamma_p \ltimes \mathbb{Z}_n)$ is a Poisson space.

This theorem will be proved in sections 6.1–6.7 below. The tangent space decomposition is derived in section 6.2, the Poisson-structure on the Γ-reduced bundle is described in section 6.6, and the symplectic structure of the bundle is described in section 6.7. That Q_N can always be chosen to have order n is proved in sections 6.4 and 6.5 and is related to the connectedness of groups of symplectic transformations.

We call N_1 the *symplectic normal space* and denote its complex structure by J_{N_1}. In [16] it is shown that every semi-invariant symplectic form on a vector space has a semiequivariant complex structure J satisfying $J^2 = -\mathrm{id}$. We will always choose J_{N_1} in such a way. If G_p is finite and so $\mathbf{J}$ is nonsingular at p, then N_1 can be identified with the intersection of the Poincaré section N with the tangent space to the energy-momentum level set through p. It can be interpreted as the space of all small *shape oscillations* near the relative periodic orbit. In a similar way, $\nu \in N_0 \simeq (\mathbf{g}_\mu/\mathbf{g}_p)^*$ parametrizes the momenta of the rigid motion, expressed in body coordinates, while E parametrizes the difference in energy from $H(p)$ (see Remark 3.4(e)).

3.2. Equations near Hamiltonian relative periodic orbits. In this subsection, we formulate the central results of this paper. These describe the form taken by a Hamiltonian vector field near a relative periodic orbit in the bundle coordinates given by Theorems 2.1 and 3.1. In the absence of any symmetries it is well known that the dynamics near a Hamiltonian periodic orbit can be described as periodically forced motion in a Poincaré section inside an energy level set [2]. Here we show how this can be generalized to Hamiltonian relative periodic orbits by combining Theorems 2.1 and 2.2 of section 2 and Theorem 3.1 with techniques used for Hamiltonian relative equilibria in [50].

First we need to recall some preliminaries.

Proposition 3.2 (see [50]).

(a) *Let G be a Lie group, let $\mathbf{g}$ be its Lie algebra, and let μ be any point in $\mathbf{g}^*$. Let, as above, $\mathbf{n}_\mu$ be a complement to $\mathbf{g}_\mu$ in $\mathbf{g}$, and let $\mathrm{P}_{\mathrm{ann}(\mathbf{g}_\mu)}$ be the projection from $\mathbf{g}^*$ to $\mathrm{ann}(\mathbf{g}_\mu)$ with kernel $\mathrm{ann}(\mathbf{n}_\mu)$. Then for each ζ sufficiently close to 0 in $\mathrm{ann}(\mathbf{n}_\mu)$ and each $\xi \in \mathbf{g}_\mu$ the equation*

$$\mathrm{P}_{\mathrm{ann}(\mathbf{g}_\mu)}\left(\mathrm{ad}^*_{\xi+\eta}(\mu+\zeta)\right) \;=\; 0 \tag{3.9}$$

has a unique solution $\eta = \eta_\mu(\xi,\zeta) \in \mathbf{n}_\mu$. The map $\eta_\mu : \mathbf{g}_\mu \oplus \mathrm{ann}(\mathbf{n}_\mu) \to \mathbf{n}_\mu$, defined on the whole of $\mathbf{g}_\mu$ and a neighborhood of $0 \in \mathrm{ann}(\mathbf{n}_\mu)$, is smooth and linear in ξ and satisfies $\eta_\mu(\xi,0) = 0$ for all $\xi \in \mathbf{g}_\mu$ and $\eta_{\lambda\mu}(\xi,\lambda\zeta) = \eta_\mu(\xi,\zeta)$ for all $\lambda \in \mathbb{R}$.

(b) *If $\mu = \mathbf{J}(p)$ and $\mathbf{n}_\mu$ is G_p-invariant, then $\eta_\mu(\xi,\zeta)$ is G_p-equivariant with respect to the adjoint action of G_p on $\mathbf{g}$ and the χ-coadjoint action of G_p on $\mathbf{g}^*$.*

(c) *Let G^0_μ denote the identity component of G_μ. If $\mathbf{n}_\mu$ is a G^0_μ-invariant complement to $\mathbf{g}_\mu$ in $\mathbf{g}$, then $\eta_\mu \equiv 0$.*

For each sufficiently small $\zeta \in \mathbf{g}^*$, we define the linear map $j_\mu : \mathbf{g}_\mu \to \mathbf{g}$ by

$$j_\mu(\zeta)\xi \;=\; \xi + \eta_\mu(\zeta,\xi). \tag{3.10}$$

Now let $p \in \mathcal{P}$ lie on a relative periodic orbit $\mathcal{P}$. If μ satisfies the condition in (c), i.e., the L_n-complement $\mathbf{n}_\mu$ to $\mathbf{g}_\mu$ in $\mathbf{g}$ can be chosen to be G^0_μ-invariant, then we say that μ is *split*; see [14, 50].

Since the linear action of the compact group G_p on N_1 is semisymplectic, there exists a momentum map $\mathbf{L}_{N_1} : N_1 \to \mathbf{g}_p^*$ which is equivariant with respect to the χ-coadjoint action of G_p on $\mathbf{g}_p^*$. Using the complement $\mathbf{m}_\mu$ to $\mathbf{g}_p$ in $\mathbf{g}_\mu$, we can identify $\mathbf{g}_p^* \simeq \mathrm{ann}_{\mathbf{g}_\mu^*}(\mathbf{m}_\mu) \subset \mathbf{g}_\mu^*$ (see (3.4)) and so embed the Poincaré section $N = N_0 \oplus N_1 \oplus N_2 \cong (\mathbf{g}_\mu/\mathbf{g}_p)^* \oplus N_1 \oplus N_2$ into the *extended Poincaré section*

$$\widetilde{N} = \mathbf{g}_\mu^* \oplus N_1 \oplus N_2 \tag{3.11}$$

by the map from $N_0 \oplus N_1$ to $\mathbf{g}_\mu^* \oplus N_1$ given by

$$(\nu, w) \;\mapsto\; (\nu + \mathbf{L}_{N_1}(w), w), \qquad \nu \in (\mathbf{g}_\mu/\mathbf{g}_p)^*,\; w \in N_1. \tag{3.12}$$

The action of L_n on $N_0 \oplus N_1$ defined by Theorem 3.1 extends to an action on $\mathbf{g}_\mu^* \oplus N_1$ by extending the action of $Q_0 = (\mathrm{Ad}^*_\alpha)^{-1}$ on $(\mathbf{g}_\mu/\mathbf{g}_p)^*$ to the whole of $\mathbf{g}_\mu^*$. The choice of $\mathbf{m}_\mu$ to be Ad_α-invariant implies that this action preserves the subspace $\mathrm{ann}_{\mathbf{g}_\mu^*}(\mathbf{m}_\mu) \simeq \mathbf{g}_p^*$. Since L_n

is compact, the momentum map $\mathbf{L}_{N_1}$ can be assumed to be L_n-equivariant by averaging. It follows that the embedding (3.12) will also be L_n-equivariant.

Let $\hat{h} = \hat{h}(\theta, \nu, w, E)$ denote the lift of the G-invariant Hamiltonian H back to the space $G \times \mathbb{R}/n\mathbb{Z} \times (N_0 \oplus N_1 \oplus N_2)$ under the map given by Theorems 2.1 and 3.1. The function $\hat{h}$ is L_n-invariant:

$$\hat{h}(\chi(g_p)\theta, \chi(g_p)(\mathrm{Ad}^*_{g_p})^{-1}\nu, g_p w, E) \;=\; \hat{h}(\theta, \nu, w, E) \quad \text{for all} \;\; g_p \in G_p,$$

and

$$\hat{h}(\theta + 1, (\mathrm{Ad}^*_{\alpha})^{-1}\nu, Q_1 w, E) \;=\; \hat{h}(\theta, \nu, w, E).$$

In particular, $\hat{h}$ is periodic in θ with period n. We can extend $\hat{h}$ to an L_n-invariant function $\hat{h}(\theta, \zeta, w, E)$ on $\mathbb{R}/n\mathbb{Z} \times \widetilde{N}$ by setting $\hat{h}(\theta, \zeta, w, E) = \hat{h}(\theta, \nu, w, E)$, where $\zeta = \nu + \zeta_p \in \mathbf{g}^*_\mu$, $\nu \in (\mathbf{g}_\mu/\mathbf{g}_p)^*$, $\zeta_p \in \mathbf{g}^*_p$.

Theorem 3.3. *Let $\mathcal{P}$ be a relative periodic orbit, and let $p = \sigma^{-1}\Phi_1(p) \in \mathcal{P}$. Assume time is parametrized so that the phase dynamics near the relative periodic orbit is given by $\dot{\theta} \equiv 1$ in the equations of Theorem 2.2. Then the Hamiltonian $\hat{h}$ in bundle coordinates is of the form*

$$\hat{h}(\theta, \nu, w, E) = h(\theta, \nu, w) + E \tag{3.13}$$

for some L_n-invariant function h on $\mathbb{R}/n\mathbb{Z} \times (N_0 \oplus N_1)$. As above, h extends to an L_n-invariant function $h(\theta, \zeta, w)$ on $\mathbb{R}/n\mathbb{Z} \times (\mathbf{g}^*_\mu \oplus N_1)$. We have $\mathrm{D}_{(\zeta,w)}h(\theta, 0, 0) = (\xi, 0)$, and the differential equations for the motion in bundle coordinates

$$(g, \theta, \zeta = \nu + \mathbf{L}_{N_1}(w), w, E) \;\in\; \Gamma \times \mathbb{R}/n\mathbb{Z} \times \widetilde{N}$$

take the form

$$\begin{aligned} \dot{g} &= g j_\mu(\zeta)\mathrm{D}_\zeta h(\theta, \zeta, w), \\ \dot{\theta} &= 1, \\ \dot{\zeta} &= \mathrm{ad}^*_{j_\mu(\zeta)\mathrm{D}_\zeta h(\theta,\zeta,w)}(\mu + \zeta), \\ \dot{w} &= J_{N_1}\mathrm{D}_w h(\theta, \zeta, w), \\ \dot{E} &= -\mathrm{D}_\theta h(\theta, \zeta, w). \end{aligned} \tag{3.14}$$

A proof of Theorem 3.3 is given in section 6.8 below.

Remarks 3.4.

(a) Note that the (θ, ζ, w) subsystem of (3.14) decouples from and forces the (g, E) equations. Hence (3.14) has a *skew-product* structure.

(b) The (ζ, w) subsystem on $\mathbf{g}^*_\mu \oplus N_1$ forms a G_p-semiequivariant Poisson system that is periodically forced with period n. Since the action of G_p on $\mathbf{g}^*_\mu \oplus N_1$ is semi-Poisson, the dynamics of this subsystem preserves a G_p momentum map $\mathbf{L}_{\mathbf{g}^*_\mu \oplus N_1}$; see section 6.6.

(c) If μ is split, for example, if G_μ is compact, then $\eta_\mu(\zeta) \equiv 0$ and $j_\mu(\zeta)\mathrm{D}_\zeta h = \mathrm{D}_\zeta h$.

(d) As in the case of relative equilibria [50], the momentum map $\mathbf{J}$ is given in bundle coordinates by $\mathbf{J}(g,\theta,\nu,w,E) = \chi(g)(\mathrm{Ad}_g^*)^{-1}(\mu+\nu+\mathbf{L}_{N_1}(w))$. This can easily be verified using the symplectic form in bundle coordinates, described in section 6.7.

(e) Because of (3.13) the energy level sets $H \equiv e$ with $e \approx 0$ are given in bundle coordinates by $E = E(\theta,\nu,w) = e - h(\theta,\nu,w)$. Since $H(p) = h(0) = 0$, the parameter E therefore parametrizes the difference in energy from $H(p)$.

The next theorem gives the equations that are obtained by projecting the $\dot\zeta$ equation on $\mathbf{g}_\mu^*$ back to $N_0 \simeq (\mathbf{g}_\mu/\mathbf{g}_p)^*$ and hence provides explicit differential equations in the bundle coordinates (g,θ,ν,w,E). First we recall some notation for the operator obtained by projecting the coadjoint action of $\mathbf{g}_\mu$ on $\mathbf{g}_\mu^*$ down to $(\mathbf{g}_\mu/\mathbf{g}_p)^*$ [50]. Let $\pi_{\mathbf{m}_\mu}$ be the L_n-equivariant projection from $\mathbf{g}_\mu$ to $\mathbf{m}_\mu \simeq \mathbf{g}_\mu/\mathbf{g}_p$ with kernel $\mathbf{g}_p$. Let $\nu \in (\mathbf{g}_\mu/\mathbf{g}_p)^*$ and $\xi,\eta \in \mathbf{m}_\mu$. Then define

$$\overline{\mathrm{ad}}_\xi(\eta) = [\xi,\eta]_{\mathbf{m}_\mu} = \pi_{\mathbf{m}_\mu}([\xi,\eta]), \quad \overline{\mathrm{ad}}_\xi^*(\nu)(\eta) = \nu\left([\xi,\eta]_{\mathbf{m}_\mu}\right). \tag{3.15}$$

Note that, in general, the bracket $[\cdot,\cdot]_{\mathbf{m}_\mu}$ and the operators $\overline{\mathrm{ad}}_\cdot$ and $\overline{\mathrm{ad}}_\cdot^*$ depend on the choice of $\mathbf{m}_\mu$. Moreover, $[\cdot,\cdot]_{\mathbf{m}_\mu}$ does not satisfy the Jacobi identity and so is not a Lie bracket. However, in the (very special) case when $\mathbf{g}_p$ is a normal subalgebra of $\mathbf{g}_\mu$ the quotient $\mathbf{g}_\mu/\mathbf{g}_p$ is again a Lie algebra, and $[\cdot,\cdot]_{\mathbf{m}_\mu}$ is equal to its natural Lie bracket for any choice of complement $\mathbf{m}_\mu$. Similarly, $\overline{\mathrm{ad}}^*$ is the usual coadjoint action of $\mathbf{g}_\mu/\mathbf{g}_p$ on its dual in this case [50].

Theorem 3.5. *Coordinates (g,θ,ν,w,E) can be chosen on $\Gamma\times\mathbb{R}/n\mathbb{Z}\times N = \Gamma\times\mathbb{R}/n\mathbb{Z}\times((\mathbf{g}_\mu/\mathbf{g}_p)^*\oplus N_1\oplus N_2)$ so that the restriction of the (G,χ)-semiequivariant Hamiltonian system (3.1) to a neighborhood of $\mathcal{P}$ can be lifted to the following system on $\Gamma\times\mathbb{R}\times N$:*

$$\begin{aligned}
\dot g &= g\left(\mathrm{D}_\nu h(\theta,\nu,w) + \hat\eta(\theta,\nu,w)\right),\\
\dot\theta &= 1,\\
\dot\nu &= \overline{\mathrm{ad}}^*_{\mathrm{D}_\nu h(\theta,\nu,w)}(\nu) + \mathrm{ad}^*_{\mathrm{D}_\nu h(\theta,\nu,w)}(\mathbf{L}_{N_1}(w)) + \mathrm{P}\left(\mathrm{ad}^*_{\hat\eta(\theta,\nu,w)}(\nu+\mathbf{L}_{N_1}(w))\right),\\
\dot w &= J_{N_1}\mathrm{D}_w h(\theta,\nu,w),\\
\dot E &= -\mathrm{D}_\theta h(\theta,\nu,w),
\end{aligned} \tag{3.16}$$

where the map $\hat\eta : \mathbb{R}\times N \to \mathbf{n}_\mu$ is given by

$$\hat\eta(\theta,\nu,w) = \eta_\mu(\mathrm{D}_\nu h(\theta,\nu,w), \nu+\mathbf{L}_{N_1}(w))$$

and P is the projection from $\mathbf{g}^*$ to $\mathrm{ann}(\mathbf{g}_p+\mathbf{n}_\mu)\cong(\mathbf{g}_\mu/\mathbf{g}_p)^*$ with kernel $\mathrm{ann}(\mathbf{m}_\mu)$.

This theorem is obtained from Theorem 3.3 in the same way as the analogous result for relative equilibria in [50].

4. Stability and bifurcations.

In this section, we outline some straightforward applications of Theorems 3.3 and 3.5. The first subsection describes the linearization of a Hamiltonian vector field at a relative periodic orbit, while the second gives two persistence theorems. The main aim of the section is to indicate potential applications of the theorems. These will be explored in greater depth in future work.

4.1. Stability of relative periodic orbits. In this subsection, we present some simple implications of Theorem 3.5 for the stability of relative periodic orbits. A relative periodic orbit $\mathcal{P}$ of the Γ-equivariant, but not necessarily Hamiltonian, differential equation (2.1) is said to be (orbitally Liapounov) stable or Γ-*stable* if $\mathcal{P}$ is a (Liapounov stable) periodic orbit for the flow on $\mathcal{M}/\Gamma$. It is said to be *exponentially unstable* if there exist solutions which start close to $\mathcal{P}$ but leave a neighborhood of $\mathcal{P}$ in $\mathcal{M}/\Gamma$ exponentially fast. Theorem 2.1 implies that stability or exponential instability of $\mathcal{P}$ is equivalent to the stability or exponential instability of the periodic solution $\{\theta \in \mathbb{R}, v = 0\}$ of the (θ, v) subsystem of (2.10).

Proposition 4.1. *Let $p = \sigma^{-1}\Phi_1(p) \in \mathcal{P}$, $\sigma \in \Gamma$, and let $M = \sigma^{-1}\mathrm{D}\Phi_1(p)$. Then the following hold.*

(a) *The map M has the following structure with respect to the decomposition $T_p\mathcal{M} = \mathbf{g}p \oplus \mathbb{R} \oplus N$:*

$$M = \begin{pmatrix} \pi_{\mathbf{m}}\mathrm{Ad}_\sigma^{-1}|_{\mathbf{m}} & 0 & D \\ 0 & 1 & \Theta \\ 0 & 0 & M_N \end{pmatrix}, \tag{4.1}$$

where $\mathbf{m} \cong \mathbf{g}p \cong \mathbf{g}/\mathbf{g}_p$ is an L_n-invariant complement to $\mathbf{g}_p$ in $\mathbf{g}$ and $\pi_{\mathbf{m}}$ is the projection from $\mathbf{g}$ to $\mathbf{m}$ with kernel $\mathbf{g}_p$.

(b) *If time is reparametrized so that $f_\Theta(\theta, v) \equiv 1$ and $\Phi^N_{1,0}$ is the time 1 map of the periodically forced system on N, then $Q_N^{-1}\Phi^N_{1,0}$ is a (symmetry reduced) Poincaré map for the periodic solution of the (θ, v) system with $v = 0$ as fixed point. The block M_N in (4.1) is the linearization of this map: $M_N = Q_N^{-1}\mathrm{D}\Phi^N_{1,0}(0)$.*

Proof. It is easily checked that $f_H(p)$ is a right eigenvector of M with eigenvalue 1. Moreover, for $\xi \in \mathbf{g}$ we have

$$\sigma^{-1}\mathrm{D}\Phi_1(p)\xi p \;=\; \sigma^{-1}\xi\Phi_1(p) \;=\; \sigma^{-1}\xi\sigma p \;=\; \left(\mathrm{Ad}_\sigma^{-1}\xi\right)p,$$

which shows that $M\xi p = \mathrm{Ad}_\sigma^{-1}\xi p$ for $\xi \in \mathbf{g}$. Therefore, M has the structure shown in (4.1). Part (b) follows from Proposition 3.1 of [55]. ■

As a consequence, $\mathcal{P}$ is exponentially unstable if and only if M_N has eigenvalues outside the unit circle.

Definition 4.2. *We call a relative periodic orbit $\mathcal{P}$* spectrally stable *if all the eigenvalues of M_N lie within or on the unit circle.*

In Hamiltonian systems, (relative) periodic orbits are typically *not* orbitally Liapounov stable. However, the above spectral stability theory for general systems remains applicable. Criteria for Liapounov stability of Hamiltonian relative periodic orbits that apply in special cases can be found in [41, 42]. In this section, we will describe the extra structure that M and M_N have for Hamiltonian systems.

As usual, let $\mu = \mathbf{J}(p)$ and $\mathbf{m}_\mu$ and $\mathbf{n}_\mu$ be as in section 3, and so $\mathbf{m} = \mathbf{n}_\mu + \mathbf{m}_\mu$. Let $\pi_{\mathbf{m}_\mu}$ be the projection from $\mathbf{g}$ to $\mathbf{m}_\mu$ with kernel $\mathbf{n}_\mu \oplus \mathbf{g}_p$, and let $\pi_{\mathbf{n}_\mu}$ be the projection from $\mathbf{g}$ to $\mathbf{n}_\mu$ with kernel $\mathbf{m}_\mu \oplus \mathbf{g}_p$. We will now define an analogue of the operators $\overline{\mathrm{ad}}_\cdot$, $\overline{\mathrm{ad}}_\cdot^*$ introduced in section 3.2 for actions of $g \in G_\mu$ on $\mathbf{m}_\mu \simeq \mathbf{g}_\mu/\mathbf{g}_p$ and $\mathrm{ann}_{\mathbf{g}_\mu}(\mathbf{g}_p) = (\mathbf{g}_\mu/\mathbf{g}_p)^*$. For $g \in G_\mu$, $\eta \in \mathbf{m}_\mu$, and $\nu \in \mathrm{ann}_{\mathbf{g}_\mu}(\mathbf{g}_p)$, let

$$\overline{\mathrm{Ad}}_g\eta = \pi_{\mathbf{m}_\mu}\mathrm{Ad}_g\eta, \quad (\overline{\mathrm{Ad}}_g^*)(\nu)(\eta) = \nu(\overline{\mathrm{Ad}}_g\eta). \tag{4.2}$$

Note that $\overline{\mathrm{Ad}}_g$ and $\overline{\mathrm{Ad}}_g^*$ depend on the choice of the complement $\mathbf{m}_\mu$ of $\mathbf{g}_p$ in $\mathbf{g}_\mu$ and vary if g is varied in $g\Gamma_p$.

Proposition 4.3. *With respect to the tangent space decomposition $T_p\mathcal{M} = T \oplus N$, where*

$$T = T_p\mathcal{P} = T_0 \oplus T_1 \oplus T_2, \quad \textit{with} \quad T_0 = \mathbf{g}_\mu p, \quad T_1 = \mathbf{n}_\mu p, \quad T_2 = \mathrm{span}(f_H(p)), \tag{4.3}$$

and $N = N_0 \oplus N_1 \oplus N_2$, the linearization M at $p \in \mathcal{P}$ has the following block structure:

$$M \;=\; \sigma^{-1}\mathrm{D}\Phi_1(p) \;=\; \begin{pmatrix} \overline{\mathrm{Ad}}_\sigma^{-1} & \pi_{\mathbf{m}_\mu}\mathrm{Ad}_\sigma^{-1}|_{\mathbf{n}_\mu} & 0 & D_0 & D_1 & D_2 \\ 0 & \pi_{\mathbf{n}_\mu}\mathrm{Ad}_\sigma^{-1}|_{\mathbf{n}_\mu} & 0 & D_3 & 0 & 0 \\ 0 & 0 & 1 & \Theta_0 & \Theta_1 & \Theta_2 \\ & & 0 & \overline{\mathrm{Ad}}_\sigma^* & 0 & 0 \\ & & 0 & M_{10} & M_1 & M_{12} \\ & & 0 & 0 & 0 & 1 \end{pmatrix}.$$

All the subblocks of M are twisted semiequivariant with respect to the appropriate actions of G_p on the subspaces T_i and N_i, $i = 0, 1, 2$. Moreover, M is symplectic and so the subblocks are related to each other by the equations given in Lemma 6.4.

This proposition is proved in section 6.3, where Lemma 6.4 is stated. Results for relative equilibria of compact group actions analogous to this and the following proposition can be found in [44, 45].

Proposition 4.4. *Consider the decomposition of M given by Proposition 4.3.*

(a) *If $1 \notin \mathrm{spec}(M_1)$, then the tangent space decomposition can be chosen so that $\Theta_1 = M_{12} = 0$.*

(b) *If $\mathrm{spec}(\overline{\mathrm{Ad}}_\sigma^*) \cap \mathrm{spec}(M_1) = \emptyset$, then the tangent space decomposition can be chosen so that $D_1 = M_{10} = 0$.*

(c) *If $\mathrm{spec}(\overline{\mathrm{Ad}}_\sigma^*) \cap \mathrm{spec}(\pi_{\mathbf{n}_\mu}\mathrm{Ad}_\sigma|_{\mathbf{n}_\mu}) = \emptyset$ or μ is split, then the tangent space decomposition can be chosen so that $\pi_{\mathbf{m}_\mu}\mathrm{Ad}_\sigma|_{\mathbf{n}_\mu} = 0$ and $D_3 = 0$.*

(d) *If time is parametrized so that $\dot{\theta} \equiv 1$, then $D_2 = \Theta_0 = \Theta_1 = \Theta_2 = M_{12} = 0$.*

Proof. Parts (a) and (b) and the first statement of (c) are linear algebra. For the second statement of part (c), observe that if μ is split, then $\mathbf{n}_\mu$ can be chosen to be both Ad_α and G_μ^0-invariant and hence also Ad_σ-invariant since $\sigma = \alpha\exp(\xi)$, $\xi \in \mathbf{g}_\mu$. That $D_3 = 0$ then follows from (6.10) below. For Part (d) note that Theorem 3.3 implies that in this case the Hamiltonian $h(\nu, w, \theta)$ in bundle coordinates does not depend on E. ■

The following corollary is a direct consequence of Proposition 4.3.

Corollary 4.5. *The relative periodic orbit $\mathcal{P}$ is spectrally stable if and only if all the eigenvalues of $M_1 : N_1 \to N_1$ lie on the unit circle, and those of $\overline{\mathrm{Ad}}_\sigma^* : N_0 \to N_0$ lie within or on the unit circle.*

Note that spectral stability does not depend on the choice of σ within the coset $\sigma\Gamma_p$. As for general systems (Proposition 4.1), we see that $\mathcal{P}$ is spectrally stable if and only if 0 lies on a spectrally stable periodic orbit of the periodically forced (ν, w) subsystem.

For many groups relevant in applications, the spectrum of $\overline{\mathrm{Ad}}_\sigma^*$ automatically lies on the unit circle. For example, this is satisfied if there is a G_μ-invariant inner product on $\mathbf{g}^*$ and so for all compact groups G. It is also satisfied by Euclidean groups and therefore in most applications.

The twisted semiequivariance of the diagonal subblocks of M may imply additional block structure for the subblocks. For example, if the twist diffeomorphism ϕ is trivial, so that M is G_p-semiequivariant, then the block M_1 maps isotypic components of N_1 with respect to the Γ_p-action into themselves [42]. The structure of equivariant symplectic linear maps is described in general in [36]. Extensions to reversible equivariant symplectic linear maps can be deduced from the results on infinitesimally symplectic linear maps in [16].

4.2. Bifurcation of relative periodic orbits. In this section, we describe some simple implications of Theorem 3.5 for bifurcations of relative periodic orbits. Relative periodic orbits which lie near $\mathcal{P}$ correspond bijectively to relative periodic orbits of the L_n-semiequivariant (θ, ν, w) subsystem of (3.16) on $\mathbb{R}/n\mathbb{Z} \times (\mathbf{g}_\mu/\mathbf{g}_p)^* \oplus N_1$ [55, 22]. The original relative periodic orbit $\mathcal{P}$ itself corresponds to the set $\{\theta \in \mathbb{R}/n\mathbb{Z},\ (\nu, w) = 0\}$. It is therefore a periodic orbit with finite phase-shift symmetry $\mathbb{Z}_n$, i.e., a *discrete rotating wave* of the (θ, ν, w) subsystem of (3.16). As L_n is compact, the problem of describing bifurcations from relative periodic orbits in systems with noncompact (reversing) symmetry groups is therefore reduced to that of describing bifurcations from discrete rotating waves in systems with compact (reversing) symmetry groups.

The description of all generic bifurcations in the (θ, ν, w) subsystem of (3.16) is a difficult problem which we will tackle in future work. In this paper, we content ourselves with describing briefly some easily obtained results for special cases. The case of “minimal” μ is considered in section 4.2.1. The case of split μ and the finite isotropy subgroup Γ_p is discussed in section 4.2.2.

4.2.1. Minimal momenta. A momentum $\mu \in \mathbf{g}^*$ is said to be *minimal* if $\dim(\mathbf{g}_\mu)$ is minimal [50, section 4.2]. It is shown in [12] that the set of minimal μ is open and dense in $\mathbf{g}^*$ and that the isotropy subgroup Γ_μ of a minimal μ is Abelian. The following result is proved in exactly the same way as the analogous result for relative equilibria [50, Proposition 4.2].

Proposition 4.6. *If μ is minimal, then $\dot{\nu} \equiv 0$ in* (3.16)*, and bifurcation from relative periodic orbits reduces to bifurcation from discrete rotating waves of the ν-dependent periodically forced w equation of* (3.16)*.*

As a corollary, we obtain a persistence result for *nondegenerate* relative periodic orbits.

Definition 4.7. *The relative periodic orbit $\mathcal{P}$ is nondegenerate if 1 is not an eigenvalue of the block M_1 in $M = \sigma^{-1}\mathrm{D}\Phi_1(\mathrm{p})$ defined in Proposition 4.3.*

The following generalizes a persistence result for relative equilibria [44, 50] to relative periodic orbits of noncompact groups.

Corollary 4.8. *Let $p = \sigma^{-1}\Phi_1(p)$ lie on a nondegenerate relative periodic orbit $\mathcal{P}$ with minimal momentum μ and energy e, and assume that Γ_p is trivial. Let $\sigma = \alpha \exp(\xi)$, $\alpha \in \Gamma_\mu$, $\alpha^n = \mathrm{id}$, and $\xi \in \mathbf{z}(\sigma) \cap \mathbf{z}^\chi(G_p)$ as in* (3.3)*. Then the following hold.*

(a) *For each momentum $\hat{\mu}$ near μ with $\mathrm{Ad}^*_\alpha \hat{\mu} = \hat{\mu}$ and each energy $\hat{e}$ near e, there exists a unique relative periodic orbit near $\mathcal{P}$ with momentum $\hat{\mu}$, energy $\hat{e}$, and relative period close to that of $\mathcal{P}$.*

(b) *Assume, in addition, that $\mathcal{P}$ is nondegenerate when considered as a relative periodic orbit of relative period n; i.e., the block M_1 in Proposition 4.3 does not have* nth *roots of unity as eigenvalues. Then, for each momentum $\hat{\mu}$ near μ and each energy $\hat{e}$ near e, there exists a unique relative periodic orbit near $\mathcal{P}$ with momentum $\hat{\mu}$, energy $\hat{e}$,* *and relative period close to some* $\ell \in \mathbb{N}$ *with* $\ell | n$*. The union of these relative periodic orbits is a symplectic submanifold of* $\mathcal{M}$ *of dimension* $\dim\Gamma + \dim\Gamma_\mu + 2$.

Proof. For simplicity, assume that time has been reparametrized such that $f_\Theta \equiv 1$. From Propositions 4.3 and 4.6 we see that $M_1 = Q_1^{-1}\mathrm{D}\Phi^{N_1}_{1,0}(0)$, where $\Phi^{N_1}_{t,t_0}(\cdot,\nu)$ is the ν-dependent time-evolution of the w equation of (3.16). Since the relative periodic orbit is nondegenerate, we conclude that 1 is not an eigenvalue of $Q_1^{-1}\mathrm{D}\Phi^{N_1}_{1,0}(0)$. So we can apply the implicit function theorem to the ν-dependent fixed point equation for $Q_1^{-1}\Phi^{N_1}_{1,0}(\cdot,\nu)$ to conclude that there is a family of periodic orbits of the w equation of (3.16) parametrized by $\nu \in \mathbf{g}^*_\mu$, with initial value $w(\nu) \in N_1$, where $w(0) = 0$. Since the $\dot{E}$ equation does not depend on E, the E-initial value provides an additional parameter. For $\nu \in \mathbf{g}^*_\mu$ with $\mathrm{Ad}^*_\alpha\nu = \nu$, we obtain a family $\mathcal{P}^1_{\nu,E}$ of relative periodic orbits of (3.1) with relative period one. Because of Remarks 3.4 (d), (e) on the momentum map in bundle coordinates and energy parametrization, this family of relative periodic orbits provides exactly one relative periodic orbit for each energy-momentum pair $(\hat{e},\hat{\mu})$ with $\hat{\mu} = \mathrm{Ad}^*_\alpha\hat{\mu}$ close to (e,μ). This proves part (a).

To prove (b) let $\mathcal{P}$ be nondegenerate as a relative periodic orbit of relative period n. Then the fixed point equation for $\Phi^{N_1}_{n,0}(\cdot,\nu)$ can be solved uniquely for any small $\nu \in \mathbf{g}^*_\mu$ giving a family $\mathcal{P}_{\nu,E}$ of relative periodic orbits of (3.1) which have relative periods ℓ for some $\ell | n$. The dimension formula then follows from the observation that each relative periodic orbit $\mathcal{P}_{\nu,E}$ has dimension $\dim(\Gamma) + 1$. Symplecticity of the submanifold formed by the union of the family $\mathcal{P}_{\nu,E}$ of relative periodic orbits is a consequence of the fact that by Theorem 3.1 a G-invariant neighborhood $\mathcal{U}$ of the relative periodic orbit $\mathcal{P}$ is symplectomorphic to the symplectic manifold $(G \times \mathbb{R}/n\mathbb{Z} \times N)/L_n$ and that the union of the family of relative periodic orbits $\mathcal{P}_{\nu,E}$ is a manifold of the form $(G \times \mathbb{R}/n\mathbb{Z} \times N_0 \oplus \{0\} \oplus N_2)/L_n$. That this is a symplectic submanifold of $(G \times \mathbb{R}/n\mathbb{Z} \times N)/L_n$ can be seen from the symplectic form in bundle coordinates given in section 6.7. ■

An extension of this result which describes nearby relative periodic orbits with the same isotropy subgroup Γ_p as $\mathcal{P}$ in the case of general nonfree actions can easily be obtained by applying the method of [40]: just replace $\mathcal{M}$ by the corresponding fixed point space $\mathrm{Fix}_{\Gamma_p}(\mathcal{M})$ and Γ by $N(\Gamma_p)/\Gamma_p$.

To study bifurcations of relative periodic orbits with less spatio-temporal symmetry (including subharmonic branching) and bifurcations from degenerate relative periodic orbits, the results in [6, 9, 10] on bifurcations from fixed points of equivariant and reversible symplectic maps can be applied to the ν-dependent symplectic G_p-semiequivariant map $\Phi^{N_1}_{1,0}(\cdot,\nu)$ on N_1 provided that $Q_1 = \mathrm{id}$.

4.2.2. Split momenta and finite isotropy subgroups.

If μ is split, i.e., if the L_n-invariant complement $\mathbf{n}_\mu$ to $\mathbf{g}_\mu$ in $\mathbf{g}$ can be chosen to be G^0_μ-invariant (cf. section 3.2), then the term $\mathrm{P}(\mathrm{ad}^*_{\hat{\eta}(\nu,w)}(\nu + \mathbf{L}_{N_1}(w)))$ in the $\dot{\nu}$ equation in (3.16) vanishes and the equation becomes

$$\dot{\nu} = \overline{\mathrm{ad}}^*_{\mathrm{D}_\nu h(\theta,\nu,w)}(\nu) + \mathrm{ad}^*_{\mathrm{D}_\nu h(\theta,\nu,w)}(\mathbf{L}_{N_1}(w)).$$

If Γ_p is finite, then $\overline{\mathrm{ad}}^*_{\mathrm{D}_\nu h} = \mathrm{ad}^*_{\mathrm{D}_\nu h}$ and $\mathbf{L}_{N_1} \equiv 0$, and so the (θ,ν,w) equations simplify to

$$\dot{\theta} = 1, \quad \dot{\nu} = \mathrm{ad}^*_{\mathrm{D}_\nu h(\theta,\nu,w)}(\nu), \quad \dot{w} = J_{N_1}\mathrm{D}_w h(\theta,\nu,w). \tag{4.4}$$

These equations define an L_n-semiequivariant Poisson system on $\mathbb{R}/n\mathbb{Z}\times N$ and a G_p-semiequivariant periodically forced Poisson system on $\mathbf{g}_\mu^* \oplus N_1$. The following persistence result for relative periodic orbits of noncompact group actions is inspired by a similar theorem of Montaldi [33] for free actions of compact groups and a result on the persistence of relative equilibria of compact groups with finite isotropy [34]. Montaldi uses the compactness of the coadjoint orbits to infer the existence of relative periodic orbits on each nearby energy-momentum level set. Without this compactness assumption, the topological techniques that he employs no longer apply; however, using the isotropy of the relative periodic orbit, we can get similar results. In contrast to the generalization of the persistence result of [33] given in [40], compactness of the normalizer of the isotropy of the relative periodic orbit is not required, and the spatio-temporal and reversing symmetries of the persisting relative periodic orbits are described.

Theorem 4.9. *Let $p = \sigma^{-1}\Phi_1(p)$ lie on a relative periodic orbit $\mathcal{P}$ with split momentum μ and finite isotropy subgroup Γ_p. Let $\sigma = \alpha \exp(\xi)$, where $\alpha^n = \mathrm{id}$ as in (3.3), and let Λ_n (resp., L_n) be the group generated by Γ_p (resp., G_p) and α. Assume that $\mathcal{P}$ is nondegenerate when considered as a relative periodic orbit of relative period n; i.e., the block M_1 in Proposition 4.3 does not have nth roots of unity as eigenvalues.*

Let $\hat\nu \in \mathbf{g}_\mu^*$, $\hat\nu \approx 0$, be such that $\mathrm{Fix}_{\mathbf{g}_\mu^*}(\hat\Gamma_p) \cap \mathbf{g}_\mu \hat\nu = \{0\}$, where $\hat\Gamma_p = \Gamma_p \cap \Gamma_{\hat\mu}$ and $\hat\mu = \mu + \hat\nu$. Then the following hold.

(a) *The group $\hat\Lambda := \Lambda_n \cap \Gamma_{\hat\mu}$ is a cyclic extension of $\hat\Gamma_p$: there exists $\ell \in \mathbb{N}$ with $\ell | n$ such that $\hat\Lambda/\hat\Gamma_p \simeq \mathbb{Z}_{n/\ell}$. Moreover, either $\hat L := L_n \cap G_{\hat\mu}$ equals $\hat\Lambda$, or $\hat\Lambda$ is a normal subgroup of $\hat L$ of index two. In the latter case, there exists $\rho \in \hat L \setminus \hat\Lambda$ such that $\hat L = \hat\Lambda_\rho$.*

(b) *There is a family $\mathcal{P}_E(\hat\nu)$ of relative periodic orbits close to $\mathcal{P}$ which is parametrized by $E \approx 0$ with points $\hat p_E \in \mathcal{P}_E(\hat\nu)$ such that*

$$\mathbf{J}(\hat p_E) = \hat\mu, \quad \Gamma_{\hat p_E} = \hat\Gamma_p, \quad G_{\hat p_E} = \hat G_p := \begin{cases} \Gamma_{\hat p_E} & \text{if} \quad \hat L = \hat\Lambda, \\ (\Gamma_{\hat p_E})_\rho & \text{if} \quad \hat L = \hat\Lambda_\rho. \end{cases}$$

The relative period of the relative periodic orbit $\mathcal{P}_E(\hat\nu)$ is $\ell \in \mathbb{N}$, where $\ell | n$ is such that $\hat\Lambda/\hat\Gamma_p \simeq \mathbb{Z}_{n/\ell}$, and we have $\hat\sigma^{-1}\Phi_\ell(\hat p_E) = \hat p_E$, where $\hat\sigma = \sigma^\ell \gamma_p \exp(\hat\xi)$, with $\alpha^\ell \gamma_p \in \hat\Lambda$ for some $\gamma_p \in \Gamma_p$, and $\hat\xi \in \mathbf{z}^\chi(\hat G_p) \cap \mathbf{g}_{\hat\mu}$ is small.

Proof. (a) Let $\gamma \in \hat\Lambda = \Lambda_n \cap \Gamma_{\hat\mu}$. Since Γ_p is normal in Λ_n, we have $\gamma\hat\gamma_p\gamma^{-1} \in \Gamma_p \cap \Gamma_{\hat\mu} = \hat\Gamma_p$ for $\hat\gamma_p \in \hat\Gamma_p$. Therefore, $\hat\Lambda \subseteq N(\hat\Gamma_p)$. We now show that $\hat\Lambda/\hat\Gamma_p$ is cyclic. Let $\gamma \in \hat\Lambda$. Then $\gamma = \gamma_p\alpha^i$ for some $\gamma_p \in \Gamma_p$, $i \in \{0, \dots, n-1\}$. Moreover, $i \neq 0$ if $\gamma_p \in \Gamma_p \setminus \hat\Gamma_p$ by definition of $\hat\Gamma_p$. As a consequence, $\gamma^n \in \hat\Gamma_p$, and, if $\tilde\gamma \in \hat\Lambda$, $\tilde\gamma = \tilde\gamma_p\alpha^i$, for some $\tilde\gamma_p \in \Gamma_p$, then $\gamma\tilde\gamma^{-1} = \gamma_p\tilde\gamma_p^{-1} \in \hat\Gamma_p$. Hence $\hat\Lambda/\hat\Gamma_p$ is a subgroup of $\mathbb{Z}_n$ and therefore cyclic, i.e., there is some $\ell | n$ with $\hat\Lambda/\hat\Gamma_p \simeq \mathbb{Z}_{n/\ell}$.

Now assume that $\hat L \neq \hat\Lambda$, and let $\rho \in \hat L \setminus \hat\Lambda$. Since Λ_n is normal in L_n, we have $\rho\hat\gamma\rho^{-1} \in \Lambda_n \cap \Gamma_{\hat\mu} = \hat\Lambda$ for $\hat\gamma \in \hat\Lambda$ and $\hat L = \hat\Lambda_\rho$.

(b) 1. We first prove that the periodically forced Poisson system on $N_0 \oplus N_1$ (the (ν, w) subsystem of (3.16)) has an n-periodic solution with isotropy $\hat\Gamma_p$. Let $\Psi_{t,t_0} = (\Psi^\nu_{t,0}, \Psi^w_{t,0})$ denote the time-evolution of this subsystem. The original relative periodic orbit corresponds to the origin: $\Psi_{t,0}(0) = 0$. Because it is Poisson, Ψ_{t,t_0} restricts to a symplectic map on the symplectic leaves $\mathcal{O} \times N_1$, where $\mathcal{O}$ is a coadjoint orbit of Γ_μ in $\mathbf{g}_\mu^*$. Moreover, $\Psi_{t,0}$ is Γ_p-equivariant. As a consequence, $\Psi_{t,0}$ maps $(\mathrm{Fix}_{\mathbf{g}_\mu^*}(\hat\Gamma_p) \cap \Gamma_\mu \hat\nu) \times \mathrm{Fix}_{N_1}(\hat\Gamma_p)$ into itself. Since by assumption $\mathrm{Fix}_{\mathbf{g}_\mu^*}(\hat\Gamma_p) \cap \mathbf{g}_\mu \hat\nu = \{0\}$, the path-connected component of the point $\hat\nu$ in $\mathrm{Fix}_{\mathbf{g}_\mu^*}(\hat\Gamma_p) \cap \Gamma_\mu \hat\nu$ is just $\{\hat\nu\}$, and so we can conclude that

$$\Psi_{t,0}^{\nu}(\hat\nu, \hat w) = \hat\nu \quad \text{for all } t \in \mathbb{R},\ \hat w \in \mathrm{Fix}_{N_1}(\hat\Gamma_p). \tag{4.5}$$

The nondegeneracy condition implies that $\mathrm{D}_w \Psi_{n,0}^w(0) - \mathrm{id}$ is invertible. So we can solve the equation $\Psi_{n,0}^w(\nu, w) = w$ uniquely for $w = w(\nu)$ if $\nu \in N_0 \simeq \mathbf{g}_\mu^*$ is small. Therefore, we have proved that $(\hat\nu, \hat w) = \Psi_{n,0}(\hat\nu, \hat w)$, with $\hat w = w(\hat\nu)$, lies on an n-periodic solution of the (ν, w) system. Moreover, since $\Psi_{t,0}$ is Γ_p-equivariant, $w(\cdot)$ is a Γ_p-equivariant map from $\mathbf{g}_\mu^*$ to N_1, and therefore $(\hat\nu, \hat w) \in \mathrm{Fix}_{N_0 \oplus N_1}(\hat\Gamma_p)$.

2. Now we investigate the spatio-temporal symmetry of this periodic solution of the (θ, ν, w) system. Let $\ell \neq 0$ be minimal with $\alpha^\ell \gamma_p \in \hat\Lambda$ for some $\gamma_p \in \Gamma_p$. Then $\hat\Lambda/\hat\Gamma_p \simeq \mathbb{Z}_{n/\ell}$ by (a). Define $\Pi(\nu, w) = \gamma_p^{-1} Q_N^{-\ell} \Psi_{\ell,0}(\nu, w)$. We want to show that $\Pi(\hat\nu, \hat w) = (\hat\nu, \hat w)$. Because of (4.5) and because $\alpha^\ell \gamma_p \in \hat\Lambda$, we conclude that $\Pi^\nu(\hat\nu, \hat w) = \hat\nu$. Since

$$\Pi^{n/\ell} \;=\; \hat\gamma_p Q_N^{-n} \Psi_{n,0} \;=\; \hat\gamma_p \Psi_{n,0},$$

where $\hat\gamma_p \in \hat\Gamma_p$ by part (a), we also have $\Pi^{n/\ell}|_{\mathrm{Fix}(\hat\Gamma_p)} = \Psi_{n,0}|_{\mathrm{Fix}(\hat\Gamma_p)}$ and, therefore, $\Pi^{\frac{n}{\ell}+1}(\hat\nu, \hat w) = \Pi(\hat\nu, \hat w)$. Since $\gamma_p^{-1}\alpha^{-\ell} \in N(\hat\Gamma_p)$ and Q_1 and α generate the same twist diffeomorphism on G_p, $\Pi^w(\hat\nu, \hat w) \in \mathrm{Fix}_{N_1}(\Gamma_p)$ also. Hence $\Psi_{n,0}(\Pi(\hat\nu, \hat w)) = \Pi(\hat\nu, \hat w)$. Since $\Pi^\nu(\hat\nu, \hat w) = \hat\nu$ and $\hat w = w(\hat\nu) = \Pi^w(\hat\nu, \hat w)$ is locally unique, we conclude that $\Pi(\hat\nu, \hat w) = (\hat\nu, \hat w)$. So $(\hat\nu, \hat w)$ lies on an n-periodic solution of the periodically forced (ν, w) system of (3.16) with isotropy $\hat\Gamma_p$ and spatio-temporal symmetry $Q_N^\ell \gamma_p$.

3. Next we study the reversing symmetries of the periodic solution of the (θ, ν, w) system. So let $\hat L \neq \hat\Lambda$ and $\rho \in \hat L \setminus \hat\Lambda$. We are looking for a brake point on the periodic solution of the (θ, ν, w) system with reversing symmetry ρ. We have $\rho = g_p \alpha^i$ for some $g_p \in G_p \setminus \Gamma_p$ and some $i \in \{0, \ldots, n-1\}$. By (2.9) $\rho = (g_p, i)$ acts on (θ, ν, w) as

$$(g_p, i)(\theta, \nu, w) \;=\; (\chi(g_p)(\theta + i), g_p Q_0^i \nu, g_p Q_1^i w).$$

By definition of $\hat L$ we have $g_p Q_0^i \hat\nu = (\mathrm{Ad}^*_{g_p \alpha^i})^{-1} \hat\nu = \hat\nu$. Moreover,

$$\chi(g_p)(\hat\theta + i) = \hat\theta \quad \text{for} \quad \hat\theta := -i/2.$$

The periodic solution of the (θ, ν, w) system is given by $\{(\theta, \hat\nu, \Psi_{\theta,0}^w(\hat\nu, \hat w)),\ \ \theta \in \mathbb{R}/n\mathbb{Z}\}$. Let Φ_t^{red} denote the L_n-semiequivariant flow of the (θ, ν, w) system. Since $\Phi_n^{\mathrm{red}}(\hat\theta, \hat\nu, \Psi_{\hat\theta,0}^w(\hat\nu, \hat w)) = (\hat\theta, \hat\nu, \Psi_{\hat\theta,0}^w(\hat\nu, \hat w))$, we have

$$\begin{aligned}(\hat\theta, \hat\nu, g_p Q_1^i \Psi_{\hat\theta,0}^w(\hat\nu, \hat w)) &= (g_p, i)\Phi_n^{\mathrm{red}}(\hat\theta, \hat\nu, \Psi_{\hat\theta,0}^w(\hat\nu, \hat w)) \\ &= \Phi_{-n}^{\mathrm{red}}(\hat\theta, \hat\nu, g_p Q_1^i \Psi_{\hat\theta,0}^w(\hat\nu, \hat w))\end{aligned}$$

so that both $(\hat\theta, \hat\nu, g_p Q_1^i \Psi^w_{\hat\theta,0}(\hat\nu, \hat w))$ and $(\hat\theta, \hat\nu, \Psi^w_{\hat\theta,0}(\hat\nu, \hat w))$ lie on a periodic solution of the (θ, ν, w) system. Since by our nondegeneracy condition the periodic solution corresponding to $\hat\nu$ is locally unique, we get $g_p Q_1^i \Psi^w_{\hat\theta,0}(\hat\nu, \hat w) = \Psi^w_{\hat\theta,0}(\hat\nu, \hat w)$. So $g_p Q_N^i$ is a reversing symmetry of the periodic solution of the (θ, ν, w) system with brake point $(\hat\theta, \hat\nu, \Psi^w_{\hat\theta,0}(\hat\nu, \hat w))$.

4. Finally, we interpret the periodic solution of the (θ, ν, w) system as a relative periodic solution of the original system. Let

$$\hat p_E \;=\; \hat p_E(\hat\nu) \;\simeq\; (\mathrm{id}, \hat\theta, \hat\nu, \Psi^w_{\hat\theta,0}(\hat\nu, \hat w), E),$$

where $E \approx 0$, and $\hat\theta$ is arbitrary in the nonreversible case and as above in the reversible case. Since the $\dot E$ equation of (3.16) does not depend on E and because of (2.9), the point $\hat p_E(\hat\nu)$ lies on a relative periodic orbit $\mathcal{P}_E(\hat\nu)$ of (3.1) with isotropy $\Gamma_{\hat p_E} = \hat\Gamma_p$ and spatio-temporal symmetry $\hat\sigma$ near $\sigma^\ell \gamma_p$. In the reversible case $\hat L \neq \hat\Lambda$, we see from (2.9) that $\rho \hat p_E = \hat p_E$ so that $G_{\hat p_E} = (\Gamma_{\hat p_E})_\rho$. The condition $\hat\sigma = \sigma^\ell \gamma_p \exp(\hat\xi)$, where $\hat\xi \in \mathbf{z}^\chi(\hat G_p) \cap \mathbf{g}_{\hat\mu}$ is small, follows from the fact that the vector field f_G on the group is G_p-semiequivariant; see (2.10). ■

5. Example: Affine rigid bodies in ideal fluids.

In this section, we illustrate how to apply the results of this paper to a specific symmetric Hamiltonian system. As our example we have chosen a finite dimensional model for the dynamics of a deformable body in an ideal irrotational fluid. The model extends the well-known Kirchhoff model for the motion of a rigid body in a fluid [18, 21, 3]. In this model, the configuration of the body, i.e., its position and orientation in $\mathbb{R}^3$, is given by the elements of the special Euclidean group SE(3). The fluid motion outside the body is assumed to be irrotational, to have normal velocity at the surface of the body equal to that of the body, and to be stationary at infinity. It is therefore determined uniquely by the motion of the body itself, and the dynamics can be described by a Hamiltonian system on the cotangent bundle $T^*\mathrm{SE}(3)$.

We extend this model by relaxing the assumption that the body is rigid to allow configurations that are obtained from orientation preserving linear deformations of a reference body. If the reference body is assumed to be a sphere, then the deformed configurations are always ellipsoids. We assume that the deformations preserve volume. The configuration space for this system is therefore the special affine group $\mathrm{SAff}(3) = \mathrm{SL}(3) \ltimes \mathbb{R}^3$ of $\mathbb{R}^3$, where SL(3) is the group of invertible linear transformations of $\mathbb{R}^3$ with determinant 1, and the semidirect product is obtained from the natural action of SL(3) on $\mathbb{R}^3$. The dynamics of the system are given by a Hamiltonian H on $T^*\mathrm{SAff}(3)$.

In addition to extending the Kirchhoff model for a rigid body in a fluid, this system also extends the “affine” or “pseudorigid” body model used in fluid dynamics and elasticity theory [8, 11, 49, 53, 54]. These models are usually invariant under a Galilean transformation group, and so the translational degrees of freedom can be ignored by using a coordinate system that moves with the center of mass [35]. This is not true for a body in a fluid that is translating relative to the fluid at infinity. In this case, the symmetry group is essentially noncompact. This is described in the next subsection.

We will use Theorem 4.9 to deduce the existence of some families of relative periodic orbits of this model in section 5.4. Since the underlying symmetry group is noncompact, the existing theories—which use compactness of the symmetry group—do not give these families of relative periodic solutions. The solutions describe simple motions of deformable bodies in fluids.

In order to construct the relative periodic solutions, we start, in section 5.2, with a spherical equilibrium and study the dynamics in a neighborhood of this equilibrium. In section 5.3, we describe some families of nonlinear normal modes close to the equilibrium, and in section 5.4 we show how these normal modes persist to relative periodic orbits.

5.1. Symmetries and conserved quantities.

In this subsection, we describe the symmetries and corresponding conserved quantities of our model of an affine rigid body in an ideal fluid. We assume that the reference body is spherically symmetric, which implies that H is invariant under the action of $\mathrm{SO}(3)$ on $T^*\mathrm{SAff}(3)$, which is induced from its natural action on the right of $\mathrm{SL}(3)$ (extended trivially to $\mathrm{SAff}(3)$):

$$B.(S,s) \;=\; (SB^{-1},s), \qquad\qquad (S,s)\in \mathrm{SAff}(3),\; B\in \mathrm{SO}(3).$$

These are the "material" or "body" symmetries of the system. We also assume that the system is invariant under rotations and translations of $\mathbb{R}^3$, i.e., the natural action of $\mathrm{SE}(3)$ on $T^*\mathrm{SAff}(3)$ induced from its action on the left of $\mathrm{SAff}(3)$:

$$(A,a).(S,s) \;=\; (AS, a+As), \qquad\qquad (S,s)\in \mathrm{SAff}(3),\; (A,a)\in \mathrm{SE}(3).$$

These are the "spatial" symmetries of the system. This assumption implies that there are no external forces such as gravity acting. In particular, the body is "neutrally buoyant" and has coincident centers of mass and buoyancy. It is natural also to assume that the system is invariant under the action of the inversion symmetry $-\mathrm{id}$ in $\mathrm{O}(3)$ acting simultaneously on the left and right of $\mathrm{SAff}(3)$. Denoting the diagonally embedded inversion operator in $\mathrm{O}(3)\times\mathrm{O}(3)$ by κ, we have

$$\kappa.(S,s) \;=\; (S,-s), \qquad (S,s)\in \mathrm{SAff}(3),\; \kappa=(-\mathrm{id},-\mathrm{id})\in \mathrm{O}(3)\times\mathrm{O}(3).$$

Note that the action of $-\mathrm{id}$ on the left or right alone does not preserve $\mathrm{SAff}(3)$. Together the body and spatial symmetries and reflection κ generate a semidirect product $\Gamma = \mathbb{Z}_2^\kappa \ltimes (\mathrm{SO}(3)\times\mathrm{SE}(3))$.

Finally, we will also assume that the system is invariant under the usual time reversal symmetry operation acting on $T^*\mathrm{SAff}(3)$. Using left translations in $\mathrm{SAff}(3)$, we identify $T^*\mathrm{SAff}(3)$ with $\mathrm{SAff}(3)\times\mathrm{saff}(3)^*$, where $\mathrm{saff}(3) = T_{(\mathrm{id},0)}\mathrm{SAff}(3) = \mathrm{sl}(3)\oplus\mathbb{R}^3$. Then the action of the time reversal symmetry becomes

$$\rho.((S,s),(\mu_S,\mu_s)) \;=\; ((S,s),(-\mu_S,-\mu_s)), \qquad (S,s)\in \mathrm{SAff}(3),\; (\mu_S,\mu_s)\in \mathrm{saff}(3)^*.$$

The full group of time preserving and time reversing symmetries is $G=\Gamma\times\mathbb{Z}_2^\rho$.

It is a straightforward exercise to write down the conserved quantities associated to these symmetries; see [1]. In body coordinates $T^*\mathrm{SAff}(3)\cong \mathrm{SAff}(3)\times\mathrm{saff}(3)^*$, the momentum generated by the material symmetry group $\mathrm{SO}(3)$ (acting from the right) is

$$\mathbf{J}_R(S,s,\mu_S,\mu_s) \;=\; -\pi(\mu_S),$$

where $\pi : \mathrm{sl}(3)^* \to \mathrm{so}(3)^*$ is the natural projection dual to the inclusion $\mathrm{so}(3) \subset \mathrm{sl}(3)$. The momentum map for the spatial symmetry group SE(3) is

$$\mathbf{J}_L(S, s, \mu_S, \mu_s) = \Pi(\mathrm{Ad}^*_{(S,s)^{-1}}(\mu_S, \mu_s)),$$

where $\Pi : \mathrm{saff}(3)^* \to \mathrm{se}(3)^*$ is the natural projection. The two components of the momentum $\mathbf{J}_L$ can be interpreted as angular and linear *impulses* of the body-fluid system (see, for example, [51]). The momentum $-\mathbf{J}_R$ is the angular impulse in body coordinates.

5.2. The spherical equilibrium.

In this subsection, we describe the dynamics near a spherical equilibrium. So assume that the spherical configuration with zero-momentum $p = ((\mathrm{id}, 0), (0, 0))$ in $\mathrm{SAff}(3) \times \mathrm{saff}(3)^*$ is an equilibrium configuration. This has conserved momenta $\mu = (\mu_L, \mu_R) = (\mathbf{J}_L, \mathbf{J}_R) = (0, 0)$, and so $G_\mu = G$. The isotropy subgroup is $G_p = \mathrm{O}(3)_D \times \mathbb{Z}_2^\rho$, where $\mathrm{O}(3)_D = \mathbb{Z}_2^\kappa \times \mathrm{SO}(3)_D$ and $\mathrm{SO}(3)_D = \{(\gamma, (\gamma, 0)) \in \mathrm{SO}(3) \times \mathrm{SE}(3) : \gamma \in \mathrm{SO}(3)\}$ is the diagonally embedded copy of SO(3) in $\mathrm{SO}(3) \times \mathrm{SE}(3)$. Let $\mathrm{so}(3)_D$ denote the Lie algebra of $\mathrm{SO}(3)_D$. A complement $\mathbf{m}_\mu$ to $\mathbf{g}_p$ in $\mathbf{g}_\mu = \mathbf{g}$ is provided by $\mathrm{so}(3)_{AD} \oplus \mathbb{R}^3$, where $\mathrm{so}(3)_{AD} = \{(-\xi, (\xi, 0)) \in \mathrm{so}(3) \oplus \mathrm{se}(3) : \xi \in \mathrm{so}(3)\}$ is the antidiagonal embedding of so(3) in $\mathrm{so}(3) \oplus \mathrm{se}(3)$. Note that $\mathrm{so}(3)_{AD}$ is not a Lie subalgebra since $[\mathrm{so}(3)_{AD}, \mathrm{so}(3)_{AD}] \subset \mathrm{so}(3)_D$.

The symplectic normal space N_1 to the group orbit through p can be identified with $V \oplus V^*$, where V is the 5-dimensional subspace of $\mathrm{sl}(3) \subset \mathrm{saff}(3)$ consisting of symmetric traceless matrices and $V^* = \ker \Pi = \mathrm{ann}(\mathrm{se}(3))$ is the dual space in $\mathrm{saff}(3)^*$. We choose the standard symplectic structure $\omega(w_1, w_2) = -\Im(\mathrm{tr}(w_1 \bar{w}_2))$, where $w_i = u_i + \mathrm{i}v_i$, $i = 0, 1$, on $V \oplus V^*$. The group $\Gamma_p = \mathrm{O}(3)_D$ acts symplectically on $V \oplus V^*$ by conjugation of matrices. Note that κ acts trivially. An equivariant momentum map for this action is given by

$$\mathbf{L}_{N_1}(w) \;=\; vu - uv, \qquad w = (u, v) \in V \oplus V^* \tag{5.1}$$

(see Lemma 5.6 of [37]).

In [50], we have presented an analogue of Theorem 3.3 for relative equilibria. In this case, there is no phase θ, the equations (3.14) of Theorem 3.3 are time-independent, and N ($\widetilde{N}$) is a slice (extended slice) transverse to the relative equilibrium. Hence the dynamics near the group orbit of equilibria through p, considered as a relative equilibrium, is given by a system of ordinary differential equations on the extended slice

$$\widetilde{N} = \mathbf{g}^*_\mu \oplus N_1 \;\cong\; \mathrm{so}(3)^* \oplus \mathrm{se}(3)^* \oplus V \oplus V^*$$

of the form

$$\dot{\zeta} = \mathrm{ad}^*_{\mathrm{D}_\zeta h(\zeta, w)}(\zeta), \quad \dot{w} = J_{N_1} \mathrm{D}_w h(\zeta, w),$$

where h is the function on $\widetilde{N}$ obtained by writing the Hamiltonian H in body coordinates and J_{N_1} is the chosen symplectic structure on $V \oplus V^*$. We have used the fact that $\mu = (0, 0)$ is split to obtain these equations. Taking $\zeta = (\zeta_R, \zeta_L, \zeta_T)$, with $\zeta_R \in \mathrm{so}(3)^*$ and $(\zeta_L, \zeta_T) \in \mathrm{se}(3)^*$, the

ζ equation takes the more concrete form (see, e.g., [50])

$$\begin{aligned}\dot\zeta_R &= \zeta_R \times \frac{\partial h}{\partial \zeta_R},\\ \dot\zeta_L &= \zeta_L \times \frac{\partial h}{\partial \zeta_L} + \zeta_T \times \frac{\partial h}{\partial \zeta_T},\\ \dot\zeta_T &= \zeta_T \times \frac{\partial h}{\partial \zeta_L},\\ \dot w &= J_{N_1}\frac{\partial h}{\partial w}.\end{aligned} \tag{5.2}$$

To get the equation on the slice N, we could use the analogue of Theorem 3.5 for relative equilibria in [50]. But instead of computing the expression $\overline{\mathrm{ad}}^*_\xi$ that occurs in (3.16) of Theorem 3.5, we prefer to project (5.2) directly from the extended slice $\widetilde N$ onto the slice N. In order to do this, we first write the differential equations for $\zeta_{AD} = \frac{1}{2}(\zeta_R - \zeta_L)$, $\zeta_D = \frac{1}{2}(\zeta_R + \zeta_L)$, and ζ_T:

$$\begin{aligned}\dot\zeta_{AD} &= \zeta_D \times \frac{\partial h}{\partial \zeta_{AD}} + \zeta_{AD} \times \frac{\partial h}{\partial \zeta_D} - \zeta_T \times \frac{1}{2}\frac{\partial h}{\partial \zeta_T},\\ \dot\zeta_D &= \zeta_D \times \frac{\partial h}{\partial \zeta_D} + \zeta_{AD} \times \frac{\partial h}{\partial \zeta_{AD}} + \zeta_T \times \frac{1}{2}\frac{\partial h}{\partial \zeta_T},\\ \dot\zeta_T &= \zeta_T \times \frac{1}{2}\left(\frac{\partial h}{\partial \zeta_D} - \frac{\partial h}{\partial \zeta_{AD}}\right).\end{aligned} \tag{5.3}$$

To obtain the equations on the slice $N = N_0 \oplus N_1 \cong \mathrm{so}(3)^*_{AD} \oplus V \oplus V^*$, we set $\zeta = \nu + \mathbf{L}_{N_1}(w)$, where $\nu \in N_0 = (\mathbf{g}_\mu/\mathbf{g}_p)^* = \mathrm{so}(3)^*_{AD}$, $\mathbf{L}_{N_1}(w) \in \mathbf{g}_p^* = \mathrm{so}(3)^*_D$, and $h(\zeta, w) = h(\nu, w)$. Setting $\nu_{AD} = \zeta_{AD}$, $\nu_T = \zeta_T$ so that $\nu = (\nu_{AD}, \nu_T) \in N_0$ and using $\zeta_D = \mathbf{L}_{N_1}(w)$ and that $h = h(\nu, w)$ is independent of ζ_D give

$$\begin{aligned}\dot\nu_{AD} &= -\frac{\partial h}{\partial \nu_{AD}} \times \mathbf{L}_{N_1}(w) + \frac{1}{2}\frac{\partial h}{\partial \nu_T} \times \nu_T,\\ \dot\nu_T &= \frac{1}{2}\frac{\partial h}{\partial \nu_{AD}} \times \nu_T,\\ \dot w &= J_{N_1}\frac{\partial h}{\partial w},\end{aligned} \tag{5.4}$$

where all the partial derivatives of h are evaluated at (ν_{AD}, ν_T, w). These equations are semiequivariant with respect to the action of $G_p = \mathrm{O}(3)_D \times \mathbb{Z}_2^\rho$ on $N_0 \oplus N_1$. It would be an interesting exercise to compute their relative equilibria for the Hamiltonians h describing the motion of the body in the fluid. However, we do not attempt to give a systematic analysis of these equations here. Instead we will describe just some of the families of periodic orbits which bifurcate from the spherical equilibrium in the next subsection.

5.3. Nonlinear normal modes.

In this subsection, we describe some families of periodic orbits near the spherical equilibrium.

The solutions of (5.4) leave invariant the subset defined by $\nu = 0$, $\mathbf{L}_{N_1} = 0$. The Hamiltonian $h(0, w)$ is $\mathrm{O}(3)_D \times Z_2^\rho$-invariant, the action factoring through that of $\mathrm{SO}(3)_D$. Families of periodic orbits that typically bifurcate from spherically symmetric linearly stable equilibria of such Hamiltonians are described and illustrated in [36, 37]. Section 5 of [37] treats the irreducible symplectic representation of SO(3) on $V \oplus V^*$, where V is the space of symmetric traceless $(3,3)$-matrices, though without taking time reversibility into account. As can be seen from Table 4 of [37], there are three different symmetry types which have $\mathbf{L}_{N_1} = 0$ (and there are two more families of "rotating wave" normal modes with $\mathbf{L}_{N_1} \neq 0$ nearby which we will not consider). Ignoring the $\mathbb{Z}_2^\kappa$ symmetry group that acts trivially on N_1 but incorporating the time reversing symmetries, these three families have symmetry group triples (L_n, G_p, Γ_p) isomorphic to

I. $(\mathrm{O}(2) \times \mathbb{Z}_2^\rho,\ \mathrm{O}(2) \times \mathbb{Z}_2^\rho,\ \mathrm{O}(2))$,
II. $(\mathbb{D}_4 \times \mathbb{Z}_2^\rho,\ \mathbb{D}_2 \times \mathbb{Z}_2^\rho,\ \mathbb{D}_2)$,
III. $(\mathbb{O}^\rho,\ \mathbb{D}_4^\rho,\ \mathbb{D}_2)$.

Here $\mathbb{D}_2$ is the subgroup of $\mathrm{SO}(3)$ consisting of rotations by π about each of three mutually perpendicular axes. The subgroup $\mathbb{D}_4$ is generated by $\mathbb{D}_2$ together with rotations by π about axes in the plane of, and bisecting, two of the $\mathbb{D}_2$ axes. The group $\mathbb{D}_4^\rho$ is the subgroup of $\mathrm{SO}(3)_D \times \mathbb{Z}_2^\rho$ obtained by composing the additional rotations by π in $\mathbb{D}_4 \setminus \mathbb{D}_2$ with ρ. The group $\mathbb{O}$ is the subgroup of order 24 in $\mathrm{SO}(3)$ consisting of all rotations which preserve a cube. It can be generated by $\mathbb{D}_4$ together with an element of order 3 corresponding to a rotation about a diagonal of the cube. The subgroup $\mathbb{O}^\rho$ in $\mathrm{SO}(3)_D \times \mathbb{Z}_2^\rho$ is similar, but with $\mathbb{D}_4$ replaced by $\mathbb{D}_4^\rho$. Finally, $\mathrm{O}(2)$ is the subgroup of $\mathrm{SO}(3)$ consisting of all rotations about one axis and rotations by π about each of the perpendicular axes. Note that the kernels of the "sign" homomorphisms $\chi : L_n \to \mathbb{Z}_2$ in the three cases are, respectively, $\mathrm{O}(2)$, $\mathbb{D}_4$, and $\mathbb{T}$, the group of all rotations which preserve a regular tetrahedron.

In the first case, the spatio-temporal symmetry σ is trivial, and $k = n = 1$. In the second case, σ can be taken to be one of the rotations by π in $\mathbb{D}_4$ that does not lie in $\mathbb{D}_2$. In this case, $k = n = 2$. In the third case, σ can be chosen to be a rotation by $2\pi/3$ about a diagonal of the cube, and $k = n = 3$.

Since for these periodic solutions the reversing isotropy subgroups G_p and isotropy subgroups Γ_p do not coincide, all of them are reversible. By construction they all have zero-momentum, i.e., $\mathbf{J}_L = \mathbf{J}_R = 0$, and all can be described as "pulsating cubes." At all times the body is ellipsoidal (which is why Γ_p always contains $\mathbb{D}_2 \times \mathbb{Z}_2^\kappa$), and its principal axes have fixed directions in both body and space. However, the lengths of the principal axes vary periodically in different ways. In the first case, the qualitative behavior is determined by the fact that the ellipsoid is always axisymmetric. In the second case, the longest axis switches periodically between two of the three, and the length of third axis varies with twice the period and a much smaller amplitude than the other two. The spatio-temporal symmetry σ corresponds to rotating by π about an axis bisecting the two principal axes with large amplitude variations. In the third case, the role of the longest principal axis is taken by each of the three in turn, with a $2\pi/3$ phase shift between them. The spatio-temporal symmetry corresponds to rotating the body by $2\pi/3$ about an axis trisecting the three principal axes. We will refer to them as the "axisymmetric," "square," and "cubic" oscillations, respectively.

We describe the (θ, ν, w) equations for each of these periodic oscillations in turn. In the last two cases, the isotropy subgroup G_p is finite, and so $N_0 = \mathbf{g}^* = \mathrm{so}(3)^* \oplus \mathrm{se}(3)^*$ with its natural χ-coadjoint action of L_n. In both cases, the symplectic normal space is a two dimensional semisymplectic representation of L_n. These representations can be read from Table 6 of [37]. For the square case, it is given by the nontrivial representation of $\mathbb{D}_4$ on $\mathbb{C}$ with kernel $\mathbb{D}_2$. The time reversal symmetry ρ acts by conjugation on $\mathbb{C}$. In the cubic case, it is the representation of $\mathbb{O}^\rho$ that factors through the two dimensional irreducible representation of $\mathbb{O}^\rho/\mathbb{D}_2 \cong \mathbb{D}_3$. The (θ, ν, w) equations in both cases have the form (5.2) with ζ replaced by ν, h an L_n-invariant function of $(\nu_R, \nu_L, \nu_T, w, \theta)$ and the addition of the equation $\dot{\theta} = 1$.

For the axisymmetric oscillations, $\mathbf{g}_p = \mathrm{so}(2)_D$, and so

$$N_0 \;=\; (\mathbf{g}/\mathbf{g}_p)^* \;=\; (\mathrm{so}(3)_D/\mathrm{so}(2)_D)^* \oplus \mathrm{so}(3)^*_{AD} \oplus (\mathbb{R}^3)^*.$$

Table 6 in [37] shows that the symplectic normal space N_1 is the four dimensional irreducible symplectic representation of $\mathrm{O}(2)$ on $\mathbb{C}^2$ with kernel $\mathbb{D}_2$ and ρ acting by conjugation. It can be identified with the subspace of $\mathrm{sl}(3) \otimes \mathbb{C} \subset \mathrm{saff}(3) \otimes \mathbb{C} \cong \mathrm{saff}(3) \oplus \mathrm{saff}(3)^*$ consisting of symmetric traceless matrices of the form

$$A = \begin{pmatrix} a & b & 0 \\ b & -a & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad a, b \in \mathbb{C}.$$

In these coordinates, the momentum map ("vibrational angular momentum") $\mathbf{L}_{N_1} : N_1 \to \mathrm{so}(2)_D$ is $\mathbf{L}_{N_1}(A) = \frac{\mathrm{i}}{2}(A\overline{A} - \overline{A}A)$ (see Lemma 5.6 of [37]). The Hamiltonian h is a function of $\nu_{AD} \in \mathrm{so}(3)^*_{AD}$, $\nu_D \in (\mathrm{so}(3)_D/\mathrm{so}(2)_D)^*$, $\nu_T \in (\mathbb{R}^3)^*$, A, $\overline{A}$, and θ. The N_0 part of the slice equations is easily obtained from (5.3) by replacing

$$\zeta_D \;\mapsto\; (\nu_D, \mathbf{L}_{N_1}(w)), \qquad \frac{\partial h}{\partial \zeta_D} \;\mapsto\; \begin{pmatrix} \frac{\partial h}{\partial \nu_D} \\ 0 \end{pmatrix}$$

and by projecting the $\dot{\zeta}_D$ equation to ν_D. To these must be added the equations

$$\dot{\theta} = 1, \qquad \dot{A} = -2\mathrm{i}\frac{\partial h}{\partial \overline{A}}, \quad \text{where} \quad A \simeq (a, b) \in \mathbb{C}^2.$$

The second of these equations is the w equation written in appropriate complex coordinates.

For all three cases, the component $M_1 : N_1 \to N_1$ of the linearizations of the (ν, w) equations at the periodic orbits will be equal to the "reduced Floquet operators" computed in [37] in terms of coefficients in the Taylor series expansion of the Hamiltonian at the spherical equilibrium. The results given there, combined with Corollary 4.5 and the fact that $\overline{\mathrm{Ad}}^*_\sigma$ is always spectrally stable for compact and Euclidean groups, imply that the cubic oscillations are always spectrally stable (since the representations of Γ_p and G_p on N_1 are cyclospectral), while typically either the axisymmetric oscillations or the square oscillations are spectrally stable but not both.

5.4. Relative periodic orbits. In this final subsection, we use Theorem 4.9 to describe some relative periodic orbits that will typically bifurcate from the square and cubic oscillations of the previous subsection as $\mathbf{J}_L$ and $\mathbf{J}_R$ are perturbed away from 0. We assume that the original normal modes are nondegenerate in the sense required in Theorem 4.9, an assumption which is generically satisfied. In both of these cases, $\Gamma_p = \mathbb{D}_2 \times \mathbb{Z}_2^\kappa$.

The coadjoint orbits $\Gamma\nu$ for the action of $\Gamma = \mathbb{Z}_2^\kappa \ltimes (\mathrm{SO}(3) \times \mathrm{SE}(3))$ on $\mathbf{g}^* = \mathrm{so}(3)^* \oplus \mathrm{se}(3)^* = \mathrm{so}(3)^*_R \oplus \mathrm{so}(3)^*_L \oplus (\mathbb{R}^3)^*$ are given by $\mathcal{O}_\nu = \mathcal{O}_{\nu_R,\nu_L,\nu_T} = \mathcal{O}_{\nu_R} \times \mathcal{O}_{\nu_L,\nu_T}$, where

$$\mathcal{O}_{\nu_R} = \{\hat{\nu}_R \in \mathrm{so}(3)^* : ||\hat{\nu}_R|| = ||\nu_R||\},$$
$$\mathcal{O}_{\nu_L,\nu_T} = \begin{cases} \{(\hat{\nu}_L, \hat{\nu}_T) \in \mathrm{se}(3)^* : \hat{\nu}_T = 0,\ ||\hat{\nu}_L|| = ||\nu_L||\} & \text{if } \nu_T = 0, \\ \{(\hat{\nu}_L, \hat{\nu}_T) \in \mathrm{se}(3)^* : \hat{\nu}_L.\hat{\nu}_T = \nu_L.\nu_T,\ ||\hat{\nu}_T|| = ||\nu_T||\} & \text{if } \nu_T \neq 0. \end{cases}$$

Thus $\mathcal{O}_{\nu_R}$ is either a point or a two-sphere, while $\mathcal{O}_{\nu_L,\nu_T}$ is a point or a two-sphere or is diffeomorphic to the tangent bundle of a two-sphere.

The isotropy subgroups $\hat{\Gamma}_p$ of the actions of $\Gamma_p = \mathbb{D}_2 \times \mathbb{Z}_2^\kappa$ on the coadjoint orbits $\Gamma\nu$ can be computed easily. The action of $\mathbb{Z}_2^\kappa$ on $\mathrm{so}(3)_R^*$ is trivial, while that of $\mathbb{D}_2$ has one dimensional fixed point subsets for each of its three $\mathbb{Z}_2$ subgroups. It follows that if τ is a nonidentity element of $\mathbb{D}_2$ and $\nu_R \neq 0$, the two-sphere $\mathcal{O}_{\nu_R}$ has precisely 2 points with isotropy subgroup $\mathbb{Z}_2^\tau \times \mathbb{Z}_2^\kappa$. The same is true for the two-sphere coadjoint orbits $\mathcal{O}_{\nu_L,0}$ in $\mathrm{se}(3)_L^*$. So if $\nu_T = 0$ and $\nu_R \neq 0 \neq \nu_L$, then $\mathcal{O}_\nu$ has precisely four points with isotropy subgroup $\mathbb{Z}_2^\tau \times \mathbb{Z}_2^\kappa$. On the $(\mathbb{R}^3)^*$ component of $\mathrm{se}(3)^*$ the operation κ acts by $-\mathrm{id}$, while $\mathbb{D}_2$ acts in the same way as on $\mathrm{so}(3)^*$. The isotropy subgroups with one dimensional fixed point spaces for $\nu_T \in (\mathbb{R}^3)^*$ are therefore equal to $\mathbb{Z}_2^\tau \times \mathbb{Z}_2^{\hat{\tau}\circ\kappa}$, where τ and $\hat{\tau}$ are any two distinct nonidentity elements in $\mathbb{D}_2$. It follows that if $\nu_T \neq 0$ but $\nu_R = 0$ and $\nu_L . \nu_T = 0$, then $\mathcal{O}_\nu$ has two points with isotropy group equal to $\mathbb{Z}_2^\tau \times \mathbb{Z}_2^{\hat{\tau}\circ\kappa}$ lying in the $\{(\nu_R, \nu_L) = 0\}$-plane. If $\nu_T \neq 0$, $\nu_R = 0$, and $\nu_L . \nu_T \neq 0$, then $\mathcal{O}_\nu$ has two points with isotropy group equal to $\mathbb{Z}_2^\tau$, while if $\nu_T \neq 0$ and $\nu_R \neq 0$, then $\mathcal{O}_\nu$ has four points with isotropy group equal to $\mathbb{Z}_2^\tau$.

Summarizing, the subgroups $\hat{\Gamma}_p$ with zero dimensional fixed point sets $\Gamma\nu \cap \mathrm{Fix}_{\hat{\Gamma}_p}(\mathbf{g}^*)$ are given in the following table. In all these fixed point spaces, $\hat{\nu}_R$, $\hat{\nu}_L$, and $\hat{\nu}_T$ are parallel to each other.

	Orbit $\Gamma\nu$	Isotropy $\hat{\Gamma}_p$	Fixed point set
1.	$\nu_T = 0,\ (\nu_R, \nu_L) \neq 0$	$\mathbb{Z}_2^\tau \times \mathbb{Z}_2^\kappa$	$\hat{\nu}_T = 0,\ \hat{\nu}_R \mid\mid \hat{\nu}_L \mid\mid \tau$
2.	$\nu_T \neq 0,\ \nu_R = 0,\ \nu_L . \nu_T = 0$	$\mathbb{Z}_2^\tau \times \mathbb{Z}_2^{\hat{\tau}\circ\kappa}$	$\hat{\nu}_T \mid\mid \tau,\ \hat{\nu}_R = \hat{\nu}_L = 0$
3.	$\nu_T \neq 0,\ \nu_R \neq 0$ or $\nu_L . \nu_T \neq 0$	$\mathbb{Z}_2^\tau$	$\hat{\nu}_T \mid\mid \hat{\nu}_R \mid\mid \hat{\nu}_L \mid\mid \tau$

In cases 1 and 3, the element τ is any of the nonidentity elements of $\mathbb{D}_2 = \mathrm{SO}(3)_D \cap \Gamma_p$, and in case 2, the elements τ and $\hat{\tau}$ are two different nonidentity elements in $\mathbb{D}_2$. The notation $\nu \mid\mid \tau$ means that ν is parallel to the axis fixed by τ.

By Theorem 4.9 the momentum $\hat{\mu}$ of the relative periodic orbits corresponding to a fixed point $\hat{\nu} = (\hat{\nu}_R, \hat{\nu}_L, \hat{\nu}_T)$ is given simply by $\hat{\mu} = \hat{\nu}$. The momentum isotropy subgroups $G_{\hat{\mu}}$ are the isotropy subgroups at $\hat{\mu}$ for the action of $G = \mathbb{Z}_2^\rho \times \mathbb{Z}_2^\kappa \ltimes (\mathrm{SO}(3) \times \mathrm{SE}(3))$ on $\mathbf{g}^* = \mathrm{so}(3)^* \oplus \mathrm{se}(3)^*$ and are easily calculated to be the following:

	Momentum $\hat{\mu}$			Momentum isotropy $G_{\hat{\mu}}$
(1a)	$\hat{\mu}_T = 0,$	$\hat{\mu}_R \neq 0,$	$\hat{\mu}_L = 0$	$\mathbb{Z}_2^\kappa \ltimes (\mathrm{O}(2)_R^\rho \times \mathrm{SE}(3))$
(1b)	$\hat{\mu}_T = 0,$	$\hat{\mu}_R = 0,$	$\hat{\mu}_L \neq 0$	$\mathbb{Z}_2^\kappa \ltimes ((\mathrm{SO}(3)_R \times \mathrm{O}(2)_L^\rho) \ltimes \mathbb{R}^3)$
(1c)	$\hat{\mu}_T = 0,$	$\hat{\mu}_R \neq 0,$	$\hat{\mu}_L \neq 0$	$\mathbb{Z}_2^\kappa \ltimes ((\mathrm{O}(2)_R \times \mathrm{O}(2)_L)^\rho \ltimes \mathbb{R}^3)$
(2)	$\hat{\mu}_T \neq 0,$	$\hat{\mu}_R = 0,$	$\hat{\mu}_L = 0$	$\mathbb{Z}_2^{\kappa\circ\rho} \ltimes (\mathrm{SO}(3)_R \times \mathrm{O}(2)_L^\rho \times \mathbb{R})$
(3a)	$\hat{\mu}_T \neq 0,$	$\hat{\mu}_R = 0,$	$\hat{\mu}_L \neq 0$	$\mathrm{SO}(3)_R \times \mathrm{O}(2)_L^\rho \times \mathbb{R}$
(3b)	$\hat{\mu}_T \neq 0,$	$\hat{\mu}_R \neq 0$		$(\mathrm{O}(2)_R \times \mathrm{O}(2)_L)^\rho \times \mathbb{R}$

In all cases, $\hat{\mu}_R$, $\hat{\mu}_L$, and $\hat{\mu}_T$ are parallel to each other. The group $\mathrm{O}(2)_R^\rho$ is the subgroup of $\mathrm{SO}(3)_R \times \mathbb{Z}_2^\rho$ consisting of all rotations about a fixed axis together with ρ composed with rotations by π about axes perpendicular to this fixed axis. The group $\mathrm{O}(2)_L^\rho$ is the analogous subgroup of $\mathrm{SE}(3) \times \mathbb{Z}_2^\rho$, and $(\mathrm{O}(2)_R \times \mathrm{O}(2)_L)^\rho$ is the group consisting of all rotations about a fixed axis in $\mathrm{SO}(3)_R$, all rotations about the same fixed axis in $\mathrm{SO}(3)_L$, together with ρ

composed with simultaneous rotations by π about axes perpendicular to this fixed axis in $\mathrm{SO}(3)_R$ and $\mathrm{SE}(3)_L$.

For each of the points with one of the isotropy subgroups $\hat{\Gamma}_p$, we can now compute $\hat{L} := L_n \cap G_{\hat{\mu}}$. The results are shown in Table 5.1 for the square case and in Table 5.2 for the cubic case. The tables also give the symmetry data of the bifurcating relative periodic orbits: the new (reversing) isotropy $\hat{G}_p$, the new relative period ℓ, the new spatio-temporal symmetry $\hat{\sigma}$, and the new drift $\hat{\xi}$, computed as in Theorem 4.9. In each case, $\hat{\sigma}$ is equal to $\sigma^\ell \gamma_p \exp(\hat{\xi})$, where σ is the spatio-temporal symmetry for the original oscillation, $\ell \in \mathbb{N}$ is minimal such that $\sigma^\ell \gamma_p \in \hat{L}$ for some $\gamma_p \in \Gamma_p$, and $\hat{\xi} = (\hat{\xi}_R, \hat{\xi}_L, \hat{\xi}_T) \approx 0$ must lie in the fixed point subspace of the χ-dual action of $\hat{G}_p$ on $\mathrm{so}(3) \oplus \mathrm{se}(3)$. For the square relative periodic orbit of type (i), we have $\ell = 2$, and so $\sigma^\ell = \mathrm{id} = \gamma_p$. For the square relative periodic orbit of type (ii), we have $\ell = 1$, γ_p is the rotation by π about one of the principal axes undergoing large amplitude oscillations, and $\sigma\gamma_p = \tau^{\frac{1}{2}}$, i.e., a rotation by $\pi/2$ about the axis of τ. For the cubic relative periodic orbits, $\ell = 3$ and $\sigma^\ell = \mathrm{id} = \gamma_p$. In all cases, the forms of the possible $\hat{\xi}$'s are shown in the final columns of the tables.

Table 5.1

Symmetries of relative periodic orbits bifurcating from the square oscillations. The group D_2^ρ is $\mathsf{Z}_2^\tau \times \mathsf{Z}_2^{\hat{\tau}\circ\rho}$, where τ and $\hat{\tau}$ are two different nonidentity elements in $\mathsf{D}_2 = \mathrm{SO}(3) \cap \Gamma_p$. The group $\hat{\mathsf{D}}_4^\rho$ is generated by the rotation $\tau^{\frac{1}{2}}$ by $\pi/2$ about the τ-axis and by $\hat{\tau} \circ \rho$.

	$\hat{\Gamma}_p$	$\hat{G}_p$	$\hat{L}$	ℓ	$\hat{\sigma}$	$\hat{\xi}$
1. (i)	$\mathsf{Z}_2^\tau \times \mathsf{Z}_2^\kappa$	$\mathsf{D}_2^\rho \times \mathsf{Z}_2^\kappa$	$\mathsf{D}_2^\rho \times \mathsf{Z}_2^\kappa$	2	$\exp(\hat{\xi})$	$\hat{\xi}_R \,\vert\vert\, \hat{\xi}_L \,\vert\vert\, \tau, \quad \hat{\xi}_T = 0$
(ii)			$\hat{\mathsf{D}}_4^\rho \times \mathsf{Z}_2^\kappa$	1	$\tau^{\frac{1}{2}} \exp(\hat{\xi})$	
2. (i)	$\mathsf{Z}_2^\tau \times \mathsf{Z}_2^{\hat{\tau}\circ\kappa}$	$\mathsf{D}_2^\rho \times \mathsf{Z}_2^{\hat{\tau}\circ\kappa}$	$\mathsf{D}_2^\rho \times \mathsf{Z}_2^{\hat{\tau}\circ\kappa}$	2	$\exp(\hat{\xi})$	$\hat{\xi}_R = \hat{\xi}_L = 0, \quad \hat{\xi}_T \,\vert\vert\, \tau$
(ii)			$\hat{\mathsf{D}}_4^\rho \times \mathsf{Z}_2^{\hat{\tau}\circ\kappa}$	1	$\tau^{\frac{1}{2}} \exp(\hat{\xi})$	
3. (i)	Z_2^τ	D_2^ρ	D_2^ρ	2	$\exp(\hat{\xi})$	$\hat{\xi}_R \,\vert\vert\, \hat{\xi}_L \,\vert\vert\, \hat{\xi}_T \,\vert\vert\, \tau$
(ii)			$\hat{\mathsf{D}}_4^\rho$	1	$\tau^{\frac{1}{2}} \exp(\hat{\xi})$	

Table 5.2

Symmetries of relative periodic orbits bifurcating from the cubic oscillations; τ and $\hat{\tau}$ are two different nonidentity elements in $\mathsf{D}_2 = \mathrm{SO}(3) \cap \Gamma_p$. The group $\widetilde{\mathsf{D}}_2^\rho$ is $\mathsf{Z}_2^\tau \times \mathsf{Z}_2^{\tilde{\tau}\circ\rho}$, where $\tilde{\tau}$ is a rotation by π about an axis perpendicular to the τ-axis and inclined at an angle of $\pi/4$ to the $\hat{\tau}$-axis. The group $\mathsf{D}_4^{\rho,\kappa}$ is generated by $\widetilde{\mathsf{D}}_2^\rho$ and $\hat{\tau} \circ \kappa$.

	$\hat{\Gamma}_p$	$\hat{G}_p$	$\hat{L}$	ℓ	$\hat{\sigma}$	$\hat{\xi}$
1.	$\mathsf{Z}_2^\tau \times \mathsf{Z}_2^\kappa$	$\widetilde{\mathsf{D}}_2^\rho \times \mathsf{Z}_2^\kappa$	$\widetilde{\mathsf{D}}_2^\rho \times \mathsf{Z}_2^\kappa$	3	$\exp(\hat{\xi})$	$\hat{\xi}_R \,\vert\vert\, \hat{\xi}_L \,\vert\vert\, \tau, \quad \hat{\xi}_T = 0$
2.	$\mathsf{Z}_2^\tau \times \mathsf{Z}_2^{\hat{\tau}\circ\kappa}$	$\mathsf{D}_4^{\rho,\kappa}$	$\mathsf{D}_4^{\rho,\kappa}$	3	$\exp(\hat{\xi})$	$\hat{\xi}_R = \hat{\xi}_L = 0, \quad \hat{\xi}_T \,\vert\vert\, \tau$
3.	Z_2^τ	$\widetilde{\mathsf{D}}_2^\rho$	$\widetilde{\mathsf{D}}_2^\rho$	3	$\exp(\hat{\xi})$	$\hat{\xi}_R \,\vert\vert\, \hat{\xi}_L \,\vert\vert\, \hat{\xi}_T \,\vert\vert\, \tau$

In the case of square oscillations, for each of the different spatial isotropy subgroups $\hat{\Gamma}_p$, the cases indicated by (i) and (ii) in Table 5.1 give two qualitatively distinct types of bifurcating relative periodic orbits. In the cases labelled by (i), the "angular velocities" $\hat{\xi}_L$, $\hat{\xi}_R$ and "linear velocity" $\hat{\xi}_T$ are all aligned with one of the axes of the pulsating cube with large amplitude oscillations, while in the cases labelled by (ii), they are aligned with the third axis with much smaller amplitude oscillations. In case (i), the relative period doubles, while in case (ii), the relative period remains approximately the same. For the relative periodic orbits of types 1(i) and 1(ii), the linear velocity is zero, but there may be both body and spatial rotations, and so the bifurcating relative periodic orbits are modulated rotating waves. For cases 2(i) and 2(ii), only the linear velocity is nonzero, and the body translates in space without rotating. For cases 3(i) and 3(ii), rotation and translation both occur. So in cases 2 and 3, the bifurcating relative periodic orbits are modulated travelling waves. All bifurcating relative periodic orbits are reversible.

The case of bifurcations from the cubic oscillations is completely analogous, except that now there is no distinction between the three principal axes of the pulsating cube, and so there is only one type of bifurcating relative periodic orbit. These may again have body and spatial rotations only, translation only, or all three, and the relative period always triples.

In future work, we will extend the bifurcation results used here and apply them to show that a number of other types of relative periodic orbits bifurcate from the square and cubic oscillations and to find relative periodic orbits bifurcating from the axisymmetric oscillations.

6. Proofs.

This section is devoted to the proofs of the main theorems. The proofs build on the construction of coordinates near relative periodic orbits of general systems that we describe in section 6.1. In the subsequent subsections, we show how to adapt this bundle construction to Hamiltonian systems. First, in subsection 6.2, the symplectic structure of the tangent space decomposition at a point p on a relative periodic orbit is described. Then, in subsection 6.3, we analyze the linearization at a point p of the relative periodic orbit as this is needed for the construction of the bundle coordinates. In subsections 6.4 and 6.5, we present the adaptations of the bundle construction of subsection 6.1 to Hamiltonian systems. In subsections 6.6 and 6.7, we describe the symplectic structure of the bundle. Finally, in subsection 6.8, we derive the differential equations in bundle coordinates.

6.1. The bundle construction for general systems.

In this section, we describe the construction of coordinates near relative periodic orbits of general systems. Most of this section summarizes results of [52, 55, 22].

As always, let Γ be algebraic, let $p = \sigma^{-1}\Phi_1(p)$ lie on a relative periodic orbit of relative period 1, and let $M = \sigma^{-1}\mathrm{D}\Phi_1(p)$. Furthermore, let P be a G_p-equivariant projection from $T_p\mathcal{M}$ to the G_p-invariant Poincaré section (or normal space) N to $\mathcal{P}$ at p with kernel $T_p\mathcal{P} = T$. According to [52, 55, 22], there is a smooth family $N(\theta)$ of Γ_p-invariant Poincaré sections to $\mathcal{P}$ at $\Phi_\theta(p)$ such that $N(0) = N$, $N(\theta) \oplus T_{\Phi_\theta(p)}\mathcal{P} = T_{\Phi_\theta(p)}\mathcal{M}$, where $T_{\Phi_\theta(p)}\mathcal{P} = \mathrm{span}(f(\Phi_\theta(p)) \oplus \mathbf{g}\Phi_\theta(p)$, and

$$N(\theta+1) = \sigma N(\theta), \quad \rho N(\theta) = N(-\theta) \ \text{ for } \ \rho \in G_p \setminus \Gamma_p.$$

Let $P(\theta)$ be the projection from $T_{\Phi_\theta(p)}\mathcal{M}$ onto $N(\theta)$ with kernel $\ker P(\theta) = T_{\Phi_\theta(p)}\mathcal{P}$. Then $P(\theta)$ is smooth in θ, $P(\theta+1)\sigma = \sigma P(\theta)$, $P(0) = P$, and $P(\theta)$ is G_p-semiequivariant:

$$P(\theta) = g_p^{-1}P(\chi(g_p)\theta)g_p, \quad g_p \in G_p.$$

Further, by [22, Lemma 5.1] (see Lemma 6.5 below), there is a Γ_p-equivariant homotopy

$I_N(\theta) \in \mathrm{GL}(N)$ depending smoothly on θ and such that

$$I_N(0) = \mathrm{id}, \quad M_N I_N(\theta+1) = I_N(\theta) Q_N^{-1}, \quad \rho I_N(\theta)\rho^{-1} = I_N(-\theta), \quad \rho \in G_p \setminus \Gamma_p, \tag{6.1}$$

where $M_N := PM|_N$ and Q_N^{-1} is twisted semiequivariant and has finite order $2n$.

The parametrization of a G-invariant neighborhood $\mathcal{U}$ of $\mathcal{P}$ in $\mathcal{M}$ is then given by a submersion $\tau : G \times N \times \mathbb{R} \to \mathcal{U}$ defined by

$$u \;=\; \tau(g,\theta,v) \;=\; g\exp(-\theta\xi)\psi(\Phi_\theta(p), P(\theta)\mathrm{D}\Phi_\theta(p)I_N(\theta)v), \tag{6.2}$$

where ψ is a G-equivariant diffeomorphism from a neighborhood of $\mathcal{P}$ in its normal bundle to $\mathcal{U}$.

In this paper, we will construct the Poincaré sections $N(\theta)$ in a slightly different way from the method used in [52, 55, 22]. We will show in section 6.5, Lemma 6.6, that there is a homotopy $I(\theta) \in \mathrm{GL}(T_p\mathcal{M})$ which is G_p-semiequivariant:

$$I(\theta) \;=\; g_p^{-1} I(\chi(g_p)\theta) g_p, \quad g_p \in G_p, \tag{6.3}$$

and such that

$$M\, I(\theta+1) = I(\theta)Q^{-1}, \quad I(0) = \mathrm{id}, \tag{6.4}$$

where $Q = \mathrm{diag}(Q_T, Q_N)$, Q_T is an orthogonal transformation of T of finite order $2n$, Q^{-1} is twisted semiequivariant, and $I(\theta)$ has block structure

$$I(\theta) = \begin{pmatrix} I_T(\theta) & I_D(\theta) \\ 0 & I_N(\theta) \end{pmatrix}$$

with $I_N(\theta)$ satisfying (6.1). We then define

$$N(\theta) := \mathrm{D}\Phi_\theta(p) I(\theta) N.$$

This gives Γ_p-invariant Poincaré sections $N(\theta)$ with the above properties, and we get

$$\mathrm{D}\Phi_\theta(p) I(\theta)|_N = P(\theta)\mathrm{D}\Phi_\theta(p) I_N(\theta). \tag{6.5}$$

6.2. Symplectic structure of the tangent space decomposition.

Again, let $p = \sigma^{-1}\Phi_1(p)$ lie on a relative periodic orbit $\mathcal{P}$, and let $T = T_0 \oplus T_1 \oplus T_2$ be the refinement of the G_p-invariant tangent space to $\mathcal{P}$ at p given in (4.3). In this subsection, we show that there is a G_p-invariant Poincaré section $N \subset T_p\mathcal{M}$ to $\mathcal{P}$ at p such that the refinement $N = N_0 \oplus N_1 \oplus N_2$ defined in (3.6) holds true, and we discuss the symplectic structure of this decomposition of the tangent space $T_p\mathcal{M} = T \oplus N$. Define the ω-orthogonal complement of any subspace $V \subset T_p\mathcal{M}$ to be

$$V^\omega = \{u \in T_p\mathcal{M} \;:\; \omega(u,v) = 0 \text{ for all } v \in V\}.$$

Lemma 6.1. *Let p lie on a relative periodic orbit with relative period different from zero. Then the following hold.*

(a) *The vector $f_H(p) \in T_2$ is G_p-semiinvariant and linearly independent of T_pGp and lies in $\ker \mathrm{D}H(p) \cap \ker \mathrm{D}\mathbf{J}(p)$.*

(b) *$\mathrm{D}H(p)$ is linearly independent of the vectors $\mathrm{D}\mathbf{J}_\xi(p)$, $\xi \in \mathbf{g}$, and there is a G_p-invariant vector $v_E \in \ker\mathrm{D}\mathbf{J}(p)$ with $\mathrm{D}H(p)v_E \neq 0$ such that $T_p\mathcal{M} = \mathrm{span}(v_E) \oplus \ker \mathrm{D}H(p)$.*

(c) *$\ker \mathrm{D}\mathbf{J}(p) = (\mathbf{g}p)^\omega$, $\ker \mathrm{D}H(p) = T_2^\omega$, $T \subset \ker \mathrm{D}H(p)$, and $T \cap \ker \mathrm{D}\mathbf{J}(p) = T_0 \oplus T_2$.*

Proof. To prove part (a) note that G_p-semi-invariance of $f_H(p)$ follows from G_p-semiequivariance of f_H. Since $\mathcal{P}$ is a relative periodic orbit and not a relative equilibrium, T_pGp and $f_H(p)$ are linearly independent. The Hamiltonian H and the momentum $\mathbf{J}$ are preserved by the Hamiltonian flow of (3.1), and so $f_H(p) \in \ker \mathrm{D}H(p) \cap \ker \mathrm{D}\mathbf{J}(p)$.

To prove part (b) observe that if $\mathrm{D}H(p) = \mathrm{D}\mathbf{J}_\xi(p)$ for some $\xi \in \mathbf{g}$, then p lies on a relative equilibrium, which we exclude. Hence there is some $v_E \neq 0$ with $v_E \in \ker \mathrm{D}\mathbf{J}(p)$, but $\mathrm{D}H(p)v_E \neq 0$. Since $\ker \mathrm{D}H(p)$ has codimension 1 in $T_p\mathcal{M}$, we conclude that $\ker \mathrm{D}H(p)$ and v_E span $T_p\mathcal{M}$. We have $g_pv_E = \pm v_E$ for each $g_p \in G_p$ because N_2 is one dimensional and G_p-invariant. Since H is G_p-invariant, $\mathrm{D}H(p)g_p = \mathrm{D}H(p)$ for all $g_p \in G_p$, and, therefore, $0 \neq \mathrm{D}H(p)g_pv_E = \mathrm{D}H(p)v_E$, which proves that v_E is G_p-invariant.

The first two equations in part (c) follow from

$$\omega(f_H(p), v) = \mathrm{D}H(p)v, \quad \omega(\xi p, v) = \mathrm{D}\mathbf{J}_\xi(p)v, \quad v \in T_p\mathcal{M}, \quad \xi \in \mathbf{g}. \tag{6.6}$$

By G-invariance of H and part (a) we have $T \subseteq \ker \mathrm{D}H(p)$, which proves the third equation of (c). Because of (a) we have $T_2 \subseteq \ker \mathrm{D}\mathbf{J}(p)$, and by G-equivariance of $\mathbf{J}$ we get $\mathrm{D}\mathbf{J}(p)\xi p = \xi\mathbf{J}(p)$, which vanishes if and only if $\xi \in \mathbf{g}_\mu$. This proves that $T_0 \subseteq \ker \mathrm{D}\mathbf{J}(p)$ and $T_1 \cap \ker \mathrm{D}\mathbf{J}(p) = \{0\}$. ■

The following proposition generalizes the usual Witt decomposition at group orbits to relative periodic orbits.

Proposition 6.1. *Let $p \in \mathcal{M}$ lie on a relative periodic orbit $\mathcal{P}$ with relative period different from zero. Then there is a G_p-invariant Poincaré section N to $\mathcal{P}$ at p such that the following are true.*

(a) *Equation (3.6) holds, and the spaces T_i, N_i, $i = 0, 1, 2$, are all G_p-invariant.*

(b) *The symplectic form ω on $T_p\mathcal{M}$ restricts to symplectic forms $\omega_{T_0 \oplus N_0}$ on $T_0 \oplus N_0$, ω_{T_1} on T_1, ω_{N_1} on N_1, and $\omega_{T_2 \oplus N_2}$ on $N_2 \oplus T_2$. The actions of G_p on these spaces are χ-semisymplectic with respect to the restricted forms. Moreover,*

$$\omega|_{T_p\mathcal{M}} = \omega_{T_0 \oplus N_0} + \omega_{T_2} + \omega_{N_1} + \omega_{T_2 \oplus N_2}.$$

(c) $\ker \mathrm{D}\mathbf{J}(p) = T_0 \oplus T_2 \oplus N_1 \oplus N_2; \ker \mathrm{D}H(p) = T \oplus N_0 \oplus N_1$.

(d) *Identify $\mathbf{g}_\mu/\mathbf{g}_p \cong T_0$ via the map $\mathbf{g} \to T_p\mathcal{M}$ given by $\xi \mapsto \xi p$. The symplectic form ω, or, equivalently, the map $v \mapsto \mathrm{D}\mathbf{J}(p)(\cdot)v$ (see (6.6)), defines a G_p-equivariant isomorphism between the induced G_p-action on N_0 and the χ-coadjoint action on $T_0^* \cong (\mathbf{g}_\mu/\mathbf{g}_p)^*$. Similarly, $\xi \mapsto \mathrm{D}\mathbf{J}_\xi(p)$ defines a G_p-equivariant isomorphism between T_0 and N_0^* such that $N_0^* = \mathrm{D}\mathbf{J}(p)(\mathbf{g}_\mu/\mathbf{g}_p)$ is the annihilator of $T \oplus N_1 \oplus N_2$. Under the first isomorphism, the symplectic form $\omega_{T_0 \oplus N_0}$ becomes the natural symplectic form on $(\mathbf{g}_\mu/\mathbf{g}_p) \oplus (\mathbf{g}_\mu/\mathbf{g}_p)^*$:*

$$\omega_{T_0 \oplus N_0}((\xi_1, \nu_1), (\xi_2, \nu_2)) \quad = \quad \nu_2(\xi_1) \ - \ \nu_1(\xi_2).$$

(e) $\mathrm{D}\mathbf{J}(p)$ *maps* T_1 *isomorphically to* $T_\mu(G\mu) \cong \mathbf{g}/\mathbf{g}_\mu$ *and* ω_{T_1} *to the Kostant–Kirillov–Souriau (KKS) form* ω_μ *(in body coordinates):*

$$\omega_{T_1}(\xi_1.p, \xi_2.p) \;=\; \omega_\mu(\xi_1, \xi_2) \;:=\; \mu\left([\xi_1, \xi_2]\right), \tag{6.7}$$

where $\xi_i \in \mathbf{g}$, $i = 1, 2$, *and* $[\cdot, \cdot]$ *is the Lie bracket on* $\mathbf{g}$.

(f) *The symplectic form* ω *defines* G_p*-equivariant isomorphisms between* T_2^* *and* N_2 *and between* T_2 *and* N_2^* *such that* $N_2^* = \mathrm{ann}(T \oplus N_0 \oplus N_1)$ *is spanned by* $\mathrm{D}H(p)$*. Under these isomorphisms, the symplectic form* $\omega_{T_2 \oplus N_2}$ *becomes the standard symplectic structure* $\omega_{T_2 \oplus N_2}((E_1, \theta_1), (E_2, \theta_2)) = E_2\theta_1 - E_1\theta_2$.

Proof. The Witt decomposition near group orbits (see, for example, [4, 36]) gives $T_p\mathcal{M} = \hat{T} \oplus \hat{N}$, where $\hat{T} = T_pGp = T_0 \oplus T_1$ and $\hat{N} = \hat{N}_0 \oplus \hat{N}_1$. Here the symplectic normal space $\hat{N}_1$ to Gp at p is a G_p-invariant complement to T_0 in $\ker \mathrm{D}\mathbf{J}(p)$, and the space $\hat{N}_0$ is a G_p-invariant complement to $T_1 + \ker \mathrm{D}\mathbf{J}(p)$, which is chosen so that $\hat{N}_1 \oplus T_1 \subset \hat{N}_0^\omega$. Now we show how to adapt this Witt decomposition to relative periodic orbits.

(a) We choose $\hat{N}_1$ to contain T_2 and v_E, which is possible by Lemma 6.1 (a), (b). Since $T_2 \subset \hat{N}_1 \subset \hat{N}_0^\omega$, we conclude from (6.6) that $\hat{N}_0 \subset \ker \mathrm{D}H(p)$ and therefore define $N_0 := \hat{N}_0$. The symplectic form $\omega_{\hat{N}_1}$ restricts to a symplectic form on $T_2 \oplus N_2$ because $\omega(f_H(p), v_E) = \mathrm{D}H(p)v_E \neq 0$. Hence $N_1 := (T_2 \oplus N_2)^\omega \cap \hat{N}_1$ is also a symplectic space which is transverse to $T_2 \oplus N_2$ and, because of (6.6), satisfies $N_1 \subset \ker \mathrm{D}H(p)$. With this construction, $\hat{N} = N_1 \oplus T_2 \oplus N_2$, and (3.6) follows.

By definition T_0 and T_1 are G_p-invariant. By Lemma 6.1 (a), (b) the spaces T_2 and N_2 are G_p-invariant, and the above construction implies that N_0 and N_1 are also G_p-invariant.

(b) This follows from the usual Witt decomposition and the proof of (a).

(c) Because of Lemma 6.1 (c) and since $N_0 \oplus N_1 = \ker \mathrm{D}H(p) \cap N$, the relation $\ker \mathrm{D}H(p) = T \oplus N_0 \oplus N_1$ holds. By definition $N \cap \ker \mathrm{D}\mathbf{J}(p) = N_1 \oplus N_2$, which proves that $\ker \mathrm{D}\mathbf{J}(p) = T_0 \oplus T_2 \oplus N_1 \oplus N_2$.

(d) From (c) we conclude that $T_0 \oplus T_2 \oplus N_1 \oplus N_2$ is annihilated by $N_0^* = \mathrm{D}\mathbf{J}(p)(\mathbf{g}_\mu/\mathbf{g}_p)$. Now let $\eta \in \mathbf{n}_\mu$, $\xi \in \mathbf{g}_\mu$. Then

$$\mathrm{D}\mathbf{J}_\xi(p)\eta p = (\eta\mathbf{J})(\xi)(p) = \mathbf{J}([\eta, \xi])(p) = -(\xi\mathbf{J})(\eta)(p) = 0,$$

which proves that $\mathrm{D}\mathbf{J}(p)(\mathbf{g}_\mu/\mathbf{g}_p)$ annihilates T_1. The other statements follow from the usual Witt decomposition near group orbits.

(e) This follows from the usual Witt decomposition.

(f) That N_2^* is spanned by $\mathrm{D}H(p)$ follows from (6.6), and that it annihilates $T \oplus N_0 \oplus N_1$ follows from (c). ■

6.3. Linearization along the relative periodic orbit.

The following three lemmas together with Proposition 4.1 prove Proposition 4.3 on the linearization near relative periodic orbits.

Lemma 6.2. *Let* $p \in \mathcal{M}$, *and let* $\mathbf{m}_\mu$, $\mathbf{n}_\mu$ *be* G_p*-invariant complements to* $\mathbf{g}_p$ *in* $\mathbf{g}_\mu$ *and to* $\mathbf{g}_\mu$ *in* $\mathbf{g}$, *respectively.*

(a) *Let* $\gamma \in N_{\Gamma_\mu}(\Gamma_p)$*. Then with respect to the decomposition* $\mathbf{g} = \mathbf{m}_\mu \oplus \mathbf{n}_\mu \oplus \mathbf{g}_p$ *the matrix*

Ad_γ *has the following block structure:*

$$\mathrm{Ad}_\gamma = \begin{pmatrix} \overline{\mathrm{Ad}}_\gamma & \pi_{\mathbf{m}_\mu}\mathrm{Ad}_\gamma|_{\mathbf{n}_\mu} & 0 \\ 0 & \pi_{\mathbf{n}_\mu}\mathrm{Ad}_\gamma|_{\mathbf{n}_\mu} & 0 \\ \pi_{\mathbf{g}_p}\mathrm{Ad}_\gamma|_{\mathbf{m}_\mu} & \pi_{\mathbf{g}_p}\mathrm{Ad}_\gamma|_{\mathbf{n}_\mu} & \mathrm{Ad}_\gamma|_{\mathbf{g}_p} \end{pmatrix}.$$

Here $\pi_{\mathbf{m}_\mu}$, $\pi_{\mathbf{n}_\mu}$, *and* $\pi_{\mathbf{g}_p}$ *are the projections from* $\mathbf{g}$ *to* $\mathbf{m}_\mu$, $\mathbf{n}_\mu$, *and* $\mathbf{g}_p$ *with kernels* $\mathbf{n}_\mu \oplus \mathbf{g}_p$, $\mathbf{m}_\mu \oplus \mathbf{g}_p$, *and* $\mathbf{m}_\mu \oplus \mathbf{n}_\mu$.

(b) *If* $\sigma \in N_{\Gamma_\mu}(\Gamma_p)$ *has the form* $\sigma = \alpha \exp(\xi)$, *where* $\xi \in \mathbf{m}_\mu \cap \mathbf{z}(\sigma) \cap \mathbf{z}(\Gamma_p)$, *and* Ad_α *leaves* $\mathbf{m}_\mu$ *invariant, then* $\overline{\mathrm{Ad}}_\sigma = \mathrm{Ad}_\alpha \exp(\overline{\mathrm{ad}}_\xi)$.

Proof. (a) is clear. To prove (b) note that since $\sigma, \alpha \in N_{\Gamma_\mu}(\Gamma_p)$ and $\exp(\xi\theta) \in N_{\Gamma_\mu}(\Gamma_p)$ for all $\theta \in \mathbb{R}$ and $\xi \in \mathrm{L}Z_{\Gamma_\mu}(\Gamma_p)$, the representations of their adjoint actions on $\mathbf{g}$ have the block structure given in part (a). The statement then follows from the fact that $\mathbf{m}_\mu$ is Ad_α-invariant. ■

Lemma 6.3. *Let* p *lie on a relative periodic orbit with minimal period* 1, *and let* $M = \sigma^{-1}\mathrm{D}\Phi_1(p)$. *Then the following hold.*

(a) $N_2^* = \mathrm{span}\{\mathrm{D}H(p)\}$ *is a left eigenspace of* M *with eigenvalue* 1.

(b) *We have* $\mathrm{D}\mathbf{J}_\xi(p)M = \mathrm{D}\mathbf{J}_{\mathrm{Ad}_\sigma \xi}(p)$ *for each* $\xi \in \mathbf{g}$ *and therefore* $M^*|_{N_0^*} = \overline{\mathrm{Ad}}_\sigma$, *where* $N_0^* = \mathrm{D}\mathbf{J}(p)(\mathbf{m}_\mu)$.

(c) *The spaces* $T_0 \oplus T_2 \oplus N_1$ *and* $T_0 \oplus T_2 \oplus N_1 \oplus N_2$ *are* M*-invariant.*

Proof. That $\mathrm{D}H(p)$ is a left eigenvector of M with eigenvalue 1 follows from the G-invariance and conservation of H. The first statement of part (b) is a direct computation which we omit. For part (c) note that T and $T_p(Gp)$ are M-invariant by Proposition 4.1. Therefore, and by the symplecticity of M,

$$T^\omega \;=\; (T_p(Gp) \oplus T_2)^\omega \;=\; \ker(\mathrm{D}\mathbf{J}(p)) \cap T_2^\omega \;=\; T_0 \oplus T_2 \oplus N_1$$

and

$$(T_p(Gp))^\omega \;=\; \ker(\mathrm{D}\mathbf{J}(p)) \;=\; T_0 \oplus T_2 \oplus N_1 \oplus N_2$$

are also M-invariant. Here again we used Lemma 6.1 (c). ■

Lemma 6.4. *Let* $M : T_p\mathcal{M} \to T_p\mathcal{M}$ *be a linear map with block structure*

$$M = \begin{pmatrix} A_0 & A_{01} & 0 & D_0 & D_1 & D_2 \\ 0 & A_1 & 0 & D_3 & 0 & 0 \\ 0 & 0 & 1 & \Theta_0 & \Theta_1 & \Theta_2 \\ 0 & 0 & 0 & M_0 & 0 & 0 \\ 0 & 0 & 0 & M_{10} & M_1 & M_{12} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.8}$$

with respect to the tangent space decomposition $T_p\mathcal{M} = T_0 \oplus T_1 \oplus T_2 \oplus N_0 \oplus N_1 \oplus N_2$. *Let* J_{N_1} *and* J_{T_1} *denote the* N_1 *and* T_1 *blocks of the skew-symmetric matrix* $J \in \mathrm{GL}(T_p\mathcal{M})$ *generating* *the symplectic form ω_p; see Proposition 6.1. Then M is symplectic if and only if*

$$A_1^T J_{T_1} A_1 = J_{T_1}, \tag{6.9}$$

$$A_{01}^T M_0 + A_1^T J_{T_1} D_3 = 0, \tag{6.10}$$

$$D_0^T M_0 - M_0^T D_0 + D_3^T J_{T_1} D_3 + M_{10}^T J_{N_1} M_{10} = 0, \tag{6.11}$$

$$D_1^T M_0 + M_1^T J_{N_1} M_{10} = 0, \tag{6.12}$$

$$M_1^T J_{N_1} M_1 = J_{N_1}, \tag{6.13}$$

$$M_0 = A_0^{-T}, \tag{6.14}$$

$$\Theta_1 - M_{12}^T J_{N_1} M_1 = 0, \tag{6.15}$$

$$\Theta_0 - D_2^T M_0 - M_{12}^T J_{N_1} M_{10} = 0. \tag{6.16}$$

The proof of this lemma is by direct computation.

6.4. Symplectic twisted semiequivariant linear maps.

We will need the following lemma for the construction of symplectic homotopies near Hamiltonian relative periodic orbits in section 6.5. Note that it is shown in [16] that every semi-invariant symplectic form on a vector space has a semiequivariant complex structure J satisfying $J^2 = -\mathrm{id}$.

Lemma 6.5. *Let G be a compact Lie group acting orthogonally and semisymplectically on a finite dimensional symplectic vector space V with complex structure J satisfying $J^2 = -\mathrm{id}$. Let $M : V \to V$ be a twisted semiequivariant linear map with twist diffeomorphim $\phi : G \to G$ of order k. Then the following hold.*

(a) *There is a twisted semiequivariant orthogonal symplectic linear map $A : V \to V$ such that $A^{2k} = \mathrm{id}$ and $A^{-1}M = \exp(-\eta)$, where η is infinitesimally G-semiequivariant ($\chi(g)g\eta = \eta g$ for all $g \in G$) and infinitesimally symplectic ($\eta^T J + J\eta = 0$) and commutes with A and M.*

(b) *We have $A^{-1} = \exp(J_-)Q$, where J_- is infinitesimally G-semiequivariant and symplectic, commutes with A, and is such that $Q^k = \mathrm{id}$. Moreover, there is a Γ-equivariant homotopy $I(\theta)$ which is smooth in θ and satisfies*

$$MI(\theta + 1) = I(\theta)Q^{-1}, \quad \rho I(\theta)\rho^{-1} = I(-\theta) \quad \textit{for all} \;\; \theta \in \mathbb{R},\, \rho \in G \setminus \Gamma.$$

Proof. Part (a) is essentially Lemma 5.2 of [22]. It is easily checked that the matrices $\exp(\theta\eta)$ defined there are symplectic if V is symplectic.

To prove (b) note that since A is symplectic and $A^{2k} = \mathrm{id}$, we have $V = V_+ \oplus V_-$, where $V_\pm$ are symplectic G-invariant subspaces of V such that $A^k|_{V_+} = \mathrm{id}$ and $A^k|_{V_-} = -\mathrm{id}$. Let $J_- : V \to V$ be the matrix defined by $J_-|_{V_+} = 0$, $J_-|_{V_-} = \frac{\pi}{k} J|_{V_-}$. Then J_- is infinitesimally G-semiequivariant and symplectic and $\exp(kJ_-) = A^k$. Moreover, $AV_+ = V_+$, $AV_- = V_-$, and, since A is symplectic and orthogonal, $AJ = JA$ so that $[A, J_-] = 0$. Defining $Q = A^{-1}\exp(-J_-)$, we get $Q^k = \mathrm{id}$, which proves the first statement of part (b).

For $\theta \in [0, 1)$, define $I(\theta) := \exp(c(\theta)\eta)\exp(c(\theta)J_-)$, where $c : [0, 1) \to \mathbb{R}_0^+$ is a C^∞ monotonically increasing function with

$$c(\theta) \equiv 0 \quad \text{for} \quad 0 \le \theta < \epsilon, \quad \epsilon < 1/2 \quad \text{fixed}, \quad c(1 - \theta) = 1 - c(\theta). \tag{6.17}$$

Then $I(1) = \exp(\eta)\exp(J_-) = M^{-1}AA^{-1}Q^{-1} = M^{-1}Q^{-1}$ so that we can smoothly extend the homotopy $I(\theta)$ to $\theta \in [n, n + 1)$, $n \in \mathbb{Z} \setminus \{0\}$, by setting $I(\theta + n) = M^{-n}I(\theta)Q^{-n}$.

It remains to prove that $\rho I(-\theta)\rho^{-1} = I(\theta)$ for $\rho \in G \setminus \Gamma$. Let $\theta \in [0, 1)$. Then by definition

$$I(-\theta) = MI(1-\theta)Q = M\exp((1-c(\theta))\eta)\exp((1-c(\theta))J_-)Q$$

so that

$$\begin{aligned}\rho I(-\theta)\rho^{-1} &= M^{-1}\exp((c(\theta)-1)\eta)\exp((c(\theta)-1)J_-)Q^{-1}\\ &= M^{-1}\exp(-\eta)I(\theta)\exp(-J_-)Q^{-1} = A^{-1}I(\theta)A = I(\theta),\end{aligned}$$

where we used that $[A, \eta] = [J_-, A] = 0$. Now let $\theta = n + \hat{\theta} \in [n, n+1)$, $n \in \mathbb{Z} \setminus \{0\}$. Then by definition $I(\theta) = M^{-n}I(\hat{\theta})Q^{-n}$ and $I(-\theta) = M^{n}I(-\hat{\theta})Q^{n}$ so that

$$\rho I(-\theta)\rho^{-1} = M^{-n}\rho I(-\hat{\theta})\rho^{-1}Q^{-n} = M^{-n}I(\hat{\theta})Q^{-n} = I(\theta). \qquad ■$$

Remark 6.2. Let G be trivial. Since $\mathrm{Sp}(V)$ is connected, we can always symplectically homotope any symplectic linear map M to the identity, and so $Q = \mathrm{id}$ for all $M \in \mathrm{Sp}(V)$. However, in general, the homotopies cannot be chosen to be exponentials. For example,

$$M = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \in \mathrm{Sp}(2)$$

is not of the form $M = \exp(\eta)$ over the reals. However, if $A = -\mathrm{id}$, there exists an exponential homotopy of $A^{-1}M$ to identity.

6.5. Symplectic homotopies. Let $p = \sigma^{-1}\Phi_1(p)$ lie on a relative periodic orbit $\mathcal{P}$ with momentum $\mu = \mathbf{J}(p)$. This subsection deals with the proof of the following lemma, which is needed for the adaptation of the bundle structure near relative periodic orbits to the Hamiltonian context. It will be used in the proof of Theorem 6.3.

Lemma 6.6. *Assume that $\sigma = \alpha\exp(\xi)$, where $\xi \in \mathbf{z}(\sigma) \cap \mathbf{z}^{\chi}(G_p) \cap \mathbf{g}_\mu$, and $\alpha \in \Gamma_\mu$ has order n, and choose the G_p-semi-invariant complex structure J_{N_1} on N_1 such that $J_{N_1}^2 = -\mathrm{id}$. Then the homotopy $I(\theta) \in \mathrm{GL}(T_p\mathcal{M})$ in (6.2), (6.4), and (6.5), which is G_p-semiequivariant in the sense of (6.3), can be chosen to be symplectic and such that the matrix Q in (6.4) has the block structure*

$$Q = \begin{pmatrix} \mathrm{Ad}_\alpha|_{\mathbf{m}_\mu \oplus \mathbf{n}_\mu} & & & & \\ & 1 & & & \\ & & Q_0 & & \\ & & & Q_1 & \\ & & & & 1 \end{pmatrix} \in \mathrm{O}(T_p\mathcal{M}) \cap \mathrm{Sp}(T_p\mathcal{M}), \tag{6.18}$$

where $Q_0 = (\mathrm{Ad}^*_\alpha|_{N_0})^{-1}$,

$$Q_1 \in \mathrm{Sp}(N_1) = \{A \in \mathrm{GL}(N_1) \mid J_{N_1} = A^T J_{N_1} A\},$$

and $Q_1^{-1} \in \mathrm{O}(N_1)$ is twisted semiequivariant of order k. Consequently, Q^{-1} is twisted semiequivariant of order n.

Since $M = \sigma^{-1}\mathrm{D}\Phi_1(p)$ is twisted semiequivariant, by Lemma 6.5 there is a symplectic homotopy $I(\theta)$ such that (6.4) and (6.3) hold provided the complex structure J on $\mathcal{M}$ is chosen such that $J^2 = -\mathrm{id}$. However, it is not clear that for the choice of homotopy of Lemma 6.5 the matrix Q has the form $Q = \mathrm{diag}(Q_T, Q_N)$ with $Q_N = \mathrm{diag}(Q_0, Q_1, 1)$, $Q_0 = (\mathrm{Ad}^*_\alpha|_{N_0})^{-1}$. The above lemma states that there always exists a G_p-semiequivariant symplectic homotopy $I(\theta)$ such that this can be achieved.

Since by Proposition 4.3 the subblock M_1 of $M = \sigma^{-1}\mathrm{D}\Phi_1(p)$ is twisted semiequivariant, by Lemma 6.5 there is a G_p-semiequivariant homotopy such that

$$M_1 I_1(\theta + 1) Q_1 = I_1(\theta). \tag{6.19}$$

Here $Q_1 \in \mathrm{Sp}(N_1) \cap \mathrm{O}(N_1)$ has order k, where k is the order of the twist diffeomorphism, and Q_1^{-1} is twisted semiequivariant. Using Lemma 6.5, we conclude that, if $Q := \mathrm{diag}(\mathrm{Ad}_\alpha|_{\mathbf{m}_\mu \oplus \mathbf{n}_\mu}, 1, (\mathrm{Ad}^*_\alpha|_{N_0})^{-1}, Q_1, 1)$, then Q^{-1} is a symplectic twisted semiequivariant map of order n.

For the construction of the homotopies in Lemma 6.6, we will first restrict ourselves to the nonreversible case, i.e., $\Gamma_p = G_p$, and we will then extend the result to reversible relative periodic orbits.

6.5.1. Equivariant symplectic homotopies. In this subsection, we construct Γ_p-equivariant symplectic homotopies $I(\theta)$ which satisfy the conditions of Lemma 6.6 . In order to do this, we will rely heavily on Lemma 6.4. Proposition 4.3 shows that $M = \sigma^{-1}\mathrm{D}\Phi_1(p)$ has the required structure for Lemma 6.4 to apply. Moreover, since time is reparametrized such that $\dot\theta \equiv 1$, we have $\Theta_i = 0$, $i = 0, 1, 2$ in (6.8), which by Lemma 6.4 implies that $D_2 = M_{12} = 0$.

We look for a homotopy $I(\theta)$ satisfying

$$M\, I(\theta + 1) = I(\theta) Q^{-1}, \quad \text{with } Q \text{ as in (6.18)}.$$

Let $I(1) = M^{-1}Q^{-1}$. Then $I(1)$ is Γ_p-equivariant, symplectic, and given by

$$I^{-1}(1) = \begin{pmatrix} \overline{\mathrm{Ad}}_{\exp(-\xi)} & \pi_{\mathbf{m}_\mu}\mathrm{Ad}_{\exp(-\xi)}|_{\mathbf{n}_\mu} & 0 & \mathrm{Ad}_\alpha D_0 & \mathrm{Ad}_\alpha D_1 & 0 \\ 0 & \pi_{\mathbf{n}_\mu}\mathrm{Ad}_{\exp(-\xi)}|_{\mathbf{n}_\mu} & 0 & \mathrm{Ad}_\alpha D_3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{\mathrm{Ad}}^*_{\exp(\xi)} & 0 & 0 \\ 0 & 0 & 0 & Q_1 M_{10} & Q_1 M_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here we used Lemma 6.2. This matrix is Γ_p-equivariantly and symplectically homotopic to the identity. To see this we first define

$$\hat I(\theta) = \begin{pmatrix} \overline{\mathrm{Ad}}_{\exp(\theta\xi)} & \pi_{\mathbf{m}_\mu}\mathrm{Ad}_{\exp(\theta\xi)}|_{\mathbf{n}_\mu} & 0 & D_0(\theta) & 0 & 0 \\ 0 & \pi_{\mathbf{n}_\mu}\mathrm{Ad}_{\exp(\theta\xi)}|_{\mathbf{n}_\mu} & 0 & D_3(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{\mathrm{Ad}}^*_{\exp(-\theta\xi)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat I_1(\theta) & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where $\hat I_1(\theta) = \exp(\theta\eta)\exp(\theta J_-)$ and η, J_- are as in Lemma 6.5. The blocks $D_0(\theta)$ and $D_3(\theta)$ are determined by the two symplecticity conditions (6.10) and (6.11) of Lemma 6.4,

$$(\pi_{\mathbf{m}_\mu}\mathrm{Ad}_{\exp(\theta\xi)}|_{\mathbf{n}_\mu})^T\overline{\mathrm{Ad}}^*_{\exp(-\theta\xi)} + (\pi_{\mathbf{n}_\mu}\mathrm{Ad}_{\exp(\theta\xi)}|_{\mathbf{n}_\mu})^T J_{T_1} D_3(\theta) = 0 \tag{6.20}$$

and

$$D_0(\theta)^T \overline{\mathrm{Ad}}^*_{\exp(-\theta\xi)} - \overline{\mathrm{Ad}}_{\exp(-\theta\xi)} D_0(\theta) + D_3(\theta)^T J_{T_1} D_3(\theta) = 0, \tag{6.21}$$

and by defining the symmetric part of $D_0(\theta)^T \overline{\mathrm{Ad}}^*_{\exp(-\theta\xi)}$ to be zero:

$$D_0(\theta)^T \overline{\mathrm{Ad}}^*_{\exp(-\theta\xi)} + \overline{\mathrm{Ad}}_{\exp(-\theta\xi)} D_0(\theta) = 0. \tag{6.22}$$

Equations (6.21) and (6.22) are equivalent to

$$2D_0(\theta)^T \overline{\mathrm{Ad}}^*_{\exp(-\theta\xi)} + D_3(\theta)^T J_{T_1} D_3(\theta) = 0. \tag{6.23}$$

Equations (6.20) and (6.23) determine $D_0(\theta)$ and $D_3(\theta)$ uniquely.

By Lemma 6.4 the homotopy $\hat{I}(\theta)$ is symplectic since (6.10)–(6.16) are satisfied, and a calculation using $\xi \in \mathbf{g}_\mu$ and $\langle \xi_1, J_{T_1}\xi_2 \rangle = \mu([\xi_1, \xi_2])$ for $\xi_1, \xi_2 \in \mathbf{n}_\mu$ shows that (6.9) is satisfied.

The Γ_p-equivariance of $\mathrm{Ad}_{\exp(\theta\xi)}$ implies that if $D_0(\theta)$ and $D_3(\theta)$ are solutions of (6.20) and (6.23), then so also are $\gamma_p D_0(\theta)\gamma_p^{-1}$ and $\gamma_p D_3(\theta)\gamma_p^{-1}$ for $\gamma_p \in \Gamma_p$. Since the solutions are unique, this means that $D_0(\theta)$ and $D_3(\theta)$ are Γ_p-equivariant, and hence so is the homotopy $\hat{I}(\theta)$. Moreover, $B = \hat{I}^{-1}(1)I(1)$ is unipotent and so symplectically and Γ_p-equivariantly homotopic to the identity by the homotopy $\exp(\theta \log(B))$.

Now we define $I(\theta) = \hat{I}(c(\theta))\exp(c(\theta)\log(B))$ for $0 \le \theta < 1$, where $c : [0,1) \to \mathbb{R}_0^+$ is the same C^∞ monotonically increasing function satisfying (6.17) as in the proof of Lemma 6.5. Since by construction $I(1) = M^{-1}Q^{-1}$, we get a smooth homotopy by defining $I(\theta)$ for $\theta = n + \hat{\theta} \in [n, n+1)$, $n \in \mathbb{Z} \setminus \{0\}$, as $I(\theta) = M^{-n}I(\hat{\theta})Q^{-n}$. Thus we obtain a Γ_p-equivariant smooth symplectic homotopy $I(\theta)$ such that (6.4) is satisfied for all θ.

Note that by construction the A_0, A_1, A_{01}, and M_0 blocks of $I(\theta)$ and $\hat{I}(c(\theta))$ coincide if we define

$$c(\theta + n) = c(\theta) + n \quad \text{for} \quad \theta \in [n, n+1), \quad n \in \mathbb{Z}. \tag{6.24}$$

Moreover, the M_1-block of $I(\theta)$ is given by the homotopy $I_1(\theta)$ of (6.19), obtained from Lemma 6.5, since we chose the same reparametrization $c(\theta)$ in the construction of both homotopies.

The D_3-blocks of $I(\theta)$ and $\hat{I}(c(\theta))$ coincide because they are uniquely defined by the corresponding A_1 and M_0-blocks; see (6.10). The other blocks of $\hat{I}(\theta)$ and $I(\theta)$ are in general not related.

6.5.2. Reversible equivariant symplectic homotopies.

In this subsection, we will extend the construction of symplectic homotopies of subsection 6.5.1 to the reversible case. So let $G_p \ne \Gamma_p$, let $I(\theta)$ be the Γ_p-equivariant symplectic homotopy satisfying (6.4) defined in subsection 6.5.1 above, and let μ_{G_p}, μ_{Γ_p} be the Haar measures of G_p and Γ_p. Since $G_p/\Gamma_p = \mathbb{Z}_2$ for any function f from G_p to a vector space, we have

$$\int_{G_p} f(g_p)\mathrm{d}\mu_{G_p} = \frac{1}{2}\int_{\Gamma_p} f(\gamma_p)\mathrm{d}\mu_{\Gamma_p} + \frac{1}{2}\int_{\Gamma_p} f(\rho\gamma_p)\mathrm{d}\mu_{\Gamma_p} \quad \text{for all } \rho \in G_p \setminus \Gamma_p.$$

As a consequence,

$$I^{\mathrm{av}}(\theta) := \int_{G_p} g_p I(\chi(g_p)\theta) g_p^{-1} \mathrm{d}\mu_{G_p} = \frac{1}{2}\left(I(\theta) + \rho I(-\theta)\rho^{-1}\right) \quad \text{for} \quad \rho \in G_p \setminus \Gamma_p.$$

This clearly defines a homotopy to the identity map, which is G_p-semiequivariant in the sense of (6.3).

We will now show that $I^{\mathrm{av}}(\theta)$ satisfies (6.4). Let $\rho \in G_p \setminus \Gamma_p$. Then

$$M(\rho I(-(\theta+1))\rho^{-1}) = \phi(\rho) M^{-1} I(-(\theta+1))\rho^{-1},$$

where $\phi : G_p \to G_p$ is the twist diffeomorphism, and we used the fact that M is twisted semiequivariant: $M g_p = \phi(g_p) M^{\chi(g_p)}$ for $g_p \in G_p$. Since $I(\theta)$ satisfies (6.4), we get

$$\phi(\rho) M^{-1} I(-(\theta+1))\rho^{-1} = \phi(\rho) I(-\theta) Q \rho^{-1},$$

and because Q^{-1} is twisted semiequivariant we altogether have

$$M(\rho I(-(\theta+1))\rho^{-1}) = \phi(\rho) I(-\theta)(\phi(\rho))^{-1} Q^{-1} \text{ for } \rho \in G_p \setminus \Gamma_p.$$

Since

$$\int_{G_p} \phi(g_p) I(\chi(g_p)\theta)(\phi(g_p))^{-1} \mathrm{d}\mu_{G_p} = \frac{1}{2}\left(I(\theta) + \rho I(-\theta)\rho^{-1}\right) = I^{\mathrm{av}}(\theta),$$

the homotopy $I^{\mathrm{av}}(\theta)$ satisfies (6.4).

Note that the A_0, A_1, A_{01}, and M_0 blocks of $I^{\mathrm{av}}(\theta)$ equal the corresponding blocks of $I(\theta)$ because these subblocks are given by $\mathrm{Ad}_{\exp(c(\theta)\xi)}$ (the A-blocks) and $\overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)}$ (the M_0-block) and are therefore G_p-semiequivariant since by (6.24) and (6.17) the function $c(\theta)$ satisfies $c(-\theta) = -c(\theta)$. Moreover, by construction we have $I_1^{\mathrm{av}}(\theta) = I_1(\theta)$.

The homotopy $I^{\mathrm{av}}(\theta)$ has the same block structure as M since all $g_p \in \mathrm{Sp}(T_p\mathcal{M})$ have the same block structure as M. As a consequence, $I^{\mathrm{av}}(\theta)$ is invertible.

The problem is that $I^{\mathrm{av}}(\theta)$ need not be symplectic in general. We modify it to obtain a G_p-semiequivariant (in the sense of (6.3)) symplectic homotopy $I^{\mathrm{rev}}(\theta)$ with the same block structure as M. We prescribe the subblocks

$$I_0^{\mathrm{rev}}(\theta) = \overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)}, \quad I_1^{\mathrm{rev}}(\theta) = I_1(\theta), \quad I^{\mathrm{rev}}(\theta)|_{T_0 \oplus T_1} = \mathrm{Ad}_{\exp(c(\theta)\xi)}.$$

Here $I_i^{\mathrm{rev}}(\theta)$ are the M_i-subblocks of $I^{\mathrm{rev}}(\theta)$, $i = 0, 1$. We define the M_{10}-block $I_{10}^{\mathrm{rev}}(\theta)$ of $I^{\mathrm{rev}}(\theta)$ to be

$$I_{10}^{\mathrm{rev}}(\theta) = I_{10}^{\mathrm{av}}(\theta),$$

and we define the symmetric part of $\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)} I_{D_0}^{\mathrm{rev}}(\theta)$ to be

$$\begin{aligned} &(I_{D_0}^{\mathrm{rev}}(\theta))^T \overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)} + \overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)} I_{D_0}^{\mathrm{rev}}(\theta) \\ &= (I_{D_0}^{\mathrm{av}}(\theta))^T \overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)} + \overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)} I_{D_0}^{\mathrm{av}}(\theta). \end{aligned} \tag{6.25}$$

The other blocks of $I^{\mathrm{rev}}(\theta)$ are defined uniquely using the symplecticity conditions of Lemma 6.4. This gives a homotopy $I^{\mathrm{rev}}(\theta)$ which is symplectic and smooth in θ. Moreover, $I^{\mathrm{rev}}(\theta)$ is G_p-semiequivariant in the sense of (6.3) with respect to the corresponding G_p-actions. This can be seen as follows. The subblocks $I^{\mathrm{rev}}(\theta)|_{T_0\oplus T_1}$ and the M_0, M_1, and M_{10} subblocks of $I^{\mathrm{rev}}(\theta)$ are G_p-semiequivariant because they equal the corresponding subblocks of the G_p-semiequivariant homotopy $I^{\mathrm{av}}(\theta)$. The D_3-subblock of $I^{\mathrm{rev}}(\theta)$ is given by (6.10). Since we know that all terms of this equation except the D_3-term are G_p-semiequivariant and the D_3 subblock of $I^{\mathrm{rev}}(\theta)$ is uniquely determined by this equation, the D_3 subblock of $I^{\mathrm{rev}}(\theta)$ is also G_p-semiequivariant. Similarly, we see that the D_1 subblock of $I^{\mathrm{rev}}(\theta)$, which is determined by (6.12), is G_p-semiequivariant. Finally, by (6.11) the antisymmetric part of the matrix $\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)}I^{\mathrm{rev}}_{D_0}(\theta)$ is G_p-semiequivariant, and by (6.25) the same holds for the symmetric part of $\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)}I^{\mathrm{rev}}_{D_0}(\theta)$. Hence $I^{\mathrm{rev}}_{D_0}(\theta)$ is also G_p-semiequivariant.

Finally, we will show that $I^{\mathrm{rev}}(\theta)$ satisfies (6.4). Due to the block structure of M, Q, and $I^{\mathrm{rev}}(\theta)$, and since the M_1, M_0, and M_{10}-subblocks of $I^{\mathrm{rev}}(\theta)$ are given by the corresponding subblocks of $I^{\mathrm{av}}(\theta)$ and $I^{\mathrm{rev}}(\theta)|_{T_0\oplus T_1} = I^{\mathrm{av}}(\theta)|_{T_0\oplus T_1}$, we see that for these subblocks (6.4) is satisfied. Moreover, since both sides of (6.4) are symplectic and therefore all subblocks of both sides of (6.4) except for the symmetric part of the $M_0^T D_0$ matrices are determined by the corresponding A and M_1, M_0 and M_{10}-subblocks by Lemma 6.4, we need only to check that the symmetric parts of the $M_0^T D_0$ matrices of both sides of (6.4) coincide.

The M_0 part of the right-hand side of (6.4) is given by $I_0^{\mathrm{rev}}(\theta)Q_0^{-1} = \overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)}Q_0^{-1}$, and the D_0 part of the right-hand side of (6.4) is $I^{\mathrm{rev}}_{D_0}(\theta)Q_0^{-1}$. So twice the symmetric part of the $M_0^T D_0$ matrices of the right-hand side of (6.4) is

$$\begin{aligned}
&(\overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)}Q_0^{-1})^T(I^{\mathrm{rev}}_{D_0}(\theta)Q_0^{-1}) + (I^{\mathrm{rev}}_{D_0}(\theta)Q_0^{-1})^T(\overline{\mathrm{Ad}}^*_{\exp(-c(\theta)\xi)}Q_0^{-1})\\
&= Q_0\left(\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)}I^{\mathrm{rev}}_{D_0}(\theta) + (\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)}I^{\mathrm{rev}}_{D_0}(\theta))^T\right)Q_0^{-1}\\
&= Q_0\left(\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)}I^{\mathrm{av}}_{D_0}(\theta) + (\overline{\mathrm{Ad}}_{\exp(-c(\theta)\xi)}I^{\mathrm{av}}_{D_0}(\theta))^T\right)Q_0^{-1}.
\end{aligned}$$

Here we used definition (6.25) of the symmetric parts of the $M_0^T D_0$ matrices of $I^{\mathrm{rev}}_{D_0}(\theta)$.

The M_0 part of the left-hand side of (6.4) is $(MI^{\mathrm{rev}}(\theta+1))_0 = M_0 I_0^{\mathrm{rev}}(\theta+1)$, and the D_0 part of the left-hand side of (6.4) is

$$(MI^{\mathrm{rev}}(\theta+1))_{D_0} = \overline{\mathrm{Ad}}_\sigma^{-1} I^{\mathrm{rev}}_{D_0}(\theta+1) + R(\theta+1),$$

where

$$R(\theta) = \pi_{\mathbf{m}_\mu}\mathrm{Ad}_\sigma^{-1}|_{\mathbf{n}_\mu} I^{\mathrm{rev}}_{D_3}(\theta) + D_0 I_0^{\mathrm{rev}}(\theta) + D_1 I_{10}^{\mathrm{rev}}(\theta).$$

So twice the symmetric part of the $M_0^T D_0$ matrices of the left-hand side of (6.4) is

$$\begin{aligned}
&(M_0 I_0^{\mathrm{rev}}(\theta+1))^T(\overline{\mathrm{Ad}}_\sigma^{-1} I^{\mathrm{rev}}_{D_0}(\theta+1) + R(\theta+1))\\
&\qquad + (\overline{\mathrm{Ad}}_\sigma^{-1} I^{\mathrm{rev}}_{D_0}(\theta+1) + R(\theta+1))^T(M_0 I_0^{\mathrm{rev}}(\theta+1))\\
&= \left(\overline{\mathrm{Ad}}_{\exp(-c(\theta+1)\xi)}I^{\mathrm{rev}}_{D_0}(\theta+1) + (\overline{\mathrm{Ad}}_{\exp(-c(\theta+1)\xi)}I^{\mathrm{rev}}_{D_0}(\theta+1))^T\right) + \tilde{R}(\theta+1)\\
&= \left(\overline{\mathrm{Ad}}_{\exp(-c(\theta+1)\xi)}I^{\mathrm{av}}_{D_0}(\theta+1) + (\overline{\mathrm{Ad}}_{\exp(-c(\theta+1)\xi)}I^{\mathrm{av}}_{D_0}(\theta+1))^T\right) + \tilde{R}(\theta+1),
\end{aligned}$$

where

$$\tilde{R}(\theta) = (M_0 I_0^{\mathrm{rev}}(\theta))^T R(\theta) + R(\theta)^T M_0 I_0^{\mathrm{rev}}(\theta).$$

Here we again used (6.25). Since $I^{\mathrm{av}}(\theta)$ satisfies (6.4) and all parts of $\tilde{R}(\theta)$ are determined by M and $I^{\mathrm{av}}(\theta)$, we conclude that the homotopy $I^{\mathrm{rev}}(\theta)$ satisfies (6.4).

6.6. Poisson structure of the Γ-reduced bundle. In this subsection, we describe the Poisson structure on the symmetry reduced bundle $\mathcal{U}/\Gamma$ near a Hamiltonian relative periodic orbit $\mathcal{P}$.

Define a bracket on the set of smooth functions on $\mathbf{g}_\mu^* \cong \mathrm{ann}(\mathbf{n}_\mu) \subset \mathbf{g}^*$ by

$$\{f_1, f_2\}^{j_\mu}(\zeta) \;=\; -(\mu+\zeta)\left([j_\mu(\zeta)\mathrm{D}_\zeta f_1(\zeta), j_\mu(\zeta)\mathrm{D}_\zeta f_2(\zeta)]\right),$$

where $j_\mu : \mathbf{g}_\mu \oplus \mathrm{ann}(\mathbf{n}_\mu) \to \mathbf{g}$ is as in (3.10) and the Lie bracket is on $\mathbf{g}$. It is straightforward to check that this is a Poisson bracket and equals the standard bracket on $\mathbf{g}_\mu^*$ if μ is split (see also [50, section 5.1]).

Extend this bracket to a Poisson structure on $\mathbf{g}_\mu^* \oplus N_1$ by defining

$$\{f_1, f_2\}(\zeta, w) \;=\; \{f_1, f_2\}^{j_\mu}(\zeta, w) \;+\; \omega_{N_1}(J_{N_1}\mathrm{D}_w f_1(\zeta, w), J_{N_1}\mathrm{D}_w f_2(\zeta, w)). \tag{6.26}$$

A straightforward calculation using the L_n-invariance of $\mathbf{n}_\mu$ shows that this Poisson bracket is L_n-semi-invariant.

This extends to a Poisson structure on $\widetilde{N} = (\mathbf{g}_\mu^* \oplus N_1 \oplus N_2)$ by making N_2 a space of Casimirs. Similarly, as the direct product of $\mathbf{g}_\mu^*$ and the symplectic manifolds N_1 and $T^*(\mathbb{R}/n\mathbb{Z}) = \mathbb{R}/n\mathbb{Z} \times N_2$, the space $\mathbb{R}/n\mathbb{Z} \times (\mathbf{g}_\mu^* \oplus N_1 \oplus N_2)$ is also naturally a Poisson space.

Let ι denote the L_n-equivariant inclusion of $\mathbf{g}_p$ into $\mathbf{g}_\mu$, and define a map

$$\mathbf{L}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}} : \mathbb{R}/n\mathbb{Z} \times \widetilde{N} \to \mathbf{g}_p^*, \qquad \mathbf{L}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}}(\theta, \zeta, w, E) = \mathbf{L}_{\mathbf{g}_\mu^*\oplus N_1}(\zeta, w),$$

where

$$\mathbf{L}_{\mathbf{g}_\mu^*\oplus N_1} : \mathbf{g}_\mu^* \oplus N_1 \to \mathbf{g}_p^*, \qquad \mathbf{L}_{\mathbf{g}_\mu^*\oplus N_1}(\zeta, w) = -\widehat{\mathrm{P}}\zeta + \mathbf{L}_{N_1}(w),$$

and $\widehat{\mathrm{P}}$ is the L_n-equivariant projection from $\mathbf{g}_\mu^*$ to $\mathbf{g}_p^*$ dual to ι. These maps are L_n-equivariant and momentum maps for the L_n-action on the Poisson spaces $\mathbb{R}/n\mathbb{Z} \times \widetilde{N}$ and $\mathbf{g}_\mu^* \oplus N_1$ (see [50, section 5.1]). It follows that the quotient variety

$$\mathcal{U}/\Gamma \equiv (\mathbb{R}/n\mathbb{Z} \times \widetilde{N})/(\Gamma_p \rtimes \mathbb{Z}_n) = \mathbf{L}^{-1}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}}(0)/(\Gamma_p \rtimes \mathbb{Z}_n),$$

where

$$\mathbf{L}^{-1}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}}(0) \equiv \mathbb{R}/n\mathbb{Z} \times N,$$

has a natural Poisson structure. The group G_p/Γ_p is isomorphic to $\mathbb{Z}_2$ if G_p contains elements that act antisymplectically on $\mathcal{M}$ and is trivial if it does not. In the first case, the action of the generator ρ of G_p/Γ_p on $\mathbf{L}^{-1}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}}(0)/(\Gamma_p \rtimes \mathbb{Z}_n)$ is "anti-Poisson."

In section 2 of [55, 22], we proved that $(\mathbb{R}/n\mathbb{Z}\times N)/(\Gamma_p \rtimes \mathbb{Z}_n)$ is diffeomorphic as a set to a neighborhood of the relative periodic orbit $\mathcal{P}$ in the orbit space $\mathcal{M}/\Gamma$. The above construction defines a Poisson structure on this neighborhood. It will follow from the proof below that this Poisson structure is isomorphic to that induced directly from $\mathcal{M}$ if we choose the homotopies $I(\theta)$ occurring in the bundle construction of section 6.1 as in Lemma 6.6.

6.7. Symplectic structure of the bundle. In this section, we describe the symplectic structure of the bundle (2.8) near a Hamiltonian relative periodic orbit.

Let the symmetry group Γ be algebraic, and let $\widetilde{\mathcal{M}}$ denote the manifold

$$\widetilde{\mathcal{M}} = G \times \mathbb{R}/n\mathbb{Z} \times \widetilde{N}, \tag{6.27}$$

where $\widetilde{N}$ is the extended Poincaré section (see (3.11)). Define a smooth action of $G \times L_n$ on $\widetilde{\mathcal{M}}$ by

$$(g, g_p, i).(\tilde{g}, \theta, \zeta, w, E) \;=\; (g\tilde{g}\alpha^{-i}g_p^{-1}, \chi(g_p)(\theta + i),\; \chi(g_p)\,(\mathrm{Ad}^*_{g_p\alpha^i})^{-1}\zeta,\; g_pQ_1^i w, E), \tag{6.28}$$

where $g, \tilde{g} \in G,\; g_p \in G_p$, and $i \in \mathbb{Z}_n$. Define a two-form $\tilde{\omega}$ on $\widetilde{\mathcal{M}}$ by

$$\tilde{\omega}(g, \theta, \zeta, w, E) \;=\; \chi(g)\left(\tilde{\omega}_G + \tilde{\omega}_\mu + \tilde{\omega}_{N_1} + \tilde{\omega}_{T_2\oplus N_2}\right), \tag{6.29}$$

where

1. $\tilde{\omega}_G$ is the pullback of the natural symplectic form ω_G on $T^*G \cong G \times \mathbf{g}^*$:

   $$\omega_G(g,\nu)\left((g\xi_1,\nu_1),(g\xi_2,\nu_2)\right) \;=\; \nu_2(\xi_1) - \nu_1(\xi_2) + \nu\left([\xi_1,\xi_2]\right), \tag{6.30}$$

   where $g \in G,\; \nu, \nu_1, \nu_2 \in \mathbf{g}^*$, and $\xi_1, \xi_2 \in \mathbf{g}$ (see [1, Proposition 4.4.1]) by the map $(g, \theta, \nu, w, E) \mapsto (g, i_\mu\nu)$, in which the inclusion $i_\mu : \mathbf{g}^*_\mu \to \mathbf{g}^*$ is induced by the G_p-invariant complement $\mathbf{n}_\mu$ to $\mathbf{g}_\mu$ in $\mathbf{g}$;
2. $\tilde{\omega}_\mu$ is the pullback of the KKS symplectic form (6.7) on the coadjoint orbit $G\mu$ by $(g, \theta, \nu, w, E) \mapsto \mathrm{Ad}^*_{g^{-1}}\mu$;
3. $\tilde{\omega}_{N_1}$ is the pullback of the symplectic form ω_{N_1} on N_1 by $(g, \theta, \nu, w, E) \mapsto w$;
4. $\tilde{\omega}_{T_2\oplus N_2}$ is the pullback of the symplectic form $\omega_{T_2\oplus N_2}$ on $\mathbb{R}/n\mathbb{Z} \times \mathbb{R}$ by $(g, \theta, \nu, w, E) \mapsto (\theta, E)$.

Then the form $\tilde{\omega}$ is a symplectic form on a $(G \times L_n)$-invariant neighborhood of $G \times \mathbb{R}/n\mathbb{Z} \times \{(0,0,0)\}$ in $\widetilde{\mathcal{M}}$. The action of G on this neighborhood is χ-semisymplectic. The action (6.28) of L_n, and, in particular, G_p, is symplectic even though the G_p-action on the symplectic slice N_1 is semisymplectic with respect to the symplectic form ω_{N_1}.

A momentum map $\mathbf{L}_{\widetilde{\mathcal{M}}} : \widetilde{\mathcal{M}} \to \mathbf{g}_p^*$ for the symplectic action of L_n on $\widetilde{\mathcal{M}}$ is given by

$$\mathbf{L}_{\widetilde{\mathcal{M}}}(g, \theta, \nu, w, E) \;= \mathbf{L}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}}(\theta, \nu, w, E).$$

The map $\mathbf{L}_{\widetilde{\mathcal{M}}}$ is L_n-equivariant with respect to the action (6.28) on $\widetilde{\mathcal{M}}$ and the usual coadjoint action of L_n on $\mathbf{g}_p^*$. Because the action of L_n on $\widetilde{\mathcal{M}}$ is free, proper, and symplectic, we can reduce $\widetilde{\mathcal{M}}$ by it to obtain a natural symplectic structure $\tilde{\omega}_0$ on a G-invariant neighborhood $\widetilde{\mathcal{U}}_0$ of $(G \times \mathbb{R}/n\mathbb{Z} \times \{0,0,0\})/L_n$ in the manifold

$$\widetilde{\mathcal{M}}_0 \;=\; \mathbf{L}^{-1}_{\widetilde{\mathcal{M}}}(0)/L_n \;=\; (G \times \mathbf{L}^{-1}_{\mathsf{R}/n\mathsf{Z}\times\widetilde{N}}(0))/L_n \;\cong\; (G \times \mathbb{R}/n\mathbb{Z} \times N)/L_n.$$

The action of G on $\widetilde{\mathcal{M}}$ drops to a χ-semisymplectic action of G on $\widetilde{\mathcal{U}}_0$.

Let $v = (\nu, w, E) \in N$. By Theorem 2.2 the differential equations on N in the new coordinates are of the form

$$\dot{\theta} = f_{\Theta}(\theta, v), \quad \dot{v} = f_N(\theta, v),$$

with $f_{\Theta}(\theta, 0) \equiv 1$. Hence ω_1 with $\tau^*\omega_1 = f_{\Theta}(\theta, v)\tau^*\omega$ is a symplectic form on a G-invariant neighborhood $\mathcal{U}$ of $\mathcal{P}$. Without loss of generality, we let $\omega = \omega_1$.

As shown in [55], $\widetilde{\mathcal{U}}_0$ is G-equivariantly diffeomorphic to a G-invariant neighborhood $\mathcal{U}$ of the relative periodic orbit $\mathcal{P}$ in $\mathcal{M}$. The following theorem says that this diffeomorphism can be chosen to be a G-equivariant symplectomorphism with respect to the symplectic form ω of $\mathcal{M}$ and the symplectic form $\tilde{\omega}_0$ on $\widetilde{\mathcal{M}}_0$. It is a generalization to relative periodic orbits of the local normal form for symplectic G-manifolds near group orbits obtained by Marle [29], Guillemin and Sternberg [15], and Bates and Lerman [4].

Theorem 6.3. *There exists a G-equivariant symplectomorphism Ψ between a G-invariant open neighborhood of $(G \times \mathbb{R}/n\mathbb{Z} \times \{0\})/L_n$ in $\widetilde{\mathcal{M}}_0 = (G \times \mathbf{L}^{-1}_{\widetilde{\mathcal{M}}}(0))/L_n \cong (G \times \mathbb{R}/n\mathbb{Z} \times N)/L_n$ and a G-invariant open neighborhood of $\mathcal{P}$ in $\mathcal{M}$.*

Proof. Because of Proposition 6.1 we have $\omega(\mathrm{id}, 0, 0) = \tilde{\omega}_0(\mathrm{id}, 0, 0)$ at $p \cong (\mathrm{id}, 0, 0)$. Moreover, since by Lemma 6.6 we can choose the G_p-semiequivariant homotopy $I(\theta)$ occurring in the parametrization (6.2) of a neighborhood $\mathcal{U}$ of the relative periodic orbit given in section 6.1 to be symplectic and such that the action of L_n on N is as in (6.28), we have that $\tilde{\omega}_0 = \omega$ on $\mathcal{P}$.

We now apply the semisymplectic relative Darboux theorem [50, Theorem 5.3] (based on [15] and [4]) to conclude that there is a diffeomorphism Ψ defined on a neighborhood $\mathcal{U}$ of $\mathcal{P}$ in $\mathcal{M}$ such that $\tilde{\omega}_0 = \Psi^*\omega$. ■

This proves Theorem 3.1.

6.8. Skew product equations. In this final subsection, we derive the skew product equations (3.14) near Hamiltonian relative periodic orbits. Again we reparametrize time so that $\dot{\theta} \equiv 1$. Let $\hat{h}(\theta, \nu, w, E)$ denote the Hamiltonian in bundle coordinates, and let $\hat{h}(\theta, \zeta, w, E) = \hat{h}(\theta, \nu, w, E)$ for $\zeta = \nu + \zeta_p$, $\zeta \in \mathbf{g}_\mu^*$, $\zeta_p \in \mathbf{g}_p^*$, and $\nu \in (\mathbf{g}_\mu/\mathbf{g}_p)^*$. The vector field $f_{\hat{h}}$ in the coordinates $(g, \theta, \zeta, w, E) \in G \times \mathbb{R}/n\mathbb{Z} \times (\mathbf{g}_\mu^* \oplus N_1 \oplus N_2)$ is determined by the equation

$$\tilde{\omega}(f_{\hat{h}}, (\hat{g}, \hat{\theta}, \hat{\zeta}, \hat{w}, \hat{E})) = \mathrm{D}_{(\theta,\zeta,w,E)}\hat{h}(\theta, \zeta, w, E)(\hat{\theta}, \hat{\zeta}, \hat{w}, \hat{E}),$$

where $\hat{g} \in g\mathbf{g}$, and, by (6.29),

$$\begin{aligned}\tilde{\omega}(f_{\hat{h}}, (\hat{g}, \hat{\theta}, \hat{\nu}, \hat{w}, \hat{E})) = &-\dot{\zeta}(g^{-1}\hat{g}) + \hat{\zeta}(g^{-1}\dot{g}) + (\zeta + \mu)[g^{-1}\dot{g}, g^{-1}\hat{g}] \\ &+\omega_{N_1}(\dot{w}, \hat{w}) + \omega_{T_2 \oplus N_2}((\dot{\theta}, \dot{E}), (\hat{\theta}, \hat{E})).\end{aligned}$$

Comparing coefficients, we obtain the differential equations

$$\dot{w} = J_{N_1}\mathrm{D}_w\hat{h}, \quad \dot{E} = -\mathrm{D}_\theta\hat{h}, \quad \dot{\theta} = \mathrm{D}_E\hat{h},$$

and, as in [50],

$$\dot{g} = gj_\mu(\zeta)\mathrm{D}_\zeta\hat{h}, \quad \dot{\zeta} = \mathrm{ad}^*_{j_\mu(\zeta)\mathrm{D}_\zeta\hat{h}}(\zeta + \mu).$$

Since $\dot{\theta} = 1$, we have $\mathrm{D}_E\hat{h} \equiv 1$ so that $h = \hat{h} - E$ is independent of E. This yields the equations of Theorem 3.3.

The equations of Theorem 3.5 are obtained as in [50].

Acknowledgment. C. Wulff thanks the University of Warwick for their hospitality during visits, when parts of this paper were written.

REFERENCES

[1] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed., Benjamin/Cummings, Reading, MA, 1978.

[2] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, Heidelberg, Berlin, 1978.

[3] V. I. Arnold and B. A. Khesin, *Topological Methods in Hydrodynamics*, Springer-Verlag, New York, Berlin, Heidelberg, 1998.

[4] L. Bates and E. Lerman, *Proper group actions and symplectic stratified spaces*, Pacific J. Math., 181 (1997), pp. 201–229.

[5] A. Blaom, *Reconstruction phases via Poisson reduction*, Differential Geom. Appl., 12 (2000), pp. 231–252.

[6] T. J. Bridges and J. E. Furter, *Singularity Theory and Equivariant Symplectic Maps*, Lecture Notes in Math. 1558, Springer-Verlag, New York, Heidelberg, Berlin, 1993.

[7] H. W. Broer and G. Vegter, *Bifurcational aspects of parametric resonance*, in Dynamics Reported: Expositions in Dynamical Systems, Springer-Verlag, Berlin, 1992, pp. 1–53.

[8] S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium*, revised ed., Dover, New York, 1987.

[9] M.-C. Ciocci and A. Vanderbauwhede, *Bifurcation of periodic orbits for symplectic mappings*, J. Differ. Equations Appl., 3 (1998), pp. 485–500.

[10] M.-C. Ciocci and A. Vanderbauwhede, *On the bifurcation and stability of periodic orbits in reversible and symplectic diffeomorphisms*, in Symmetry and Perturbation Theory, World Scientific, River Edge, NJ, 1999, pp. 159–166.

[11] H. Cohen and R. G. Muncaster, *The Theory of Pseudo-Rigid Bodies*, Springer-Verlag, New York, 1988.

[12] M. Duflo and M. Vergne, *Une proprieté de la représentation coadjointe d'une algébre de Lie*, C. R. Acad. Sci. Paris Sér. A-B, 268 (1969), pp. 583–585.

[13] F. Fassò and D. Lewis, *Stability properties of the Riemann ellipsoids*, Arch. Ration. Mech. Anal., 158 (2001), pp. 259–292.

[14] V. Guillemin, E. Lerman, and S. Sternberg, *Symplectic Fibrations and Multiplicity Diagrams*, Cambridge University Press, Cambridge, UK, 1996.

[15] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, UK, 1990.

[16] I. Hoveijn, J. Lamb, and R. M. Roberts, *Reversible Equivariant Linear Hamiltonian Systems*, in preparation.

[17] P. Jensen, G. Osmann, and I. N. Kozin, *The formation of four-fold rovibrational energy clusters in H_2S, H_2Se, and H_2Te*, in Vibration-Rotational Spectroscopy and Molecular Dynamics, D. Papousek, ed., World Scientific, Singapore, 1997, pp. 298–351.

[18] G. R. Kirchhoff, *Vorlesungen über Mathematische Physik. Mechanik.*, Teubner, Leipzig, Germany, 1876.

[19] I. N. Kozin, R. M. Roberts, and J. Tennyson, *Symmetry and structure of rotating* H_3^+, J. Chem. Phys., 111 (1999), pp. 140–150.

[20] I. N. Kozin, R. M. Roberts, and J. Tennyson, *Relative equilibria of D_2H^+ and H_2D^+*, Mol. Phys., 98 (2000), pp. 295–307.

[21] H. Lamb, *Hydrodynamics*, Cambridge University Press, Cambridge, UK, 1932.

[22] J. S. W. Lamb and C. Wulff, *Reversible relative periodic orbits*, J. Differential Equations, 178 (2002), pp. 60–100.

[23] J. S. W. Lamb and I. Melbourne, *Bifurcation from discrete rotating waves*, Arch. Ration. Mech. Anal., 149 (1999), pp. 229–270.

[24] N. E. Leonard and J. E. Marsden, *Stability and drift of underwater vehicle dynamics: Mechanical systems with rigid motion symmetry*, Phys. D, 105 (1997), pp. 130–162.

[25] E. Lerman and T. Tokieda, *On relative normal modes*, C. R. Acad. Sci. Paris Sér. I Math., 328 (1999), pp. 413–418.

[26] D. Lewis and T. S. Ratiu, *Rotating n-gon / kn-gon vortex configurations*, J. Nonlinear Sci., 6 (1996), pp. 385–414.

[27] D. Lewis and J. C. Simo, *Nonlinear stability of rotating pseudo-rigid bodies*, Proc. Roy. Soc. London, 427 (1990), pp. 281–319.

[28] D. Lewis, T. Ratiu, J. C. Simo, and J. E. Marsden, *The heavy top: A geometric treatment*, Nonlinearity, 5 (1992), pp. 1–48.

[29] C.-M. Marle, *Modèle d'action Hamiltonienne d'un groupe de Lie sur une variété symplectique*, Rend. Sem. Mat. Univ. Politec. Torino, 43 (1985), pp. 227–251.

[30] J. E. Marsden, R. Montgomery, and T. S. Ratiu, *Reduction, symmetry and phases in mechanics*, Mem. Amer. Math. Soc., 436 (1990), pp. 1–110.

[31] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Springer-Verlag, New York, Berlin, Heidelberg, 1994.

[32] K. R. Meyer and G. R. Hall, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Springer-Verlag, New York, 1992.

[33] J. Montaldi, *Persistance d'orbites périodiques relatives dans les systèmes Hamiltoniens symétriques*, C. R. Acad. Sci. Paris Sér. I Math., 324 (1997), pp. 553–558.

[34] J. Montaldi and R. M. Roberts, *Relative equilibria of molecules*, J. Nonlinear Sci., 9 (1999), pp. 53–88.

[35] J. Montaldi and R. M. Roberts, *Note on semisymplectic group actions*, C. R. Acad. Sci. Paris Sér. I Math., 330 (2000), pp. 1079–1084.

[36] J. Montaldi, R. M. Roberts, and I. N. Stewart, *Periodic solutions near equilibria of symmetric Hamiltonian systems*, Philos. Trans. Roy. Soc. London Ser. A, 325 (1988), pp. 237–293.

[37] J. Montaldi, R. M. Roberts, and I. N. Stewart, *Stability of nonlinear normal modes of symmetric Hamiltonian systems*, Nonlinearity, 3 (1990), pp. 731–772.

[38] P. Newton, *The N-Vortex Problem*, Appl. Math. Sci. 145, Springer-Verlag, New York, 2001.

[39] J.-P. Ortega, *Relative Normal Modes for Nonlinear Hamiltonian Systems*, preprint, University of Lausanne, Lausanne, Switzerland, 2000.

[40] J.-P. Ortega and T. S. Ratiu, *Persistence and smoothness of critical relative elements in Hamiltonian systems with symmetry*, C. R. Acad. Sci. Paris Sér. I Math., 325 (1997), pp. 1107–1111.

[41] J.-P. Ortega and T. S. Ratiu, *A Dirichlet criterion for the stability of periodic and relative periodic orbits in Hamiltonian systems*, J. Geom. Phys., 32 (1999), pp. 131–159.

[42] J.-P. Ortega and T. S. Ratiu, *Non-linear stability of singular relative periodic orbits in Hamiltonian systems with symmetry*, J. Geom. Phys., 32 (1999), pp. 160–188.

[43] J.-P. Ortega and T. S. Ratiu, *The Dynamics Around Stable Hamiltonian Relative Equilibria*, preprint, University of Lausanne, Lausanne, Switzerland, 2000.

[44] G. Patrick, *Relative equilibria of Hamiltonian systems with symmetry: Linearization, smoothness and drift*, J. Nonlinear Sci., 5 (1995), pp. 373–418.

[45] G. Patrick, *Dynamics near relative equilibria: Nongeneric momenta at a* 1:1 *group-reduced resonance*, Math. Z., 232 (1999), pp. 747–788.

[46] S. Pekarsky and J. E. Marsden, *Point vortices on a sphere: Stability of relative equilibria*, J. Math. Phys., 39 (1998), pp. 5894–5906.

[47] G. E. Roberts, *Spectral instability of relative equilibria in the planar n-body problem*, Nonlinearity, 12 (1999), pp. 757–769.

[48] R. M. Roberts and M. E. R. de Sousa Dias, *Bifurcations from relative equilibria of Hamiltonian systems*, Nonlinearity, 10 (1997), pp. 1719–1738.

[49] R. M. Roberts and M. E. R. de Sousa Dias, *Symmetries of Riemann ellipsoids*, Resenhas, 4 (1999), pp. 183–221.

[50] R. M. Roberts, C. Wulff, and J. S. W. Lamb, *Hamiltonian systems near relative equilibria*, J. Differential Equations, to appear.

[51] P. G. Saffman, *Vortex Dynamics*, Cambridge University Press, Cambridge, UK, 1992.

[52] B. Sandstede, A. Scheel, and C. Wulff, *Bifurcations and dynamics of spiral waves*, J. Nonlinear Sci., 9 (1999), pp. 439–478.

[53] J. J. Slawianowski, *Affinely-rigid bodies and Hamiltonian systems on $gl(n, r)$*, Rep. Math. Phys., 26 (1988), pp. 73–119.

[54] M. E. R. de Sousa Dias, *Pseudo-rigid bodies: A geometric Lagrangian approach*, Acta Appl. Math., 70 (2002), pp. 209–230.

[55] C. Wulff, J. S. W. Lamb, and I. Melbourne, *Bifurcations from relative periodic solutions*, Ergodic Theory Dynam. Systems, 21 (2001), pp. 605–635.

[56] C. Wulff, *Persistence of Relative Equilibria in Hamiltonian Systems with Noncompact Symmetry*, preprint, Free University of Berlin, Berlin, Germany, 2001.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 1, pp. 44–64

Stepwise Precession of the Resonant Swinging Spring*

Darryl D. Holm† and Peter Lynch‡

Abstract. The swinging spring, or elastic pendulum, has a 2:1:1 resonance arising at cubic order in its approximate Lagrangian. The corresponding modulation equations are the well-known three-wave equations that also apply, for example, in laser-matter interaction in a cavity. We use Hamiltonian reduction and pattern evocation techniques to derive a formula that describes the characteristic feature of this system's dynamics, namely, the stepwise precession of its azimuthal angle.

Key words. classical mechanics, variational principles, averaged Lagrangian, elastic pendulum, nonlinear resonance

AMS subject classifications. Primary, 37J15; Secondary, 37J35, 70H06, 37N05, 37N15

PII. S1111111101388571

1. Introduction.

1.1. Problem statement, approach, and summary of results.

The elastic pendulum or swinging spring is a simple mechanical system that exhibits complex dynamics. It consists of a heavy mass suspended from a fixed point by a light spring which can stretch but not bend, moving under gravity. We investigate the 2:1:1 resonance dynamics of this system in three dimensions and study its characteristic feature—the regular stepwise precession of its azimuthal angle.

When the Lagrangian is approximated to cubic order and averaged over the fast dynamics, the resulting modulation equations have three independent constants of motion and are completely integrable. These modulation equations are identical to the three-wave equations for resonant triad interactions in fluids and plasmas and in laser-matter interaction. We reduce the system to a form amenable to analytical solution and show how the full solution may be reconstructed. We examine the geometry of the solutions in phase-space and develop a number of simple qualitative descriptions of the motion.

We compare solutions of the exact and modulation equations and show that they are remarkably similar. A characteristic stepwise precession occurs as the motion cycles between quasi-vertical and quasi-horizontal. That is, during each quasi-vertical phase, the azimuth of the swing plane precesses by a constant angular increment. This stepwise azimuthal precession occurs in bursts when the motion is nearly vertical. By transforming to nonuniformly rotating coordinates and assuming a geometric constraint (essentially the method of pattern evocation), we find a formula for the rotation of the swing plane. This formula gives a highly accurate

*Received by the editors April 27, 2001; accepted for publication (in revised form) by M. Golubitsky November 13, 2001; published electronically April 9, 2002.

http://www.siam.org/journals/siads/1-1/38857.html

†Theoretical Division and CNLS, Los Alamos National Laboratory, MS B284, Los Alamos, NM 87545 (dholm@lanl.gov). The work of this author was supported by US DOE, under contract W-7405-ENG-36.

‡Met Éireann, Glasnevin Hill, Dublin 9, Ireland (Peter.Lynch@met.ie).

description of the stepwise precession of the azimuthal angle of the motion. It is a striking result that the stepwise precession can be described so accurately by assuming a geometric constraint to hold, and this invites investigation on a deeper level.

We restrict our attention in this study to the pure 2:1:1 resonance. However, the analysis may easily be generalized to allow for frequency ratios which are not precisely in resonance. As the amplitude of the motion increases, energy exchange may be expected to occur for increasingly larger detuning of the frequencies. Aničin, Davidović, and Babović [4] investigated how the parameter range for energy exchange depends on the amplitude. Ultimately, the assumptions underlying perturbation analysis break down, and chaotic behavior is found. Georgiou [9] has studied the global geometric structure of the dynamics of the planar elastic pendulum. The phenomenon of precession has been noticed in a number of other contexts [26, 9].

1.2. History of the problem. The first comprehensive analysis of the elastic pendulum appeared in [28]. These authors were inspired by the analogy between this system and the Fermi resonance of a carbon-dioxide molecule. We make connections in this paper with other physical systems of current interest. For example, we show that the modulation equations for the averaged motion of the swinging spring may be transformed into the equations for three-wave interactions that appear in analyzing fluid and plasma systems and in laser-matter interaction. These three complex equations are also identical to the Maxwell–Schrödinger envelope equations for the interaction between radiation and a two-level resonant medium in a microwave cavity [11]. The three-wave equations also govern the envelope dynamics of light-waves in an inhomogeneous material [8, 2, 3]. For the special case where the Hamiltonian takes the value zero, the equations reduce to Euler's equations for a freely rotating rigid body. Finally, the equations are also equivalent to a complex (unforced and undamped) version of the Lorenz [17] three-component model, which has been the subject of many studies [27]. Thus the simple spring pendulum, which was first studied to provide a classical analogue to the quantum phenomenon of Fermi resonance, now provides a concrete mechanical system which simulates a wide range of physical phenomena.

All of the previous studies of the spring pendulum known to us have considered motion in two dimensions. To our knowledge, only Cayton [6] discussed three-dimensional solutions and observed the curious rotation of the swing plane between successive cycles when the horizontal energy is maximum. This particular aspect of the behavior of the swinging spring in three dimensions is its most striking difference from two-dimensional motions. Suppose the system is excited initially near its vertical oscillation mode. Since purely vertical motion is unstable, horizontal motion soon develops. The horizontal oscillations grow to a maximum and then subside again. An alternating cycle of quasi-vertical and quasi-horizontal oscillations recurs indefinitely. Seen from above, during each horizontal excursion of several oscillations, the projected motion is approximately elliptical. Experimentally and numerically one observes that between any two successive horizontal excursions the orientation of the projected ellipse rotates by the same angle, thereby causing a *stepwise precession of the swing plane.* In principle, the precession angle between successive horizontal excursions can be deduced from the complete solution of the integrable envelope equations. We seek a simple approximate expression for the precession of the swing plane in terms of the solution of the reduced dynamics.

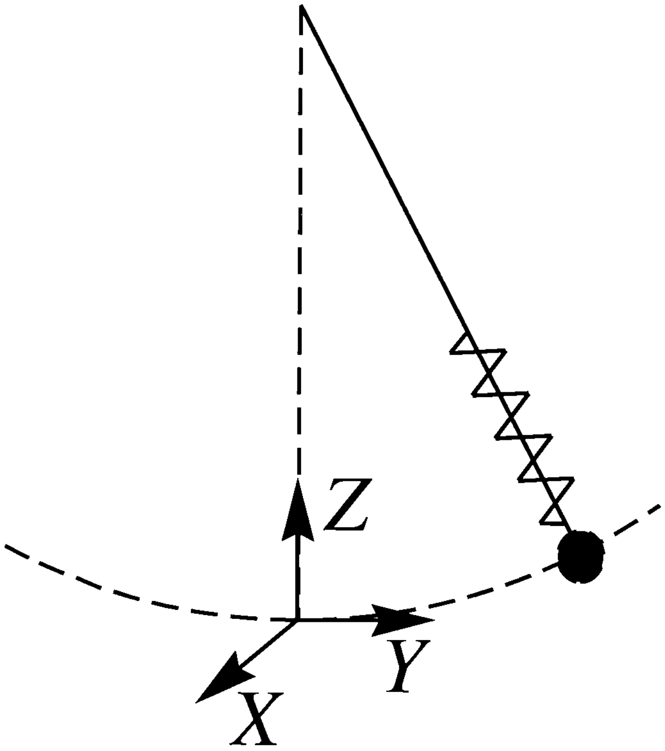


Figure 2.1. *Schematic diagram of the elastic pendulum, or swinging spring. Cartesian coordinates centered at the position of equilibrium are used.*

Lynch [19] found a particular solution for the rate of precession of the swing plane by using the method of multiple time scales in rotating coordinates and introducing a certain angular solution Ansatz. We recover Lynch's particular solution among a family of other solutions for the swing plane precession rate. This family is obtained via the method of averaged Lagrangians by seeking solutions of the modulation equations that satisfy a geometrical constraint of being "instantaneously elliptical." We apply the method of *pattern evocation in shape space* [21, 22]. Using this process, one identifies patterns by viewing the dynamics relative to rotating frames with certain critical angular velocities. Our numerical integrations show that the solution resulting from this geometrical postulate estimates the precession of the swing plane with surprisingly high accuracy.

2. Equations of motion. The physical system under investigation is an elastic pendulum, or swinging spring, consisting of a heavy mass suspended from a fixed point by a light spring which can stretch but not bend, moving under gravity, g. We assume an unstretched length ℓ_0, a length ℓ at equilibrium, a spring constant k, and a unit mass $m = 1$. The corresponding Lagrangian, approximated to cubic order in the amplitudes, is

$$L = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - \frac{1}{2}\left(\omega_R^2(x^2 + y^2) + \omega_Z^2 z^2\right) + \frac{1}{2}\lambda(x^2 + y^2)z\,, \tag{2.1}$$

where x, y, and z are Cartesian coordinates centered at the point of equilibrium, $\omega_R = \sqrt{g/\ell}$ is the frequency of linear pendular motion, $\omega_Z = \sqrt{k/m}$ is the frequency of its elastic oscillations, and $\lambda = \ell_0\omega_Z^2/\ell^2$. The system is illustrated schematically in Figure 2.1. The Euler–Lagrange equations of motion may be written

$$\ddot{x} + \omega_R^2 x = \lambda x z\,, \tag{2.2}$$

$$\ddot{y} + \omega_R^2 y = \lambda y z\,, \tag{2.3}$$

$$\ddot{z} + \omega_Z^2 z = \frac{1}{2}\lambda(x^2 + y^2)\,. \tag{2.4}$$

There are two constants of the motion, the total energy E and the angular momentum h given by

$$E = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) + \frac{1}{2}\left(\omega_R^2(x^2+y^2) + \omega_Z^2 z^2\right) - \frac{1}{2}\lambda(x^2+y^2)z\,, \quad h = (x\dot{y} - y\dot{x})\,.$$

The system is not integrable. Its chaotic motions have been studied by many authors (see, e.g., references in [20]). Previous studies have considered the two-dimensional case, for which the angular momentum h vanishes.

We confine our attention to the resonant case $\omega_Z = 2\omega_R$ and apply the averaged Lagrangian technique [30]. The solution of (2.2)–(2.4) is assumed to be of the form

$$x = \Re[a_0(t)\exp(i\omega_R t)]\,, \tag{2.5}$$

$$y = \Re[b_0(t)\exp(i\omega_R t)]\,, \tag{2.6}$$

$$z = \Re[c_0(t)\exp(2i\omega_R t)]\,. \tag{2.7}$$

(The zero-subscripts in a_0, b_0, and c_0 are introduced to distinguish from the variables a, b, and c in a rotating frame, introduced below.) The coefficients $a_0(t)$, $b_0(t)$, and $c_0(t)$ are assumed to vary on a time scale which is much longer than the time scale of the oscillations, $\tau = 1/\omega_R$. The Lagrangian is averaged over time τ to give

$$\langle L\rangle = \frac{1}{2}\omega_R\left[\Im\{\dot{a}_0 a_0^* + \dot{b}_0 b_0^* + 2\dot{c}_0 c_0^*\} + \Re\{\kappa(a_0^2 + b_0^2)c_0^*\}\right]\,,$$

where $\kappa = \lambda/(4\omega_R)$. We regard the quantities a_0, b_0, c_0 as generalized coordinates. The averaged Lagrangian equations of motion are then

$$i\dot{a}_0 = \kappa a_0^* c_0\,, \tag{2.8}$$

$$i\dot{b}_0 = \kappa b_0^* c_0\,, \tag{2.9}$$

$$i\dot{c}_0 = \frac{1}{4}\kappa(a_0^2 + b_0^2). \tag{2.10}$$

Equations (2.8)–(2.10) are the complex versions of (68)–(73) in [19]. These were derived using the method of multiple time-scale analysis, where the small parameter ϵ for the analysis was the amplitude of the dependent variables so that terms quadratic in the unknowns were second order whereas linear terms were first order in ϵ. Thus the averaged Lagrangian technique yields results completely equivalent to the results using more standard averaging theory.

We now transform variables as follows:

$$A = \frac{1}{2}\kappa(a_0 + ib_0)\,, \quad B = \frac{1}{2}\kappa(a_0 - ib_0)\,, \quad C = \kappa c_0\,.$$

Consequently, the equations of motion take the form

$$i\dot{A} = B^*C\,, \tag{2.11}$$
$$i\dot{B} = CA^*\,, \tag{2.12}$$
$$i\dot{C} = AB\,. \tag{2.13}$$

These three complex equations are well known as the *three-wave interaction equations*, which govern quadratic wave resonance in fluids and plasmas.

The three-wave interaction equations (2.11)–(2.13) may be written in canonical form with Hamiltonian $H = \Re(ABC^*)$ and Poisson brackets $\{A, A^*\} = \{B, B^*\} = \{C, C^*\} = -2i$ as

$$i\dot{A} = i\{A, H\} = 2\partial H/\partial A^*\,, \tag{2.14}$$
$$i\dot{B} = i\{B, H\} = 2\partial H/\partial B^*\,, \tag{2.15}$$
$$i\dot{C} = i\{C, H\} = 2\partial H/\partial C^*\,. \tag{2.16}$$

These equations conserve the following three quantities:

$$H = \frac{1}{2}(ABC^* + A^*B^*C) = \Re(ABC^*)\,, \tag{2.17}$$
$$N = |A|^2 + |B|^2 + 2|C|^2\,, \tag{2.18}$$
$$J = |A|^2 - |B|^2\,. \tag{2.19}$$

Thus the modulation equations for the swinging spring are transformed into the three-wave equations, which are known to be completely integrable. See [2] for references to the three-wave equations and an extensive elaboration of their properties as a paradigm for Hamiltonian reduction.

The following positive-definite combinations of N and J are physically significant:

$$N_+ \equiv \frac{1}{2}(N + J) = |A|^2 + |C|^2\,, \qquad N_- \equiv \frac{1}{2}(N - J) = |B|^2 + |C|^2\,.$$

These combinations are known as the *Manley–Rowe relations* in the extensive literature about three-wave interactions. The quantities H, N_+, and N_- provide three independent constants of the motion.

2.1. A brief history of the three-wave equations.

Fluids and plasmas. The three-wave equations model the nonlinear dynamics of the amplitudes of three waves in fluids or plasmas [5]. The equations result from a perturbation analysis of the barotropic potential vorticity equation

$$\frac{\partial}{\partial t}[\nabla^2\psi - F\psi] + \left(\frac{\partial\psi}{\partial x}\frac{\partial\nabla^2\psi}{\partial y} - \frac{\partial\psi}{\partial y}\frac{\partial\nabla^2\psi}{\partial x}\right) + \beta\frac{\partial\psi}{\partial x} = 0 \tag{2.20}$$

(see, e.g., [25] for theory and notation). This equation is equivalent to the Hasegawa–Mima equation describing drift-waves in an inhomogeneous plasma in a magnetic field [10]. Longuet-Higgins and Gill [16] examined the interactions between planetary Rossby waves in the atmosphere and derived detailed conditions for three-wave resonance. The correspondence between Rossby waves in the atmosphere and drift-waves in plasma has been thoroughly explored in [15]. Resonant wave-triad interactions play an essential role in the generation of turbulence and in determining the statistics of the power spectrum. Both energy and enstrophy are conserved in fluid systems governed by the potential vorticity equation (2.20).

Laser-matter interaction. Equations (2.11)–(2.13) are also equivalent to the Maxwell–Schrödinger envelope equations for the interaction between radiation and a two-level resonant medium in a microwave cavity. Holm and Kovačič [11] show that perturbations of this system lead to homoclinic chaos, but we shall not explore that issue here. Wersinger, Finn, and Ott [29] used a forced and damped version of the three-wave equations to study instability saturation by nonlinear mode coupling and found irregular solutions indicating the presence of a strange attractor. See also [1, 12, 13, 24] for more detailed studies of the perturbed three-wave system.

Nonlinear optics. The three-wave system also describes the dynamics of the envelopes of light-waves interacting quadratically in nonlinear material. The system has been examined in a series of recent papers [2, 3, 18] using a geometrical approach which allowed the reduced dynamics for the wave intensities to be represented as motion on a closed surface in three dimensions—the three-wave surface. Information about the corresponding reconstruction phases was recovered using the theory of connections on principal bundles.

In the special case when $H = 0$, the system (2.11)–(2.13) reduces to three real equations. Let

$$A = iX_1\exp(i\phi_1)\,, \quad B = iX_2\exp(i\phi_2)\,, \quad C = iX_3\exp(i(\phi_1+\phi_2)),$$

where X_1, X_2, and X_3 are real and the phases ϕ_1 and ϕ_2 are constants. The modulation equations become

$$\dot{X}_1 = -X_2X_3\,, \quad \dot{X}_2 = -X_3X_1\,, \quad \dot{X}_3 = +X_1X_2\,. \tag{2.21}$$

We note that these equations are rescaled versions of the Euler equations for the rotation of a free rigid body. The dynamics in this special case is expressible as motion on $\mathbf{R}^3$, namely,

$$\dot{\mathbf{X}} = \frac{1}{8}\nabla J \times \nabla N = \frac{1}{4}\nabla N_+ \times \nabla N_-\,. \tag{2.22}$$

Considering the constancy of J and N, we can describe a trajectory of the motion as an intersection between a hyperbolic cylinder (J constant; see (2.19)) and an oblate spheroid (N

constant; see (2.18)). Equation (2.22) provides an alternative description. Here we have used the freedom in the $\mathbf{R}^3$ Poisson bracket exploited by [7] to represent the equations of motion on the intersection of two orthogonal circular cylinders, the level surfaces of the Manley–Rowe quantities, N_+ and N_-. The invariance of the trajectories means that while the level surfaces of J and N differ from those of N_+ and N_-, their intersections are precisely the same. For this particular value of $H = 0$, the motion may be further reduced by expressing it in the coordinates lying on one of these two cylinders. See [14] for the corresponding transformation of rigid body motion into pendular motion. See [2, 3, 7, 23] for discussions of geometric phases in this situation.

2.2. Reduction of the system and reconstruction of the solution. To reduce the system for $H \neq 0$, we employ a further canonical transformation introduced in [11]. The goal is to encapsulate complete information about the Hamiltonian in a single variable Z by using the invariants of the motion. Once Z is found, the Manley–Rowe relations yield the remaining variables. We set

$$A = |A| \exp(i\xi) \,, \tag{2.23}$$

$$B = |B| \exp(i\eta) \,, \tag{2.24}$$

$$C = Z \exp(i(\xi + \eta)) \,. \tag{2.25}$$

This transformation is canonical—it preserves the symplectic form

$$dA \wedge dA^* + dB \wedge dB^* + dC \wedge dC^* = dZ \wedge dZ^* \,.$$

In these variables, the Hamiltonian is a function of only Z and Z^*:

$$H = \frac{1}{2}(Z + Z^*) \cdot \sqrt{N_+ - |Z|^2} \cdot \sqrt{N_- - |Z|^2} \,.$$

The Poisson bracket is $\{Z, Z^*\} = -2i$, and the canonical equations reduce to

$$i\dot{Z} = i\{Z, H\} = 2\frac{\partial H}{\partial Z^*} \,.$$

This provides the slow dynamics of both the amplitude and phase of $Z = |Z|e^{i\zeta}$.

The amplitude $|Z| = |C|$ is obtained in closed form in terms of Jacobi elliptic functions as the solution of

$$\left(\frac{d\mathcal{Q}}{d\tau}\right)^2 = \left[\mathcal{Q}^3 - 2\mathcal{Q}^2 + (1 - \mathcal{J}^2)\mathcal{Q} + 2\mathcal{E}\right] , \tag{2.26}$$

where $\mathcal{Q} = 2|Z|^2/N$, $\mathcal{J} = J/N$, $\mathcal{E} = -4H^2/N^3$ and $\tau = \sqrt{2N}t$. This is equivalent to (75) in [19]. (An explicit expression for the solution in terms of elliptic functions is given in that paper.) Once $|Z|$ is known, $|A|$ and $|B|$ follow immediately from the Manley–Rowe relations:

$$|A| = \sqrt{N_+ - |Z|^2} \,, \qquad |B| = \sqrt{N_- - |Z|^2} \,.$$

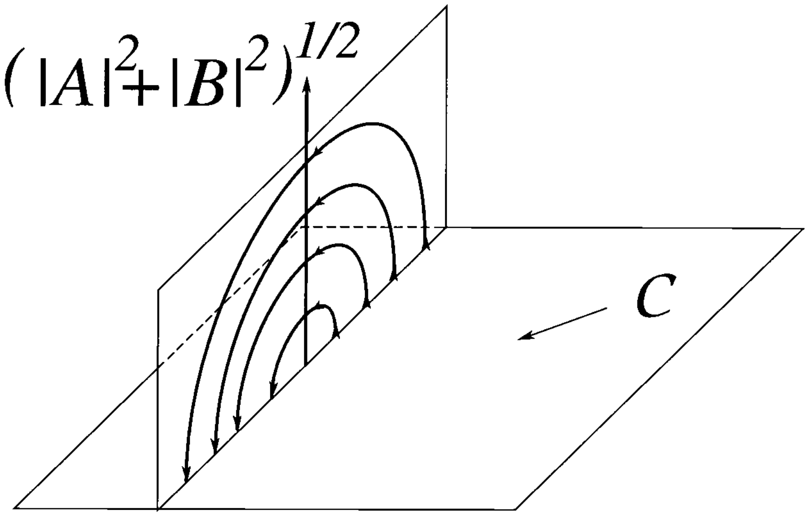


Figure 3.1. *$\mathcal{C}$ is the plane of critical points, $A = 0 = B$. The vertical axis is $R = \sqrt{|A|^2 + |B|^2}$. The vertical plane contains heteroclinic semiellipses passing from c_0 to $-c_0$.*

The phases ξ and η may now be determined. Using the three-wave equations (2.11)–(2.13) together with (2.23)–(2.25), one finds

$$\dot{\xi} = -\frac{H}{|A|^2}\,, \qquad \dot{\eta} = -\frac{H}{|B|^2} \tag{2.27}$$

so that ξ and η can be obtained by quadratures. Finally, the phase ζ of Z is determined unambiguously by

$$\frac{d|Z|^2}{dt} = -2H\tan\zeta \qquad \text{and} \qquad H = |A||B||Z|\cos\zeta\,. \tag{2.28}$$

Hence we can now reconstruct the full solution as

$$A = |A|\exp(i\xi)\,, \quad B = |B|\exp(i\eta)\,, \quad C = |Z|\exp\big(i(\xi+\eta+\zeta)\big)\,.$$

3. Phase portraits. Consider the plane $\mathcal{C}$ in phase-space defined by $A = B = 0$. This is a plane of unstable equilibrium points, representing purely vertical oscillations of the spring. The Hamiltonian vanishes identically on this plane, as does the angular momentum J. Each point c_0 in $\mathcal{C}$ has a heteroclinic orbit linking it to its antipodal point $-c_0$. Thus the plane $\mathcal{C}$ of critical points is connected to itself by heteroclinic orbits. In Figure 3.1, the horizontal plane is $\mathcal{C}$, and the vertical plane contains heteroclinic orbits from c_0 to $-c_0$. The vertical axis is $R = \sqrt{|A|^2 + |B|^2}$. Since $N = R^2 + 2|C|^2$ is constant, each heteroclinic orbit is a semiellipse. Motion starting on one of these semiellipses will move toward an end-point, taking infinite time to reach it.

In Figure 3.2 taken from [11], we present another view of the trajectories for $J = 0$. The Hamiltonian is

$$H = \frac{1}{2}(Z + Z^*)\cdot\left(\frac{1}{2}N - |Z|^2\right)\,.$$

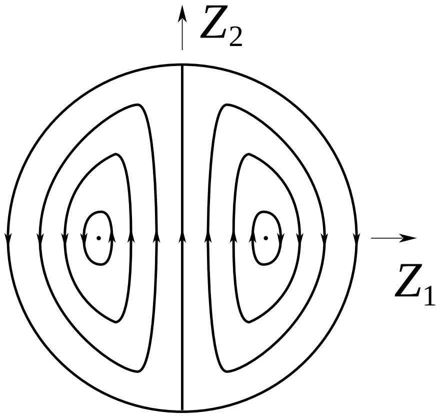


Figure 3.2. *Phase portrait in the Z-plane for $J = 0$. The motion is confined within the circle $|Z|^2 = \frac{1}{2}N$. The segment of the imaginary axis within this circle is the homoclinic orbit.*

Accessible points lie on or within the circle $|Z|^2 = N/2$. For $H = 0$, the trajectory is the segment of the imaginary axis within the circle. This is the homoclinic orbit. For $H \neq 0$, we solve for the imaginary part of $Z = Z_1 + iZ_2$,

$$Z_2 = \pm\sqrt{-Z_1^2 + \frac{1}{2}N - (H/Z_1)}.$$

This allows us to plot the trajectories for the range of H for which real solutions exist. There are two equilibrium points, at $Z = \pm\sqrt{N/6}$, corresponding to solutions for which there is no exchange of energy between the vertical and horizontal components. These are the cup-like and cap-like solutions first discussed by Vitt and Gorelik [28].

3.1. Geometry of the motion for fixed J. The vertical amplitude is governed by (2.26), which we write as

$$\frac{1}{2}\left(\frac{d\mathcal{Q}}{d\tau}\right)^2 + \mathcal{V}(\mathcal{Q}) = \mathcal{E}\,, \tag{3.1}$$

with the potential $\mathcal{V}(\mathcal{Q})$ given by

$$\mathcal{V}(\mathcal{Q}) = -\frac{1}{2}\left[\mathcal{Q}^3 - 2\mathcal{Q}^2 + (1 - \mathcal{J}^2)\mathcal{Q}\right]\,. \tag{3.2}$$

We note that $\mathcal{V}(\mathcal{Q})$ has three zeros: $\mathcal{Q} = 0$, $\mathcal{Q} = 1 - \mathcal{J}$, and $\mathcal{Q} = 1 + \mathcal{J}$. Equation (3.1) is an energy equation for a particle of unit mass, with position $\mathcal{Q}$ and energy $\mathcal{E}$, moving in a cubic potential field $\mathcal{V}(\mathcal{Q})$. In Figure 3.3 we plot $\dot{\mathcal{Q}}$, given by (3.1), against $\mathcal{Q}$ for the cases $\mathcal{J} = 0$ (left panel) and $\mathcal{J} = 0.25$ (right panel), for a range of values of $\mathcal{E}$. Each curve represents the projection onto the reduced phase-space of the trajectory of the modulation envelope. The centers are relative equilibria, corresponding to the elliptic-parabolic solutions of [19], which are generalizations of the cup-like and cap-like solutions of [28]. The case when $\mathcal{J} = 0$ includes the homoclinic trajectory, for which $H = 0$.

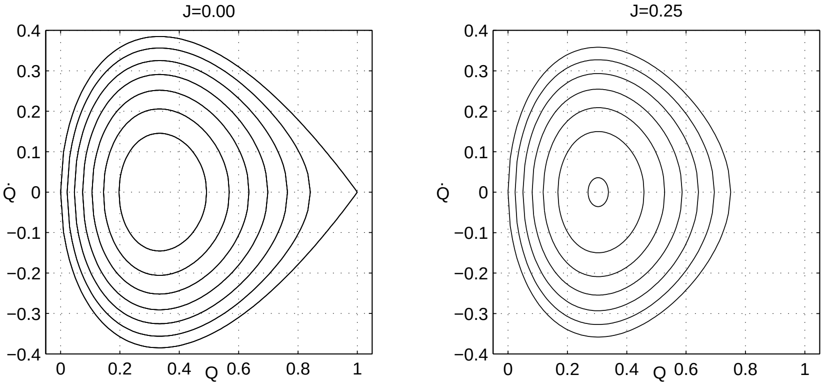


Figure 3.3. *Plots of $\dot{\mathcal{Q}}$ versus $\mathcal{Q}$ for $\mathcal{J} = 0$ and $\mathcal{J} = 0.25$ for a range of values* $E \in \{-0.0635, -0.0529, -0.0423, -0.0317, -0.0212, -0.0106, 0\}$.

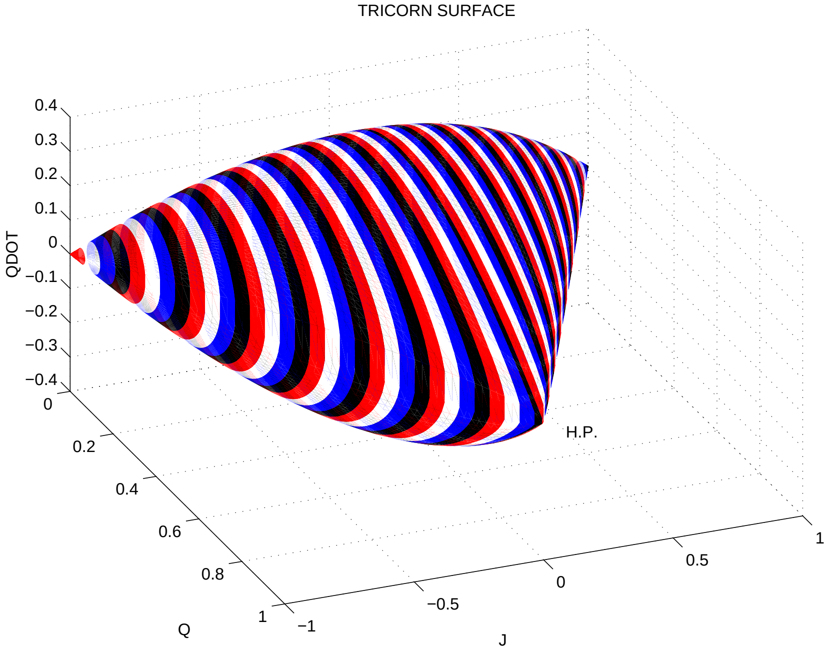
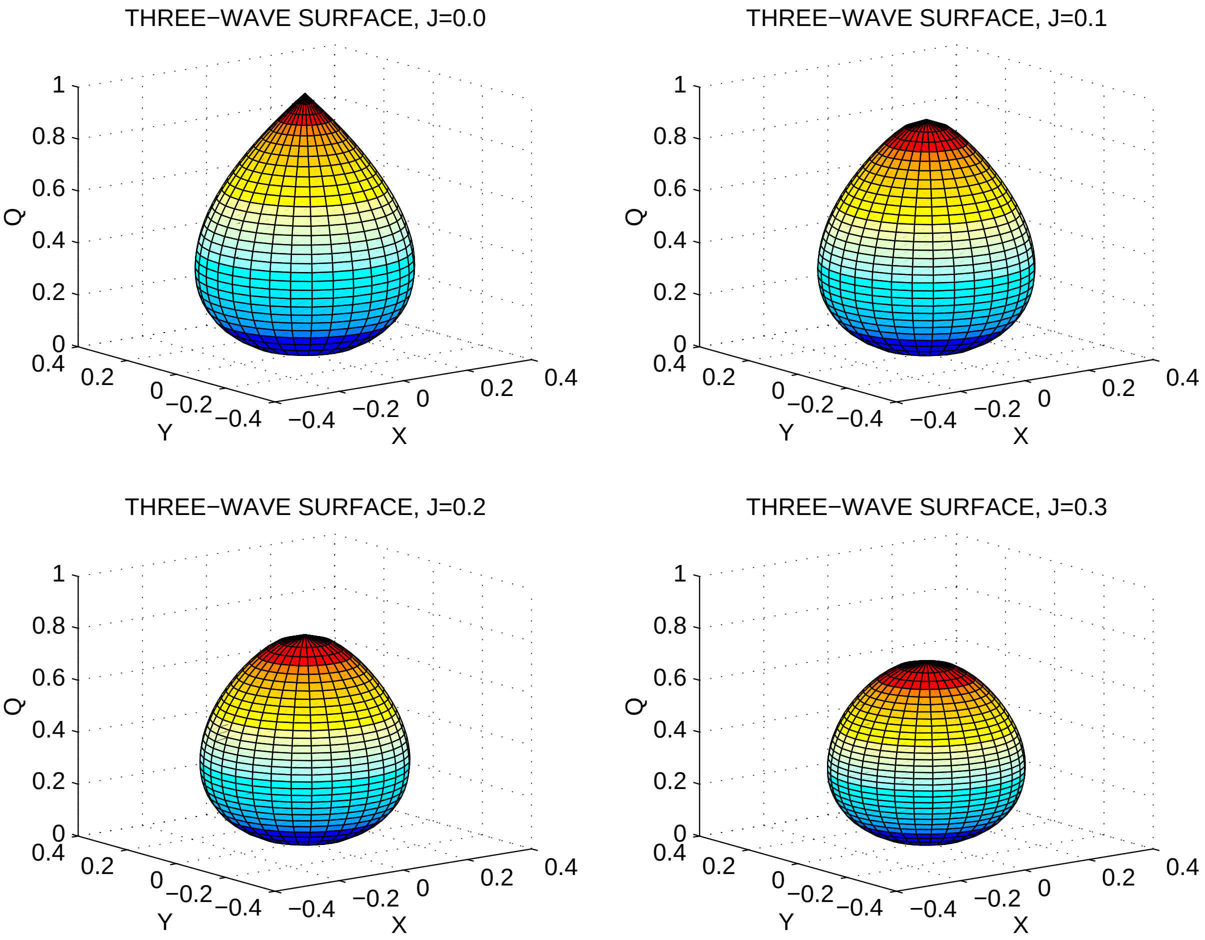


Figure 3.4. *Tricorn surface, upon which motion takes place when $H = 0$. The coordinates are $\mathcal{J}$, $\mathcal{Q}$, $\dot{\mathcal{Q}}$. The motion takes place on the intersections of this surface with a plane of constant $\mathcal{J}$ (such planes are indicated by the stripes). This surface has three singular points. The homoclinic point is marked H.P.*

3.2. Geometry of the motion for $H = 0$. For arbitrary $\mathcal{J}$, the $H = 0$ motions are on a surface in the space with coordinates $(\mathcal{Q}, \dot{\mathcal{Q}}, \mathcal{J})$. This surface is depicted in Figure 3.4. It has three singular points (i.e., it is equivalent to a sphere with three pinches), and its shape is similar to a tricorn hat. The motion takes place on an intersection of this surface with a plane of constant $\mathcal{J}$. There are three equilibrium solutions: that with $\mathcal{J} = 0$ (marked H.P. in Figure 3.4) is at the extremity of the homoclinic trajectory and corresponds to purely vertical oscillatory motion; those with $\mathcal{J} = \pm 1$ correspond to purely horizontal motion, clockwise or


THREE−WAVE SURFACE, J=0.0

THREE−WAVE SURFACE, J=0.1

THREE−WAVE SURFACE, J=0.2

THREE−WAVE SURFACE, J=0.3

Figure 3.5. *Surfaces of revolution about the $\mathcal{Q}$-axis for $\mathcal{J} \in \{0.0, 0.1, 0.2, 0.3\}$. The radius for given $\mathcal{Q}$ is given by the square-root of the cubic $-\mathcal{V}(\mathcal{Q})$. For given $\mathcal{J}$, the motion takes place on the intersection of the corresponding surface with a plane of constant X.*

anticlockwise, with the spring tracing out a cone. The purely vertical motion is unstable; the conical motions are stable. (Perturbations about conical motion were investigated by Lynch [19].) The dynamics on the tricorn is similar to the motion of a free rigid body. The three singular points correspond to the steady states of rotation about the three principal axes.

3.3. Three-wave surfaces. There is another way to depict the motion in reduced phase-space. Let us consider a reduced phase-space with x- and y-axes $X = \Re\{ABC^*\}$ and $Y = \Im\{ABC^*\}$ and z-axis $\mathcal{Q} = 2|Z|^2/N$. We note that $X \equiv H$. It follows from (2.17)–(2.19) that

$$X^2 + Y^2 = |A|^2|B|^2|C|^2 = \frac{1}{4}|Z|^2\left[(2|Z|^2 - N)^2 - J^2\right].$$

We define $\mathcal{X} = (2/N^{3/2})X$ and $\mathcal{Y} = (2/N^{3/2})Y$ and can write

$$\mathcal{X}^2 + \mathcal{Y}^2 = \frac{1}{2}\left[\mathcal{Q}^3 - 2\mathcal{Q}^2 + (1 - \mathcal{J}^2)\mathcal{Q}\right] = -\mathcal{V}(\mathcal{Q}) \tag{3.3}$$

where $\mathcal{V}$ is as defined in (3.2). We note that $\mathcal{X}^2 = -\mathcal{E}$ and $\mathcal{Y}^2 = \frac{1}{2}(d\mathcal{Q}/d\tau)^2$. Equation (3.3) implies that the motion takes place on a surface of revolution about the $\mathcal{Q}$-axis. The radius for a given value of $\mathcal{Q}$ is the square-root of the cubic $-\mathcal{V}(\mathcal{Q})$. The physically assessable region is $0 \leq \mathcal{Q} \leq 1 - |\mathcal{J}|$. Several such surfaces (for $\mathcal{J} \in \{0.0, 0.1, 0.2, 0.3\}$) are shown in Figure 3.5. Since $\mathcal{X}^2 = \mathcal{H}^2 = 4H^2/N^3$, the motion for given $\mathcal{J}$ takes place on the intersection of the corresponding surface of revolution with a plane of constant $\mathcal{X}$.

We can relate the tricorn surface to the surface of revolution. The former is appropriate for $H = 0$; the $H \neq 0$ case is represented by trajectories inside this surface. If we slice the tricorn surface in a plane of fixed $\mathcal{J}$, we get a set of closed trajectories—the outside one for $H = 0$ and the others for $H \neq 0$. (The cases $\mathcal{J} = 0$ and $\mathcal{J} = 0.25$ are plotted in Figure 3.3.) If we now distort the $\mathcal{J}$-section into a cup-like surface, by taking H as a vertical coordinate and plotting each trajectory at a height depending on its H value, we get half of a closed surface. Each trajectory is selected by an H-plane section. Alteration of the sign of H corresponds to reversal of time. Completing the surface by reflection in the plane $H = 0$ gives the surface generated by rotating the root-cubic graph $\sqrt{-\mathcal{V}(\mathcal{Q})}$ about the $\mathcal{Q}$-axis, i.e., the surface given by (3.3). These surfaces are what Alber et al. [2] call the three-wave surfaces. They foliate the volume contained within the surface for $\mathcal{J} = 0$.

4. The precession of the swing plane.

The characteristic feature of the behavior of the physical spring is its stepwise precession, which we shall now analyze. As the oscillations change from horizontal to vertical and back again, it is observed that each successive horizontal excursion departs in a different direction. The only reference to this phenomenon of which we are aware, prior to Lynch [19], is Cayton [6]. Cayton briefly discussed three-dimensional solutions and mentioned the precession of the swing plane but did not analyze its dynamics. Surprisingly, the characteristic stepwise precession of the swinging spring has been largely ignored, although it is immediately obvious upon observation of a physical elastic pendulum with $\omega_Z \approx 2\omega_R$. Indeed, this precession is almost impossible to suppress experimentally when the initial motion is close to vertical.

4.1. Qualitative description.

If the horizontal projection of the motion is an ellipse of high eccentricity, the motion is approximately planar. We call the vertical plane through the major axis of this ellipse the *swing plane*. When the initial oscillations are quasi-vertical, the motion gradually develops into an essentially horizontal swinging motion. This horizontal swinging does not persist but soon passes again into nearly vertical springing oscillations similar to the initial motion. Subsequently, a horizontal swing again develops but now in a different direction. The stepwise precession of this exchange between springing and swinging motion continues indefinitely in the absence of dissipation and is the characteristic experimental feature of the swinging spring. We shall seek an expression for the change in direction of the swing plane from one horizontal excursion to the next.

A full knowledge of the solutions of the three equations of motion would of course suffice to determine the swing plane at each moment in time. In [19] the equations were expressed in rotating coordinates, and a particular solution for the slow rotation of the swing plane was posited as a function of the vertical amplitude $|C|$ by assuming a certain angular relation. Following this assumption, the angle of the swing plane could be expressed as an integral involving elliptic functions.

4.2. Pattern evocation in shape space.

We shall approach the precession problem using pattern evocation in shape space. Pattern evocation seeks a relative equilibrium (in shape space) in which a phase relationship between the variables (the shape) is preserved [21, 22]. We track the pattern by moving to a nonuniformly rotating frame in which the orientation of the shape is fixed. This is a generalization of the idea of tracking a satellite orbit by evoking constancy of the areal velocity required to conserve angular momentum.

Our particular geometric assumption is that the angle between the complex amplitudes a and b remains constant in an appropriately rotating frame. Writing these amplitudes in vector form as $\mathbf{a} = (|a|\cos\alpha, |a|\sin\alpha, 0)$, $\mathbf{b} = (|b|\cos\beta, |b|\sin\beta, 0)$ and taking $\mathbf{k} = (0,0,1)$ yield

$$J = -\,\mathbf{k}\cdot\mathbf{a}\times\mathbf{b} = |ab|\sin(\alpha-\beta)\,, \qquad \mathbf{a}\cdot\mathbf{b} = |ab|\cos(\alpha-\beta)\,.$$

Consequently, our geometric pattern evocation assumption that the phase difference $\alpha-\beta$ remains constant immediately implies that $|ab|$ is also constant. The conservation of angular momentum J means that the area of the parallelogram formed by the vectors $\mathbf{a}$ and $\mathbf{b}$ is constant. The requirement of constant $\alpha-\beta$ imposes an additional geometric constraint on the possible shape of the orbits. For example, when $\alpha-\beta = \pi/2 \pmod{\pi}$, the orbits are elliptical.

4.3. Modulation equations in rotating coordinates.

We shall transform to rotating coordinates and seek an expression for the (slow) rotation frequency $\Omega(t)$ that allows us to estimate the stepwise precession of the swinging spring by imposing the pattern evocation constraint that $\alpha-\beta$ remains constant.

In a *rotating frame*, the approximate Lagrangian (2.1) at cubic order in the coordinate displacements becomes, with $\mathbf{x} = (x,y,z)$,

$$L = \frac{1}{2}|\dot{\mathbf{x}} + \Omega(t)\,\hat{\mathbf{z}}\times\mathbf{x}|^2 - \frac{1}{2}\left(\omega_R^2(x^2+y^2) + \omega_Z^2 z^2\right) + \frac{1}{2}\lambda(x^2+y^2)z\,. \tag{4.1}$$

Now x, y, and z are Cartesian coordinates centered at the point of equilibrium *in the rotating frame*, $\omega_R = \sqrt{g/\ell}$ is the frequency of linear pendular motion, $\omega_Z = \sqrt{k/m}$ is the frequency of its elastic oscillations, and $\lambda = \ell_0\omega_Z^2/\ell^2$. The corresponding Euler–Lagrange equations of motion (2.2)–(2.4) may be written in rotating coordinates as

$$\ddot{x} - \dot{\Omega}(t)y - 2\Omega(t)\dot{y} + \big(\omega_R^2 - \Omega^2(t)\big)x = \lambda xz\,, \tag{4.2}$$

$$\ddot{y} + \dot{\Omega}(t)x + 2\Omega(t)\dot{x} + \big(\omega_R^2 - \Omega^2(t)\big)y = \lambda yz\,, \tag{4.3}$$

$$\ddot{z} + \omega_Z^2 z = \frac{1}{2}\lambda(x^2+y^2)\,. \tag{4.4}$$

The vertical component of angular momentum is

$$h = \hat{\mathbf{z}}\cdot\mathbf{x}\times\big(\dot{\mathbf{x}} + \Omega(t)\,\hat{\mathbf{z}}\times\mathbf{x}\big) = (x\dot{y} - \dot{x}y) + \Omega(t)(x^2+y^2)$$

and is a constant of the motion for these equations. However, upon Legendre-transforming, one finds that the time-dependent Hamiltonian satisfies

$$\dot{H} = -\,\dot{\Omega}(t)h\,.$$

Thus, perhaps not unexpectedly, exact conservation of energy breaks down, to the extent that the rotation frequency is nonuniform.

4.4. Averaged Lagrangian and modulation equations for slow rotation. The modulation equations in axes rotating with angular velocity $\Omega(t)$ about the vertical are obtained in the resonant case $\omega_Z = 2\omega_R$ by applying the averaged Lagrangian technique [30]. Accordingly, the solution of (4.2)–(4.4) is assumed to be of the form

$$x = \Re[a(t)\exp(i\omega_R t)]\,, \tag{4.5}$$

$$y = \Re[b(t)\exp(i\omega_R t)]\,, \tag{4.6}$$

$$z = \Re[c(t)\exp(2i\omega_R t)]\ . \tag{4.7}$$

(Note that subscript zeros are dropped for these modulation amplitudes in the rotating frame.) In these variables, the averaged Lagrangian corresponding to (4.1) may be written as

$$\begin{aligned}\langle L\rangle = \frac{1}{2}\omega_R[\Im\{\dot a a^* + \dot b b^* + 2\dot c c^*\} + \Re\{\kappa(a^2+b^2)c^*\} + 2\Omega\Im\{ab^*\}]\\ + \frac{1}{2}\Omega\,\Re\,[a^*\dot b - \dot a^* b] + \frac{1}{4}\Omega^2\Big[|a|^2+|b|^2\Big]\,.\end{aligned} \tag{4.8}$$

On assuming that the rotation frequency is sufficiently slow that $\Omega/\omega_R \ll 1$, we shall *neglect* all terms in the averaged Lagrangian (4.8) that are not multiplied by ω_R. In this approximation of slow rotation, the averaged Lagrangian is given by the simpler expression,

$$\langle L\rangle = \frac{1}{2}\omega_R[\Im\{\dot a a^* + \dot b b^* + 2\dot c c^*\} + \Re\{\kappa(a^2+b^2)c^*\} + 2\Omega J]\,. \tag{4.9}$$

Here $J = \Im\{ab^*\}$ is the angular momentum, a conserved quantity at this level of approximation and formally identical to the expression in nonrotating coordinates. The Euler–Lagrange modulation equations in this approximation may be written as

$$i\dot a = \kappa a^* c + i\Omega b\,, \tag{4.10}$$

$$i\dot b = \kappa b^* c - i\Omega a\,, \tag{4.11}$$

$$i\dot c = \frac{1}{4}\kappa(a^2+b^2)\,. \tag{4.12}$$

We may also write these leading order equations in Hamiltonian form. When $\langle H\rangle$ is defined by

$$\langle H\rangle = \Re\{\kappa(a^2+b^2)c^*\} + 2\Omega\Im\{ab^*\}\,,$$

with coordinates (a,b,c), conjugate momenta $(a^*,b^*,2c^*)$, and Poisson brackets defined by $\{a,a^*\} = \{b,b^*\} = 2\{c,c^*\} = -i$, the modulation equations (4.10)–(4.12) are expressible in canonical Hamiltonian form as

$$i\dot a = i\{a,\langle H\rangle\} = \frac{\partial\langle H\rangle}{\partial a^*}\,,\quad i\dot b = i\{b,\langle H\rangle\} = \frac{\partial\langle H\rangle}{\partial b^*}\,,\quad i\dot c = i\{c,\langle H\rangle\} = \frac{\partial\langle H\rangle}{\partial\, 2c^*}\,.$$

The three constants of the motion for these equations are

$$H_0 = \frac{1}{2}\kappa[(a^2+b^2)c^* + (a^{*2}+b^{*2})c] = \Re\{\kappa(a^2+b^2)c^*\}\,, \tag{4.13}$$

$$N = |a|^2 + |b|^2 + 4|c|^2\,, \tag{4.14}$$

$$J = (ab^* - a^*b)/2i = \Im\{ab^*\}\,. \tag{4.15}$$

We now introduce the pattern evocation assumption that $\alpha-\beta$ is constant, which also implies that $|ab|$ is constant. Noting that

$$|ab|^2 = (\Re\{ab^*\})^2 + (\Im\{ab^*\})^2$$

and that the second term is just J^2 implies constancy of $\Re\{ab^*\}$. Using (4.10) and (4.11), it follows that

$$\frac{d}{dt}|ab|^2 = -2\Re\{ab^*\}\left[2\kappa\Im\{abc^*\} + \Omega(|a|^2-|b|^2)\right] = 0\,. \tag{4.16}$$

Either factor may vanish so there appears to be two possibilities for the solution. We first assume that the factor in square brackets in (4.16) vanishes. This implies

$$\Omega = -\frac{2\kappa\Im\{abc^*\}}{|a|^2-|b|^2} = -\frac{2\kappa|abc|\sin(\alpha+\beta-\gamma)}{|a|^2-|b|^2} \tag{4.17}$$

(where $c=|c|e^{i\gamma}$). The precession angle $\Theta=\int_0^t\Omega(t')dt'$ can be ascertained by integrating Ω over the time interval of the motion. In the special case when $\alpha-\beta=\frac{\pi}{2}(\operatorname{mod}\pi)$, one finds by using the constants of motion that

$$\Omega = \frac{2\kappa J H_0}{(N-4|c|^2)^2-4J^2}\,. \tag{4.18}$$

This case also corresponds to the vanishing of the first factor in (4.16) so that $\Re\{ab^*\}=0$ and $\mathbf{a}$ and $\mathbf{b}$ are 90° out of phase. This was the Ansatz introduced by Lynch [19]. He showed that, in this case, the rotation rate is given by (4.18). We now see that the result in [19] is a special case of the general result (4.17). In this special case, Ω can be computed as soon as $|c|$ is known. We will examine this case numerically below.

4.5. The instantaneous ellipse.

In order to define precisely the precession angle, we introduce an ellipse which approximates the horizontal projection of the trajectory of the pendulum. Recall that the full solution for the horizontal components is

$$x=\Re\{a\exp(i\omega_R t)\}=|a|\cos(\omega_R t+\alpha)\,,\qquad y=\Re\{b\exp(i\omega_R t)\}=|b|\cos(\omega_R t+\beta)\,,$$

where α and β are the phases of a and b. The amplitudes and phases are assumed to vary slowly. If they are regarded as constant over a period $\tau=1/\omega_R$ of the fast motion, these equations describe a central ellipse,

$$Px^2+2Qxy+Ry^2=S, \tag{4.19}$$

where $P=|b|^2$, $Q=-|ab|\cos(\alpha-\beta)$, $R=|a|^2$, and $S=J^2$. The area of the ellipse is easily calculated and is found to have the constant value πJ. Its orientation is determined by eliminating the cross-term in (4.19). This is achieved as usual by rotating the axes through an angle θ, given by

$$\tan 2\theta = \frac{2Q}{P-R} = \frac{2|ab|\cos(\alpha-\beta)}{|a|^2-|b|^2}\,. \tag{4.20}$$

The semiaxes of the ellipse are given by

$$A_1 = \frac{J}{\sqrt{P\cos^2\theta + Q\sin 2\theta + R\sin^2\theta}}\,, \quad A_2 = \frac{J}{\sqrt{P\sin^2\theta - Q\sin 2\theta + R\cos^2\theta}}\,. \tag{4.21}$$

The area is $\pi A_1 A_2 = \pi J$, and the eccentricity can be calculated immediately. In the case of unmodulated motion, such as the elliptic-parabolic modes described in [19], the instantaneous ellipse corresponds to the trajectory, which is a precessing ellipse. In general, it is only an approximation to the trajectory, but we may define the orientation or azimuth at any time to be the angle θ given by (4.20). This angle will be compared to the precession angle Θ calculated by integrating (4.18) and shown to give almost identical results.

5. Numerical results.

We examine the results of numerical integrations of the modulation equations (2.8)–(2.10) and compare them to the solutions of the exact equations (2.2)–(2.4). It will be seen that the modulation equations provide an excellent description of the envelope of the rapidly varying solution of the full equations. We then compare the stepwise precession angle predicted by a formula based on constancy of the angle $\alpha - \beta$ with the numerical simulation of this quantity and show that the two values track each other essentially exactly.

The parameter values chosen for all numerical integrations are $m = 1\,\mathrm{kg}$, $\ell = 1\,\mathrm{m}$, $g = \pi^2\,\mathrm{m\,s^{-2}}$, and $k = 4\pi^2\,\mathrm{kg\,s^{-2}}$ so that $\omega_R = \pi$, $\omega_Z = 2\pi$, and the resonance condition $\omega_Z = 2\omega_R$ holds. The linear rotational mode has period $\tau_R = 2\,\mathrm{s}$, and the vertical mode has period $\tau_Z = 1\,\mathrm{s}$. The initial conditions are set as follows:

$$(x_0, y_0, z_0) = (0.006, 0, 0.012), \qquad (\dot{x}_0, \dot{y}_0, \dot{z}_0) = (0, 0.00489, 0)\,.$$

(The value of $\dot{y}_0$ was chosen to tune the precession angle to be an even fraction of 180°, making the amplitudes, though not the phases, periodic.) The corresponding initial values for the modulation equations (2.8)–(2.10) are given by

$$\alpha_0 = \arctan\left(\frac{-\dot{x}_0}{\omega_R x_0}\right)\,, \quad \beta_0 = \arctan\left(\frac{-\dot{y}_0}{\omega_R y_0}\right)\,, \quad \gamma_0 = \arctan\left(\frac{-\dot{z}_0}{2\omega_R z_0}\right)\,,$$

$$|a_0| = \left(\frac{x_0}{\cos\alpha_0}\right)\,, \quad |b_0| = -\left(\frac{\dot{y}_0}{\omega_R \sin\beta_0}\right)\,, \quad |c_0| = \left(\frac{z_0}{\cos\gamma_0}\right)\,,$$

giving the values $(|a_0|, |b_0|, |c_0|) = (0.006, 0.002, 0.012)$ and $(\alpha_0, \beta_0, \gamma_0) = (0, -\pi/2, 0)$. The constants of the motion take the following values:

$$H = 4.03 \times 10^{-7}\,, \qquad J = 9.34 \times 10^{-6}\,, \qquad N = 6.14 \times 10^{-4}\,.$$

The integration was extended over a period of 1000 seconds (i.e., 1000 vertical oscillations). As a check on numerical accuracy, the changes in these quantities, which should remain constant, were calculated, with the following results:

$$\left(\frac{H_{\mathrm{Final}}}{H_{\mathrm{Initial}}}\right) = 100.04\%\,, \quad \left(\frac{J_{\mathrm{Final}}}{J_{\mathrm{Initial}}}\right) = 99.997\%\,, \quad \left(\frac{N_{\mathrm{Final}}}{N_{\mathrm{Initial}}}\right) = 100.00\%\,.$$

We now directly compare the solutions of the "exact" equations (2.2)–(2.4) and the "approximate" or modulation equations (2.8)–(2.10). Once the modulation equations have been

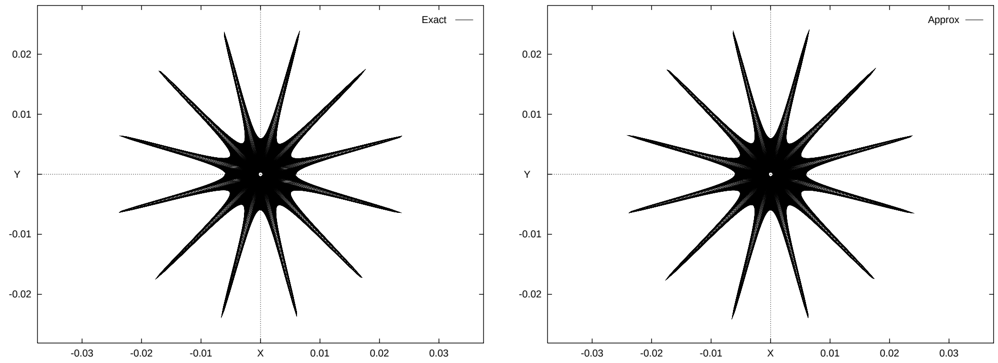


Figure 5.1. *Horizontal projection of the solution for an integration of* 1000 *seconds. Left: Solution of the "exact" equations. Right: Solution of the "approximate" equations.*

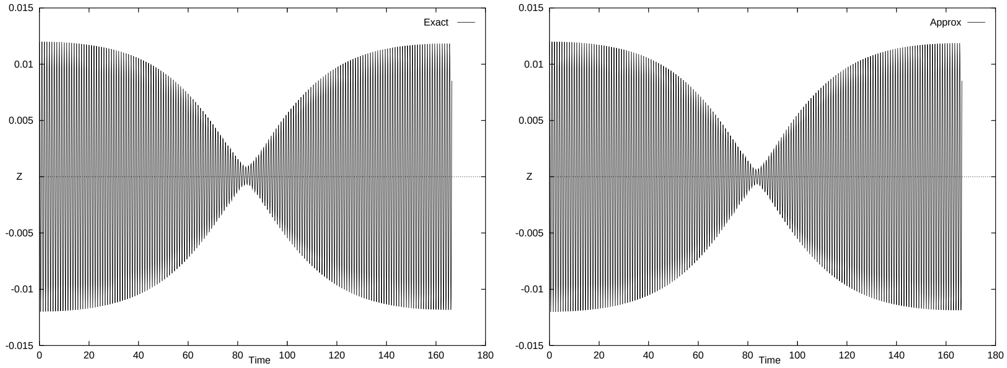


Figure 5.2. *Vertical amplitude of the solution for the first modulation cycle (first* 167 *seconds). Left: Solution of the "exact" equations. Right: Solution of the "approximate" equations.*

solved for the envelope amplitudes and phases, the full approximate solution is given by (2.5)–(2.7). We first consider the horizontal projection of the solution for the 1000-second integration. This is the period required for the solution to precess through approximately 180°. In Figure 5.1 (left panel) we plot x versus y for the exact solution. In Figure 5.1 (right panel) we plot the corresponding solution from the modulation equations. It is clear that there is great similarity between the two solutions; indeed, the two plots are indistinguishable. The precession angle between horizontal excursions is close to 30°. (The value of $\dot{y}_0$ was chosen to ensure this.) The modulation period is approximately 167 seconds; thus the instantaneous ellipse rotates through six cycles and 180° in 1000 seconds.

The vertical structure of the solution is displayed in Figure 5.2, where z for the exact solution (left panel) and $\Re\{c_0(t)\exp(2i\omega_R t)\}$ for the approximate solution (right panel) are seen to be virtually identical. For clarity, the solutions are plotted only for the first modulation cycle of 167 seconds. The character of the solution—rapid oscillations with a slowly varying amplitude envelope—is clear from the figure. The vertical amplitude is close to zero when horizontal excursions are at their peak. This is confirmed in Figure 5.3 (left panel), where the horizontal modulation amplitude $S = \sqrt{|a|^2 + |b|^2}$ and vertical modulation amplitude $C = |c|$ are plotted against time.

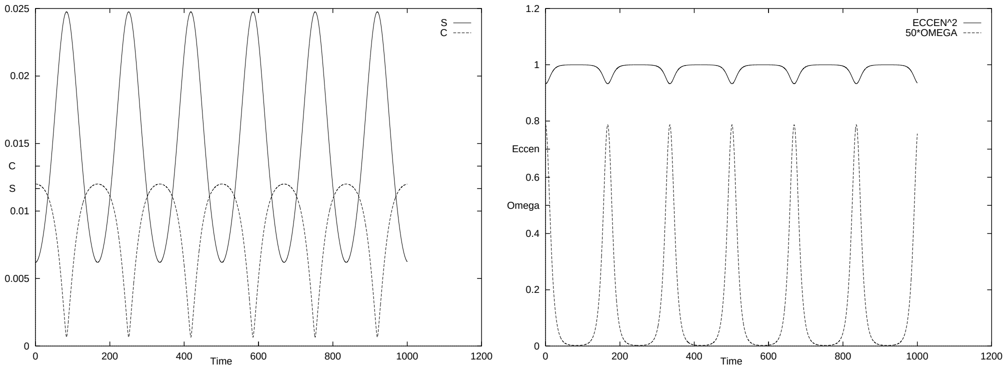


Figure 5.3. *Left panel: Envelope amplitude of the approximate solution.* $S = \sqrt{|a|^2 + |b|^2}$ *(solid line) and* $C = |c|$ *(dashed line). Right panel: Square of the eccentricity (solid line) and angular velocity* Ω *(scaled by* 50*) of the instantaneous ellipse (dashed line).*

In Figure 5.3 (right panel) we plot the squared eccentricity $e^2 = (1 - A_{\text{min}}^2 / A_{\text{maj}}^2)$ of the envelope of the horizontal projection of the approximate solution, where the semiaxes A_{maj} and A_{min} are calculated from (4.21). The eccentricity is close to unity for most of the integration. Horizontal excursions of the pendulum occur during this time. For short periods, when the horizontal amplitude is minimum, the value of e drops significantly (solid line). During this time, the angular velocity, calculated as the rate of change of the azimuth given by (4.20), reaches a maximum (dashed line). Thus the precession occurs in bursts near the times when the vertical amplitude is maximum and the horizontal amplitude is minimum.

The stepwise nature of the precession is clearly illustrated in Figure 5.4. The azimuthal angle ϑ of the numerical solution of the exact equations may be calculated by fitting a central conic to every three consecutive points on the trajectory. Assuming a solution of the form

$$\tilde{P}x^2 + 2\tilde{Q}xy + \tilde{R}y^2 = 1$$

and requiring that the three points lie on this curve, we obtain three equations for the coefficients $(\tilde{P}, \tilde{Q}, \tilde{R})$. From these, the azimuth ϑ and the semiaxes are obtained from equations analogous to (4.20) and (4.21). This is compared in Figure 5.4 to the corresponding value θ resulting from integration of the modulation equations. It is noteworthy that ϑ and θ remain quasi-constant for most of the modulation cycle, changing rapidly only over short intervals around the times when C is maximum and S is minimum. The advances in phase are very similar for the exact and approximate solutions. However, there are small differences: $\theta - \vartheta$ is also plotted in Figure 5.4 (dotted line). This sensitive quantity reaches its maximum value of 4.35° at the end of the integration.

Comparing Figures 5.3 (left panel) and 5.4, it is clear that the azimuthal angle remains close to a constant value during horizontal excursions (when S is large) and changes rapidly when the vertical oscillation amplitude is close to a maximum and the energy of the horizontal component is small. Thus the stepwise precession takes place in sudden bursts when the motion of the system is quasi-vertical. The variations of the azimuth appear to occur symmetrically about the times of vertical energy maxima.

The azimuthal change between successive horizontal excursions is very close to 30° for both

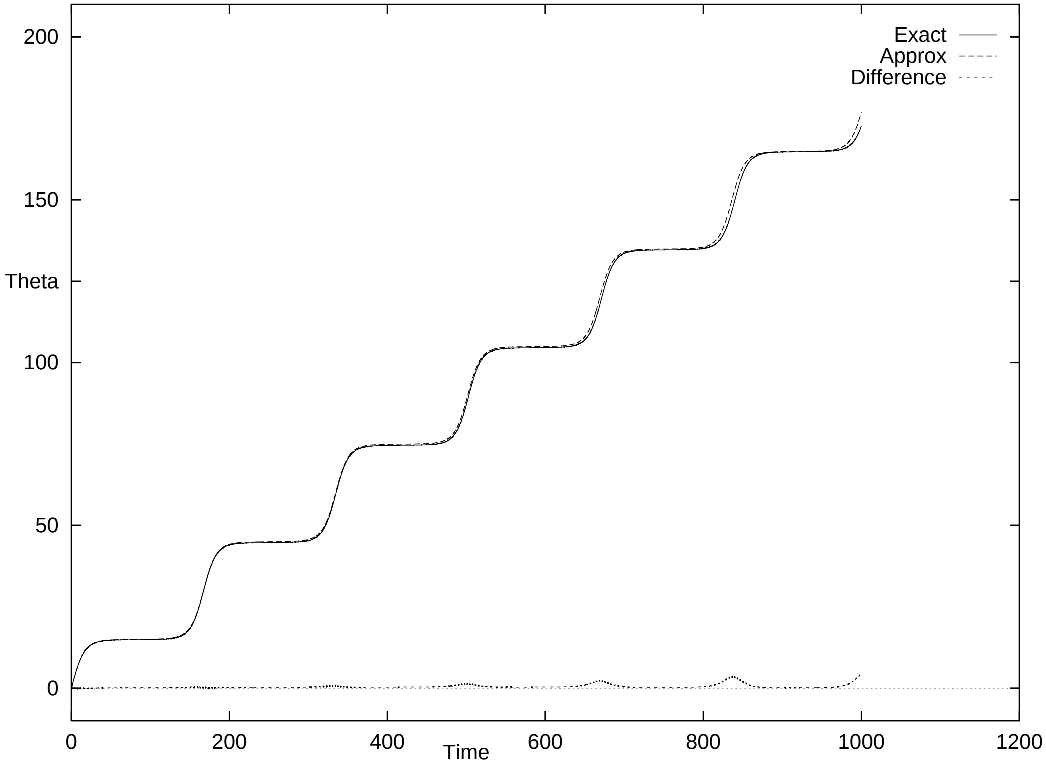


Figure 5.4. *Azimuth angle (in degrees) for the "exact" solution (ϑ, solid line) and the "approximate" solution (θ, dashed line). The difference $\theta - \vartheta$ is plotted as a dotted line. The azimuth Θ resulting from integration of* (4.18) *(not plotted) is indistinguishable from the values θ of the approximate solution.*

exact and approximate solutions. We also calculated the angle Θ resulting from an integration of (4.18). The graphs of θ and Θ (not plotted) are indistinguishable. The maximum difference $|\theta - \Theta|$ was only 0.0063°. This is remarkable: the value Θ derived from (4.18) involves an assumption that $\alpha - \beta$ is constant in a particular rotating frame. The azimuth θ from the modulation equations makes no such assumption, yet the two solutions are practically identical. This confirms that the pattern evocation assumption which yields the result (4.18) is sound.

Numerous other integrations of the exact and modulation equations were also carried out. They confirm that the stepwise precession of the azimuthal angle is a distinct characteristic of the swinging spring. This is also in complete agreement with simple experiments with a physical pendulum, where the periodic exchange of energy between horizontal and vertical and the precession of the swing plane between horizontal excursions are the main observable properties of the motion.

The perturbation methods used in this study are valid only for small values of the system energy. However, the numerical solution of the exact equations gives insight into the dynamics for larger amplitudes. Although we do not consider chaotic motion here, the transition from regular to chaotic motion of the planar elastic pendulum has been studied elsewhere (see [20] and references therein).

Acknowledgments. The authors would like to acknowledge the facilities provided by the Isaac Newton Institute at Cambridge University, where this work began. Peter Lynch would like to thank the Center for Nonlinear Studies, Los Alamos National Laboratory, for support and hospitality. Darryl Holm is grateful to Gregor Kovačič and Jerry Marsden for many fruitful discussions and ongoing collaborations in dynamical systems analysis.

REFERENCES

[1] A. B. Aceves, D. D. Holm, G. Kovačič, and I. Timofeyev, *Homoclinic orbits and chaos in a second-harmonic generating optical cavity*, Phys. Lett. A, 233 (1997), pp. 203–208.

[2] M. S. Alber, G. G. Luther, J. E. Marsden, and J. M. Robbins, *Geometric phases, reduction and Lie-Poisson structure for the resonant three-wave interaction*, Phys. D, 123 (1998), pp. 271–290.

[3] M. S. Alber, G. G. Luther, J. E. Marsden, and J. M. Robbins, *Geometry and control of three-wave interactions*, in The Arnoldfest (Toronto, ON, 1997), Fields Inst. Commun. 24, AMS, Providence, RI, 1999, pp. 55–80.

[4] B. A. Aničin, D. M. Davidović, and V. M. Babović, *On the linear theory of the elastic pendulum*, European J. Phys., 14 (1993), pp. 132–135.

[5] F. P. Bretherton, *Resonant interactions between waves: The case of discrete oscillations*, J. Fluid Mech., 20 (1964), pp. 457–479.

[6] Th. E. Cayton, *The laboratory spring-mass oscillator: An example of parametric instability*, Amer. J. Phys., 45 (1977), pp. 723–732.

[7] D. David and D. D. Holm, *Multiple Lie-Poisson structures, reductions, and geometric phases for the Maxwell-Bloch traveling-wave equations*, J. Nonlinear Sci., 2 (1992), pp. 241–262.

[8] D. David, D. D. Holm, and M. Tratnick, *Hamiltonian chaos in nonlinear optical polarization dynamics*, Phys. Rep., 187 (1990), pp. 281–367.

[9] I. T. Georgiou, *On the global geometric structure of the dynamics of the elastic pendulum*, Nonlinear Dynam., 18 (1999), pp. 51–68.

[10] A. Hasegawa and K. Mima, *Pseudo-three-dimensional turbulence in magnetized nonuniform plasmas*, Phys. Fluids, 21 (1977), pp. 87–92.

[11] D. D. Holm and G. Kovačič, *Homoclinic chaos in a laser-matter system*, Phys. D, 56 (1992), pp. 270–300.

[12] D. D. Holm, G. Kovačič, and T. A. Wettergren, *Near integrability and chaos in a resonant-cavity laser model*, Phys. Lett. A, 200 (1995), pp. 299–307.

[13] D. D. Holm, G. Kovačič, and T. A. Wettergren, *Homoclinic orbits in the Maxwell-Bloch equations with a probe*, Phys. Rev. E (3), 54 (1996), pp. 243–256.

[14] D. D. Holm and J. E. Marsden, *The rotor and the pendulum*, in Symplectic Geometry and Mathematical Physics, P. Donato, C. Duval, J. Elhadad, and G. M. Tuynman, eds., Progr. Math. 99, Birkhäuser Boston, Boston, 1991, pp. 189–203.

[15] W. Horton and A. Hasegawa, *Quasi-two-dimensional dynamics of plasmas and fluids*, Chaos, 4 (1994), pp. 227–251.

[16] M. S. Longuet-Higgins and A. E. Gill, *Resonant interactions between planetary waves*, Proc. Roy. Soc. Edinburgh Sect. A., 299 (1967), pp. 120–140.

[17] E. N. Lorenz, *Deterministic non-periodic flow*, J. Atmospheric Sci., 20 (1963), pp. 130–141.

[18] G. G. Luther, M. S. Alber, J. E. Marsden, and J. M. Robbins, *Geometric analysis of optical frequency conversion and its control in quadratic nonlinear media*, J. Opt. Soc. Amer. B Opt. Phys., 17 (2000), pp. 932–941.

[19] P. Lynch, *Resonant motions of the three-dimensional elastic pendulum*, Internat. J. Non-Linear Mech., 37 (2002), pp. 345–367.

[20] P. Lynch, *The swinging spring: A simple model for atmospheric balance*, in Large-Scale Atmosphere-Ocean Dynamics: Vol II: Geometric Methods and Models, Cambridge University Press, Cambridge, UK, 2002, pp. 64–108.

[21] J. E. Marsden, J. Scheurle, and J. M. Wendlandt, *Visualization of orbits and pattern evocation for the double spherical pendulum*, in ICIAM 95 (Hamburg, 1995), Math. Res. 87, Akademie Verlag, Berlin, 1996, pp. 213–232.

[22] J. E. Marsden and J. Scheurle, *Pattern evocation and geometric phases in mechanical systems with symmetry*, Dynam. Stability Systems, 10 (1995), pp. 315–338.

[23] R. Montgomery, *How much does the rigid body rotate? A Berry's phase from the* 18*th century*, Amer. J. Phys., 59 (1991), pp. 394–398.

[24] E. Ott, *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge, UK, 1993, p. 385.

[25] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer-Verlag, New York, 1987, p. 710.

[26] R. D. Peters, *Chaotic motion from support constraints of a nondriven rigid spherical pendulum*, Phys. Rev. A (3), 38 (1988), pp. 5352–5359.

[27] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors*, Springer-Verlag, New York, 1982, p. 269.

[28] A. Vitt and G. Gorelik, *Kolebaniya uprugogo mayatnika kak primer kolebaniy dvukh parametricheski svyazannykh linejnykh sistem*, Zh. Tekh. Fiz. (J. Tech. Phys.), 3 (1933), pp. 294–307. Available in English translation: *Oscillations of an Elastic Pendulum as an Example of the Oscillations of Two Parametrically Coupled Linear Systems*, Translated by Lisa Shields with an Introduction by Peter Lynch, Historical Note 3, Met Éireann, Dublin, 1999.

[29] J.-M. Wersinger, J. M. Finn, and E. Ott, *Bifurcations and strange behavior in instability saturation by nonlinear mode coupling*, Phys. Rev. Lett., 44 (1980), pp. 453–456.

[30] G. B. Whitham, *Linear and Nonlinear Waves*, John Wiley and Sons, New York, 1974.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 1, pp. 65–104

Homoclinic Stripe Patterns[*]

Arjen Doelman[†] and Harmen van der Ploeg[†]

Abstract. In this paper, we study homoclinic stripe patterns in the two-dimensional generalized Gierer–Meinhardt equation, where we interpret this equation as a prototypical representative of a class of singularly perturbed monostable reaction-diffusion equations. The structure of a stripe pattern is essentially one-dimensional; therefore, we can use results from the literature to establish the existence of the homoclinic patterns. However, we extend these results to a maximal domain in the parameter space and establish the existence of a bifurcation that forms a new upper bound on this domain. Beyond this bifurcation, the Gierer–Meinhardt equation exhibits self-replicating pulse, respectively, stripe patterns in one, respectively, two dimension(s). The structure of the self-replication process is very similar to that in the Gray–Scott equation.

We investigate the stability of the homoclinic stripe patterns by an Evans function analysis of the associated linear eigenvalue problem. We extend the recently developed nonlocal eigenvalue problem (NLEP) approach to two-dimensional systems. Except for a region near the upper bound of the domain of existence in parameter space, this method enables us to get explicit information on the spectrum of the linear problem. We prove that, in this subregion, all homoclinic stripe patterns must be unstable as solutions on $\mathbf{R}^2$. However, stripe patterns can be stable on domains of the type $\mathbf{R} \times (0, L_y)$. Our analysis enables us to determine an upper bound on L_y; moreover, the analysis indicates that stripe patterns can become stable on $\mathbf{R^2}$ near the upper bound of the existence domain. This is confirmed numerically: it is shown by careful simulations that there can be stable homoclinic stripe patterns on $\mathbf{R}^2$ for parameter values near the self-replication bifurcation.

Key words. reaction-diffusion equations, pattern formation, stability analysis, Evans function, singular perturbation theory

AMS subject classifications. 35K57, 35B25, 35B32, 35B35, 35B40, 34C37, 92C15

PII. S1111111101392831

1. Introduction.

Stripe patterns can be observed in many (bio)chemical reactions and appear frequently in numerical simulations of reaction-diffusion equations. Moreover, these patterns are quite robust in the sense that they exist for large "open" sets of parameter combinations (see the review [5] and the references therein). However, the mathematical theory of stripe patterns in reaction-diffusion systems is largely restricted to systems at near-critical conditions, which means that the system parameter values are close to a Turing bifurcation. Near such a bifurcation, the (Turing) patterns necessarily are of small amplitude, where the smallness is related to the distance to the Turing bifurcation in parameter space. Under these "weakly nonlinear" conditions, the patterns generated by the system can be analyzed by a normal form or a Ginzburg–Landau approach. (See [22] for a (formal) application of this approach to reaction-diffusion systems and [23] for a survey of the general mathematical

[*] Received by the editors July 26, 2001; accepted (in revised form) by B. Sandstede December 11, 2001; published electronically May 8, 2002. This research was supported by the Dutch Science Foundation NWO.

http://www.siam.org/journals/siads/1-1/39283.html

[†] Korteweg-de Vries Instituut, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands (doelman@science.uva.nl, hvdploeg@science.uva.nl).

theory.)

In this paper, we study two-dimensional stripe patterns in the strongly nonlinear regime; i.e., we consider systems that are not close to a Turing bifurcation (Remark 1.4). As a consequence, the amplitudes of the solutions cannot be assumed to be small. In this regime, there is no equivalent of the general Ginzburg–Landau theory. However, when one restricts oneself to two-component reaction-diffusion systems and assumes that the ratio of the two diffusion constants is small, then one can use singular perturbation theory to study the existence, stability, and dynamics of patterns "far from equilibrium"; see, for instance, [35], [29], [12], [28], [20], [39], [10].

Most of these recent papers on pattern formation in singularly perturbed reaction-diffusion equations consider spike, pulse, or spot patterns (see Remark 1.1). Similar to these patterns, a homoclinic stripe pattern is isolated in the sense that both components $U(x, y, t)$ and $V(x, y, t)$ are close to a trivial homogeneous background state outside a neighborhood of the stripe (see Figure 1.1). A Turing bifurcation imposes a spatial periodicity on the pattern; hence, in general, there cannot be homoclinic patterns at near-critical conditions. The stripe patterns we consider are assumed to be stationary, linear (or straight), and essentially one-dimensional. Since reaction-diffusion equations in two space dimensions are invariant with respect to rotations in the plane, we can define x as the direction perpendicular to the stripe and y as the coordinate along the stripe. The assumption that the homoclinic stripe pattern is "essentially one-dimensional" implies that the stripes have no structure in the y-direction; i.e., the stripe patterns are of the form $(U_{\text{stripe}}(x, y, t), V_{\text{stripe}}(x, y, t)) = (U_0(x), V_0(x))$. As a consequence, the stripe patterns correspond to homoclinic pulse solutions as function of the variable perpendicular to the stripe, x. This enables us to refer for the existence to the literature on stationary homoclinic pulse solutions of systems in one spatial variable. (We will, in particular, use [10].)

We study the existence, stability, and bifurcations of homoclinic stripe patterns in the generalized Gierer–Meinhardt equation:

$$\left\{ \begin{array}{rclll} U_t & = & \Delta U & - \mu U & + U^{\alpha_1} V^{\beta_1}, \\ V_t & = & \varepsilon^2 \Delta V & - V & + U^{\alpha_2} V^{\beta_2} \end{array} \right. \tag{1.1}$$

for $(x, y) \in \mathbf{R}^2$ [20], [28], [29], [19], [39], where we assume that the ratio of the diffusion coefficients of the two "species" V and U, d_V, and d_U is asymptotically small: $\varepsilon^2 = d_V/d_U \ll 1$. Throughout this paper, the parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ and μ are assumed to satisfy

$$\alpha_1 > 1 + \frac{\alpha_2 \beta_1}{\beta_2 - 1}, \quad \alpha_2 < 0,\ \beta_1 > 1,\ \beta_2 > 1,\ \mu > 0 \tag{1.2}$$

(compare to [20], [28], [29], [19], [39]). This model can be regarded as the leading order part of a "normal form" that appears from a scaling analysis in a large class of singularly perturbed reaction-diffusion equations (see Remark 1.2). Equation (1.1) can also be seen as a generalization of the original model introduced by Gierer and Meinhardt [15] in the context of morphogenesis. The special case $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = 2, \beta_2 = 2$ in (1.1) corresponds to the original (biological) values of the parameters in [15].

Localized solutions of a singularly perturbed equation as (1.1) will, in general, have asymptotically large amplitudes. Following [19], [10], and [20], we therefore introduce the $\mathcal{O}(1)$

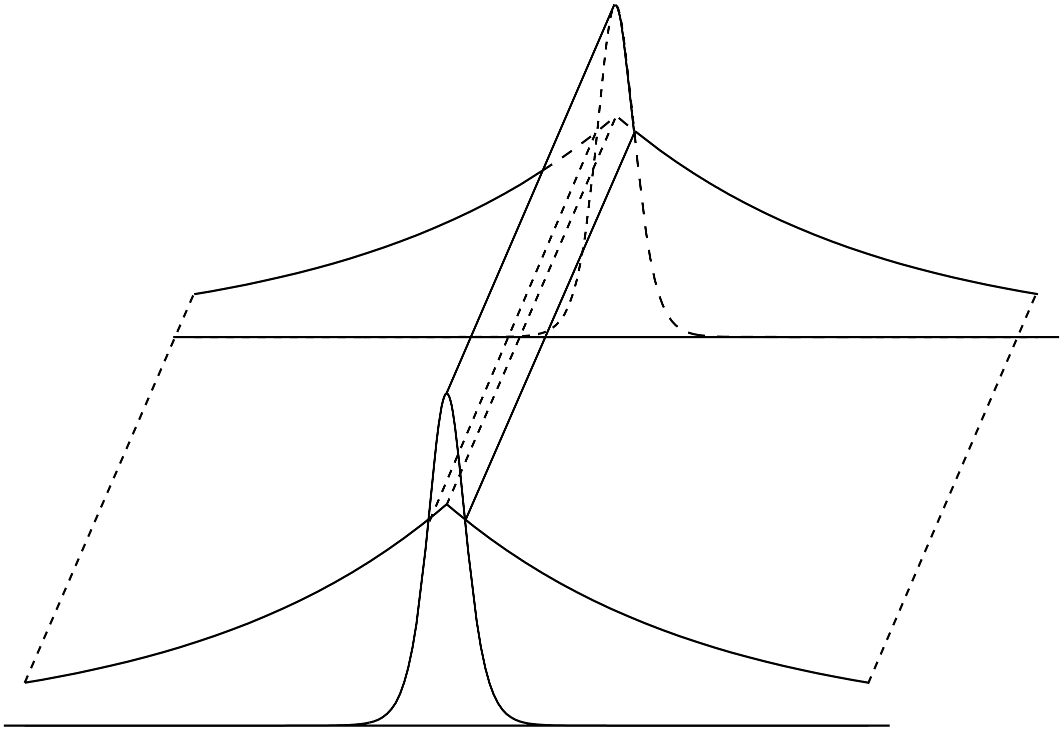


Figure 1.1. *The homoclinic stripe pattern. The V-component is strongly localized, and the U-component decays to the limit state $U \equiv 0$ on a long spatial scale.*

functions $\tilde{U}(x,t)$ and $\tilde{V}(x,t)$ and rescale x, y, and ε:

$$\begin{array}{c}\tilde{U}(x,t) = \varepsilon^r U(x,t),\ r = \dfrac{\beta_2 - 1}{D} > 0,\ \tilde{V}(x,t) = \varepsilon^s V(x,t),\ s = -\dfrac{\alpha_2}{D} > 0, \\ D = (\alpha_1 - 1)(\beta_2 - 1) - \alpha_2\beta_1 > 0,\ x = \sqrt{\varepsilon}\tilde{x},\ y = \sqrt{\varepsilon}\tilde{y},\ \tilde{\varepsilon} = \sqrt{\varepsilon}.\end{array} \tag{1.3}$$

Equation (1.1) can thus be written as

$$\left\{\begin{array}{rclll} \varepsilon^2 U_t & = & \Delta U & -\varepsilon^2\mu U & +\ U^{\alpha_1}V^{\beta_1}, \\ V_t & = & \varepsilon^2\Delta V & -V & +\ U^{\alpha_2}V^{\beta_2}. \end{array}\right. \tag{1.4}$$

This equation is equivalent to (1.1), but it has the advantage that the stripe patterns in (1.1) have an $\mathcal{O}(1)$ amplitude (with respect to ε) as solution of (1.4); therefore, we will consider the Gierer–Meinhardt equation in the form of (1.4) in this paper.

It is shown in [10] that the one-dimensional Gierer–Meinhardt equation, (1.4) without y-dependence, has stationary, homoclinic pulse solutions for parameters satisfying (1.2) and $\mu = \mathcal{O}(1)$. This result establishes the existence of homoclinic stripe patterns to (1.4); see Theorem 2.1. By a careful re-examination of the construction of the one-dimensional homoclinic pulses, we are able to determine analytically an upper bound in μ on the existence domain of the stripe patterns: we establish that homoclinic stripe patterns exist in (1.4) up to $\mu = \mu_{\text{split}} = \mathcal{O}(1/\varepsilon^4)$ and not beyond this value (Theorem 2.2); see also Remark 1.3. This bifurcation is, in essence, a bifurcation of the one-dimensional problem. The nonexistence result is similar to the proof of the existence of a "disappearance bifurcation" in the one-dimensional Gray–Scott model in [8], [7]. Note that numerical continuations and simulations in [31] for the Gray–Scott system indicate that this "disappearance bifurcation" is, in fact, a saddle-node bifurcation of homoclinic orbits. It is natural to expect that the same is true for the "disappearance" or "splitting" bifurcations in the systems studied in this paper. However, as in the Gray–Scott case, the bifurcation takes place in the region in parameter space, where the existence problem is no longer singularly perturbed. When $\mu = \mathcal{O}(1/\varepsilon^m) \gg 1$, i.e., $m > 0$, one needs to rescale (1.4) since U and V can no longer be assumed to be $\mathcal{O}(1)$ (see section 2). The rescaled system is singularly perturbed in the scaled parameter $\tilde{\varepsilon} = \mathcal{O}(\varepsilon^{1-m/4})$ so that one can no longer use the ideas of geometric singular perturbation theory [21] for $m = 4$. As a consequence, the analytic "control" of the homoclinic orbits decreases significantly. Therefore, the identification of the "disappearance bifurcation" as a saddle-node bifurcation of homoclinic orbits has become a challenging task.

It was shown in [8], [7] that the "disappearance bifurcation" initiates the well-known pulse splitting, or self-replication, process in the Gray–Scott model [33], [35], [34], [12], [31]. Exactly the same behavior can be observed in numerical simulations of the one-dimensional Gierer–Meinhardt model (1.4) for $\mu > \mu_{\text{split}}$; see section 2. This implies that the self-replicating process is a "generic phenomenon" in singularly perturbed reaction-diffusion equations and that it thus is not special to the Gray–Scott equation model (see Remarks 1.2 and 2.3). Increasing μ through μ_{split} induces a *stripe splitting bifurcation* in the two-dimensional system (1.4). The end-product of the self-replication process is a spatially periodic stripe pattern; see Figure 5.3.

This splitting bifurcation is related to the existence problem of the stripe patterns. Other bifurcations, such as the bifurcations from stripes to spots, are associated to the (in)stability of the stripe pattern $(U_{\text{stripe}}(x,y,t), V_{\text{stripe}}(x,y,t)) = (U_0(x), V_0(x))$. The stability of the two-dimensional stripes again relies heavily on insights in the stability of one-dimensional homoclinic pulse patterns. Since we assume that (1.4) is defined on the unbounded plane, i.e., $(x,y) \in \mathbf{R}^2$, it follows that the linearized stability problem reduces to the study of a one-parameter family of eigenvalue problems in the one-dimensional variable x. This family is parametrized by a wave number in the y-direction, l (see section 3). We show that these eigenvalue problems can be studied by the recently developed extension of the Evans function method, the so-called nonlocal eigenvalue problem (NLEP) approach [8], [9], [10]. This method enables us to determine the spectrum of the linear stability problem associated to the stripe pattern *explicitly* as function of l.

We again find that the magnitude of μ with respect to ε is crucial for the stability of the stripes; therefore, we again introduce m and set $\mu = \mathcal{O}(1/\varepsilon^m)$. We show in section 3 that the eigenvalues $\lambda(l)$ all are stable and real; i.e., $\lambda(l) < 0$ for $|l| > l_{\text{R,stab}} = \sqrt{(\beta_2+1)^2/4 - 1} = \mathcal{O}(1)$ and $\lambda(l) = \lambda(0)$, at leading order, for $|l| \ll \varepsilon^{2-m/2}$ (Lemma 3.8). The latter result implies that the possible stability of the stripe pattern strongly depends on the stability of the associated pulse solution of the one-dimensional equation (that does not depend on y). We find, in the case that the one-dimensional pulse pattern is stable, that there are two symmetrical bands of unstable wave numbers l: $\mathcal{O}(\varepsilon^{2-m/2}) \le |l| \le \mathcal{O}(1)$. This implies that all homoclinic stripe patterns are unstable as solution of (1.4) on $\mathbf{R}^2$ if $m < 4$, i.e., $\mu \ll \mathcal{O}(1/\varepsilon^m)$

(Theorem 3.9).

Nevertheless, these results also indicate that the homoclinic stripe patterns can be stable on $\mathbf{R}^2$ for $m = 4$ since the bands of unstable wave numbers might disappear for $m = 4$. A necessary ingredient is, of course, the stability of the one-dimensional pulse pattern. Therefore, we first follow [10] and consider the classical case $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = 2, \beta_2 = 2$, and $0 < \mu \ll 1/\varepsilon^4$. In this case, the one-dimensional pulse is stable for $\mu > \mu_{\text{Hopf}}(0) = 0.36\ldots (= \mathcal{O}(1))$ (Theorem 4.3 or [10]). Furthermore, we use the NLEP machinery to explicitly determine the band of unstable wave numbers in this case: we deduce that, for $0 < \mu_{\text{Hopf}}(0) \leq \mu = \tilde{\mu}/\varepsilon^m$ and $m < 4$, there is one (unique) real unstable eigenvalue $\lambda(l) > 0$ for $|l| \in (\varepsilon^{2-m/2}\sqrt{3\tilde{\mu}}, \sqrt{5/4})$, at leading order in ε; there are more unstable wave numbers l with $|l| \leq \varepsilon^{2-m/2}\sqrt{3\tilde{\mu}}$ for $\mu \leq \mu_{\text{Hopf}}(0)$ (Theorem 4.5).

Next we consider the general problem in more detail. We focus, for simplicity, on stability analysis of the one-dimensional problem (i.e., $l = 0$). We distinguish two open, unbounded domains, $\mathcal{V}_{\text{large}}$ and $\mathcal{V}_{\text{singular}}$, in the $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ parameter space (with boundaries given by (1.2)) in which the homoclinic pulse pattern cannot be stable. The region $\mathcal{V}_{\text{large}}$ includes the case $\alpha_1 > 1$. We show that, for $\alpha_1 > 1$ and μ large enough (but not necessarily $\gg 1$ with respect to ε), there is always (at least) one unstable real eigenvalue $\lambda^0(\mu, 0)$ that grows linearly with μ (Theorem 4.9). In Theorem 4.10, we establish the existence of the region $\mathcal{V}_{\text{singular}}$, which includes $\beta_2 > 2\beta_1 + 1$, in which the stability problem has at least one bounded unstable real eigenvalue for all $\mu > 0$.

Finally, we consider the stability of the stripes for $m = 4$, i.e., $\mu = \mathcal{O}(1/\varepsilon^4)$, by numerical simulations. This is necessary since the stability analysis is based on the same rescaled value of ε as the existence analysis: like the existence problem, the linear stability is no longer singularly perturbed for $m = 4$. The simulations are performed on bounded domains; therefore, we first interpret some of our results in terms of cylindrical domains, or strips, of the form $\mathbf{R} \times (0, L_y)$. (The length, L_x, of the domain in the x-coordinate is taken so long as it has no leading order influence; see [12], [8], [6], and section 5.) We conclude that a homoclinic stripe solution that is stable as a one-dimensional pattern is automatically stable on a strip $\mathbf{R} \times (0, L_y)$ if $L_y < \pi\varepsilon/\sqrt{(\beta_2 + 1)^2/4 - 1}$ (Corollary 5.1). This critical value of L_y is confirmed by the numerical simulations for $\mu \ll 1/\varepsilon^4$ (where we considered the classical case, i.e., $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = \beta_2 = 2$). Moreover, it is found that the stripe pattern can be stable as a solution on $\mathbf{R}^2$ for $\mu > \mu_{\text{stripe}} = \mu_{\text{stripe}}(\alpha_1, \alpha_2, \beta_1, \beta_2) = \mathcal{O}(1/\varepsilon^4)$. This bifurcation is followed by the splitting bifurcation predicted by the existence analysis of section 2: $\mu_{\text{split}} > \mu_{\text{stripe}}$ or, more explicitly,

$$\mu_{\text{split}}(0, -1, 2, 2) \approx \frac{0.14}{\varepsilon^4} > \frac{0.12}{\varepsilon^4} \approx \mu_{\text{stripe}}(0, -1, 2, 2).$$

Beyond the splitting bifurcation value of μ, a *self-replicating stripe pattern* is observed; i.e., the homoclinic stripe splits into two repelling stripes, which split again (etc., depending on the length L_x of the domain) so that eventually an asymptotically stable spatially periodic stripe pattern appears (see Figure 5.3). The dynamics are completely homogeneous in the y-direction so that the bifurcation is, in essence, a one-dimensional process. Note that these simulations also imply that (numerically) stable spatially periodic stripe patterns exist on $\mathbf{R}^2$ (see also Remark 1.4).

The numerical investigations are concluded with an analysis of the "fate" of the stripe pattern as $\mu < \mu_{\text{stripe}}$ and L_y increases through a bifurcation value. This confirms the varicose type of the instability [32], [18]: the stripe, in general, bifurcates into half a spot at either one of the boundaries $y = 0$ or $y = L_y$ in the middle of the $(0, L_x)$ interval. For values of μ close to μ_{stripe}, it is possible to have stripe patterns on a strip of width L_y that is larger than that given by Corollary 5.1. Such stripes bifurcate into a full spot, two half spots, one and a half spots, etc. (see Figure 5.4). This behavior is typical for systems near a bifurcation of Turing type; it can be explained qualitatively by the analysis.

Remark 1.1. As mentioned in the literature, i.e., [35], [29], [12], [28], [20], [39], [10], we consider so-called monostable systems. There is much literature on the existence and stability of (multi)front patterns in (singularly perturbed) bistable systems. (Unlike in monostable systems, there are (at least) two different stable trivial solutions/patterns in bistable systems.) See, for instance, [30] and the references therein. We refer to [32] for the stability analysis of a double front, i.e., homoclinic, stripe pattern in a bistable system with a piecewise linear nonlinear term and to [37], [38] for the stability analysis of stripes/planar fronts in more general bistable systems. An essential difference between the stability of (multi)front solutions in bistable systems and that of the monostable homoclinic solutions studied here is that all potentially unstable eigenvalues are asymptotically small, i.e., approach 0 in the limit $\varepsilon \to 0$, in the bistable case. The limits of these eigenvalues correspond to "marginally stable" eigenvalues of the so-called fast reduced limit systems [30]. The relation between the stability in the full $\varepsilon \neq 0$ system and the $\varepsilon = 0$ limit systems is completely different in the monostable equations studied here. In fact, the fast reduced limit systems have $\mathcal{O}(1)$ unstable eigenvalues (see section 3). This is called "the NLEP paradox" in [9], [10].

Remark 1.2. The following general class of singularly perturbed reaction-diffusion equations was considered in [10]:

$$\left\{ \begin{array}{lllll} U_t & = & d_U \Delta U & + a_{11} U + a_{12} V & + H_1(U, V), \\ V_t & = & d_V \Delta V & + a_{21} U + a_{22} V & + H_2(U, V), \end{array} \right. \tag{1.5}$$

where $0 < d_V \ll d_U$, a_{ij} is such that the trivial pattern $(U(x,t), V(x,t)) \equiv (0,0)$ of the linear equation (i.e., $H_i(U,V) \equiv 0$) is asymptotically stable, and $H_i(U,V)$ is such that certain growth conditions are satisfied. In [10], it has been shown by a scaling analysis that, under general conditions, the existence and stability of "large" localized pulse solutions to the one-dimensional equivalent of (1.5) is governed by a "normal form" of which the generalized Gierer–Meinhardt equation, in the form of (1.4), is the leading order part. As a consequence, all results on the stability and bifurcations of homoclinic stripe patterns obtained in this paper for the generalized Gierer–Meinhardt equation can be reformulated in terms of this general class of reaction-diffusion equations (see, however, Remark 2.3).

Remark 1.3. The critical magnitude of μ, $\mu = \mathcal{O}(1/\varepsilon^4)$, that appears throughout this paper is in terms of the *rescaled* parameter ε of the (truncated) "normal form" (1.4). It follows from (1.3) that this corresponds to $\mu = \mathcal{O}(1/\varepsilon^2)$ in terms of the *original* parameter ε of the unscaled generalized Gierer–Meinhardt equation (1.1). This parameter has been defined as the square root of the ratio between the diffusion coefficients of V and U, $\varepsilon^2 = d_V/d_U \ll 1$, which implies that the splitting bifurcation (and the Turing bifurcation—see Remark 1.4) will occur at $\mu = \mathcal{O}(d_U/d_V)$.

Remark 1.4. Numerical simulations show that the amplitude of the periodic stripe pattern that occurs through the splitting bifurcation decreases as μ approaches the value $\mu_{\text{Turing}} > \mu_{\text{split}}$. At $\mu_{\text{Turing}} = \mu_{\text{Turing}}(\alpha_1, \alpha_2, \beta_1, \beta_2) = \mathcal{O}(1/\varepsilon^4)$, a Turing bifurcation takes place. (μ_{Turing} can be determined by a straightforward linear analysis; see, for instance, [27].) Thus the homoclinic stripe patterns are indeed "far from equilibrium." We refer to [24] for a detailed analysis of the connection between homoclinic pulse patterns and Turing bifurcations in the context of the one-dimensional Gray–Scott equation.

Remark 1.5. An important subtheme of this paper is the competition between stripe and spot patterns, far from equilibrium. There is much literature on the interactions between patterns—most of it on systems close to bifurcation/equilibrium. We refer to [3] and [16] and the references therein.

2. The existence problem.

The following result on the existence of singular, stationary, multipulse homoclinic solutions of the generalized Gierer–Meinhardt equation was proved in [10].

Theorem 2.1. *Let $(\alpha_1, \alpha_2, \beta_1, \beta_2, \mu)$ satisfy (1.2), and let μ be $\mathcal{O}(1)$. Then, for any $N \geq 1$ with $N = \mathcal{O}(1)$ and $\varepsilon > 0$ small enough, there is a stationary N-pulse homoclinic stripe solution $(U(x,y,t), V(x,y,t)) = (U_N^{\text{hom}}(x), V_N^{\text{hom}}(x))$ to (1.4) so that $\lim_{|x|\to\infty} U_N^{\text{hom}}(x) = \lim_{|x|\to\infty} V_N^{\text{hom}}(x) = 0$, and $\lim_{|x|\to\infty} U_N^{\text{hom}}(x) e^{\varepsilon\sqrt{\mu}|x|}$, $\lim_{|x|\to\infty} V_N^{\text{hom}}(x) e^{|x|}$ exist and are nonzero. The V-component $V_N^{\text{hom}}(x)$ has a sequence of N consecutive narrow pulses of the same height (at leading order) that are $\mathcal{O}(\varepsilon|\log\varepsilon|)$ close to each other; $V_N^{\text{hom}}(x)$ decreases to $\mathcal{O}(\sqrt{\varepsilon})$ in between two adjacent pulses. Both $U_N^{\text{hom}}(x)$ and $V_N^{\text{hom}}(x)$ are monotonous functions of x outside the region of pulses. Moreover, the amplitudes $U_N^{\max}$ and $V_N^{\max}$ of the $U_N^{\text{hom}}(x)$ and $V_N^{\text{hom}}(x)$ pulses are, at leading order, given by*

$$U_N^{\max} = \left[\frac{2\sqrt{\mu}}{NW(\beta_1,\beta_2)}\right]^{\frac{\beta_2-1}{D}}, \quad V_N^{\max} = \left[\frac{\beta_2+1}{2}\right]^{\frac{1}{\beta_2-1}} \left[\frac{2\sqrt{\mu}}{NW(\beta_1,\beta_2)}\right]^{-\frac{\alpha_2}{D}}, \tag{2.1}$$

where D is given in (1.3) and

$$W(\beta_1,\beta_2) = \int_{-\infty}^{\infty} (w_h(\xi;\beta_2))^{\beta_1} d\xi, \tag{2.2}$$

with $w_h(\xi)$ the (positive) homoclinic solution of

$$\ddot{w} = w - w^{\beta_2}. \tag{2.3}$$

It is clear from the formulation of this theorem that only the case $\mu = \mathcal{O}(1)$ (with respect to ε) has been considered in [10]. In this section, we consider the existence problem for $\mu = \mathcal{O}(1/\varepsilon^m)$ for $m \geq 0$. We do not pay attention to the multipulse solutions with $N \geq 2$ since these solutions cannot even be stable on the one-dimensional (unbounded) domain, as has been proved in [10]. In this paper, we denote $(U_1^{\text{hom}}(x), V_1^{\text{hom}}(x))$ by $(U_0(x), V_0(x))$.

2.1. Scaling analysis.

We introduce $\tilde{\mu} = \mathcal{O}(1)$ and $m \geq 0$ by

$$\mu = \frac{\tilde{\mu}}{\varepsilon^m}. \tag{2.4}$$

It follows from (2.1) that $U_0(x)$ and $V_0(x)$ can no longer be considered as $\mathcal{O}(1)$ for $\mu \gg 1$. Therefore, we have to scale $U(x,y,t)$ and $V(x,y,t)$ in (1.4) and thus introduce the $\mathcal{O}(1)$ quantities $\tilde{U}(x,y,t)$ and $\tilde{V}(x,y,t)$ by

$$U = \varepsilon^{-\frac{(\beta_2-1)m}{2D}}\tilde{U}, \;\; V = \varepsilon^{\frac{\alpha_2 m}{2D}}\tilde{V}. \tag{2.5}$$

Inserting these scalings into (1.4) yields

$$\left\{\begin{array}{rcllcl} \varepsilon^{2+\frac{m}{2}}\tilde{U}_t & = & \varepsilon^{\frac{m}{2}}\Delta\tilde{U} & -\varepsilon^{2-\frac{m}{2}}\tilde{\mu}\tilde{U} & + & \tilde{U}^{\alpha_1}\tilde{V}^{\beta_1}, \\ \tilde{V}_t & = & \varepsilon^2\Delta\tilde{V} & -\tilde{V} & + & \tilde{U}^{\alpha_2}\tilde{V}^{\beta_2}. \end{array}\right. \tag{2.6}$$

Hence we can introduce $\tilde{x}$, $\tilde{y}$, and $\tilde{\varepsilon}$ by

$$\tilde{x} = \varepsilon^{-\frac{m}{4}}x, \;\; \tilde{y} = \varepsilon^{-\frac{m}{4}}y, \;\; \tilde{\varepsilon} = \varepsilon^{1-\frac{m}{4}} \tag{2.7}$$

so that

$$\left\{\begin{array}{rcllcl} \tilde{\varepsilon}^{2+\frac{4m}{4-m}}\tilde{U}_t & = & \tilde{\Delta}\tilde{U} & -\tilde{\varepsilon}^2\tilde{\mu}\tilde{U} & + & \tilde{U}^{\alpha_1}\tilde{V}^{\beta_1}, \\ \tilde{V}_t & = & \tilde{\varepsilon}^2\tilde{\Delta}\tilde{V} & -\tilde{V} & + & \tilde{U}^{\alpha_2}\tilde{V}^{\beta_2}. \end{array}\right. \tag{2.8}$$

Except for the factor in front of the term $\tilde{U}_t$, this equation is identical to (1.4). In this section, we are interested in the existence of solutions to (2.8) that depend neither on t nor on $\tilde{y}$. Thus we write (2.8) as a (four-dimensional) ODE in $\tilde{x}$:

$$\left\{\begin{array}{rlcll} \tilde{U}_{\tilde{x}\tilde{x}} & -\tilde{\varepsilon}^2\tilde{\mu}\tilde{U} & + & \tilde{U}^{\alpha_1}\tilde{V}^{\beta_1} & = 0, \\ \tilde{\varepsilon}^2\tilde{V}_{\tilde{x}\tilde{x}} & -\tilde{V} & + & \tilde{U}^{\alpha_2}\tilde{V}^{\beta_2} & = 0. \end{array}\right. \tag{2.9}$$

Except for the tildes, this equation is identical to the existence problem for stationary solutions that do not depend on y of the original equation (1.4). Hence we can immediately apply Theorem 2.1 to system (2.9) and conclude that there exist N-pulse patterns $(\tilde{U}_N^{\mathrm{hom}}(\tilde{x}), \tilde{V}_N^{\mathrm{hom}}(\tilde{x}))$ in (2.8) and thus in (1.4) for $\mu \gg 1$. However, there is one crucial condition in Theorem 2.1 that cannot be satisfied for all $\mu \gg 1$: $\tilde{\varepsilon}$ must be small enough. Hence, by (2.7), Theorem 2.1 can only be applied for $m < 4$ (see Remark 1.3). The condition on ε is essential to the proof of Theorem 2.1 since it is based on geometric singular perturbation theory [21] and it exploits the fact that $\tilde{U}$ and $\tilde{U}_{\tilde{x}}$ vary slowly compared to $\tilde{V}$ and $\tilde{V}_{\tilde{x}}$; i.e., $\tilde{U}, \tilde{U}_{\tilde{x}} = \mathcal{O}(\tilde{\varepsilon})$, while $\tilde{V}, \tilde{V}_{\tilde{x}} = \mathcal{O}(1)$. This approach can no longer be used when ε becomes $\mathcal{O}(1)$, i.e., when $m = 4$ (2.7).

On the other hand, when m becomes > 4 or, equivalently, when $\tilde{\varepsilon}$ becomes $\gg 1$, it might be possible to use geometric singular perturbation theory to construct homoclinic solutions to the saddle point $(\tilde{U}, \tilde{U}_{\tilde{x}}, \tilde{V}, \tilde{V}_{\tilde{x}}) = (0,0,0,0)$ by reversing the roles of $\tilde{U}$ and $\tilde{V}$. Therefore, we introduce $\hat{U}$, $\hat{V}$, $\hat{\varepsilon}$, $\hat{\mu}$, and $\hat{x}$ by

$$\hat{U} = (\tilde{\varepsilon}^2\tilde{\mu})^{\frac{1+\beta_1-\beta_2}{D}}\tilde{U}, \; \hat{V} = (\tilde{\varepsilon}^2\tilde{\mu})^{\frac{1-\alpha_1+\alpha_2}{D}}\tilde{V} \; \hat{\varepsilon} = \frac{1}{\tilde{\varepsilon}}, \; \hat{\mu} = \frac{1}{\tilde{\mu}}, \; \hat{x} = \sqrt{\tilde{\mu}}\tilde{x} \tag{2.10}$$

so that (2.9) can be written as

$$\left\{\begin{array}{rlcll} \hat{\varepsilon}^2\hat{U}_{\hat{x}\hat{x}} & -\hat{U} & + & \hat{U}^{\alpha_1}\hat{V}^{\beta_1} & = 0, \\ \hat{V}_{\hat{x}\hat{x}} & -\hat{\varepsilon}^2\hat{\mu}\hat{V} & + & \hat{U}^{\alpha_2}\hat{V}^{\beta_2} & = 0. \end{array}\right. \tag{2.11}$$

This equation is identical to (2.9) after the following substitutions:

$$\tilde{U} \to \hat{V},\ \tilde{V} \to \hat{U},\ \alpha_1 \to \beta_2,\ \beta_1 \to \alpha_2,\ \alpha_2 \to \beta_1,\ \beta_2 \to \alpha_1, \tag{2.12}$$

and, of course, $\tilde{\varepsilon} \to \hat{\varepsilon}$, $\tilde{\mu} \to \hat{\mu}$.

2.2. The existence and disappearance of homoclinic solutions. The "symmetry" (2.12) between the two scaled systems (2.9) and (2.11) can be used to obtain the following extension of Theorem 2.1.

Theorem 2.2. *Let $(\alpha_1, \alpha_2, \beta_1, \beta_2, \mu)$ satisfy (1.2), and let $(U_1^{\rm hom}(x), V_1^{\rm hom}(x)) = (U_0(x), V_0(x))$ be the (1-pulse) homoclinic stripe solution of (1.4) described in Theorem 2.1 (for $\mu = \mathcal{O}(1)$). Then there exists a critical value $\mu_{\rm split}$ of μ with $\mu_{\rm split} = \frac{\tilde{\mu}_{\rm split}}{\varepsilon^4}$ and $0 < \tilde{\mu}_{\rm split} = \mathcal{O}(1)$ such that $(U_0(x), V_0(x))$ exists for $0 < \mu < \mu_{\rm split}$. Equation (1.4) does not have a homoclinic solution $(U_0(x), V_0(x))$ for $\mu > \mu_{\rm split}$.*

A similar result has been proved for the one-dimensional Gray–Scott model in [9] (although the proof is based on a topological shooting approach in the Gray–Scott context). For the Gray–Scott model, the upper boundary on the existence domain of the homoclinic pulse solutions has been identified numerically in [31] as a saddle-node bifurcation of homoclinic orbits. A similar behavior is expected here. Moreover, there is a very natural candidate that can act as the "partner" of the pulse pattern $(U_0(x), V_0(x))$ in the saddle-node bifurcation. It is the 2-pulse homoclinic orbit $(U_2^{\rm hom}(x), V_2^{\rm hom}(x))$ (see Theorem 2.1) since it has the same structure as the unstable "partner" of the homoclinic pulse that has been found numerically in [31] (for the Gray–Scott model) near the saddle-node/disappearance bifurcation. (Furthermore, $(U_2^{\rm hom}(x), V_2^{\rm hom}(x))$ is unstable; see [10].) Note that it is quite a challenge to prove this conjecture, especially since the bifurcation occurs in a parameter region where the system can no longer be treated as a singularly perturbed problem.

It was shown by numerical simulations that this "disappearance" of the homoclinic pulse solution marks the boundary of the region (in parameter space) in which a solitary pulse "splits" into two slowly traveling "copies" of the initial pulse (with opposite speeds) [9]. Thus the equivalent of Theorem 2.2 for the Gray–Scott model gave an analytical foundation of the origin of the so-called self-replication process. This phenomenon has been a challenging topic of research in recent years (see [33], [35], [34], [12], [31], and the references therein; we refer to [7] for a discussion on the literature on this subject).

By analogy to the Gray–Scott model, the "disappearance" result of Theorem 2.2 at $\mu = \mathcal{O}(\frac{1}{\varepsilon^4})$ provides a strong motivation to run simulations of the generalized Gierer–Meinhardt model for $\mu \gg 1$. It is shown in Figure 2.1 that the critical value $\mu_{\rm split}$ defines the boundary of a domain in parameter space in which the (generalized) Gierer–Meinhardt model also exhibits self-replication of pulses. This yields a strong indication that the self-replication phenomenon is not a special feature of the Gray–Scott model but that it will occur in a large family of reaction-diffusion equations [31]; see also Remark 1.2.

Proof of Theorem 2.2. This proof is based on the proof of Theorem 2.1 in [10]. Here we present the main arguments and refer to [10] for the details.

As was already noted in the previous section, Theorem 2.1 applied to (2.11) establishes the existence of $(U_0(x), V_0(x))$ for $0 < \mu \ll \frac{1}{\varepsilon^4}$. Using the "symmetry" (2.12), we can apply Theorem 2.1 to (2.11) for $\mu \gg \frac{1}{\varepsilon^4}$ since $\tilde{\varepsilon} \gg 1$ and thus $\hat{\varepsilon} \ll 1$ (2.10). This yields the existence

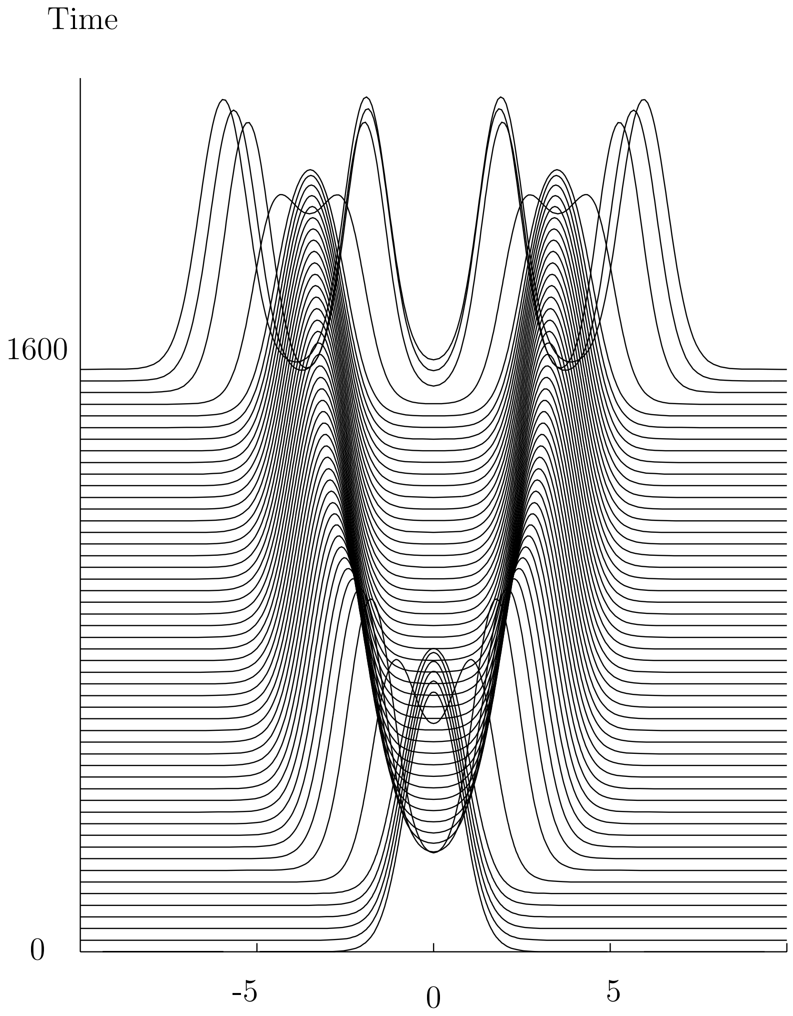


Figure 2.1. *The self-replication process in the classical Gierer–Meinhardt problem (i.e.,* $\alpha_1 = -1, \alpha_2 = 0, \beta_1 = \beta_2 = 2$*). This simulation was done for* $\mu = 56$*,* $\varepsilon^2 = 0.05$*. Note that only the* V*-components of the solutions are shown.*

of the homoclinic solution $(U_0(x), V_0(x))$ of (1.4) if $\alpha_1, \alpha_2, \beta_1, \beta_2$ satisfy a condition that is the equivalent of (1.2) under (2.12):

$$\alpha_1 > 1,\ \alpha_2 > 1,\ \beta_1 < 0,\ \beta_2 > 1 + \frac{\alpha_2 \beta_1}{\alpha_1 - 1}. \tag{2.13}$$

The nonexistence of a homoclinic $(U_0(x), V_0(x))$ pattern follows from the obvious conflict between (1.2) and (2.13). However, one cannot, of course, obtain a nonexistence result from the fact that an existence result cannot be applied. As is explained in detail in [10], the conditions on the α_i's in (1.2) are not essential to the existence of homoclinic solutions. The conditions on the α_i's are determined by our decision to study *large* pulses in (1.1), i.e., our decision to impose $r, s > 0$ in (1.3)—see also Remark 2.4.

On the other hand, the condition on β_2 is sharp in the sense that there cannot be homoclinic solutions for $\beta_2 \leq 1$. This can be seen immediately by writing (2.9) as a four-dimensional system in the "fast" scaling; i.e., we introduce the fast variable $\xi = x/\varepsilon$ and obtain

$$(2.14) \qquad \left\{ \begin{array}{lcl} \dot{u} &=& \varepsilon p, \\ \dot{p} &=& \varepsilon[-u^{\alpha_1} v^{\beta_1} + \varepsilon^2 \mu u], \\ \dot{v} &=& q, \\ \dot{q} &=& v - u^{\alpha_2} v^{\beta_2}, \end{array} \right.$$

where we have neglected the tildes and where $\dot{}$ denotes the derivative with respect to ξ. For $\beta_2 \leq 1$, there is no homoclinic solution in the fast reduced limit $u \equiv u_0$, $p \equiv p_0$, and $\ddot{v} = v - u_0^{\alpha_2} v^{\beta_2}$ so that it is impossible to construct a homoclinic solution to $(0,0,0,0)$—see [10] for the details. The condition on β_1 might be relaxed to $\beta_1 > 0$ (see Remarks 2.3 and 3.2 in [10] and Remark 2.4 below); however, the lower boundary on β_1 cannot be decreased beyond 0: there cannot be homoclinic solutions to $(0,0,0,0)$ in (2.14) for $\beta_1 < 0$. This follows especially from the equation for $\dot{p}$: the accumulated change Δp in p over an orbit that is homoclinic to $(0,0,0,0)$ cannot be bounded for $\beta_1 < 0$ since the integral over $\dot{p}$ (from $-\infty$ to ∞) diverges in this case; see Remark 2.7 in [10].

By the "symmetry" (2.13), we thus conclude that there cannot be homoclinic solutions to $(0,0,0,0)$ in (2.11) for $\alpha_1 < 1$ or $\alpha_2 < 0$. Since we assumed that $\alpha_2 < 0$ in Theorems 2.1 and 2.2, it follows from (2.7) and (2.10) that the homoclinic stripe pattern $(U_0(x), V_0(x))$ cannot exist as a solution of (1.4) for $\mu \gg \frac{1}{\varepsilon^4}$.

For the intermediate case, $m = 4$ in (2.4), we set $\tilde{\varepsilon} = 1$ in (2.9) and once more use (2.1) to scale $(\tilde{U}, \tilde{V})$ in a similar fashion as (U, V) was scaled in (2.5):

$$(2.15) \qquad \tilde{U} = \tilde{\mu}^{\frac{\beta_2 - 1}{2D}} \check{U}, \quad \tilde{V} = \tilde{\mu}^{\frac{-\alpha_2}{2D}} \check{V}.$$

This way, (2.9) becomes

$$(2.16) \qquad \left\{ \begin{array}{rlll} \check{U}_{\check{x}\check{x}} & -\sqrt{\tilde{\mu}}\check{U} & + \ \check{U}^{\alpha_1} \check{V}^{\beta_1} & = 0, \\ \sqrt{\tilde{\mu}}\check{V}_{\check{x}\check{x}} & -\check{V} & + \ \check{U}^{\alpha_2} \check{V}^{\beta_2} & = 0, \end{array} \right.$$

where $\check{x} = (\tilde{\mu})^{1/4} \tilde{x}$. This equation is identical to (2.9) when we set $\tilde{\mu} = 1$ and $\tilde{\varepsilon}^2 = \sqrt{\tilde{\mu}}$ in (2.9). Hence we can apply Theorem 2.1 and conclude that the homoclinic pattern $(U_0(x), V_0(x))$ exists for $\tilde{\mu} < \tilde{\mu}_0$ small enough. Note that we have only introduced the $\check{U}$-, $\check{V}$-scaling (2.15) to validate the intuitively clear observation that the limit $\tilde{\mu} \to 0$ with $m = 4$ corresponds to the case $m < 4$, i.e., $\tilde{\varepsilon} \ll 1$. Similarly, one can scale (2.9) with $\tilde{\varepsilon} = 1$ as in (2.10) so that the new system can be identified with (2.11) with $\hat{\mu} = 1$ and $\hat{\varepsilon}^2 = 1/\sqrt{\tilde{\mu}}$. Hence the nonexistence result for (2.11) can be applied for $\tilde{\mu} > \hat{\mu}_0$ large enough.

We conclude that must be a value $\tilde{\mu}_{\text{split}}$ in between $\tilde{\mu}_0$ and $\hat{\mu}_0$ so that $(U_0(x), V_0(x))$ exists for $m = 4$ and $\tilde{\mu} < \tilde{\mu}_{\text{split}}$ but not (immediately) beyond this value. ■

Remark 2.3. The proof of Theorem 2.2 cannot be applied directly to the more general system (1.5) of Remark 1.2. Equation (1.4) appears from (1.5) as *the leading order part* of a normal form. Thus (1.5) can only be approximated by (1.4) for $\varepsilon \ll 1$. The proof of Theorem 2.2 is based on the "symmetry" between $\varepsilon \ll 1$ and $\varepsilon \gg 1$ in (1.4) and is thus special for the

Gierer–Meinhardt model. However, the existence result for $\varepsilon \ll 1$ is based on a combination of properties of the singular perturbed model. In general, it can be expected that such a combination no longer exists for $\varepsilon = \mathcal{O}(1)$ and/or $\varepsilon \gg 1$. Hence it is natural to suspect that there is an equivalent of Theorem 2.2 for the general model (1.5) if the model satisfies some additional conditions. The topological shooting method employed for the proof of this result in the Gray–Scott context [9] seems to be the most suitable method for the proof of such a general result.

Remark 2.4. As is explained in the proof of Theorem 2.2, the conditions on the α's in (1.2) are based on our preference to have (asymptotically) large solutions in (1.1). Only the conditions $\beta_1 > 0$, $\beta_2 > 1$, and $D \neq 0$ (1.3) are necessary for the existence proof in [10]. However, we note that it has been necessary to assume that $\beta_1 > 1$ in the proof of Theorem 2.1 in [10] in order to be able to apply the standard persistence results of Fenichel; see [21]. Thus it is not completely straightforward to relax the condition $\beta_1 > 1$ to $\beta_1 > 0$ in Theorem 2.1. (The case $\beta_1 = 0$ is special; see Remark 2.7 in [10].) Nevertheless, it can thus be expected, by combining (1.2) and (2.13), that there exist homoclinic solutions in (1.1) of the type described by Theorems 2.1 and 2.2 for $\varepsilon \ll 1$ and for $\varepsilon \gg 1$ if $\alpha_1 > 1$, $\alpha_2 > 0$, $\beta_1 > 0$, $\beta_2 > 1$, and $D \neq 0$. Thus one does not expect "splitting dynamics" [7] in this case. However, we will find in section 4 that the homoclinic pulse pattern cannot be stable when $\alpha_1 > 1$ and μ is large (Theorem 4.9) so that this (possible) persistence result will not be relevant for the dynamics of (1.1).

3. Stability analysis.

In this section, we consider the stability of the stationary homoclinic stripe pattern $(U(x,y,t), V(x,y,t)) = (U_0(x), V_0(x))$ on the unbounded domain, i.e., with $(x,y) \in \mathbf{R}^2$. Due to the lack of structure in the y-direction, we can study the linearized stability of the stripe by introducing a wave number $l \in \mathbf{R}$ and set

$$\begin{aligned} U(x,y,t) &= U(\xi,\eta,t) = U_0(\xi) + u(\xi)e^{\lambda t}\, e^{il\eta}, \\ V(x,y,t) &= V(\xi,\eta,t) = V_0(\xi) + v(\xi)e^{\lambda t}\, e^{il\eta}. \end{aligned} \tag{3.1}$$

Thus we consider the stability problem in the fast spatial variables, defined by

$$(x,y) = (\varepsilon\xi, \varepsilon\eta). \tag{3.2}$$

Moreover, it will be convenient to introduce a scaled version $\hat{l}$ of the wave number l:

$$l = \varepsilon^2 \hat{l}. \tag{3.3}$$

By construction, the wave number l, or $\hat{l}$, appears as a parameter in the linear stability analysis. In this and the following sections, we will determine, for any fixed $l \in \mathbf{R}$, the spectrum of the associated ξ-dependent linear operator. The stripe is spectrally stable when the union of the spectra (over all $l \in \mathbf{R}$) has no intersection with the unstable half plane $\{\mathrm{Re}(\lambda) > 0\}$; see Remark 1.3. In this paper, we will not consider the nonlinear stability of the stripes. We refer to [10] for some remarks about the nonlinear theory for the one-dimensional case. In this section, we present an extension of the Evans function method, the NLEP approach, as it has been developed in [8], [9], [10]. Here, we sketch the main ideas behind this method. The statements in this section can all be proved by the methods developed in [9], [10]. We refer to these papers for the analytical details.

3.1. The linearized equations.

Inserting (3.1) into (1.4) yields, after linearization:

$$\left\{\begin{array}{l} u_{\xi\xi} = -\varepsilon^2[\alpha_1 U_0^{\alpha_1-1} V_0^{\beta_1} u + \beta_1 U_0^{\alpha_1} V_0^{\beta_1-1} v] + \varepsilon^4[\mu + \lambda + \hat{l}^2]u, \\ v_{\xi\xi} + [\beta_2 U_0^{\alpha_2} V_0^{\beta_2-1} - (1 + \lambda + \varepsilon^4 \hat{l}^2)]v = -[\alpha_2 U_0^{\alpha_2-1} V_0^{\beta_2}]u, \end{array}\right. \tag{3.4}$$

where we have used (3.2) and (3.3). Note that $u(\xi)$ remains constant to leading order on ξ-intervals of (at least) $\mathcal{O}(1)$ length, as long as $\mu, \lambda \ll \frac{1}{\varepsilon^4}$ and $|\hat{l}| \ll \frac{1}{\varepsilon^2}$, i.e., $|l| \ll 1$ (3.3). System (3.4) can be written as a four-dimensional linear equation,

$$\dot{\phi} = A(\xi; \lambda, \hat{l}, \varepsilon)\phi \;\text{ with }\; \phi(\xi) = (u(\xi), p(\xi), v(\xi), q(\xi))^t, \tag{3.5}$$

so that

(3.6)

$$A(\xi; \lambda, \hat{l}, \varepsilon) = \begin{pmatrix} 0 & \varepsilon & 0 & 0 \\ -\varepsilon\alpha_1 U_0^{\alpha_1-1} V_0^{\beta_1} + \varepsilon^3(\mu + \lambda + \hat{l}^2) & 0 & -\varepsilon\beta_1 U_0^{\alpha_1} V_0^{\beta_1-1} & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_2 U_0^{\alpha_2-1} V_0^{\beta_2} & 0 & -\beta_2 U_0^{\alpha_2} V_0^{\beta_2-1} + (1 + \lambda + \varepsilon^4 \hat{l}^2) & 0 \end{pmatrix}.$$

We know by Theorems 2.1 and 2.2 that $V_0(\xi)$ decays much faster than $U_0(\xi)$, as function of $|\xi|$, for any $\mu \ll \frac{1}{\varepsilon^4}$. Thus, using (1.2), we can take the limit $|\xi| \to \infty$ in $A(\xi)$:

$$A_\infty(\lambda, \hat{l}, \varepsilon) = \begin{pmatrix} 0 & \varepsilon & 0 & 0 \\ \varepsilon^3(\mu + \lambda + \hat{l}^2) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & (1 + \lambda + \varepsilon^4 \hat{l}^2) & 0 \end{pmatrix}. \tag{3.7}$$

Matrix $A(\xi)$ converges exponentially fast to A_∞ for any $\hat{l} \in \mathbf{R}$ and any $\mu \ll \frac{1}{\varepsilon^4}$; i.e., there exist positive $\mathcal{O}(1)$ constants $C_{1,2}$ such that

$$\|A(\xi; \lambda, \hat{l}, \varepsilon) - A_\infty(\lambda, \hat{l}, \varepsilon)\| \leq C_1 e^{-C_2|\xi|} \;\text{ for }\; |\xi| > \frac{1}{\varepsilon^\sigma}, \;\text{ for any } \sigma > 0. \tag{3.8}$$

As we shall see, this implies that the NLEP approach can be applied to the eigenvalue problem (3.6). The essential spectrum $\sigma_{\text{ess}}(l)$ associated to the matrix operator $A(\xi; \lambda, \hat{l}, \varepsilon)$ (3.6) is for any fixed l determined by A_∞ [17]–see also Remark 3.1. Since the eigenvalues Λ_i and eigenvectors E_i $(i = 1, \dots, 4)$ of A_∞ are given by

$$\begin{array}{ll} \Lambda_{1,4}(\lambda, \hat{l}, \varepsilon) = \pm\sqrt{1 + \lambda + \varepsilon^4 \hat{l}^2}, & \Lambda_{2,3}(\lambda, \hat{l}, \varepsilon) = \pm\varepsilon^2\sqrt{\mu + \lambda + \hat{l}^2}, \\ E_{1,4}(\lambda, \hat{l}, \varepsilon) = (0, 0, 1, \pm\sqrt{1 + \lambda + \varepsilon^4 \hat{l}^2})^t, & E_{2,3}(\lambda, \hat{l}, \varepsilon) = (1, \pm\varepsilon\sqrt{\mu + \lambda + \hat{l}^2}, 0, 0)^t, \end{array} \tag{3.9}$$

with $\text{Re}(\Lambda_1) \geq \text{Re}(\Lambda_2) \geq \text{Re}(\Lambda_3) \geq \text{Re}(\Lambda_4)$, it follows immediately that

$$\sigma_{\text{ess}}(\hat{l}) = \{\lambda \in \mathbf{R} : \lambda \leq \max(-\mu - \hat{l}^2, -1 - \varepsilon^4 \hat{l}^2)\}. \tag{3.10}$$

Hence the essential spectrum has no influence on the stability of the stripe; the (in)stability is completely determined by the discrete spectrum, i.e., the eigenvalues, of (3.6); see Remark 3.1. We introduce the complement of a δ-neighborhood of the essential spectrum:

$$\mathcal{C}_\delta(\hat{l}) = \mathbf{C}\backslash \begin{array}{l} \{(\mathrm{Re}(\lambda) < \max(-\mu - \hat{l}^2, -1 - \varepsilon^4\hat{l}^2), |\mathrm{Im}(\lambda)| < \delta) \text{ or} \\ \|\lambda - \max(-\mu - \hat{l}^2, -1 - \varepsilon^4\hat{l}^2)\| < \delta\}, \end{array} \tag{3.11}$$

where $0 < \delta \ll 1$ is a second asymptotically small parameter that is independent of ε. By restricting λ to $\mathcal{C}_\delta$, we cannot run into problems with the definition and analysis of the Evans function near the essential spectrum (see below).

Remark 3.1. The spectral stability of the *two-dimensional* stripe (i.e., $(x, y) \in \mathbf{R}^2$) is determined by the eigenvalues $\lambda(\hat{l})$ of the $\hat{l}$-family of *one-dimensional* systems (3.5) (in $\xi \in \mathbf{R}$). It is clear form the "ansatz" (3.1) that a curve of eigenvalues $\{\lambda(\hat{l}), \hat{l} \in \mathbf{R}\}$ in the $(l,\ \mathrm{Re}(\lambda))$-plane is a component, or branch, of the essential spectrum of the "full" two-dimensional stability problem associated to the stripe. (Perturbations of the type (3.1) are, by definition, not integrable over $\mathbf{R}^2$.) The full problem cannot have point spectrum, plotted in the $(l,\ \mathrm{Re}(\lambda))$-plane; it consists of a two-dimensional region, $\{\sigma_{\mathrm{ess}}(\hat{l}), \hat{l} \in \mathbf{R}\}$ (3.10), and a number of one-dimensional curves $\{\lambda(\hat{l}), \hat{l} \in \mathbf{R}\}$. In this paper, we refer to the elements of these curves as eigenvalues (since they are in the point spectrum of (3.6) for a certain value of $\hat{l}$).

3.2. The Evans function.

The eigenvalues $\lambda(\hat{l}, \varepsilon)$ of (3.6), i.e., those values of λ for which (3.6) has an exponentially decaying eigenfunction solution $\phi(\xi)$, correspond to zeros of the Evans function $\mathcal{D}(\lambda; \hat{l}, \varepsilon)$ associated to (3.6). Here we give a brief sketch of the construction of the Evans functions $\mathcal{D}(\lambda; \hat{l})$ associated to (3.6), following [10]. We refer to [1], [13], [9], [10] for the details. The definition of $\mathcal{D}(\lambda, \hat{l})$ and its decomposition is based on the following results.

Lemma 3.2 (see [1], [13], [9], [10]). *For all $\lambda \in \mathcal{C}_\delta(\hat{l})$, there exist four independent solutions $\phi_j(\xi; \lambda, \hat{l}, \varepsilon)$ of (3.6), $j = 1, \ldots, 4$, such that*

(i) $\lim_{\xi\to-\infty} \phi_1(\xi; \lambda, \hat{l}) e^{-\Lambda_1(\lambda,\hat{l})\xi} = E_1(\lambda, \hat{l})$,

(ii) $\lim_{\xi\to-\infty} \phi_2(\xi; \lambda, \hat{l}) e^{-\Lambda_2(\lambda,\hat{l})\xi} = E_2(\lambda, \hat{l})$,

(iii) $\lim_{\xi\to\infty} \phi_3(\xi; \lambda, \hat{l}) e^{-\Lambda_3(\lambda,\hat{l})\xi} = E_3(\lambda, \hat{l})$,

(iv) $\lim_{\xi\to\infty} \phi_4(\xi; \lambda, \hat{l}) e^{-\Lambda_4(\lambda,\hat{l})\xi} = E_4(\lambda, \hat{l})$,

where $\Lambda_j(\lambda, \hat{l})$ and $E_j(\lambda, \hat{l})$, $j = 1, \ldots, 4$ have been defined in (3.9); $\phi_1(\xi; \lambda, \hat{l})$ and $\phi_2(\xi; \lambda, \hat{l})$ span the two-dimensional family $\Phi_-(\xi; \lambda, \hat{l})$ of solutions to (3.6) that approach $(0, 0, 0, 0)^t$ as $\xi \to -\infty$, $\phi_3(\xi; \lambda, \hat{l})$, and $\phi_4(\xi; \lambda, \hat{l})$ span the two-dimensional family $\Phi_+(\xi; \lambda, \hat{l})$ of solutions to (3.6) that approach $(0, 0, 0, 0)^t$ as $\xi \to \infty$. Furthermore, there exist two transmission functions $t_1(\lambda, \hat{l}, \varepsilon)$ and $t_2(\lambda, \hat{l}, \varepsilon)$ such that

$$\lim_{\xi\to\infty} \phi_1(\xi; \lambda, \hat{l}) e^{-\Lambda_1(\lambda,\hat{l})\xi} = t_1(\lambda, \hat{l}) E_1(\lambda, \hat{l}), \quad \lim_{\xi\to\infty} \phi_2(\xi; \lambda, \hat{l}) e^{-\Lambda_2(\lambda,\hat{l})\xi} = t_2(\lambda, \hat{l}) E_2(\lambda, \hat{l}); \tag{3.12}$$

$t_1(\lambda, \hat{l}, \varepsilon)$ is analytic as a function of $\lambda \in \mathcal{C}_\delta$, and $t_2(\lambda, \hat{l}, \varepsilon)$ is only defined for $t_1(\lambda) \neq 0$. The solutions $\phi_1(\xi; \lambda, \hat{l})$ and $\phi_2(\xi; \lambda, \hat{l})$ are determined uniquely: $\phi_1(\xi; \lambda, \hat{l})$ by (i) *and $\phi_2(\xi; \lambda, \hat{l})$ by* (ii) *and the existence of $t_2(\lambda, \hat{l})$.*

Note that these results are quite natural. By the limit behavior of the matrix A (3.8), one expects that there are two independent solutions to (3.6) that behave in the limit $\xi \to -\infty$ as the two independent unstable solutions of the constant coefficients problem associated to the limit matrix A_∞. These are the solutions ϕ_1 and ϕ_2. The solutions ϕ_3 and ϕ_4 correspond to the stable solutions associated to A_∞. Note that neither ϕ_2 nor ϕ_3 is determined uniquely by these requirements; ϕ_1 and ϕ_4 are selected by the normalizations in (i) and (iv). The existence of the transmission function t_1 confirms the intuitive idea that a general solution of (3.6) will grow as the most unstable eigenfunction $e^{\Lambda_1 \xi}$ as $\xi \to \infty$. The existence of t_2 follows from the observation that ϕ_1 and ϕ_2 are independent: there will be orbits in $\Phi_-(\xi; \lambda, \hat{l})$ that do not grow as $e^{\Lambda_1 \xi}$ if $\xi \to \infty$. The growth of these orbits will then be determined by $e^{\Lambda_2 \xi}$. Thus the solution $\phi_2(\xi; \lambda, \hat{l})$ is selected as one of these orbits. Note that this construction implies that t_2 cannot be defined "automatically" for $t_1 = 0$.

The Evans function $\mathcal{D}(\lambda)$ is defined by

$$\mathcal{D}(\lambda, \hat{l}, \varepsilon) = \det[\phi_1(\xi; \lambda, \hat{l}), \phi_2(\xi; \lambda, \hat{l}), \phi_3(\xi; \lambda, \hat{l}), \phi_4(\xi; \lambda, \hat{l})]. \tag{3.13}$$

The determinant $\mathcal{D}(\lambda)$ does not depend on ξ (since the trace of $A(\xi)$ is 0 [1]) and is analytic as function of λ for $\lambda \in \mathcal{C}_\delta$ (or, in general, for λ outside the essential spectrum [1]). It follows from the general results of [1] that the zeros of $\mathcal{D}(\lambda)$ coincide with the eigenvalues of (3.6), counting multiplicities. Intuitively, this can be made clear by observing that an eigenfunction ϕ of (3.6), associated to an eigenvalue λ, must approach $(0,0,0,0)^t$ for both $\xi \to -\infty$ and $\xi \to +\infty$. This implies that $\phi \in \Phi_- \cap \Phi_+$ so that $\mathcal{D}(\lambda) = 0$. At the same time, $\Phi_- \cap \Phi_+ \neq \emptyset$ at a zero of $\mathcal{D}$. Hence there must be an eigenfunction $\phi \in \Phi_- \cap \Phi_+$.

The Evans function $\mathcal{D}(\lambda)$ can be decomposed into a product of $t_1(\lambda)$, $t_2(\lambda)$ and a nonzero component:

$$\begin{array}{rcll} \mathcal{D}(\lambda, \varepsilon) & = & \lim_{\xi \to \infty} & \det[\phi_1(\xi), \phi_2(\xi), \phi_3(\xi), \phi_4(\xi)] \\ & = & \lim_{\xi \to \infty} & \det[\phi_1(\xi)e^{-\Lambda_1 \xi}, \phi_2(\xi)e^{-\Lambda_2 \xi}, \phi_3(\xi)e^{-\Lambda_3 \xi}, \phi_4(\xi)e^{-\Lambda_4 \xi}] \\ & = & & \det[t_1 E_1, t_2 E_2, E_3, E_4] \\ & = & & 4\varepsilon t_1(\lambda, \hat{l}, \varepsilon) t_2(\lambda, \varepsilon, \hat{l}) \sqrt{(\mu + \lambda + \hat{l}^2)(1 + \lambda + \varepsilon^4 \hat{l}^2)} \end{array} \tag{3.14}$$

since $\sum_{i=1}^4 \Lambda_i(\lambda) \equiv 0$ (3.9). Thus the zeros of $\mathcal{D}(\lambda, \varepsilon)$ are determined by solving $t_1(\lambda) = 0$ and $t_2(\lambda) = 0$. Since $t_1(\lambda)$ is associated to the "fast" solution ϕ_1, i.e., the solution that behaves as $E_1 e^{\Lambda_1 \xi}$ as $\xi \to -\infty$, it is relatively straightforward to determine the zeros of $t_1(\lambda, \hat{l})$.

Lemma 3.3. *Let $\lambda_f^j(l) \in \mathbf{R}$ be an eigenvalue of the reduced limit problem*

$$(\mathcal{L}_f(\xi; l) - \lambda)v = v_{\xi\xi} + [\beta_2 u_h^{\alpha_2}(v_h(\xi))^{\beta_2 - 1} - (1 + \lambda + l^2)]v = 0, \tag{3.15}$$

where $u_h = U_1^{\max}$ (2.1) and $v_h(\xi) =$ the leading order approximation of $V_0(\xi)$, i.e., the (positive) homoclinic solution of $\ddot{v} = v - u_h^{\alpha_2} v^{\beta_2}$. Then there exists a unique $\lambda^j(l, \varepsilon)$ such that $t_1(\lambda^j(l, \varepsilon), l, \varepsilon) = 0$ and $\lim_{\varepsilon \to 0} \lambda^j(l, \varepsilon) = \lambda_f^j(l)$.

The proof is based on a winding number argument applied to a contour around $\lambda_f^j(l)$; see [9], [10] for the details. The result is, once again, quite natural. It follows from the structure of E_1 (3.9) and the approximation (3.8) that the u-component of $\phi_1(\xi)$ is asymptotically small for $\xi \ll -1$. Moreover, $u_{\xi\xi} = \mathcal{O}(\varepsilon^2)$ (3.4) for $\hat{l}$ not "too large" (see below). This yields that the u-component of $\phi_1(\xi)$ is also asymptotically small for $\xi = \mathcal{O}(1/\varepsilon^\sigma)$, for some $\sigma > 0$. Hence, as $\varepsilon \to 0$, the v-component of the solution ϕ_1 of (3.4) merges with a solution v_1 of the reduced fast limit problem (3.15) that converges to 0 as $\xi \to -\infty$. The assumption that the transmission function $t_1(\lambda)$ has a zero implies that $\phi_1(\xi)$ does not grow as $e^{\Lambda_1 \xi}$ for $\xi \to \infty$. In the limit $\varepsilon \to 0$, this is equivalent to assuming that the solution v_1 of (3.15) does not grow exponentially. Hence v_1 is an eigenfunction of (3.15), and λ must be asymptotic to an eigenvalue. This argument cannot be applied when $\hat{l}$ becomes "too large." However, if the u-component of ϕ_1 does not remain asymptotically small, it will grow exponentially in this case, which implies that t_1 cannot be 0 (see section 3.4).

3.3. The NLEP approach. The NLEP approach has been developed to determine the zeros of the slow transmission function t_2. Here we will sketch this method by assuming that $|l| \ll 1$, i.e., $|\hat{l}| \ll 1/\varepsilon^2$ (3.3). Furthermore, we assume that $\mu = \mathcal{O}(1)$. In section 3.4, we will consider the case $l \in \mathbf{R}$ and $\mu \gg 1$.

We introduce $\gamma \in (0, 2]$ and assume that $|\hat{l}| \ll 1/\varepsilon^{2-\gamma}$. It is clear from (3.4) that $u(\xi)$ remains constant (at leading order) on intervals of length $\le \mathcal{O}(1/\varepsilon^{\tilde{\gamma}})$, where $\tilde{\gamma} = \frac{1}{2}\min\{1, \gamma\}$. Therefore, we introduce the interval

$$I_\gamma = [-1/\varepsilon^{\tilde{\gamma}}, 1/\varepsilon^{\tilde{\gamma}}],\ \tilde{\gamma} = \frac{1}{2}\min\{1, \gamma\}. \tag{3.16}$$

By (3.8), we know that outside I_γ, the behavior of the solutions of (3.6) is dominated by the constant coefficients matrix $A_\infty(\lambda, \hat{l})$ (3.7). Thus we know by Lemma 3.2 that there are $\mathcal{O}(1)$ constants $C_-, C_+ > 0$ such that

(3.17)
$$\phi_2(\xi; \lambda, \hat{l}) = \begin{cases} E_2(\lambda, \hat{l}) e^{\Lambda_2(\lambda,\hat{l})\xi} + \mathcal{O}(e^{+C_-\xi}), & \xi < -\varepsilon^{-\tilde{\gamma}}, \\ t_2(\lambda, \hat{l}) E_2(\lambda, \hat{l}) e^{\Lambda_2(\lambda,\hat{l})\xi} + t_3(\lambda, \hat{l}) E_3(\lambda, \hat{l}) e^{\Lambda_3(\lambda,\hat{l})\xi} + \mathcal{O}(e^{-C_+\xi}), & \xi > +\varepsilon^{-\tilde{\gamma}}. \end{cases}$$

Here $t_3(\lambda, \hat{l}, \varepsilon)$ is a third meromorphic transmission function that determines the component of ϕ_2 that is associated to the Λ_3 eigenvalue of A_∞. The u-components of $E_{2,3} = 1$ (3.9); thus it follows that

$$t_2(\lambda, \hat{l}, \varepsilon) + t_3(\lambda, \hat{l}, \varepsilon) = 1 + \mathcal{O}(\varepsilon^{\tilde{\gamma}}) \ \text{ for } \ \xi \in I_\gamma \tag{3.18}$$

(see [9], [10] for all details). We can obtain a second relation between t_2 and t_3 by imposing a "matching condition" that couples the slow evolution, i.e., the (almost) linear flow dominated by A_∞, to the fast field. This idea was first developed outside the context of the Evans function in [8].

Since $u = 1$, at leading order in I_γ we observe that the v-equation decouples from the full system (3.4). For $\xi \in I_\gamma$, we can furthermore approximate U_0 by $u_h = U_1^{\max}$ and $V_0(\xi)$ by $v_h(\xi)$ (Lemma 3.3). Thus we obtain, at leading order,

$$\begin{aligned}(\mathcal{L}_f(\xi; \varepsilon^2\hat{l}) - \lambda)v &= v_{\xi\xi} + [\beta_2 u_h^{\alpha_2}(v_h(\xi))^{\beta_2-1} - (1 + \lambda + \varepsilon^4\hat{l}^2)]v \\ &= -\alpha_2 u_h^{\alpha_2-1}(v_h(\xi))^{\beta_2}, \quad \xi \in I_\gamma.\end{aligned} \tag{3.19}$$

Note that we could have neglected the $\varepsilon^4\hat{l}^2 v$ term; it does not have any leading order influence. However, we prefer to keep the presence of $\hat{l}$ explicit. The inhomogeneous problem (3.19) is of Sturm–Liouville type and thus has a unique bounded solution $v_{\text{in}}(\xi;\lambda)$ for $\lambda \in \mathcal{C}_\delta(\varepsilon^4\hat{l}^2) = \mathcal{C}_\delta(0)+$ h.o.t. (3.11) and $\lambda \neq \lambda_f^j(\varepsilon^4\hat{l}^2) = \lambda_f^j(0)+$ h.o.t. (Lemma 3.3); see also [10]. By construction, we know that $v_{\text{in}}(\xi)$ is the leading order approximation of $v_2(\xi)$, the v-component of $\phi_2(\xi) = (u_2(\xi), p_2(\xi), v_2(\xi), q_2(\xi))$. The evolution of $u_2(\xi)$ is thus at leading order governed by

$$u_{\xi\xi} = -\varepsilon^2[\alpha_1 u_h^{\alpha_1-1}(v_h(\xi))^{\beta_1} + \beta_1 u_h^{\alpha_1}(v_h(\xi))^{\beta_1-1}v_{\text{in}}(\xi)] \tag{3.20}$$

(3.4). From this we obtain a leading order approximation of the total change in the u-component of $\phi_2(\xi)$ through I_γ:

$$\Delta_{\text{fast}}u_\xi = -\varepsilon^2\int_{-\infty}^{\infty}\left[\alpha_1 u_h^{\alpha_1-1}(v_h(\xi))^{\beta_1} + \beta_1 u_h^{\alpha_1}(v_h(\xi))^{\beta_1-1}v_{\text{in}}(\xi)\right]d\xi, \tag{3.21}$$

where we have replaced the integration over I_γ by an integration over $\mathbf{R}$ since this does not have a leading order effect. This expression should match the leading order slow "jump" in u_ξ over I_γ, as is described by (3.17):

$$\Delta_{\text{slow}}u_\xi = u_\xi(1/\varepsilon^{\tilde{\gamma}}) - u_\xi(-1/\varepsilon^{\tilde{\gamma}}) = [\Lambda_2 t_2(\lambda,\hat{l}) + \Lambda_3 t_3(\lambda,\hat{l})] - \Lambda_2. \tag{3.22}$$

Thus we can solve $t_2(\lambda,\hat{l})$ by combining (3.18) with the "matching condition" $\Delta_{\text{fast}}u_\xi = \Delta_{\text{slow}}u_\xi$:

$$t_2(\lambda,\hat{l}) = 1 - \frac{1}{2\sqrt{\mu+\lambda+\hat{l}^2}}\int_{-\infty}^{\infty}\left[\alpha_1 u_h^{\alpha_1-1}v_h^{\beta_1} + \beta_1 u_h^{\alpha_1}v_h^{\beta_1-1}v_{\text{in}}\right]d\xi. \tag{3.23}$$

The first order corrections to (3.23) are $\mathcal{O}(\varepsilon^{\tilde{\gamma}})$ (3.16). Note that $t_2(\lambda;\hat{l})$ depends almost trivially on $\hat{l}$. "Slow" eigenvalues to (3.6) are now determined by solving $t_2(\lambda,\hat{l}) = 0$. The nonlocal eigenvalue problem, i.e., the combination of (3.19) and the equation $t_2(\lambda,\hat{l}) = 0$ (3.23), gets its name from the nonlocal term in (3.23); see Remark 3.6.

We can use the explicit information on u_h and $v_h(\xi)$ (Theorem 2.1 and Lemma 3.3) to obtain a somewhat less involved expression for $t_2(\lambda,\hat{l})$. We first note that the reduced fast limit problem (3.15) is equivalent to

$$(\mathcal{L}_f(\xi;l) - \lambda)w = w_{\xi\xi} + [\beta_2(w_h(\xi))^{\beta_2-1} - (1+\lambda+l^2)]w = 0. \tag{3.24}$$

Using hypergeometric functions, it is possible to determine the eigenvalues and eigenfunctions of this equation explicitly.

Lemma 3.4. *Let $J = J(\beta_2) \in \mathbf{N}$ be such that $J < (\beta_2+1)/(\beta_2-1) \leq J+1$. The eigenvalue problem* (3.24) *has $J+1$ eigenvalues given by*

$$\lambda_f^j(l) = \frac{1}{4}[(\beta_2+1) - j(\beta_2-1)]^2 - 1 - l^2 \;\text{ for }\; j = 0,1,\dots,J \tag{3.25}$$

so that $-1-l^2 < \lambda_f^J(l) < \lambda_f^{J-1}(l) < \cdots < \lambda_f^1(l) = -l^2 < \lambda_f^0(l)$. *The eigenfunctions* $w_f^j(\xi)$ *can be expressed explicitly in terms of* $w_h(\xi)$ *and* $\dot{w}_h(\xi)$; $w_f^j(\xi)$ *is even, respectively, odd, as function of* ξ *for* j *even, respectively, odd.*

See Remark 4.2 for the main ideas behind the proof of this result. Note that the eigenvalues $\lambda_f^j(\hat{l})$ are at leading order given by $\lambda_f^j(0)$ for $|\hat{l}| \ll 1/\varepsilon^2$ (3.3). Using (1.3) and (2.1), we can rewrite (3.23) as

$$t_2(\lambda,\hat{l}) = 1 - \frac{\sqrt{\mu}}{\sqrt{\mu+\lambda+\hat{l}^2}}\left[\alpha_1 - \frac{\alpha_2\beta_1}{W(\beta_1,\beta_2)}\int_{-\infty}^{\infty} w_{\text{in}} w_h^{\beta_1-1} d\xi\right], \tag{3.26}$$

where $W(\beta_1,\beta_2)$ is defined in (2.2) and $w_{\text{in}}(\xi) = w_{\text{in}}(\xi;\lambda)$ is the unique bounded solution of

$$(\mathcal{L}_f(\xi;\varepsilon^2\hat{l}) - \lambda)w = w_{\xi\xi} + [\beta_2(w_h(\xi))^{\beta_2-1} - (1+\lambda+\varepsilon^4\hat{l}^2)]w = (w_h(\xi))^{\beta_2}. \tag{3.27}$$

We know by Lemma 3.3 that the zeros of $t_1(\lambda)$ are at leading order given by (3.25). However, we now see by (3.26) that we should expect $t_2(\lambda)$ to have a pole (of order one) near these same eigenvalues (since $\mathcal{L}_f(\xi) - \lambda$ will, in general, not be invertible at an eigenvalue). The Evans function $\mathcal{D}(\lambda)$ is analytic in $\mathcal{C}_\delta$ [1], which implies that a pole of $t_2(\lambda)$ must coincide with a zero of $t_1(\lambda)$. Thus the eigenvalues of the fast reduced limit problem (3.15)/(3.24) do not automatically appear as eigenvalues of the full problem (3.6)! Nevertheless, some of the zeros of $t_1(\lambda)$ persist as zeros of $\mathcal{D}(\lambda)$ (and are thus eigenvalues of (3.6)). The slow transmission function $t_2(\lambda)$ does not have a pole near $\lambda_f^j(l)$, i.e., the solution $w_{\text{in}}(\xi)$ of (3.27) exists, if the following solvability condition is satisfied:

$$\int_{-\infty}^{\infty} (w_h(\xi))^{\beta_2} w_f^j(\xi) d\xi = 0.$$

Since $w_h(\xi)$ is even as a function of ξ (Lemma 3.4), we conclude by that $\mathcal{D}(\lambda)$ must have a zero near $\lambda_f^j(l)$ for j odd. We can now give a full characterization of the eigenvalues of (3.6).

Lemma 3.5. *Let* $\lambda \in \mathcal{C}_\delta$ *be a zero of* $\mathcal{D}(\lambda)$. *Then either* $t_2(\lambda) = 0$ *or* $\lambda \to \lambda_f^j$ (3.25) *with* j *odd as* $\varepsilon \to 0$.

Since the most unstable eigenvalue of (3.24), $\lambda_f^0(l)$, is cancelled by a pole of t_2, we may conclude that the stability of the homoclinic stripe pattern is determined by the zeros of the slow transmission function $t_2(\lambda, l)$.

Remark 3.6. The combination of (3.19) with $t_2(\lambda,\hat{l}) = 0$ (3.23) can, at leading order, be written in an equivalent but more standard and compact way:

$$v_{\xi\xi} + [\beta_2 u_h^{\alpha_2} v_h^{\beta_2-1} - (1+\lambda)]v = \frac{\alpha_2\beta_1 u_h^{\alpha_2-1} v_h^{\beta_2}}{\alpha_1 \int_{-\infty}^{\infty} u_h^{\alpha_1-1} v_h^{\beta_1} d\xi - 2\sqrt{\mu+\lambda+\hat{l}^2}} \int_{-\infty}^{\infty} u_h^{\alpha_1} v_h^{\beta_1-1} v d\xi,$$

where v must decay exponentially as $\xi \to \pm\infty$ (note that $v(\xi) = v_{\text{in}}(\xi)$ up to a multiplication factor). The NLEP equation was originally introduced in a form similar to this in [8].

3.4. $|\hat{l}| \gg 1$ and $\mu \gg 1$. It follows from the explicit expression (3.26) that $t_2(\lambda)$ must have zeros near its poles for μ small enough. In that case, $t_2(\lambda)$ is close to 1 on a contour K encircling the pole so that the winding number of t_2 over K must be zero. Since t_2 has a pole inside K (that gives a contribution of -1 to the winding number), it must also have a zero inside K. Complex eigenvalues come in pairs, and so this argument also implies that the eigenvalue is real. This result has been established in [10, Theorem 5.1]. The essence of the proof is the derivation of sufficiently small (and uniform) upper bounds on $|t_2 - 1|$ and $|\frac{d}{d\lambda}t_2|$ over the contour K. Due to the factor $\sqrt{\mu}/\sqrt{\mu + \lambda + \hat{l}^2}$ in (3.26), that can be made as small as necessary by decreasing μ; this is a straightforward procedure. (The integrals appearing in (3.26) can be estimated uniformly for λ on a contour that is bounded away from the pole [10].) Since $\hat{l}$ does appear only in $t_2(\lambda, \hat{l})$ (at leading order) through this same factor, we can immediately conclude by this winding number argument that $t_2(\lambda, \hat{l})$ must have a unique zero near $\lambda_f^j(l)$ with j even for $|\hat{l}| \gg 1$.

We have developed the NLEP procedure under the assumption that $|\hat{l}| \ll 1/\varepsilon^{2-\gamma}$ for some $\gamma \in (0, 2]$. It follows from (3.26) and the above argument that all possible zeros of $t_2(\lambda)$ must be asymptotically close to a pole of $t_2(\lambda)$ for $1 \ll |\hat{l}| \ll 1/\varepsilon^{2-\gamma}$. By Lemma 3.5, this implies that all zeros of $\mathcal{D}(\lambda)$ are asymptotically close to an eigenvalue of the reduced limit problem (3.24). The NLEP procedure cannot be used for larger $\hat{l}$; however, it is clear from (3.4) that this statement must also be valid for these $\hat{l}$. As soon as $|\hat{l}|$ becomes $\gg 1/\varepsilon$ (i.e., $\gamma < 1$), the u-equation decouples, at leading order, from the system (3.4):

$$u_{\xi\xi} = \varepsilon^4[\mu + \lambda + \hat{l}^2]u + \text{ h.o.t.}$$

This equation cannot have a nontrivial bounded solution. This either implies that u must be asymptotically small at an eigenvalue of (3.4) when $|\hat{l}| \gg 1/\varepsilon$ or, equivalently, that v must be asymptotically large. (Note that u has been scaled to be close to 1 in the development of the NLEP approach.) Hence v must be a solution of (3.15) at leading order (Lemma 3.3).

Lemma 3.7. *Assume that $\mu = \mathcal{O}(1)$ and that $|\hat{l}| \gg 1$, i.e., $|l| \gg \varepsilon^2$ (3.3). All eigenvalues $\lambda(l)$ of (3.6) are real and asymptotically close to an eigenvalue $\lambda_f^j(l)$ of the fast reduced limit problem (3.4)/(3.24).*

This result also implies that (3.6) has only nontrivial eigenvalues for $\hat{l} = \mathcal{O}(1)$.

So far we have not considered the case $\mu \gg 1$. The extension of the analysis to $0 < \mu = \tilde{\mu}/\varepsilon^m \ll 1/\varepsilon^4$, i.e., $m \in [0, 4)$, is straightforward: we have seen in section 2.1 that (1.4) can be scaled to (2.8). The only difference between these two equations is the magnitude of the factors in front of the U_t and the $\tilde{U}_t$, ε^2 and $\tilde{\varepsilon}^2 \times \tilde{\varepsilon}^{4m/(4-m)}$. This introduces an extra factor $\tilde{\varepsilon}^{4m/(4-m)}$ in the Evans function/NLEP analysis that is asymptotically small (for $m > 0$) so that (3.26) reduces to

$$\tilde{t}_2(\lambda, \hat{\tilde{l}}) = 1 - \frac{\sqrt{\tilde{\mu}}}{\sqrt{\tilde{\mu} + \hat{\tilde{l}}^2}} \left[\alpha_1 - \frac{\alpha_2\beta_1}{W(\beta_1, \beta_2)} \int_{-\infty}^{\infty} w_{\text{in}}(\tilde{\xi}; \lambda)(w_h(\tilde{\xi}))^{\beta_1 - 1} d\tilde{\xi}\right], \tag{3.28}$$

with $\hat{\tilde{l}}$ the $m > 0$ equivalent of $\hat{l}$ (2.7), (3.3):

$$l = \tilde{\varepsilon}^2 \hat{\tilde{l}} = \varepsilon^{2-m/2} \hat{\tilde{l}}. \tag{3.29}$$

Thus, for $\mu \gg 1$, $\tilde{t}_2$ depends only on λ through w_{in} (3.27). To understand the evolution of the eigenvalues of (3.6) as μ is increased from $\mathcal{O}(1)$ to $\mathcal{O}(1/\varepsilon^4)$, we need to formulate an equivalent of Lemma 3.7 for $\mu = \mathcal{O}(1/\varepsilon^m)$ in terms of the original, unscaled, wave number l.

Lemma 3.8. *Assume that $\mu = \mathcal{O}(1/\varepsilon^m)$ with $m \in [0, 4)$. All eigenvalues $\lambda(l)$ of (3.6) are real and asymptotically close to a fast reduced eigenvalue $\lambda_f^j(l)$ for $|l| \gg \varepsilon^{2-m/2}$; $\lambda(l) = \lambda(0)$ are at leading order for $|l| \ll \varepsilon^{2-m/2}$; i.e., for $|l| \ll \varepsilon^{2-m/2}$, the eigenvalues of (3.6) are at leading order given by those that determine the stability of the one-dimensional pulse problem.*

The proof follows immediately from (2.7), (3.2), and (3.3). Note that $\tilde{l} = l$ but that the analysis for $m > 0$ is done in terms of $\tilde{\varepsilon}$ so that $\hat{\tilde{l}} \neq \hat{l}$ (3.29): $|l| \ll \varepsilon^{2-m/2}$ corresponds to $|\hat{\tilde{l}}| \ll 1$, which simplifies (3.28) even further. As was the case for the geometric singular perturbation approach to the existence problem in section 2, we cannot use the NLEP approach to study the stability of the stripes when m is increased to 4 (since $\tilde{\varepsilon}$ becomes $\mathcal{O}(1)$ for $m = 4$ (2.7)). This case will be considered in section 5.

Lemmas 3.7 and 3.8, of course, have an immediate impact on the stability of the homoclinic stripe pattern $(U_0(x), V_0(x))$. In [10], it has been shown that the one-dimensional pulse pattern can be (spectrally) stable, i.e., that all eigenvalues $\lambda(l)$ can satisfy $\text{Re}(\lambda(0)) \leq 0$ (depending on the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$, and μ; see section 4). However, for $|l| \gg \varepsilon^{2-m/2}$, there is a (real) eigenvalue of (3.6) near $\lambda_f^0(l)$, and this eigenvalue can be positive (3.25). We define the critical value $l_{\text{R,stab}}$ of l by

$$l_{\text{R,stab}} = \sqrt{\frac{1}{4}(\beta_2 + 1)^2 - 1} \tag{3.30}$$

so that $\lambda_f^0(l) < 0$ for $|l| > l_{\text{R,stab}}$. Lemma 3.8 can be applied to $\pm l$ near $l_{\text{R,stab}}$ if $\varepsilon^{2-m/2} \ll 1$ (3.30), i.e., if $2 - m/2 > 1$. Thus we recover the same critical boundary $m < 4$ as in the existence analysis (Theorem 2.2)! Since $\lambda_f^0(l)$ becomes positive as $|l|$ decreases through $l_{\text{R,stab}}$, we conclude the following theorem.

Theorem 3.9. *The homoclinic stripe pattern $(U(x, y, t), V(x, y, t)) = ((U_0(x), V_0(x))$ is unstable as a solution of (1.4) for $(x, y) \in \mathbf{R}^2$, for $0 < \mu \ll 1/\varepsilon^4$.*

This result does not exclude the possibility of asymptotically stable homoclinic stripe patterns on $\mathbf{R}^2$, as we shall see in section 5. By Theorem 2.2 we know that $(U_0(x), V_0(x))$ exist up to $\mu = \tilde{\mu}_{\text{split}}/\varepsilon^4$, and the theorem does not include case $m = 4$. Moreover, this result does not give any insight into the transition from the (possibly stable and complex) eigenvalues in the one-dimensional case ($l = 0$) to the unstable real eigenvalues at $|l| \gg \varepsilon^{2-m/2}$. This will be discussed in detail in section 4. For $0 < m < 4$, it follows from Lemma 3.8 that the transition occurs at $|l| = \mathcal{O}(\varepsilon^{2-m/2})$, and we expect two symmetrical "bands" of unstable wave numbers—the one with $l > 0$ being bounded from below by an $\mathcal{O}(\varepsilon^{2-m/2})$ expression and from above by $l_{\text{R,stab}} = \mathcal{O}(1)$. As $m \uparrow 4$, the width of the unstable bands decreases, and all boundaries become $\mathcal{O}(1)$. This implies that the bands of unstable wave numbers might disappear as μ becomes $\mathcal{O}(1/\varepsilon^4)$, i.e., that the homoclinic stripe pattern might become stable. See also Remark 4.7 for an explicit, quantitative, but formal refinement of this argument. Note, however, that the stripe patterns can only be stabilized when the band of unstable wave numbers disappears before μ reaches $\mu_{\text{split}} = \mathcal{O}(1/\varepsilon^4)$, the upper boundary on the existence domain of the homoclinic stripes (Theorem 2.2). Since $m = 4$ corresponds to $\tilde{\varepsilon} = 1$ (2.7), we cannot study this case by the asymptotic NLEP analysis. In section 5, we will consider the stability of the homoclinic stripe patterns for μ up to μ_{split} by numerical simulations.

Remark 3.10. The instability associated to $l_{\text{R,stab}}$ is of so-called varicose type, since the associated eigenfunction, $w_f^0(\xi)$, is even as a function of ξ (Lemma 3.4); see [32], [18]. This is confirmed by the numerical simulations of section 5: there, it is shown that $l_{\text{R,stab}}$ marks the transition of a stripe pattern to a spot pattern.

4. The eigenvalues as function of l.

In this section, we determine all eigenvalues $\lambda(l)$ of (3.6) explicitly as a function of $l \in \mathbf{R}$. Lemma 3.7 provides all relevant information of $\lambda(l)$ for $l \gg \varepsilon^2$, i.e., $\hat{l} \gg 1$, so that we need only to consider $\hat{l} = \mathcal{O}(1)$ here. The case $\mu = \mathcal{O}(1/\varepsilon^m)$, $m \in (0,4)$, can be considered as a subcase of $\mu = \mathcal{O}(1)$ since (3.28) is equivalent to considering $\mu \gg 1$ and $\hat{l} \gg 1$ in (3.26), as we have seen above.

4.1. Hypergeometric functions.

The NLEP equation, i.e., the system of $t_2(\lambda, l) = 0$ (3.26) coupled to (3.27), can be solved explicitly (with the aid of Mathematica) by transforming (3.27) into a hypergeometric differential equation. As in [10], we first consider (3.24) and introduce $P = P(\lambda, l)$, $F = F(\xi; P)$, and the new independent variable z by

$$P(\lambda, l) = \sqrt{1 + \lambda + l^2}, \; w(\xi) = F(\xi)(w_h(\xi))^P, \; z = \frac{1}{2}\left(1 - \frac{\dot{w}_h(\xi)}{w_h(\xi)}\right), \tag{4.1}$$

where $w_h(\xi) = w_h(\xi, \beta_2)$ is the homoclinic solution defined by (2.3) that can be expressed explicitly by

$$w_h(\xi, \beta_2) = \left(\frac{\beta_2 + 1}{2\cosh^2 \frac{1}{2}(\beta_2 - 1)\xi}\right)^{\frac{1}{\beta_2 - 1}} \tag{4.2}$$

[4]. Note that $P = P(\lambda, 0)$ at leading order in the region where the zeros of $t_2(\lambda)$ are not prescribed by Lemma 3.7. The eigenvalue problem (3.24) can now be written as a hypergeometric differential equation [26] for F as function of z:

$$z(1-z)F'' + (1-2z)\frac{\beta_2 + 2P - 1}{\beta_2 - 1}F' + 2\frac{(\beta_2 - P)(\beta_2 + 1) - 2P(P-1)}{(\beta_2 - 1)^2}F = 0 \tag{4.3}$$

(see Remark 4.2). The inhomogeneous problem (3.27) can be transformed in precisely the same fashion. Following (4.1), we introduce $G(z)$ by

$$F(z; P, \beta_2) = \frac{[2(\beta_2 + 1)]^{\frac{\beta_2 - P}{\beta_2 - 1}}}{(\beta_2 - 1)^2} G(z; P, \beta_2) \tag{4.4}$$

so that (3.19) can be written as an inhomogeneous hypergeometric differential equation:

$$\begin{aligned} z(1-z)G'' &+ (1-2z)\frac{\beta_2 + 2P - 1}{\beta_2 - 1}G' \\ &+ 2\frac{(\beta_2 - P)(\beta_2 + 1) - 2P(P-1)}{(\beta_2 - 1)^2}G = [z(1-z)]^{\frac{1-P}{\beta_2 - 1}}. \end{aligned} \tag{4.5}$$

This equation can be solved explicitly in terms of the independent hypergeometric functions that are solutions of the associated homogeneous problem (4.3); $G(z;P)$ is determined uniquely by the condition that $w_{\text{in}}(\xi)$ is bounded for $\xi \to \pm\infty$; see [10] for the details. We introduce

$$\begin{aligned} \mathcal{R}(P;\beta_1,\beta_2) &= \tfrac{1}{B(\beta_1,\beta_2)} \int_0^1 G(z;P,\beta_2)[z(1-z)]^{\frac{P+\beta_1-\beta_2}{\beta_2-1}}\,dz, \\ \text{with }\ B(\beta_1,\beta_2) &= \tfrac{(\beta_2-1)^2}{2\beta_1(\beta_2+1)} \int_0^1 [z(1-z)]^{\frac{\beta_1-\beta_2+1}{\beta_2-1}}\,dz, \end{aligned} \tag{4.6}$$

where the integral in $B(\beta_1,\beta_2)$ comes from the term $W(\beta_1,\beta_2)$ (2.2) so that the equation for $t_2(\lambda)$ (3.26) can be written as

$$t_2(\lambda,\hat{l}) = 1 - [\alpha_1 - \alpha_2\mathcal{R}(P;\beta_1,\beta_2)]\sqrt{\frac{\mu}{\mu+\lambda+\hat{l}^2}}, \tag{4.7}$$

with $P = P(\lambda,l)$ (4.1). Thus we conclude that the equation $t_2(\lambda,\hat{l}) = 0$ can be solved explicitly by (4.7) (with the use of Mathematica). In general, its solutions, i.e., the eigenvalues of (3.6), will be complex. The real eigenvalues can be found through the graph of $\mu_{\text{real}}(\lambda)$, where μ_{real} solves $t_2(\lambda) = 0$:

$$\mu_{\text{real}}(\lambda,\hat{l}) = \frac{\lambda+\hat{l}^2}{\left[\alpha_1 - \alpha_2\mathcal{R}(P(\lambda,\hat{l});\beta_1,\beta_2)\right]^2 - 1}, \tag{4.8}$$

with the extra condition

$$\alpha_1 - \alpha_2\mathcal{R}(P;\beta_1,\beta_2) \geq 0. \tag{4.9}$$

In section 4.3, we will use (4.8) to derive some general (in)stability results. It is clear that understanding $\mathcal{R}(P;\beta_1,\beta_2)$ is crucial to the analysis of $\mu_{\text{real}}(\lambda,\hat{l})$. The following (technical) lemma extends the results on $\mathcal{R}(P;\beta_1,\beta_2)$ in [10].

Lemma 4.1. (i) *$\mathcal{R}(P;\beta_1,\beta_2)$ and its derivative can be determined explicitly at $P=1$:*

$$\mathcal{R}(1,\beta_1,\beta_2) = \frac{\beta_1}{\beta_2-1}, \quad \frac{\partial}{\partial P}\mathcal{R}(1,\beta_1,\beta_2) = \frac{2\beta_1-\beta_2+1}{(\beta_2-1)^2}, \tag{4.10}$$

where $P=1$ corresponds to $\lambda = \lambda_f^1 (= 0$, at leading order, for $|l| \ll 1$ (4.1)).

(ii) *The leading order behavior of $\mathcal{R}(P;\beta_1,\beta_2)$ is, for $P \gg 1$, given by*

$$\mathcal{R}(P;\beta_1,\beta_2) = -\frac{\beta_1^2(\beta_2+1)}{2\beta_1+\beta_2-1}\frac{1}{P^2} + \mathcal{O}\left(\frac{1}{P^4}\right). \tag{4.11}$$

These results give only leading order approximations. It is possible to determine higher corrections (as can be seen from the proof below). Moreover, the equivalents of *(i)* can be obtained for all (eigen)values of P that correspond to a λ_f^j with j odd.

Proof. The key to the derivation of (4.10) is the observation that (4.5) or, equivalently, (3.27), has a simple solution, i.e., not in terms of hypergeometric functions, at $P = 1$, or $\lambda = \lambda_f^1 = 0$ at leading order; see also [10]. The solution to (3.27) at $\lambda = 0$ is given by

$$w_{\text{in}}(\xi; 0) = \frac{1}{\beta_2 - 1} w_h(\xi) + C\dot{w}_h(\xi), \tag{4.12}$$

where $C \in \mathbf{R}$ is a free constant. (The solution to (3.27) cannot be unique since λ_f^1 is an eigenvalue; $\dot{w}_h(\xi)$ is the eigenfunction $w_f^1(\xi)$ associated to λ_f^1; recall that $w_h(\xi)$ has been defined in (2.3).) The function $\mathcal{R}(P; \beta_1, \beta_2)$ can, of course, also be expressed in terms of $w_{\text{in}}(\xi)$ instead of $G(z)$ (4.6):

$$\mathcal{R}(\lambda; \beta_1, \beta_2) = \frac{\beta_1}{W(\beta_1, \beta_2)} \int_{-\infty}^{\infty} w_{\text{in}}(\xi; \lambda)(w_h(\xi))^{\beta_2 - 1} d\xi \tag{4.13}$$

(see (2.2)). Substituting (4.12) into (4.13) yields the first part of (4.10). Note that the C drops out of the equation since $\dot{w}_h(\xi)$ is odd as a function of ξ. The second identity in (4.10) can be obtained in a similar fashion.

The leading order behavior of $\mathcal{R}(P; \beta_1, \beta_2)$ for $P^2 = 1 + \lambda + \hat{l}^2 \gg 1$ (4.1) can again be obtained from (3.27) and (4.13). We decompose $w_{\text{in}}(\xi; P^2)$ into

$$w_{\text{in}}(\xi; P) = \frac{1}{P^2} w_1(\xi) + \frac{1}{P^4} w_r(\xi; P)$$

and substitute this into (3.27). It follows that $w_1(\xi) = -(w_h(\xi))^{\beta_2}$ and that $w_r(\xi; P) = \mathcal{O}(1)$ with respect to the small parameter $1/P^2$. Together with the explicit expression (4.2), the expansion of w_{in} can now be used to evaluate the leading order behavior of $\mathcal{R}(P; \beta_1, \beta_2)$ by (4.13). ■

Remark 4.2. An eigenfunction of (3.24) is a solution that decays for $\xi \to \pm\infty$ and therefore corresponds by (4.1) to a solution $F(z)$ of (4.3) that is regular at both $z = 0$ and $z = 1$. Lemma 3.4 can now be proved using the classical theory on hypergeometric functions (see [26]), by which the local expansions of $F(z)$ near $z = 0$ and $z = 1$ can be studied (see [10] for all details).

4.2. The classical Gierer–Meinhardt equation.

In this subsection, we consider the special case $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = 2, \beta_2 = 2$ in full detail. However, some of our results extend immediately to the general case and will thus be presented in the general setting. The classical parameter combination corresponds to the original (biological) values of the parameters in [15]. Since $\beta_2 = 2$, there are three eigenvalues to the fast reduced equation (3.24): $\lambda_f^0(l) = \frac{5}{4} - l^2$, $\lambda_f^1(l) = 0 - l^2$, and $\lambda_f^2(l) = -\frac{3}{4} - l^2$ (Lemma 3.4).

We first consider the stability of the one-dimensional pulse pattern (that was already established in [10]); i.e., we search for eigenvalues $\lambda^j(\mu, l)$ of (3.6) with $l = 0$. For μ small enough, there are three eigenvalues: $\lambda^0(\mu, 0)$, $\lambda^1(\mu, 0)$, and $\lambda^2(\mu, 0)$. Two of them already occurred in the general analysis of section 3. There is a solution $\lambda^0(\mu, 0)$ of $t_2(\lambda, 0) = 0$ close to the pole $\frac{5}{4}$ (see section 3.4 or Theorem 5.1 in [10]); $\lambda^2(\mu, 0) \equiv 0$ is the "trivial" solution of $t_1(\lambda, 0) = 0$, i.e., the eigenvalue that corresponds to the translation symmetry in (1.4). Note

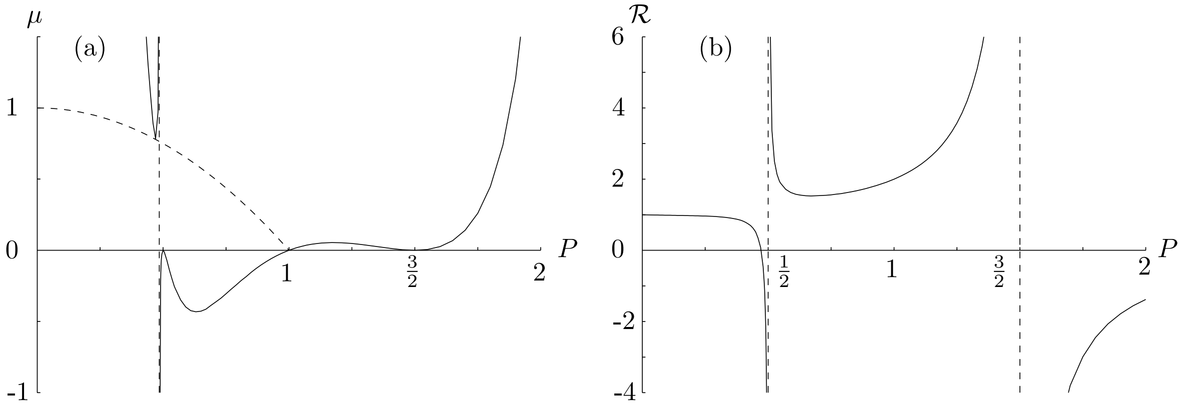


Figure 4.1. *The curves* (a) $\mu_{\mathrm{real}}(P, 0)$ *and* (b) $\mathcal{R}(P; \beta_1, \beta_2)$ *for the classical Gierer–Meinhardt equation.*

that there is no eigenvalue (yet) near the pole of $t_2(\lambda, 0)$ at $-\frac{3}{4}$ since this pole is imbedded in the essential spectrum $\sigma_{\mathrm{ess}}(0)$ for $\mu > 0$ small enough (3.10); this eigenvalue will appear from the essential spectrum as an edge bifurcation when μ increases (see also Remark 4.8). However, for $0 < \mu \ll 1$, there is a second zero of $t_1(\lambda, 0)$ close to $\lambda = 0$ that can be determined explicitly by using (4.10) to solve $t_2(\lambda, 0) = 0$ (4.7) for $\lambda = \mathcal{O}(\mu)$: $\lambda^1(\mu, 0) = 3\mu + \mathcal{O}(\mu^2)$.

Of course, this is not special for the classical case: all three eigenvalues $\lambda^j(\mu, 0)$ also exist for the general homogeneous case [10], and $\lambda^1(\mu, 0)$ can be determined explicitly for all parameters that satisfy (1.2):

$$\lambda^1(\mu, 0) = \frac{D^2 + 2(\beta_2 - 1)D}{(\beta_2 - 1)^2}\mu + \mathcal{O}(\mu^2) \ \text{ for } \ 0 < \mu \ll 1, \tag{4.14}$$

with D as in (1.3)—see also the proof of Theorem 4.10.

We use Mathematica to compute $\mathcal{R}(P; \beta_1, \beta_2)$, $t_2(\lambda, 0)$, and $\mu_{\mathrm{real}}(\lambda, 0)$ so that we can follow $\lambda^0(\mu, 0)$, $\lambda^1(\mu, 0)$, and $\lambda^2(\mu, 0)$ as μ is increased. Condition (4.9) is met for all λ between the poles $\lambda^0_f = \frac{5}{4}$ and $\lambda^2_f = -\frac{3}{4}$; therefore, $\mu_{\mathrm{real}}(\lambda, 0)$ is defined on this interval. However, $\mu_{\mathrm{real}}(\lambda, 0)$ is only positive on the interval $(0, \lambda^0_f)$, and $\mu_{\mathrm{real}}(0, 0) = \mu_{\mathrm{real}}(\frac{5}{4}, 0) = 0$ (see (4.8); recall that $\mathcal{R}(P)$ has a pole at $\lambda = \frac{5}{4}$). Thus we recover for μ small enough the eigenvalues $\lambda^0(\mu, 0)$ and $\lambda^1(\mu, 0)$. The graph of $\mu_{\mathrm{real}}(\lambda, 0)$ has a maximum value of $\mu_{\mathrm{complex}} = \mu_{\mathrm{complex}}(0) = 0.053\ldots$, and at this value the two real eigenvalues $\lambda^0(0)$ and $\lambda^1(0)$ merge and become a pair of complex conjugated eigenvalues; see Figure 4.1(a). If μ is increased further, the real part of the two complex eigenvalues decreases and eventually becomes negative. Thus the homoclinic pulse is stabilized by a Hopf bifurcation at $\mu = \mu_{\mathrm{Hopf}} = \mu_{\mathrm{Hopf}}(0) = 0.36\ldots$. All eigenvalues remain in the stable half plane for all $\mu_{\mathrm{Hopf}}(0) < \mu \ll 1/\varepsilon^4$. The upper bound on μ reflects the fact that we cannot analyze the case $\mu = \mathcal{O}(1/\varepsilon^4)$ by the NLEP method (see section 3.4). In section 5, we will see that the pulses are (numerically) stable up to $\mu_{\mathrm{split}} = \mathcal{O}(1/\varepsilon^4)$. The following result is a (straightforward) extension to $\mu \gg 1$ of Theorem 5.11 in [10].

Theorem 4.3 (see [10]). *Let $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = 2, \beta_2 = 2$, and $0 < \mu \ll 1/\varepsilon^4$. The pulse solution $(U_0(x), V_0(x))$ is (spectrally) stable as a solution of the one-dimensional Gierer–Meinhardt equation, i.e., (1.4), in which there is no y-dependence, for $\mu > \mu_{\mathrm{Hopf}} = \mu_{\mathrm{Hopf}}(0) =$ $0.36\cdots + \mathcal{O}(\varepsilon)$ and unstable for $\mu < \mu_{\mathrm{Hopf}}$.*

As was already noted, $\lambda^0(\mu,0)$, $\lambda^1(\mu,0)$, and $\lambda^2(\mu,0)$ are not the only eigenvalues: at points where $\alpha_1 - \alpha_2\mathcal{R}(P;\beta_1,\beta_2) = \mathcal{R}(P;2,2) = 0$, so that the graph of $\mu_{\text{real}}(\lambda,\hat{l})$ is tangent to the edge of the essential spectrum (3.10), there is an edge bifurcation at which a fourth eigenvalue $\lambda^3(\mu,0) \in \mathbf{R}$ "pops" out of the essential spectrum. This occurs at $\mu = \mu_{\text{edge}} = \mu_{\text{edge}}(0) = 0.77\cdots + \mathcal{O}(\varepsilon) = -\lambda^3(\mu_{\text{edge}}(0),0)$. Note that $\lambda^3(\mu,0)$ appears just below the second pole $\lambda_f^2 = -\frac{3}{4}$. The eigenvalue $\lambda^3(\mu,0)$ exists for all $\mu_{\text{edge}}(0) < \mu \ll 1/\varepsilon^4$; it moves slowly toward $\lambda = -1$ as μ increases. Since $\lambda^3(\mu,0)$ remains real and negative for all μ, it plays no role in the stability analysis (see also Remark 4.8).

The function $\mu_{\text{real}}(\lambda,\hat{l})$ and the bifurcation values μ_{Hopf}, μ_{complex}, and μ_{edge} are functions of $\hat{l}$. Therefore, the bifurcations will vary with $\hat{l}$ and may even disappear. Moreover, when $\hat{l} \neq 0$, it is possible that the eigenvalue $\lambda = 0$ is no longer isolated: one of the real eigenvalues can move through 0. The influence of $\hat{l}$ on $\mu_{\text{real}}(\lambda,\hat{l})$ is eminent from (4.8). It is possible to express $\mu_{\text{real}}(\lambda,\hat{l})$ in terms of $\mu_{\text{real}}(\lambda,0)$ by using the fact that all quantities involved are real:

$$\mu_{\text{real}}(\lambda,\hat{l}) = \mu_{\text{real}}(\lambda,0)\left(1 + \frac{\hat{l}^2}{\lambda}\right). \tag{4.15}$$

Note that this identity holds for any (allowed) combination of $\alpha_1, \alpha_2, \beta_1, \beta_2$; it is not special for the classical case. For positive λ, the multiplication shifts $\mu_{\text{real}}(\lambda,\hat{l})$ upward with respect to $\mu_{\text{real}}(\lambda,0)$. Furthermore, the multiplication has a stronger effect for lower λ, which results in a shift of the maximum of the graph of $\mu_{\text{real}}(\lambda,\hat{l})$ toward lower λ: the value of $\mu_{\text{complex}}(\hat{l})$ will increase with increasing $\hat{l}$, whereas the corresponding eigenvalue will decrease. Between λ_f^2 and $\lambda_f^1 = 0$, $\mu_{\text{real}}(\lambda,0)$ is defined (condition (4.9)) but negative (4.8); see Figure 4.1(a) and (b). (Note that (4.9) reduces to $\mathcal{R}(P;2,2) \geq 0$.) Since the multiplication factor is negative when $-\hat{l}^2 < \lambda < 0$, $\mu_{\text{real}}(\lambda,\hat{l})$ is positive for these λ. Therefore, there can be negative real eigenvalues for $\hat{l} \neq 0$; see Figure 4.2. The function $\mu_{\text{real}}(\lambda,\hat{l})$ is always zero at $\lambda = \lambda_f^2(\hat{l})$; thus, by the pole in $\mathcal{R}(P)$, the limit $\mu \to 0$ of $\lambda^1(\mu,\hat{l})$ is given by $\max(-\hat{l}^2, -\lambda_f^2(\hat{l}))$ (recall that $\lambda^1(\mu,0)$ is the positive real eigenvalue near 0 for μ small enough). Thus, $\lambda^1(\mu,\hat{l}) < 0$ for μ small enough, and it increases with μ. It crosses through zero at some critical value of μ, called $\mu_{\text{double}}(\hat{l})$; see Figure 4.2. Note that for $\hat{l} = \mathcal{O}(1)$ there are two eigenvalues at zero (at leading order) for this value of μ: $\lambda^1(\mu,\hat{l})$ and $\lambda^2(\mu,\hat{l})$ (Lemmas 3.4 and 3.5). The value of $\mu_{\text{double}}(\hat{l})$ can by construction be found by evaluating $\mu_{\text{real}}(\lambda,\hat{l})$ (4.7) at $\lambda = 0$, i.e., $P = 1$, at leading order (4.1). Since this argument is neither special for the classical Gierer–Meinhardt case nor for $\mu = \mathcal{O}(1)$, we use (4.9) and (1.3) and formulate the outcome in its most general setting.

Corollary 4.4. *Assume that $\mu = \tilde{\mu}/\varepsilon^m = \mathcal{O}(1/\varepsilon^m)$ with $0 \leq m < 4$ and that $|l| \ll 1$. Then, for any given l, there is a uniquely determined $\mu_{\text{double}}(l)$, with*

$$\mu_{\text{double}}(l) = \frac{(\beta_2 - 1)^2}{D^2 + 2(\beta_2 - 1)D}\frac{l^2}{\varepsilon^4} \ll \frac{1}{\varepsilon^4}, \quad \tilde{\mu}_{\text{double}} = \frac{(\beta_2 - 1)^2}{D^2 + 2(\beta_2 - 1)D}\hat{\tilde{l}}^2 \tag{4.16}$$

(3.29), at leading order, such that eigenvalue problem (3.6) has a double eigenvalue at $\lambda = 0$.

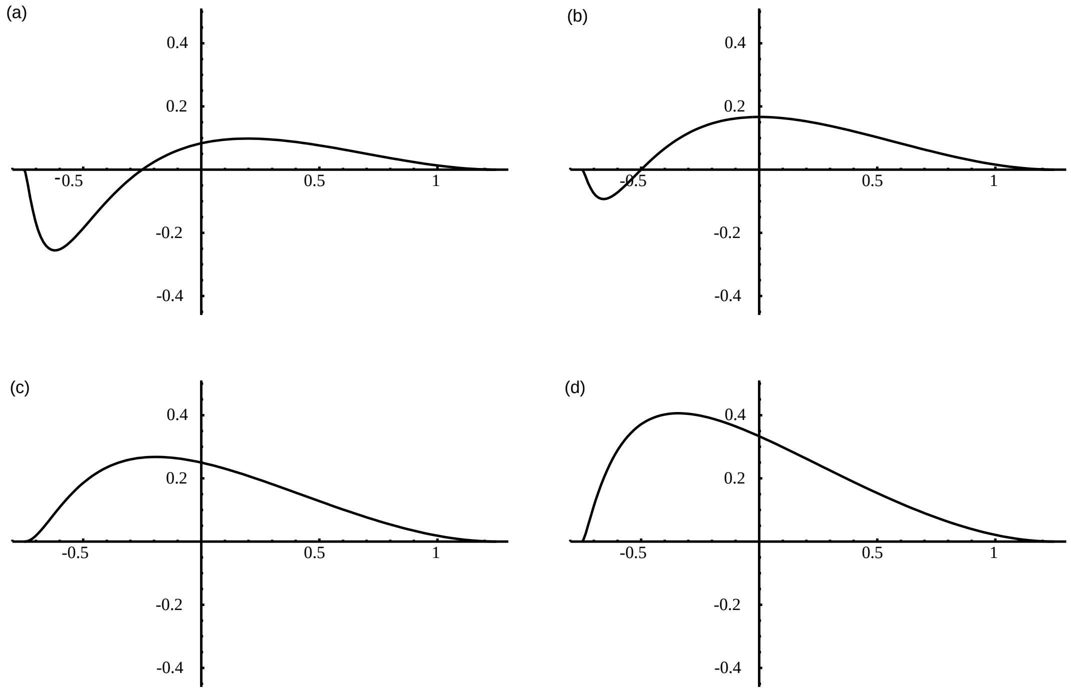


Figure 4.2. *The function* $\mu_{\mathrm{real}}(\lambda, \hat{l})$ *for various choices of* $\hat{l} > 0$*:* $\hat{l} = \frac{1}{2}$*,* $\hat{l} = \frac{1}{2}\sqrt{2}$*,* $\hat{l} = \frac{1}{2}\sqrt{3}$*,* $\hat{l} = 1$*.*

Equivalently, for any μ*, there is a uniquely determined value* $l_{\mathrm{L,stab}} > 0$ *of* l*, with*

$$(4.17) \qquad l_{\mathrm{L,stab}} = \varepsilon^2 \frac{\sqrt{D^2 + 2(\beta_2 - 1)D}}{\beta_2 - 1}\sqrt{\mu} \ll 1, \quad \hat{\hat{l}}_{\mathrm{L,stab}} = \frac{\sqrt{D^2 + 2(\beta_2 - 1)D}}{\beta_2 - 1}\sqrt{\tilde{\mu}},$$

at leading order, such that the nonlocal eigenvalue problem $t_2(\lambda; l) = 0$ *has a solution at* $\lambda = 0$*.*

Thus, although the eigenvalue problem (3.6) can have many eigenvalues (Lemmas 3.4 and 3.5), there is only one value of μ for which an eigenvalue can cross through $\lambda = 0$ (for a given $|l| \ll 1$). Note that in the classical case with $\mu = \mathcal{O}(1)$, (4.16) and (4.17) reduce, at leading order, to

$$(4.18) \qquad \mu_{\mathrm{double}}(l) = \frac{l^2}{3\varepsilon^4} = \frac{\hat{l}^2}{3} \text{ and } l_{\mathrm{L,stab}} = \varepsilon^2\sqrt{3\mu} \text{ or } \hat{l}_{\mathrm{L,stab}} = \sqrt{3\mu}.$$

In Figure 4.3, we have used the expression for $t_2(\lambda, \hat{l}, \mu)$ in terms of hypergeometric functions to plot the "orbits" of the eigenvalues $\lambda^0(\mu, \hat{l})$ and $\lambda^1(\mu, \hat{l})$ as function of μ for several values of $\hat{l}$. There is a critical value of $\hat{l}$ at which the Hopf bifurcation disappears in the sense that both eigenvalues $\lambda^{0,1}(\mu, \hat{l})$ become (or remain) real and negative before they merge. At this bifurcation value of $\hat{l}$, $\hat{l} = \hat{l}_{\mathrm{triple}}$, $\mu_{\mathrm{complex}}(\hat{l})$, and $\mu_{\mathrm{double}}(\hat{l})$ merge so that, by definition, $\mu_{\mathrm{complex}}(\hat{l}_{\mathrm{triple}}) = \mu_{\mathrm{double}}(\hat{l}_{\mathrm{triple}}) = \mu_{\mathrm{Hopf}}(\hat{l}_{\mathrm{triple}}) \overset{\mathrm{def}}{=} \mu_{\mathrm{triple}}$; see Figures 4.2(b), 4.3(c), and 4.4. This value of $\hat{l}$ can determined explicitly by taking the derivative of $\mu_{\mathrm{real}}(\lambda, \hat{l})$ with respect to

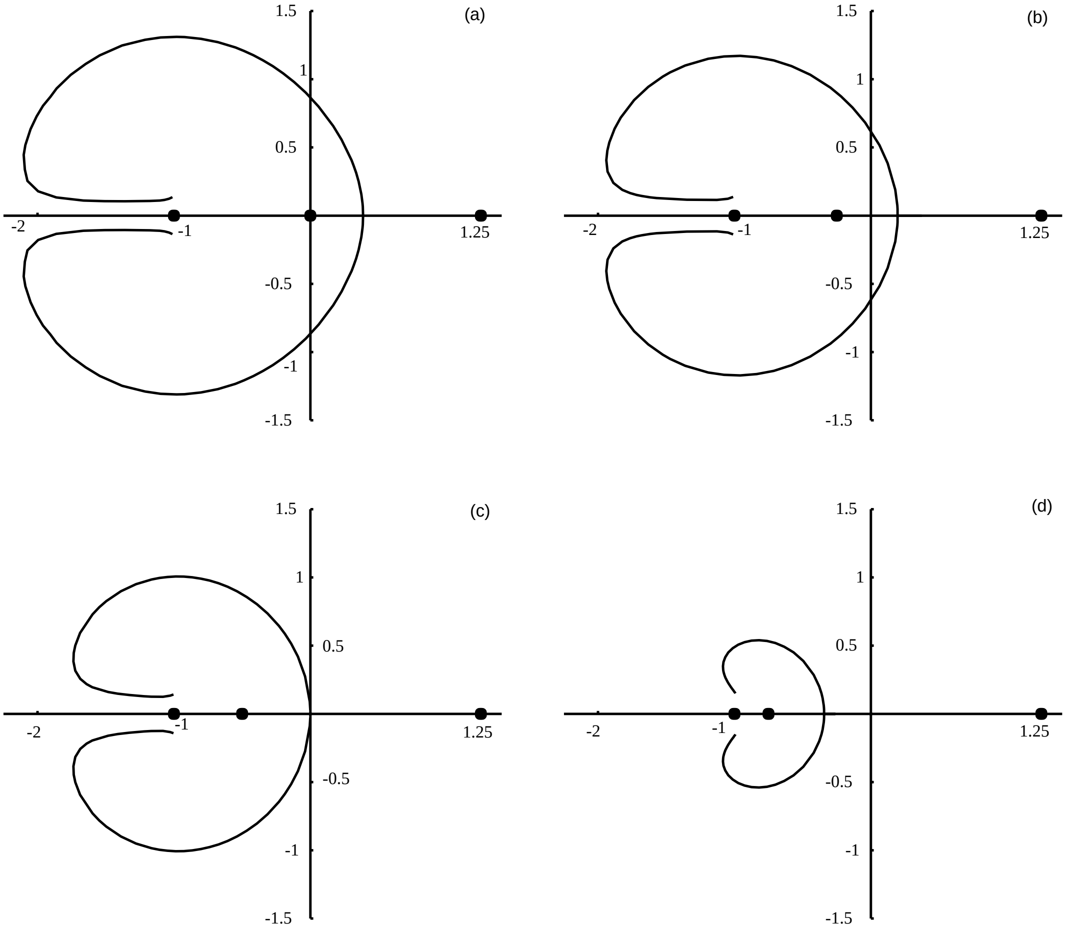


Figure 4.3. *The orbits through the complex plane of the eigenvalues $\lambda^0(\mu, \hat{l})$ and $\lambda^1(\mu, \hat{l})$ as function of μ, for several values of $\hat{l}$: $\hat{l} = 0$, $\hat{l} = \frac{1}{2}$, $\hat{l} = \frac{1}{2}\sqrt{2}$, $\hat{l} = \frac{1}{2}\sqrt{3}$. Note that $\lambda^0(0, \hat{l}) = \frac{5}{4}$ and $\lambda^1(\mu, \hat{l}) = 0$ at leading order for these ($\mathcal{O}(1)$) values of $\hat{l}$.*

λ. This expression must be equal to 0 at $\lambda = 0$ and $\hat{l} = \hat{l}_{\text{triple}}$. It follows from Lemma 4.1 that $\hat{l}_{\text{triple}} = \pm\frac{1}{2}\sqrt{2}$ and $\mu_{\text{triple}} = \frac{1}{6}$ (at leading order) in the classical Gierer–Meinhardt case (see Remark 4.6 for the general case).

For the stability analysis, it is more natural to fix μ and to determine the behavior of the eigenvalues $\lambda^j(\mu, l)$ as function of l or $\hat{l}, \hat{\tilde{l}}$. The following result follows directly from the above analysis and gives a precise description of the unstable eigenvalues of (3.6) in the classical Gierer–Meinhardt case.

Theorem 4.5. *Let $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = 2, \beta_2 = 2$, and $0 < \mu \ll 1/\varepsilon^4$. There is one unique unstable eigenvalue $\lambda^0(\mu, l) > 0$ for any $l \in (l_{\text{L,stab}}, l_{\text{R,stab}}) = (\varepsilon^2\sqrt{3\mu}, \sqrt{5/4})$ at leading order. There are no unstable eigenvalues for $l > l_{\text{R,stab}}$. Except for the symmetrical eigenvalues with $l < 0$, these are the only unstable eigenvalues for $\mu > \mu_{\text{Hopf}}(0) = \mathcal{O}(1)$. For $\mu_{\text{triple}} = 1/6 < \mu < \mu_{\text{Hopf}}(0)$, there is an additional interval $(0, l_{\text{C,stab}}(\mu)) \subset (0, l_{\text{L,stab}})$ in which (3.6) has a pair of complex conjugate unstable eigenvalues; for $0 < \mu < \mu_{\text{triple}}$, there are always two*

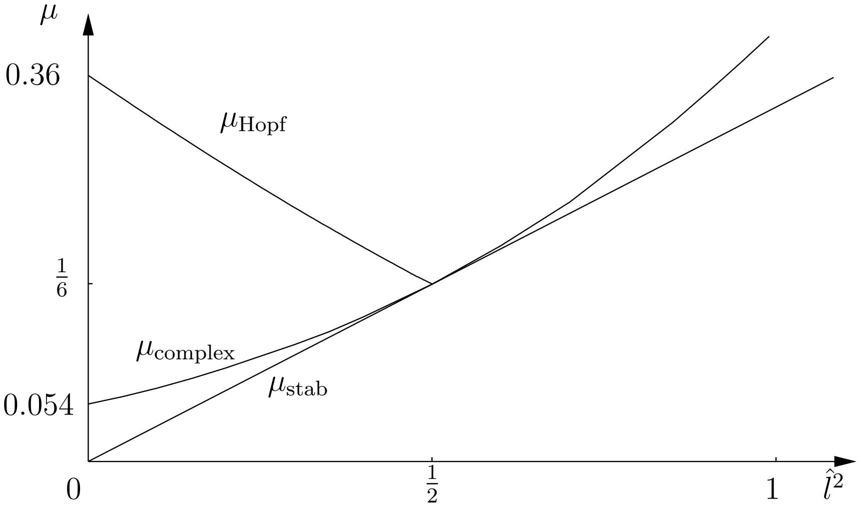


Figure 4.4. *The graphs of the bifurcation values* $\mu_{\text{double}}(\hat{l})$, $\mu_{\text{complex}}(\hat{l})$, *and* $\mu_{\text{Hopf}}(\hat{l})$ *as functions of* $\hat{l}^2$. *Note that* $(\hat{l}^2_{\text{triple}}, \mu_{\text{triple}}) = (\frac{1}{2}, \frac{1}{6})$.

unstable eigenvalues in the interval $(0, l_{\text{L,stab}})$, *both of which are real when* $0 < \mu < \mu_{\text{complex}}(0)$.

Note that only the eigenvalues $\lambda^0(\mu, l)$ and $\lambda^1(\mu, l)$ are important for the stability of the homoclinic stripe pattern. We know that there can be a fourth eigenvalue, $\lambda^3(\mu, l)$, for μ large enough, but this eigenvalue always remains negative; see Remark 4.8 (recall that $\lambda^2(\mu, l) = -l^2$ at leading order). In Figure 4.5, the graphs of $\lambda^j(\mu, l)$, $j = 0, 1$, are plotted for several values of μ.

Remark 4.6. The critical values $\hat{l}_{\text{triple}}$ and μ_{triple} are defined by $\mu_{\text{complex}}(\hat{l}_{\text{triple}}) = \mu_{\text{double}}(\hat{l}_{\text{triple}}) = \mu_{\text{Hopf}}(\hat{l}_{\text{triple}}) = \mu_{\text{triple}}$. Since

$$\frac{\partial}{\partial \lambda}\mu_{\text{real}}(\lambda, \hat{l})|_{\lambda=0} = \frac{([\alpha_1 - \alpha_2 \mathcal{R}(1)]^2 - 1) + 2(\lambda + \hat{l}^2)\,[\alpha_1 - \alpha_2\mathcal{R}(1)]\,\alpha_2\mathcal{R}'(1)\frac{\partial P}{\partial \lambda}|_{\lambda=0}}{([\alpha_1 - \alpha_2\mathcal{R}(1)]^2 - 1)^2}$$

(at leading order), it follows by (4.10) and (1.3) that

$$\hat{l}^2_{\text{triple}} = -\frac{[\alpha_1 - \alpha_2\mathcal{R}(1; \beta_1, \beta_2)]^2 - 1}{\alpha_2\mathcal{R}'(1; \beta_1, \beta_2)\,(\alpha_1 - \alpha_2\mathcal{R}(1; \beta_1, \beta_2))} = \frac{D(\beta_2 - 1)(D + 2(\beta_2 - 1))}{|\alpha_2|(2\beta_1 - (\beta_2 - 1))(D + (\beta_2 - 1))}$$

at leading order (recall that $\alpha_2 < 0$ (1.2)). Note that $\hat{l}_{\text{triple}}$ does not exist for $2\beta_1 - (\beta_2 - 1) < 0$; see also section 4.3 and Theorem 4.10. Finally, we see from (4.16) that

$$\mu_{\text{triple}} = \frac{(\beta_2 - 1)^3}{|\alpha_2|(2\beta_1 - (\beta_2 - 1))(D + (\beta_2 - 1))}.$$

Note that we have considered only the case $\mu = \mathcal{O}(1)$, the case $\mu = \mathcal{O}(1/\varepsilon^m)$ follows immediately.

Remark 4.7. As μ becomes $\mathcal{O}(1/\varepsilon^4)$, the derivations of the expression for $l_{\text{L,stab}}$ (4.17) and $l_{\text{R,stab}}$ (3.30), both of which have become $\mathcal{O}(1)$, can no longer be valid. This is of course so,

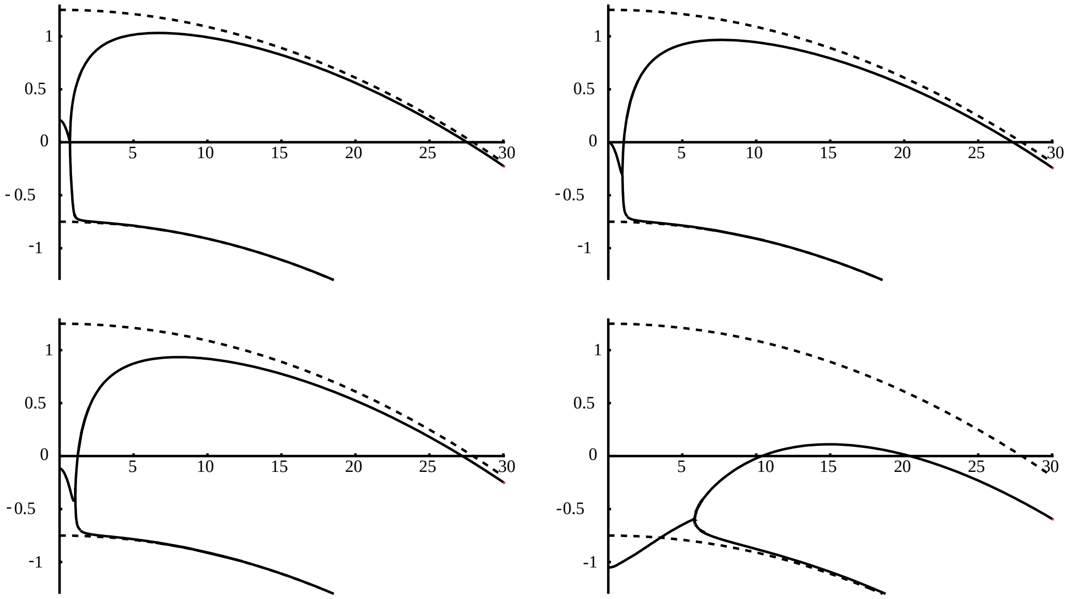


Figure 4.5. *The (real parts of) the two critical eigenvalues $\lambda^0(\mu, \hat{l})$ and $\lambda^1(\mu, \hat{l})$ as function of $\hat{l}$ for $\hat{l} \in (0, \mathcal{O}(1/\varepsilon^2))$ or, equivalently, $l \in (0, \mathcal{O}(1))$: $\mu = \mu_{\text{triple}} = \frac{1}{6} < \mu_{\text{Hopf}}(0)$, $\mu = \mu_{\text{Hopf}}(0)$, $\mu = 0.5 > \mu_{\text{Hopf}}(0)$, and $\mu = 25 \gg \mu_{\text{Hopf}}(0)$.*

since the asymptotically small parameter $\tilde{\varepsilon}$ has become $\mathcal{O}(1)$ (2.7), but if we proceed *formally* and assume that all results of the preceding sections can be extrapolated to this case, we see two other reasons why these derivations break down. The former is no longer valid since $\lambda = 0$ no longer corresponds to $P = 1$ at leading order for $l = \mathcal{O}(1)$ (4.1), and the latter is no longer valid since its derivation is based on the fact that all eigenvalues of (3.6) are close to the eigenvalues of the fast reduced system for l large enough, but $l = \mathcal{O}(1)$ is no longer large enough if $m = 4$ (see Lemma 3.8). For a given value of μ, both $l_{\text{L,stab}}(\mu)$ and $l_{\text{R,stab}}(\mu)$ are defined as these values of l for which there is an eigenvalue $\lambda = 0$. Thus, if we again assume *formally* that the expression (4.8) for $\mu_{\text{real}}(\lambda, \hat{l})$ can be extrapolated to $\mu = \mathcal{O}(1/\varepsilon^4)$, we see that both l_{stab}'s must solve the equation

$$\mu = \mu_{\text{real}}(0, l) = \frac{l^2}{\left[\alpha_1 - \alpha_2 \mathcal{R}(P(0, l); \beta_1, \beta_2)\right]^2 - 1} \tag{4.19}$$

(recall (3.3)). This expression has exactly the same form as $\mu_{\text{real}}(\lambda, 0)$ (4.8) after the interchange $\lambda \leftrightarrow l^2$! Thus $\mu_{\text{real}}(0, l)$ is in essence given by Figure 4.1(a). For μ small enough, we can indeed formally determine two zeros, $l_{\text{L,stab}}(\mu)$ and $l_{\text{R,stab}}(\mu)$. However, there is a maximum, explicitly given by $\mu = \mu_{\text{complex}}(0)/\varepsilon^4$, at which these zeros come together. For higher values of $\mu > \mu_{\text{complex}}(0)/\varepsilon^4$, there are no solutions so that there is no interval in which (3.6) has unstable eigenvalues (Theorem 4.5): the homoclinic stripe pattern must be stable. Of course, this is a completely formal argument; however, we shall see in the numerical simulations of section 5 that this "analysis" is qualitatively, although not quantitatively, correct.

Remark 4.8. Just as in the $\hat{l} = 0$ case, there can be edge bifurcations for $\hat{l} \neq 0$. We first consider the case when $\mu = \mathcal{O}(1)$. Edge bifurcations occur as $\alpha_1 - \alpha_2 \mathcal{R}(P; \beta_1, \beta_2) = 0$. For these critical values, the graph of $\mu_{\text{real}}(\lambda, \hat{l})$ is tangent to the edge of the essential spectrum (3.10)—see Figure 4.1(a) for the case when $\hat{l} = 0$. It follows from (3.10) that $\mu_{\text{edge}}(\hat{l}) = -\lambda_{\text{edge}} - \hat{l}^2$, with $-1 < \lambda^3(\mu_{\text{edge}}(\hat{l}), \hat{l}) = \lambda_{\text{edge}} < \lambda_f^2 = -3/4$, which implies that the edge bifurcation value $\mu_{\text{edge}}(\hat{l})$ decreases with increasing $\hat{l}$. For $1 > \hat{l} > \sqrt{-\lambda_{\text{edge}}}$, $\mu_{\text{edge}}(\hat{l})$ is negative so that eigenvalue $\lambda^3(\mu, \hat{l})$ is present for all $\mu > 0$. However, when $\hat{l} > 1$, the multiplication factor in front of the $\mu_{\text{real}}(\lambda, 0)$ in (4.15) is negative for all $\lambda \in (-1, 0)$ so that $\mu_{\text{real}}(\lambda, \hat{l})$ must be negative for all allowed (4.9) with $-1 < \lambda < \lambda_f^2(\hat{l}) = -3/4$ at leading order; see Figure 4.2. Thus, for these values of $\hat{l}$, the edge eigenvalue $\lambda^3(\mu, \hat{l})$ has again disappeared back into the essential spectrum. A similar argument can be applied to the case when $\mu = \mathcal{O}(1/\varepsilon^m)$ with $m \in (0, 4)$. It follows that the edge eigenvalues do not play a role in the stability question.

4.3. Stability analysis for general α_1, α_2, β_1, β_2.

So far, we have obtained a number of general results that seem to suggest that it is possible to obtain an equivalent of Theorem 4.5 for a large (and open) set $\mathcal{V}_{\text{classical}}$ of parameter values $(\alpha_1, \alpha_2, \beta_1, \beta_2)$. For instance, there is the result of [10] on the existence of two positive eigenvalues, $\lambda^0(\mu, 0)$ (near $\lambda_f^0(0)$) and $\lambda^1(\mu, 0)$ (near $\lambda_f^1(0)$), for $0 < \mu$ small enough and $\alpha_1, \alpha_2, \beta_1, \beta_2$ satisfying (1.2); see also sections 3.4 and 4.2 (4.14). Moreover, we note that, for $\hat{l} = 0$, eigenvalues cannot cross through $\lambda = 0$, i.e., that $\lambda = 0$ is always a simple eigenvalue (for $\hat{l} = 0$) so that (in)stability can only set in by a Hopf bifurcation. This observation was already made in [10], and it follows from Lemma 4.1 (4.10) and (4.7) that

$$t_2(0,0) = 1 - \left[\alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2 - 1}\right] = -\frac{D}{\beta_2 - 1} < 0$$

by (1.3). Hence $\mathcal{D}(\lambda, 0)$ always has a simple zero (associated to $t_1(\lambda, 0)$) at $\lambda = 0$. The combination of these results with the transparent relation (4.15) between $\mu_{\text{real}}(\lambda, 0)$ and $\mu_{\text{real}}(\lambda, \hat{l})$ and the results of Corollary 4.4. and Remark 4.6 suggest that a generalization of the classical case considered in Theorem 4.5 must be feasible, i.e., that there indeed is a large and open subset $\mathcal{V}_{\text{classical}}$ of the $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ parameter space in which a result similar to Theorem 4.5 can be proved.

However, we will not pursue such a general result here. Instead we will formulate two results in which the behavior of the eigenvalues (as functions of μ) differs significantly from that in the classical case considered in the previous section. For simplicity, we will mainly focus on the one-dimensional case; i.e., we consider homogeneous perturbations with $l = \hat{l} = 0$.

We know that there must be two real unstable eigenvalues, $\lambda^0(\mu, 0)$ and $\lambda^1(\mu, 0)$, for μ small enough. We have seen in the previous section that $\lambda^{0,1}(\mu, 0)$ can merge at a critical value $\mu_{\text{complex}}(0)$ and become a pair of complex conjugated eigenvalues. However, such a pair of complex conjugated eigenvalues does not necessarily remain complex for all $\mu > \mu_{\text{complex}}(0)$. In Figure 4.6 it is shown that, for $\alpha_1 = 5/4$, $\alpha_2 = -3$, and $\beta_1 = \beta_2 = 2$, the complex conjugate pair $\lambda^{0,1}(\mu, 0)$ crosses to the stable side of the Im(λ)-axis at a certain value $\mu_{\text{Hopf}}(0)$ so that the one-dimensional pulse becomes stable. Nevertheless, the pair returns to the Re$(\lambda) > 0$ half plane at a $\mu_{\text{Hopf},2}(0)$ so that the pulse pattern again destabilizes. This bifurcation is followed (for increasing μ) by another bifurcation, at $\mu = \mu_{\text{complex},2}(0)$, at which the complex pair splits again into a pair of (unstable) real eigenvalues. The (re)occurrence of unstable real eigenvalues for μ large enough is a general phenomenon.

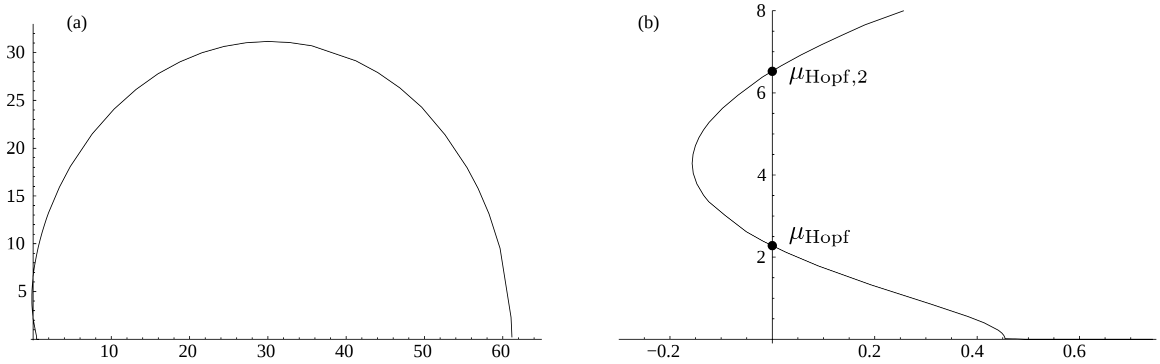


Figure 4.6. (a) *The orbits of the eigenvalues $\lambda^0(\mu,\hat{l})$ and $\lambda^1(\mu,\hat{l})$ through the upper half plane as function of μ for $\alpha_1 = 5/4$, $\alpha_2 = -3$, $\beta_1 = \beta_2 = 2$, and $\hat{l} = 0$;* (b) *zooms in near the Im-axis. Note that there are two (Hopf-) bifurcation values: $\mu_{\text{Hopf}}(0)$ and $\mu_{\text{Hopf},2}(0) > \mu_{\text{Hopf}}(0)$.*

Theorem 4.9. *Let $(\alpha_1,\alpha_2,\beta_1,\beta_2)$ satisfy (1.2), and assume that $\alpha_1 > 1$. Then there is a critical $\mathcal{O}(1)$ value $\mu_{\text{destab}}(0;\alpha_1,\alpha_2,\beta_1,\beta_2)$ of μ so that, for all $\mu > \mu_{\text{destab}}(0)$ (and $\mu \ll 1/\varepsilon^4$), the eigenvalue problem (3.6), with $l = 0$, has (at least) one unstable, real eigenvalue $\lambda^0(\mu,0)$; for $\mu \gg 1$, $\lambda^0(\mu,0)$ is given by*

$$\lambda^0(\mu,0) = (\alpha_1^2 - 1)\mu + \mathcal{O}(1). \tag{4.20}$$

Proof. We know from (4.11) that $\lim_{\lambda\to\infty}\mathcal{R}(P(\lambda),\beta_1,\beta_2) = 0$. Thus existence condition (4.9) will be satisfied for λ or P large enough if $\alpha_1 > 0$. We see by (4.11) and (4.8) that

$$\mu_{\text{real}}(\lambda,0) = \frac{\lambda}{\alpha_1^2 - 1} + \mathcal{O}(1)$$

for large λ (recall (4.1)), which implies that there must be an unstable eigenvalue $\lambda^0(\mu,0)$, given by (4.20), for $\alpha_1 > 1$ and μ beyond a certain critical value $\mu_{\text{destab}}(0)$. ■

Note that we have not necessarily proved the reoccurrence and thus the existence of *a pair* of unstable eigenvalues, as appears in the example of Figure 4.6. It is, a priori, also possible that the $\lambda^0(\mu,0)$ of Theorem 4.9 "just" decreases toward $\lambda_f^0(0)$ as μ decreases to 0, without ever becoming complex. In this case, the unstable eigenvalue $\lambda^0(\mu,0)$, which must exist close to $\lambda_f^0(0)$ for μ small enough, always remains larger then $\lambda_f^0(0)$. This implies that $\lambda^0(\mu,0)$ cannot merge with $\lambda^1(\mu,0)$, as happened in the classical case. It is shown in [10] that the homoclinic pulse pattern is unstable for any $\mu > 0$ if $\lambda^0(\mu,0) > \lambda_f^0(0)$ for μ small enough. This implies in terms of Theorem 4.9 that $\mu_{\text{destab}}(0) = 0$.

Theorem 4.9 also indicates that the two-dimensional homoclinic pulse patterns will be unstable for all $0 < \mu < \mu_{\text{split}}$: it is unstable against homogeneous perturbations (i.e., perturbations with $l = 0$) for any μ "large enough," and we know from section 3 (especially Theorem 3.9) that stable stripe patterns can only be expected for large μ. (However, Theorem 4.9 is of course only valid for $\mu \ll 1/\varepsilon^4$, i.e., not up to μ_{split}.)

Thus, with Theorem 4.9, we have shown that there is a large (unbounded, open) domain $\mathcal{V}_{\text{large}}$ in the $(\alpha_1,\alpha_2,\beta_1,\beta_2)$ parameter space in which (3.6), with $l = 0$, always has a "large" unstable real eigenvalue for μ large enough. Our next result shows that there is another (unbounded, open) subset $\mathcal{V}_{\text{singular}}$ of the parameter space in which there is always, i.e., for any $\mu > 0$, a real and unstable eigenvalue between $\lambda_f^1(0) = 0$ and $\lambda_f^0(0)$.

Theorem 4.10. *Let $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ satisfy (1.2), and assume that $\beta_2 > 2\beta_1 + 1$. Then, for any $\alpha_2 < 0$, there is a critical value $\alpha_{\text{singular}} > 1 + \alpha_2\beta_1/(\beta_2 - 1)$ of α_1 such that, for all $\alpha_1 \in (1 + \alpha_2\beta_1/(\beta_2 - 1), \alpha_{\text{singular}})$, the eigenvalue problem (3.6), with $l = 0$, has (at least) one unstable, real eigenvalue $\lambda^1(\mu, 0) \in (0, \lambda_f^0(0))$ for all $\mu > 0$ (and $\mu \ll 1/\varepsilon^4$).*

Proof. As was the case for Theorem 4.9, the proof is again based on the asymptotic results of Lemma 4.1. We know that there always exists an unstable eigenvalue $\lambda^1(\mu, 0)$ near 0 for $\mu \ll 1$. This eigenvalue is given by (4.14), a result that has been obtained by plugging (4.10) into (4.8). As an intermediate step in the derivation of (4.14), we find that the denominator of $\mu_{\text{real}}(\lambda, 0)$ is at $\lambda = 0$ given by

$$[\alpha_1 - \alpha_2\mathcal{R}(1; \beta_1, \beta_2)]^2 - 1 = \frac{D}{\beta_2 - 1}\left(\frac{D}{\beta_2 - 1} + 2\right) > 0.$$

This denominator will decrease as λ increases when $\partial\mathcal{R}/\partial P$ is negative at $P = 1$ (since $\alpha_2 < 0$). This is the case when $\beta_2 > 2\beta_1 + 1$ (4.10). Thus the denominator of $\mu_{\text{real}}(\lambda, 0)$ will decrease through 0 for λ increasing from 0 if $D > 0$ is small enough and $\beta_2 > 2\beta_1 + 1$. Note that the condition on α_1 in the statement of the theorem causes D to be "small enough." A change in sign of the denominator implies that $\mu_{\text{real}}(\lambda, 0)$ has a singularity at a certain value $\lambda_{\text{singular}}$ of λ that is in between 0 and $\lambda_f^0(0)$. As a consequence, there must be an unstable eigenvalue $\lambda^1(\mu, 0) \in (0, \lambda_{\text{singular}})$ for any $\mu > 0$. ■

Under the conditions of Theorem 4.10, one expects two singularities in $\mu_{\text{real}}(\lambda, 0)$ between 0 and $\lambda_f^0(0)$ in the case that $\lambda^0(\mu, 0) < \lambda_f^0(0)$ for μ small enough. Thus, in such a case, there will be two unstable real eigenvalues for all $\mu > 0$. We will not consider this in more detail here. As was the case for parameter combinations in $\mathcal{V}_{\text{large}}$ (Theorem 4.9), one does not expect stable homoclinic stripe patterns for $(\alpha_1, \alpha_2, \beta_1, \beta_2) \in \mathcal{V}_{\text{singular}}$; i.e., we already know that it is impossible for $0 < \mu \ll 1/\varepsilon^4$ (Theorem 3.9), and Theorem 4.10 strongly suggests that the pattern is unstable with respect to homogeneous ($l = 0$) perturbations up to $\mu = \mu_{\text{split}}$.

We have thus distinguished three regions in $(\alpha_1, \alpha_2, \beta_1, \beta_2)$-space: $\mathcal{V}_{\text{classical}}$, in which the eigenvalues of (3.6) behave as in the classical case (Theorem 4.5), $\mathcal{V}_{\text{large}}$, in which there is always a large real unstable eigenvalue for μ large enough (Theorem 4.9 and Figure 4.6), and $\mathcal{V}_{\text{singular}}$, in which there is an unstable real eigenvalue between 0 and $\lambda_f^0(0)$ for any $\mu > 0$ (Theorem 4.10). These results give only indications of the possible behavior of the eigenvalues of (3.6); we do not consider other types of behavior in this paper.

5. Numerical simulations. In this section, we confirm and extend the analysis of the previous sections by performing numerical simulations on the full PDE (1.4) for (x, y) on a *bounded* domain with homogeneous Neumann boundary conditions: $(x, y) \in (0, L_x) \times (0, L_y) \subset \mathbf{R}^2$. We choose $L_x \gg 1$ large enough so that it has no leading order effect on the existence or stability of the one-dimensional homoclinic pulse pattern. We refer to [12], [8], [6] for more details on the influence of L_x and the type of boundary conditions. We use L_y as an additional bifurcation parameter.

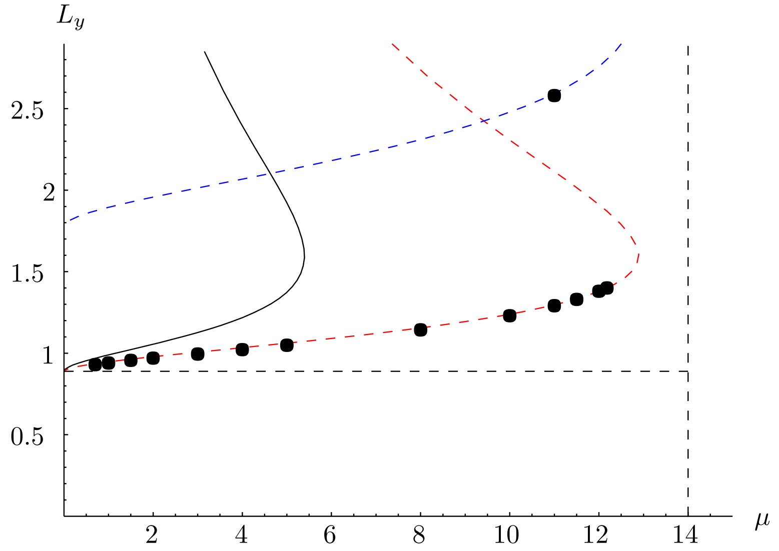


Figure 5.1. *The critical width L_y of the domain $\mathbf{R} \times (0, L_y)$ for the classical Gierer–Meinhardt equation with $\varepsilon^2 = 0.1$ for several values of μ. The horizontal dotted line represents the critical value of L_y given by Corollary 5.1 and (5.2), and the vertical dotted lines stands for the stripe splitting bifurcation. The other curves are all based on (formal) relation (4.19)—see text.*

Due to the homogeneous Neumann conditions at $y = 0$ and $y = L_y$, the wave number l of the perturbation can now attain only discrete values:

$$l = l_m = \frac{m\pi\varepsilon}{L_y}, \quad m = 0, \pm 1, \pm 2, \dots \tag{5.1}$$

(recall (3.1), (3.2)). The analysis of section 3 immediately implies the following corollary.

Corollary 5.1. *Let $0 < \mu \ll 1/\varepsilon^4$. Assume that the homoclinic solution $(U_0(x), V_0(x))$ is stable as a one-dimensional pattern (i.e., with respect to homogeneous, $l = 0$ perturbations). Then it is stable as a stripe pattern on the strip $\mathbf{R} \times (0, L_y)$ for all*

$$0 < L_y < \frac{\pi\varepsilon}{\sqrt{\frac{1}{4}(\beta_2 + 1)^2 - 1}}.$$

Thus, on strips that are "narrow enough," the entire ($l > 0$)-band of unstable eigenvalues that exist for $\hat{l}$ large enough (Lemma 3.8, Theorem 4.5) is between $l = 0$ and the first nonhomogeneous wave number $l = \pi\varepsilon/L_y$. This result provides us with a simple method to check the analysis of the preceding sections numerically (Remark 5.2).

In Figure 5.1, we fixed $\varepsilon^2 = 0.1$ (note that ε is thus "quite large") and considered the classical Gierer–Meinhardt case ($\alpha_1 = 0, \alpha_2 = -1, \beta_1 = \beta_2 = 2$) for various choices of μ. We made a homoclinic stripe pattern by taking the one-dimensional stable homoclinic pulse $(U_0(x), V_0(x))$ as the initial condition on the two-dimensional domain $(0, L_x) \times (0, L_y)$ and extending it homogeneously in the y-direction. We took L_x so large that the boundaries in the x-direction are not relevant (at leading order) and L_y initially "very small" (the maximum

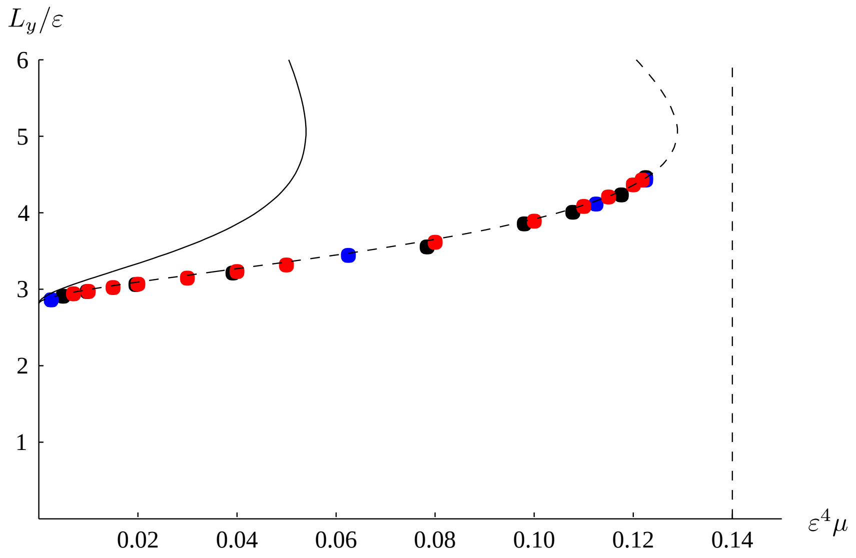


Figure 5.2. *The critical width L_y of the domain $\mathbf{R} \times (0, L_y)$ for the classical Gierer–Meinhardt equation for various values of μ and $\varepsilon^2 = 0.1, 0.07$ and 0.05. Note that the x-axis is scaled by $\varepsilon^4\mu$ and the y-axis by L_y/ε so that the simulations confirm the obtained scaling relations. The curves are again based on relation* (4.19).

of the stripe is situated in the middle of the domain, i.e., at $x = L_x/2$). Next we increased L_y up to the point at which the homoclinic stripe became unstable. Corollary 5.1 predicts that the bifurcation should take place at

$$L_y = \pi\sqrt{\frac{0.1}{1.25}} \approx 0.888\ldots \tag{5.2}$$

(at leading order in $\varepsilon \approx 0.316\ldots$ (!)). This critical value is confirmed by the simulations for μ up to, say, 10 (taking into account the $\mathcal{O}(\varepsilon)$ correction term). For $\mu < 2$, the prediction of Corollary 5.1 is extremely accurate. Figure 5.2 shows the outcome of similar experiments, now for $\varepsilon^2 = 0.1, 0.07$ and 0.05; here the vertical axis represents L_y/ε so that by Corollary 5.1 the bifurcation should take place at $\pi/\sqrt{1.25} \approx 2.80\ldots$; the horizontal axis is measured in $\varepsilon^4\mu$. All observed bifurcation values are very close to a (dotted) line, which confirms the scaling properties of the Gierer–Meinhardt equation established in this paper. Again, the critical value of L_y/ε is recovered with remarkably high accuracy.

The curves in Figure 5.1 are based on the *formal* relation (4.19) between μ and l obtained in Remark 4.7; the same curves, but now in the form of $\varepsilon^4\mu$ as function of L_y/ε, are plotted in Figure 5.2. The bifurcation values are all very close to a stretched version of curve (4.19), i.e., one of the dotted curves (thus the dotted curves are obtained from (4.19) by multiplication of μ with a well-chosen (and numerically determined) factor). Hence, although the relation (4.19) is obtained by a formal extrapolation of the asymptotic results into the regime $\mu = \mathcal{O}(1/\varepsilon^4)$, it seems to give a reasonably accurate qualitative description of the behavior of the critical value of L_y as a function of μ for $\mu = \mathcal{O}(1/\varepsilon^4)$ However, it should be noted that the quantitative

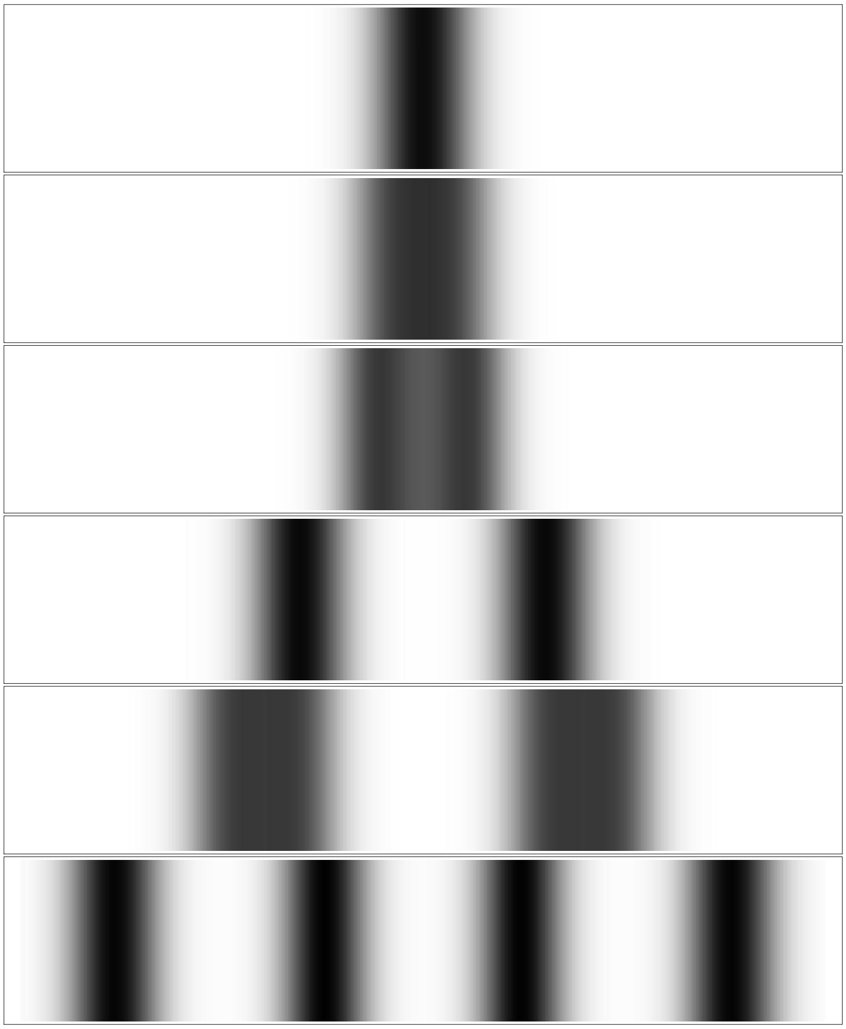


Figure 5.3. *The self-replication of stripes process. The gray shades represent the magnitude of the V-components. The simulation was run for the classical Gierer–Meinhardt case with $\mu = 14$ and $\varepsilon^2 = 0.1$. Time increases in the downward direction. The dynamics converge to a stable spatially periodic stripe pattern.*

error is certainly "not small" for μ/ε^4 "not small." Beyond

$$\mu = \mu_{\text{stripe}} = \mu_{\text{stripe}}(\alpha_1 = 0, \alpha_2 = -1, \beta_1 = \beta_2 = 2) \approx \frac{0.12}{\varepsilon^4},$$

the stripe pattern appeared to be stable on $(0, L_x) \times (0, L_y)$ for any L_y (and any L_x). Thus we conclude by the numerical simulations that, for $\mu > \mu_{\text{stripe}}$, stable homoclinic stripe patterns exist for $(x, y) \in \mathbf{R}^2$. However, the simulations also show that the *stripe splitting bifurcation*, predicted by the analysis of section 2, sets in at

$$\mu = \mu_{\text{split}} = \mu_{\text{split}}(\alpha_1 = 0, \alpha_2 = -1, \beta_1 = \beta_2 = 2) \approx \frac{0.14}{\varepsilon^4};$$

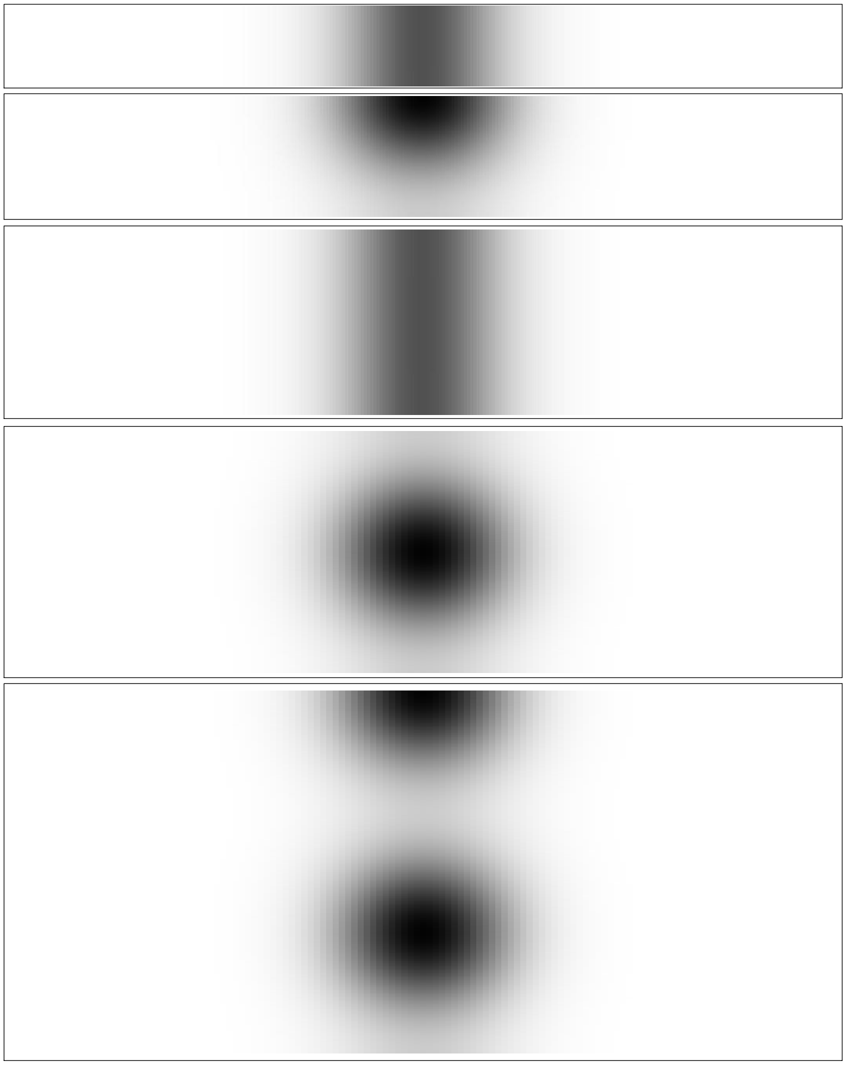


Figure 5.4. *The fate of the homoclinic stripe patterns on various domains* $\mathbf{R} \times (0, L_y)$ *for the classical Gierer–Meinhardt case with* $\mu = 11$ *and* $\varepsilon^2 = 0.1$; (a) $L_y = 1.0$: *the homoclinic stripe is stable;* (b) $L_y = 1.5$: *the wave number* l_1 (5.1) *has become unstable, i.e.,* $l_1 \in (l_{\mathrm{L,stab}}, l_{\mathrm{R,stab}})$*, and the stripe bifurcates into half a spot;* (c) $L_y = 2.3$: *the homoclinic stripe is again stable:* l_1 *is no longer in the instability interval* $(l_{\mathrm{L,stab}}, l_{\mathrm{R,stab}})$. (d) $L_y = 3.0$: $l_2 \in (l_{\mathrm{L,stab}}, l_{\mathrm{R,stab}})$*, and the stripe bifurcates into a full spot centered in the middle of the domain;* (e) $L_y = 4.5$: $l_3 \in (l_{\mathrm{L,stab}}, l_{\mathrm{R,stab}})$ *(and neither one of the other wave numbers), and the stripe bifurcates into one and a half spot.*

see Figures 5.1 and 5.2. It is observed that beyond μ_{split}, there can be no homoclinic stripe patterns, which confirms the analysis of section 2. The self-replication process is completely equivalent to the self-replication of pulses in the one-dimensional case (see Figure 2.1 and [33], [35], [34], [12], [31] for the Gray–Scott equation): the stripe splits into two stripes that slowly move away from each other, and these stripes split again (etc., depending on the length L_x of the domain), until the system reaches an asymptotically stable spatially periodic stripe pattern (Figure 5.3). Hence the stripe replication process is in essence a one-dimensional phenomenon.

Thus we may conclude that the classical Gierer–Meinhardt equation has stable homoclinic stripe patterns on $\mathbf{R}^2$ for $\mu \in (\mu_{\text{stripe}}, \mu_{\text{split}}) \approx (0.12/\varepsilon^4, 0.14/\varepsilon^4)$. Moreover, there exist spatially periodic stripe patterns that are (numerically) stable on $\mathbf{R}^2$ for $\mu > \mu_{\text{per-stripe}} = \mathcal{O}(1/\varepsilon^4)$, where $\mu_{\text{per-stripe}}$ depends on the distance L_λ between the stripes (i.e., the wave length); $\mu_{\text{per-stripe}}(L_\lambda) < \mu_{\text{split}}$ if L_λ is large enough and $\lim_{L_y \to \infty} \mu_{\text{per-stripe}}(L_\lambda) = \mu_{\text{stripe}}$. Once again, this behavior is completely similar to the one-dimensional case (see [24] for a detailed existence and stability analysis of one-dimensional spatially periodic patterns in the Gray–Scott equation). Note that the existence of spatially periodic patterns in the (generalized) Gierer–Meinhardt system (1.1)/(1.4) has been established in [11].

Finally, we consider the "fate" of the homoclinic stripe pattern on $\mathbf{R} \times (0, L_y)$ as L_y increases through its critical value. Except for one measurement in Figure 5.1, all critical values of L_y shown in Figures 5.1 and 5.2 correspond to the situation in which the smallest allowed nonhomogeneous eigenvalue l_1 (5.1) decreases through $l_{\text{R,stab}}$ (3.30). It is shown in Figure 5.4(b) that the stripe bifurcates into half a spot (at either one of the y-boundaries in the middle of the domain with respect to x) at these bifurcations. This fully agrees with the varicose type of the associated unstable eigenfunction (see Remark 3.10). Note that this indicates that the bifurcation is subcritical. Figure 5.4 shows the end-product of the homoclinic stripe pattern for various choices of L_y (with μ fixed at $11 < \mu_{\text{stripe}}$, $\varepsilon^2 = 0.1$, and $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = \beta_2 = 2$). Thus, as is shown in Figure 5.4(c), the stripe "returns" as a stable object for an additional interval of L_y values. In this region, the entire band of unstable eigenvalues, $(l_{\text{L,stab}}, l_{\text{R,stab}})$ (Theorem 4.5), is in between the first and second nonhomogeneous eigenvalues (5.1); i.e., $(l_{\text{L,stab}}, l_{\text{R,stab}}) \subset (l_1, l_2)$. The stripe again destabilizes as l_2 enters $(l_{\text{L,stab}}, l_{\text{R,stab}})$. This bifurcation is of course again of varicose type. Due to the spatial structure in the y-direction associated to l_2, the stripe now bifurcates either into a (full) stable spot in the middle of the domain or to two half spots at both y-boundaries; see Figure 5.4. The curve on which this bifurcation occurs has been indicated with the second dotted line in Figure 5.1. This l_2-curve is just a translated and stretched version (in the L_y-direction) of the l_1-curve. The above-mentioned special measurement is on this l_2-curve and indicates the bifurcation of the stripe pattern into a (full) spot. In Figure 5.4(e), the one-and-a-half spot pattern that occurs as a stable state from the homoclinic stripe pattern as l_3 enters $(l_{\text{L,stab}}, l_{\text{R,stab}})$ after l_2 has decreased through $l_{\text{L,stab}}$. See Remark 5.3.

Remark 5.2. The two-dimensional numerical simulations were performed using the VLUGR2 code of Blom, Trompert, and Verwer [2]. This adaptive mesh code was developed especially for two-dimensional PDEs that generate steep spatio-temporal gradients.

Remark 5.3. The accumulation of the l_m-bifurcation curves, where an l_m-curve indicates that the stripe bifurcates into $m/2$ spots, is typical for systems that have a bifurcation of Turing or Ginzburg–Landau type. Here the background pattern, which destabilizes, is the homoclinic stripe pattern. We refer to [27] for a description of this phenomenon with a homogeneous state as the background pattern.

6. Conclusion.

In this paper, we have shown that the stability of homoclinic stripe patterns in the generalized Gierer–Meinhardt equations can be studied in full analytical detail as long as the parameter μ is not "too large," i.e., as long as $\mu \ll 1/\varepsilon^2$ in the original equation (1.1) or $\mu \ll 1/\varepsilon^4$ in the rescaled equation (1.4); see Remark 1.3. The stability analysis is based on an extension of the NLEP method to two-dimensional problems. The NLEP method follows the Evans function approach to the linear eigenvalue problem that is associated to the stability question. This method has recently been developed in the context of the stability of pulse solutions in monostable (Remark 1.1) one-dimensional reaction-diffusion equations [8], [9], [10]. By transforming the reduced nonlocal eigenvalue problem into a hypergeometric differential equation, it is possible to obtain an explicit description of the spectrum associated to the stability of the homoclinic stripe pattern (see Theorem 4.5 and Figure 4.5).

However, the NLEP analysis establishes that the homoclinic stripe patterns cannot be stable as long as $\mu \ll 1/\varepsilon^4$ in (1.4), or $\mu \ll 1/\varepsilon^2$ (1.1); see Theorem 3.9. Nevertheless, formal extrapolation of the asymptotic analysis (see especially Remark 4.7) strongly suggests that homoclinic stripe patterns can be stable for $\mu = \mathcal{O}(1/\varepsilon^4)$ (in (1.4)). This is confirmed by numerical simulations: it is found in the case of the classical Gierer–Meinhardt equation (i.e., $\alpha_1 = 0, \alpha_2 = -1, \beta_1 = \beta_2 = 2$) that the homoclinic stripe pattern is stable for $\mu > \mu_{\text{stripe}} = \mathcal{O}(1/\varepsilon^4)$. Nevertheless, this will not be the case for all allowed parameter combinations (1.2): in section 4.3, several (open, unbounded) regions in parameter space have been determined in which there cannot be stable homoclinic stripe patterns (Theorems 4.9 and 4.10).

The prediction of Theorem 2.2, which is an extension of the existence results in the literature, is also confirmed by the numerical simulations: at $\mu_{\text{stripe}} > \mu > \mu_{\text{split}} = \mathcal{O}(1/\varepsilon^4)$, a self-replication of stripe patterns takes place. This implies that the homoclinic stripe pattern evolves into a (numerically) stable spatially periodic stripe pattern (Figure 5.3). See Remark 6.1.

Remark 6.1. The stability of the spatially periodic stripes patterns, whose existence has been shown in [11], can also be studied by the NLEP approach, in combination with the ideas presented in [14] and [36]. We refer to [8] for a formal stability analysis of spatially periodic pulse patterns in the one-dimensional Gray–Scott equation using the NLEP method. The structure of the spectrum of the stability problem associated to the homoclinic stripe pattern strongly suggests that the long-wave-length, nearly homoclinic, spatially periodic stripe patterns will also be unstable for $\mu \ll 1/\varepsilon^4$ in (1.4); see [14], [36]. However, it should be noted that the distances between the stripes in the periodic patterns are in general not large enough for the application of the ideas of the analysis of weakly interacting pulses, i.e., pulses that are so far away from each other that all U_j-components are exponentially close to the background state; here $(U_1, U_2) = (U, V) = (0, 0)$ in between the pulses (see [8], [6], [7], and [24] for detailed discussions of several aspects of this issue in the context of the one-dimensional Gray–Scott model). Hence the stripes in the spatially periodic patterns observed in this paper are so "close" to each other that the instability of the stripe pattern for $\mu \ll 1/\varepsilon^4$ does not automatically follow from the "homoclinic" analysis in this paper, combined with [14] and [36]. The stability of spatially periodic structures is the subject of work in progress.

Remark 6.2. In this paper, we paid attention only to completely linear, or straight, homoclinic stripe patterns. Reaction-diffusion equations also exhibit stripe, or "volcano," patterns in a circular shape [32], [25]. The NLEP approach can also be used to study the stability of such circular structures and the bifurcation of these patterns into rings of spots, as is shown in [25].

REFERENCES

[1] J. Alexander, R. A. Gardner, and C. K. R. T. Jones, *A topological invariant arising in the stability of traveling waves*, J. Reine Angew. Math., 410 (1990), pp. 167–212.

[2] J. G. Blom, R. A. Trompert, and J. G. Verwer, *Algorithm* 758 : *VLUGR*2 : *A vectorizable adaptive grid solver for PDEs in* 2*D*, ACM Trans. Math. Software, 22 (1996), pp. 302–328.

[3] F. H. Busse and S. C. Müller, eds., *Evolution of Spontaneous Structures in Dissipative Continuous Systems*, Lecture Notes in Phys. 55, Springer-Verlag, Berlin, 1998.

[4] P. De Groen, private communication.

[5] A. De Wit, *Spatial patterns and spatiotemporal dynamics in chemical systems*, in Adv. Chem. Phys. 109, Wiley, New York, 1999, pp. 435–513.

[6] A. Doelman, W. Eckhaus, and T. J. Kaper, *Slowly modulated two-pulse solutions in the Gray–Scott model* I: *Asymptotic construction and stability*, SIAM J. Appl. Math., 61 (2000), pp. 1080–1102.

[7] A. Doelman, W. Eckhaus, and T. J. Kaper, *Slowly modulated two-pulse solutions in the Gray–Scott model* II: *Geometric theory, bifurcations, and splitting dynamics*, SIAM J. Appl. Math., 61 (2001), pp. 2036–2062.

[8] A. Doelman, R. A. Gardner, and T. J. Kaper, *Stability analysis of singular patterns in the* 1-*D Gray–Scott model: A matched asymptotics approach,* Phys. D, 122 (1998), pp. 1–36.

[9] A. Doelman, R. A. Gardner, and T. J. Kaper, *A stability index analysis of* 1-*D patterns of the Gray–Scott model*, Mem. Amer. Math. Soc., 155 (2002).

[10] A. Doelman, R. A. Gardner, and T. J. Kaper, *Large stable pulse solutions in reaction-diffusion equations*, Indiana Univ. Math. J., 50 (2001), pp. 443–507.

[11] A. Doelman, T. J. Kaper, and H. van der Ploeg, *Spatially periodic and aperiodic multi-pulse patterns in the one-dimensional Gierer-Meinhardt equation*, Methods Appl. Anal., 8 (2001).

[12] A. Doelman, T. J. Kaper, and P. Zegeling, *Pattern formation in the one-dimensional Gray-Scott model*, Nonlinearity, 10 (1997), pp. 523–563.

[13] R. A. Gardner and C. K. R. T. Jones, *Stability of the travelling wave solutions of diffusive predator-prey systems*, Trans. Amer. Math. Soc., 327 (1991), pp. 465–524.

[14] R. A. Gardner, *Spectral analysis of long wavelength periodic waves and applications*, J. Reine Angew. Math., 491 (1997), pp. 149–181.

[15] A. Gierer and H. Meinhardt, *A theory of biological pattern formation*, Kybernetik, 12 (1972), pp. 30–39.

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

[17] D. Henry, *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Math. 840, Springer-Verlag, New York, 1981.

[18] P. Hirschberg and E. Knobloch, *Zigzag and varicose instabilities of a localized stripe*, Chaos, 3 (1993), pp. 713–721.

[19] D. Iron and M. J. Ward, *A metastable spike solution for a nonlocal reaction-diffusion model*, SIAM J. Appl. Math., 60 (2000), pp. 778–802.

[20] D. Iron, M. J. Ward, and J. Wei, *The stability of spike solutions of the one-dimensional Gierer-Meinhardt model*, Phys. D, 150 (2001), pp. 25–62.

[21] C. K. R. T. Jones, *Geometric singular perturbation theory*, in Dynamical Systems (Montecatibi Terme, 1994), Lecture Notes in Math. 1609, R. Johnson, ed., Springer-Verlag, New York, 1995, pp. 44–118.

[22] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer-Verlag, Berlin, 1984.

[23] A. Mielke, *The Ginzburg-Landau equation in its role as a modulation equation*, in Handbook for Dynamical Systems, Vol. 2, Elsevier, New York, 2002, pp. 759–834.

[24] D. S. Morgan, A. Doelman, and T. J. Kaper, *Stationary periodic patterns in the* 1*D Gray-Scott model*, Methods Appl. Anal., 7 (2000), pp. 105–150.

[25] D. S. Morgan and T. J. Kaper, *Annular Ring Solutions in the* 2-*D Gray-Scott Model and Their Destabilization into Spots*, in preparation.

[26] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw–Hill, New York, 1953.

[27] J. D. Murray, *Mathematical Biology*, Biomathematics Texts 19, Springer-Verlag, New York, 1989.

[28] W.-M. Ni, *Diffusion, cross-diffusion, and their spike-layer steady states*, Notices Amer. Math. Soc., 45 (1998), pp. 9–18.

[29] W.-M. Ni and I. Takagi, *Point condensation generated by a reaction-diffusion system in axially symmetric domains*, Japan J. Indust. Appl. Math., 12 (1995), pp. 327–365.

[30] Y. Nishiura, *Coexistence of infinitely many stable solutions to reaction-diffusion systems*, Dynamics Reported (New Series), 3 (1994), pp. 25–103.

[31] Y. Nishiura and D. Ueyama, *A skeleton structure for self-replication dynamics*, Phys. D, 130 (1999), pp. 73–104.

[32] T. Ohta, M. Mimura, and R. Kobayashi, *Higher-dimensional localized patterns in excitable media*, Phys. D, 34 (1989), pp. 115–144.

[33] J. E. Pearson, *Complex patterns in a simple system*, Science, 261 (1993), pp. 189–192.

[34] V. Petrov, S. K. Scott, and K. Showalter, *Excitability, wave reflection, and wave splitting in a cubic autocatalysis reaction-diffusion system*, Phil. Trans. Roy. Soc. London Ser. A, 347 (1994), pp. 631–642.

[35] W. N. Reynolds, J. E. Pearson, and S. Ponce-Dawson, *Dynamics of self-replicating patterns in reaction diffusion systems*, Phys. Rev. Lett., 72 (1994), pp. 2797–2800.

[36] B. Sandstede and A. Scheel, *On the stability of periodic travelling waves with large spatial period*, J. Differential Equations, 172 (2001), pp. 134–188.

[37] M. Taniguchi and Y. Nishiura, *Instability of planar interfaces in reaction-diffusion systems*, SIAM J. Math. Anal., 25 (1994), pp. 99–134.

[38] 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, 125 (1996), pp. 117–145.

[39] J. Wei, *On single interior spike solutions of the Gierer-Meinhardt system: Uniqueness and spectrum estimates*, European J. Appl. Math., 10 (1999), pp. 353–378.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 1, pp. 105–114

Chaos in the Hodgkin–Huxley Model[*]

John Guckenheimer[†] and Ricardo A. Oliva[†]

Abstract. The Hodgkin–Huxley model was developed to characterize the action potential of a squid axon. It has served as an archetype for compartmental models of the electrophysiology of biological membranes. Thus the dynamics of the Hodgkin–Huxley model have been extensively studied both with a view to their biological implications and as a test bed for numerical methods that can be applied to more complex models. This note demonstrates previously unobserved dynamics in the Hodgkin–Huxley model, namely, the existence of chaotic solutions in the model with its original parameters. The solutions are found by displaying rectangles in a cross-section whose images under the return map produce a Smale horseshoe. The chaotic solutions are highly unstable, but they are significant as they lie in the basin boundary that establishes the threshold of the system.

Key words. Hodgkin–Huxley, chaos, action potential, horseshoe

AMS subject classifications. 92C20, 37G35

PII. S1111111101394040

1. Introduction. The Hodgkin–Huxley model [15] for the action potential of a space-clamped squid axon is defined by the four dimensional vector field

$$\begin{aligned}
\dot{v} &= I - \Big[120m^3h\,(v+115) + 36n^4\,(v-12) + 0.3\,(v+10.599)\Big],\\
\dot{m} &= (1-m)\,\Psi\left(\frac{v+25}{10}\right) - m\left(4\exp\frac{v}{18}\right),\\
\dot{n} &= (1-n)\,0.1\Psi\left(\frac{v+10}{10}\right) - n\left(0.125\exp\frac{v}{80}\right),\\
\dot{h} &= (1-h)\,0.07\exp\left(\frac{v}{20}\right) - \frac{h}{1+\exp\frac{v+30}{10}},\\
\Psi(x) &= \frac{x}{\exp(x)-1}
\end{aligned}$$

with variables (v, m, n, h) that represent membrane potential, activation of a sodium current, activation of a potassium current, and inactivation of the sodium current and a parameter I that represents injected current into the space-clamped axon. Recall that the Hodgkin–Huxley convention for membrane potential reverses the sign from modern conventions, and so the voltage spikes of action potentials are negative in the Hodgkin–Huxley model. While improved models for the membrane potential of the squid axon [3] have been formulated, the Hodgkin–Huxley model remains the paradigm for conductance-based models of neural

[*] Received by the editors August 21, 2001; accepted for publication (in revised form) by M. Golubitsky February 19, 2002; published electronically May 14, 2002. This research was partially supported by the National Science Foundation, Department of Energy, and the ARO-EPRI Program.

http://www.siam.org/journals/siads/1-1/39404.html

[†] Mathematics Department, Cornell University, Ithaca, NY 14853 (gucken@cam.cornell.edu, rao@cornell.edu).

systems. From a mathematical viewpoint, varied properties of the dynamics of the Hodgkin–Huxley vector field have been studied [13, 8, 11, 14, 16, 19, 5]. Nonetheless, we remain far from a comprehensive understanding of the dynamics displayed by this vector field. It has become conventional wisdom that the qualitative properties of the Hodgkin–Huxley model can be reduced to a two dimensional flow such as the Fitzhugh–Nagumo model [7]. Rinzel and Miller [19] first gave evidence that this is not always the case. Hassard [13] and Labouriau [16] also studied the Hopf bifurcation that plays an important role in locating regions of bistability in the Hodgkin–Huxley model. Doi and Kumagai [5] recently showed the existence of chaotic attractors in a modified Hodgkin–Huxley model that changes the time constant of one of the currents by a factor of 100. This note extends the work of Rinzel and Miller, demonstrating the existence of chaotic solutions in the Hodgkin–Huxley model with the "standard" parameters used by Hodgkin and Huxley.

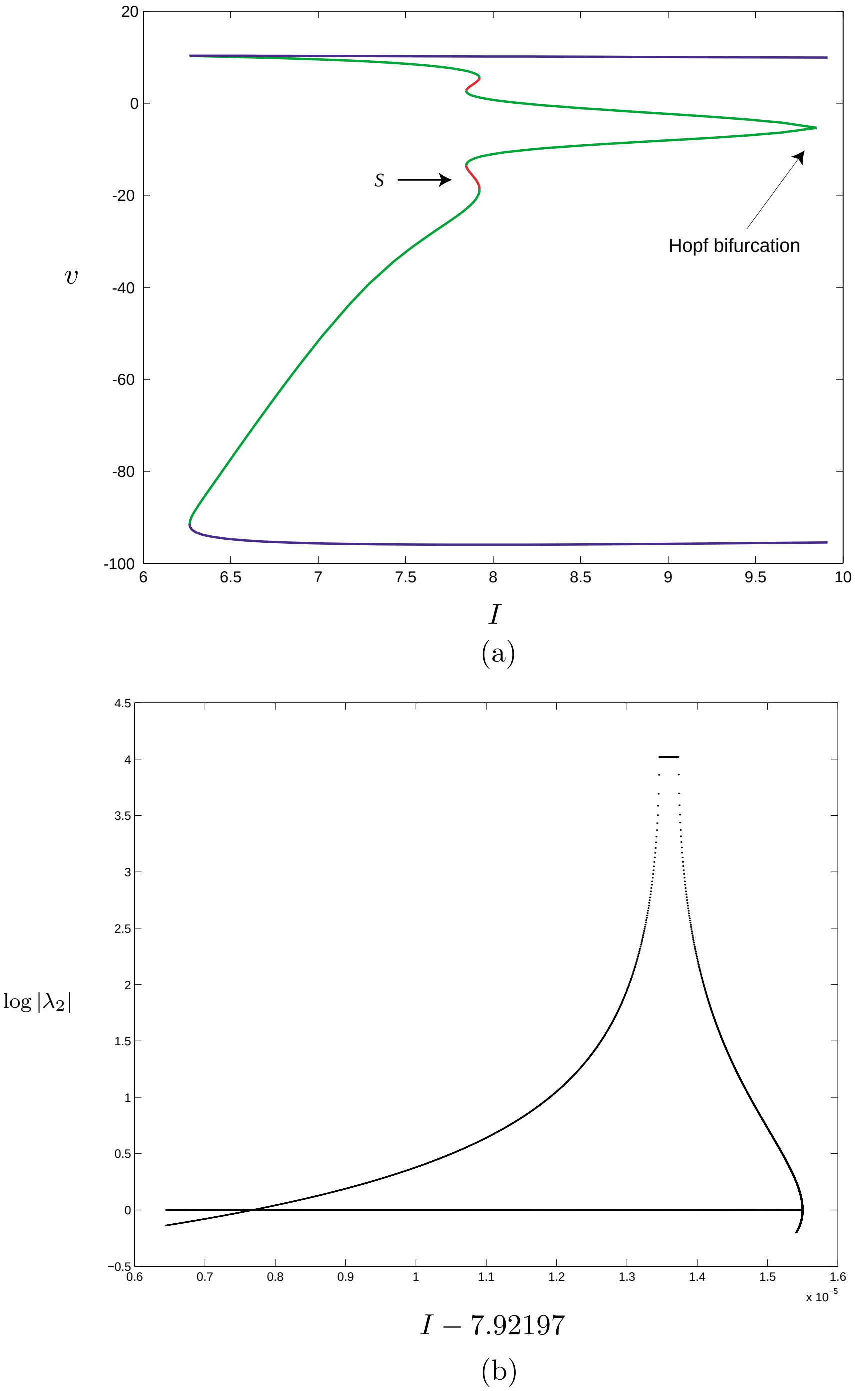
Extensive efforts have been made to discover chaos in many physical and biological systems, including neural systems [2]. Chaotic solutions to the Hodgkin–Huxley equations with periodic forcing [1] and greatly altered parameters [5] have been discovered but not in the original Hodgkin–Huxley model with its original parameters. The chaotic solutions we exhibit are highly unstable. We employed systematic methods to find them as described below. Note that, while we find our numerical evidence for the existence of chaos in the Hodgkin–Huxley model compelling, we do not give a rigorous proof that chaos exists in this system. The biological significance of chaos in the Hodgkin–Huxley system is related to the character of the threshold that separates states leading to repetitive firing from states that lead to a stable steady state. We return to this issue at the end of this note. (The implications of long-time unpredictability in deterministic chaotic systems have been widely discussed in a broad context by Stewart [21].)

2. Evidence for chaos in the Hodgkin–Huxley system.

A stringent definition of chaos in a discrete dynamical system is that there is an invariant subset on which the transformation is hyperbolic and topologically equivalent to a subshift of finite type [10, 20]. Continuous time dynamical systems are reduced to discrete time maps through the introduction of cross-sections and Poincaré return maps [10]. We utilize the cross-section V given by a suitably chosen value $v = -4.5$ to define a Poincaré return map f for the Hodgkin–Huxley model. Specifically, $(\bar{m}, \bar{n}, \bar{h}) = f(m, n, h)$ if the trajectory beginning at $(-4.5, m, n, h)$ next intersects the cross-section V (with v increasing) at the point $(-4.5, \bar{m}, \bar{n}, \bar{h})$. To demonstrate that f has a chaotic invariant set, we follow the strategy described by Moser [18]. We find two subsets R_1 and R_2 of V and approximate splittings of the tangent bundles into stable and unstable directions on these sets so that the following hold:

1. The derivative of f maps expanding directions close to themselves, stretching the lengths of these vectors.
2. The sets $f(R_1)$ and $f(R_2)$ each intersect R_1 and R_2 so that their images stretch across R_1 and R_2 in the unstable directions and intersect the boundaries of R_1 and R_2 only on sets transverse to the unstable directions.

Using the more precise concept of invariant cone fields, Moser proved that a map f satisfying these properties has a "Smale horseshoe," a hyperbolic invariant set on which f is topologically equivalent to the shift on two symbols.



(a)

(b)

Figure 2.1. (a) *The amplitude of periodic orbits in the Hodgkin–Huxley model as a function of input current. Maximum and minimum values of v are plotted for each periodic orbit. Stable periodic orbits are shown in blue; those shown in green have a single positive unstable eigenvalue, while those shown in red have either two unstable eigenvalues or a negative unstable eigenvalue.* (b) *The magnitude of the second largest eigenvalue λ_2 of the return map for the periodic orbits along the vertical branch for a small interval centered at $I = 7.92197$.*

Table 2.1

	m	n	h
R_1	0.08510711565266	0.37702513977759	0.43770368051793
	0.08511751929147	0.37708261653418	0.43799786108786
	0.08506722811171	0.37702247883827	0.43770500525944
	0.08507763175053	0.37707995559486	0.43799918582936
	0.08506795096894	0.37673045670586	0.43526697935397
	0.08507180274587	0.37675541738823	0.43543828382985
	0.08502806342799	0.37672779576653	0.43526830409547
	0.08503191520493	0.37675275644891	0.43543960857136
R_2	0.08500054963158	0.37635307899354	0.43231650435083
	0.08500126298925	0.37635224747250	0.43225555564349
	0.08499057463645	0.37635239912448	0.43231667147632
	0.08499128799412	0.37635156760344	0.43225572276897
	0.08500090973955	0.37635423427118	0.43231211561955
	0.08500146426647	0.37635361157621	0.43226513798843
	0.08499093474442	0.37635355440212	0.43231228274504
	0.08499148927134	0.37635293170715	0.43226530511391

As a parameter of a dynamical system is varied, sets satisfying the above conditions are frequently created through the "period doubling route" to chaos [6]. Rinzel and Miller [19] located period doubling bifurcations in the Hodgkin–Huxley model by computing eigenvalues along a family of periodic orbits. As the parameter corresponding to external current is varied in the model, there is a Hopf bifurcation of steady states at $I \approx 9.78$. The bifurcation is subcritical, with a family of unstable periodic orbits collapsing to the equilibrium at the bifurcation. Beginning at this bifurcation, we followed the family of periodic orbits using continuation methods. Some of the observed phenomena along this family of periodic orbits are numerically sensitive, and so we used two continuation algorithms: a collocation method as implemented by Doedel in the program AUTO [4] and a multiple shooting method that employs Taylor series methods and automatic differentiation [12]. The results of the two methods agreed with one another qualitatively.

Figure 2.1(a) plots the amplitude of the periodic orbits, showing the maximum and minimum values of v for each orbit versus the parameter I for this family of periodic orbits. The color coding corresponds to the number of eigenvalues with magnitude larger than one (blue $= 0$, green $= 1$, red $= 2$ or 1 unstable negative and 1 stable negative). The orbits grow in amplitude as I decreases from its value at the Hopf bifurcation. The branch has three turning points, all of which are saddle-node (or fold) bifurcations of periodic orbits. The first two turning points reached from the Hopf bifurcation bound a short branch of periodic orbits with two unstable eigenvalues. At the third turning point, the periodic orbits meet a family of stable periodic orbits. The amplitude of the stable periodic orbits does not change much. On the (red) branch S of periodic orbits, there is a parameter interval near $I \approx 7.92197$ inside which the unstable eigenvalues of the periodic orbit change dramatically. Figure 2.1(b) plots the log of the magnitude of the real part of one of these eigenvalues. Along the flat portion at the top of this graph, the eigenvalue is complex. This allows the eigenvalue to become real

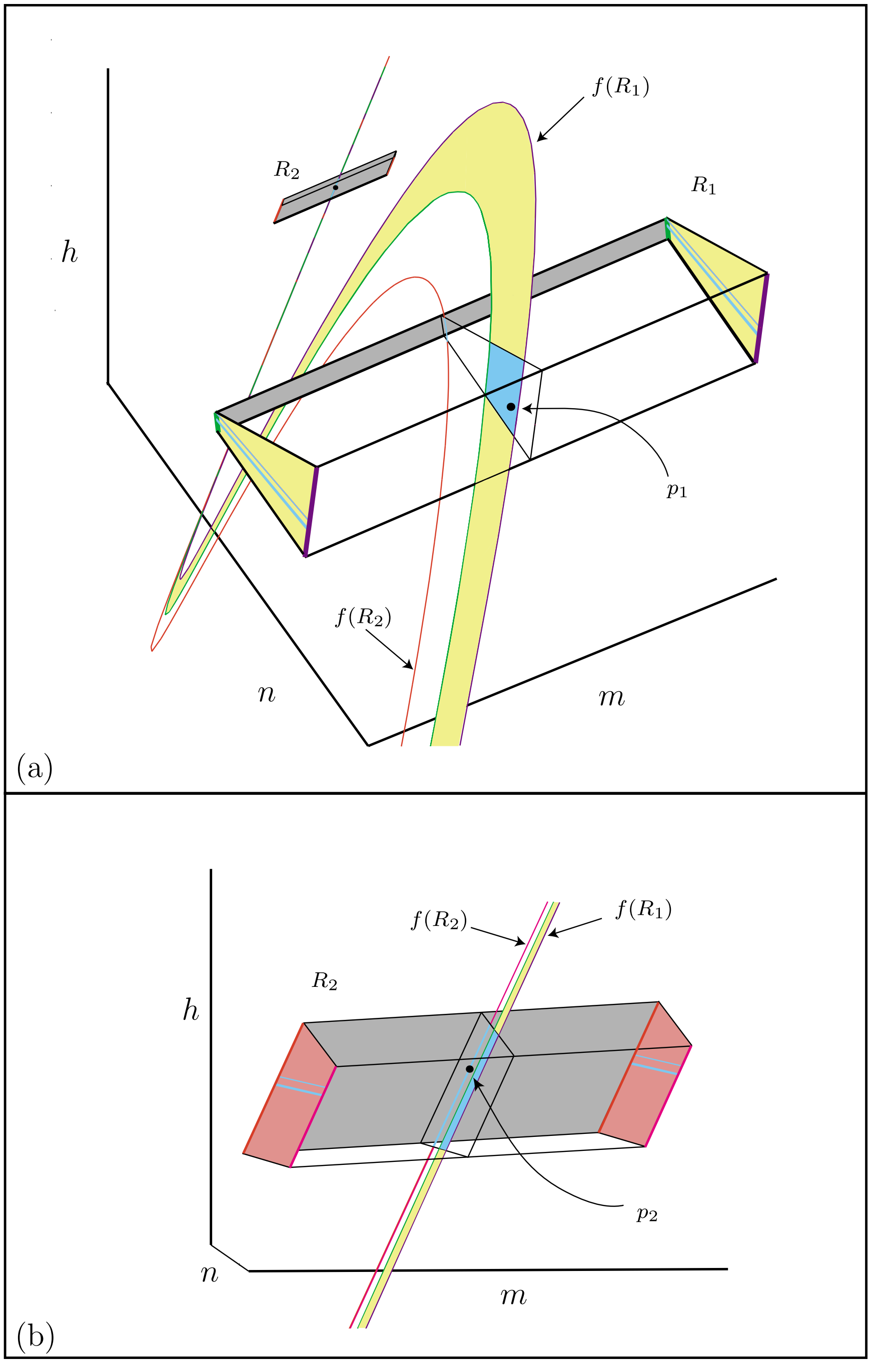


Figure 2.2. *Three dimensional boxes, and their images, which produce a horseshoe for the return map of the Hodgkin–Huxley model.* (a) *The ends of box R_1 transverse to the strong stable direction, and their images are shaded in yellow with the strips mapping into R_1 and R_2 shaded in light blue.* (b) *An expanded view of the region around R_2. The ends of box R_2 transverse to the strong stable direction, and their images are shaded in red with the strips mapping into R_1 and R_2 shaded in light blue.*

and negative without becoming stable along S. A period doubling bifurcation occurs on the part of S where the eigenvalue is negative and takes the value -1. We followed the doubled period orbits bifurcating from S and observed that they too undergo period doubling bifurcations. These computations suggest the existence of a period doubling cascade; we followed the cascade to its third period doubling. Nonetheless, these results prompted us to look for horseshoes in the return map f for values of I slightly larger than those at which the period doubling was found.

In the regime near the period doubling bifurcation, the periodic orbit of the family computed with continuation has a negative unstable characteristic multiplier, a negative stable characteristic multiplier, and a third positive characteristic multiplier of very small magnitude. We investigated these periodic orbits in more detail for the parameter $I = 7.8617827403$. The return map f has a fixed point p_1 at the intersection of the periodic orbit with the cross-section V. We find that

$$p_1 \approx (-4.5, 0.08508337639787, 0.37698374610906, 0.43727279295129).$$

The sets R_1 and R_2 of our construction intersect the unstable manifold of p_1 for the return map f. We found, as expected, that the unstable manifold has a bend. Starting in the unstable manifold, we used a shooting method to find a second fixed point of f:

$$p_2 \approx (-4.5, 0.08499590453730, 0.37635277095981, 0.43229451177364).$$

The unstable manifold of p_1 passes near p_2 and vice-versa. Near p_1, both unstable manifolds lie close to the plane of p_1 perpendicular to the strongly contracting eigenvector. To define R_1, we begin with two segments of the unstable manifolds of p_1 and p_2 near p_1. We enlarge the convex hull of these two segments in the plane perpendicular to the strongly contracting eigenvector of p_1 to form a quadrilateral. Finally, we construct a prism with this quadrilateral as base with edges parallel to the strongly contracting eigenvector of p_1. The set R_2 is obtained by an analogous procedure near p_2. Table 2.1 gives the coordinates of the vertices of R_1 and R_2. Figure 2.2 shows pictures of the sets R_1 and R_2 and their images under the return map f. (The scales of the coordinate axes in this figure are not uniform.) Each of $f(R_1)$ and $f(R_2)$ maps across both R_1 and R_2 in the unstable direction intersecting the boundaries of R_1 and R_2 in sets that are transverse to the unstable direction. These properties are evidence for the existence of a Smale horseshoe in V for the return map f.

3. Biological significance. We turn now to the significance of these chaotic dynamics in the Hodgkin–Huxley model. The chaotic invariant set located above is a highly unstable structure associated with the "threshold" for action potentials. Action potentials of neurons are large all-or-nothing voltage spikes. We specify a minimum amplitude of voltage (say, $v = -50$) that must be reached to make the definition of action potential precise. In axons that are not space-clamped, like those represented by the Hodgkin–Huxley model, action potentials propagate along axons as traveling waves and stimulate synaptic currents in adjacent postsynaptic neurons. Threshold is the magnitude of an input that must be exceeded for an action potential to fire. This definition of threshold is based upon the assumption that there is a critical current input I_c above which the axon will fire an action potential when given a brief stimulus of magnitude I_c (with fixed duration) and below which it will not. The chaotic invariant set of the Hodgkin–Huxley model casts strong doubt on the validity of this assumption. It suggests that the boundary between initial states that lead to action potentials and those that do not is a fractal set. Related observations about fractal basin boundaries have been made for periodically forced neural models, for example, by Gong and Xu [9], but the firing of an action potential does not coincide with lying in one basin of attraction or another. There may be initial conditions leading to the firing of one or more action potentials followed by a decay in the quiescent equilibrium state.

The concept of threshold is rarely given a precise mathematical meaning, even in the context of models. We propose the following definition for the Hodgkin–Huxley model: $v = v_t(m, n, h)$ is a threshold function if initial states with $v > v_t(m, n, h)$ yield action potentials, while initial states with $v < v_t(m, n, h)$ do not produce action potentials. We conjecture that no such function exists over a small range of steady input currents, all other model parameters taking the values assigned by Hodgkin and Huxley. Note that this definition is formulated in terms of initial conditions of the model with fixed parameters. All of the trajectories that appear to have local minima near the cutoff $v = -50$ for action potentials tend rapidly to the stable periodic orbit on their next oscillation.

The membrane oscillations within the horseshoe are intermediate in amplitude between action potentials and the resting membrane potential of the axon. The stable branch of periodic solutions in Figure 2.1(a) displays that the minimum of the action potentials is in the range $[-100, -90]$ mv for a Hodgkin–Huxley model that fires repetitively. The resting membrane potential is near 0. (With now standard conventions for membrane potential, v should be replaced by $v_o - v$ for a value of $v_o \approx -55$ in the Hodgkin–Huxley model.) Figure 3.1(a) displays two periods of the oscillations of the periodic orbits through points p_1 (blue) and p_2 (red) and the stable periodic orbit (black). Figure 3.1(b) shows the projections of these periodic orbits onto the (v, h) plane. Other trajectories in the horseshoe oscillate in an irregular pattern but with amplitudes that are approximated by the amplitudes of the orbits through p_1 and p_2.

The stable manifolds of unstable, chaotic invariant sets often form fractal basin boundaries of attractors in a dynamical system [17]. We conjecture that this is the case in the Hodgkin–Huxley model for the parameters used in this study; namely, the basin boundary between the basins of attraction of the stable equilibrium and stable periodic orbit is a fractal set that contains the stable manifold of the chaotic invariant we have discovered. Similar observations about fractal basin boundaries have been made for periodically forced neural models—for example, in the Fitzhugh–Nagumo model by Gong and Xu [9]. Initial states that lead to the stable rest state and those that lead to the periodic firing state are interleaved. Instead of a single sheet given by the graph of the threshold function v_t, we expect an infinite number of layers that lead to action potentials interspersed with layers that lead to a stable steady state. The trajectories tending to the stable steady state may show a transition with oscillations of smaller amplitude than the action potentials. There are also uncountable sheets that lie at threshold in the sense that they lead neither to the stable rest state nor to fully formed action potentials, but every neighborhood contains states that lead to action potentials and states that lead to the stable steady state. Due to the stiffness of the model and the fine scales on which the fractals appear, numerical calculation of the fractal basin boundaries of the Hodgkin–Huxley model with its standard parameter values appears difficult.

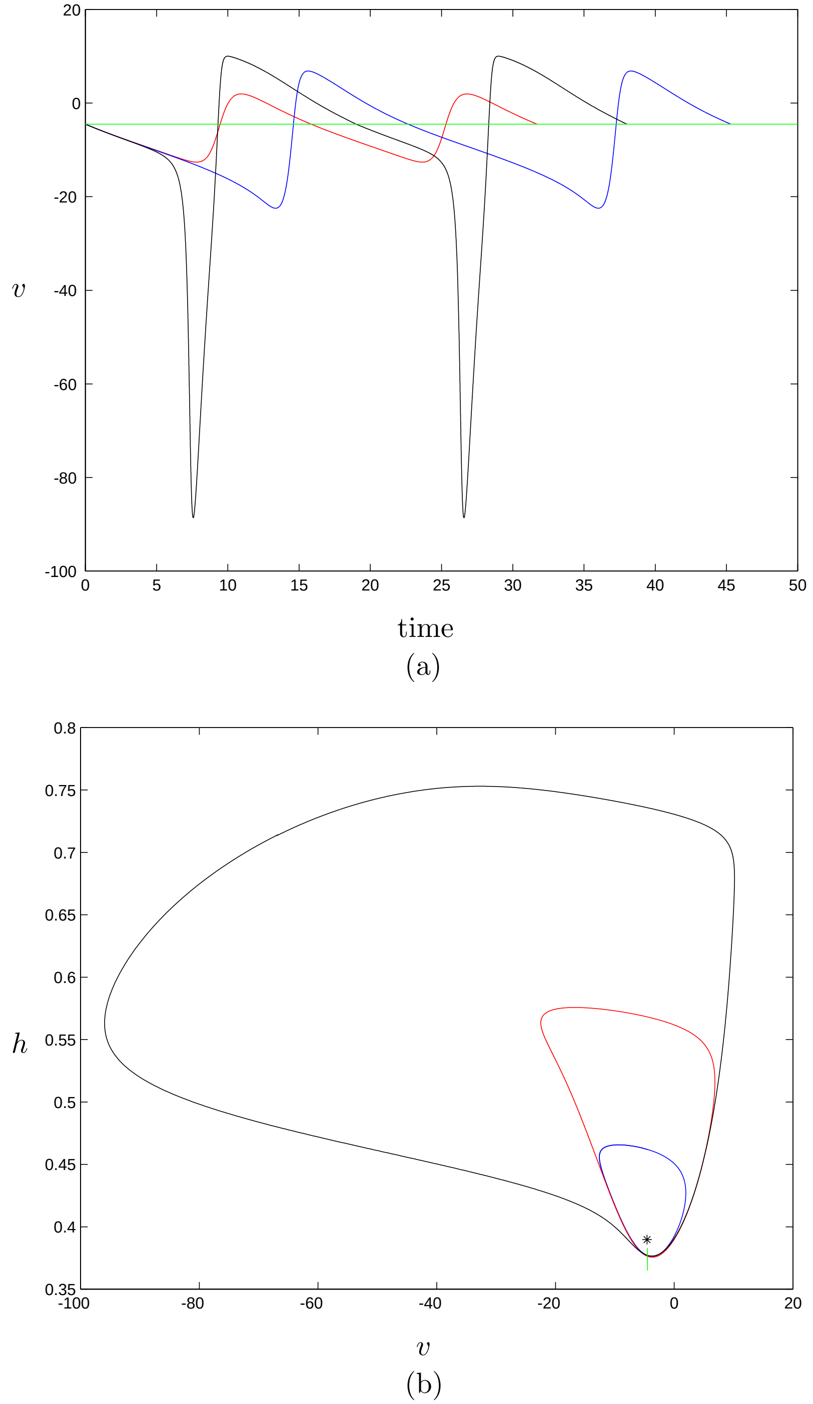


Figure 3.1. (a) *Two cycles of three periodic orbits of the Hodgkin–Huxley model with external current* $I = 14.2211827403$. *The stable periodic orbit is shown in black, and two unstable periodic orbits are shown in red and blue. The green horizontal line is at the value* $v = -4.5$ *of the cross-section used in our computations of a return map.* (b) *Projection of the three periodic orbits onto the* (v, h) *plane. The green segment is the cross-section used to obtain return maps.*

If our conjectural description of the phase space of the Hodgkin–Huxley model is correct, then there is a degree of unpredictability about how the system will respond to stimulation. Brief current inputs to the axon that evolve to the stable steady state and those that evolve to the firing state are finely interleaved with each other as the amplitude of the current input is varied. The nonlinear dynamics underlying action potentials yield an inherent lack of predictability in determining how large an input is required to cross the threshold for firing action potentials. Due to the extreme Lyapunov exponents associated to the chaotic invariant set, the fractal, interleaved structure of the membrane threshold is hardly observable in the Hodgkin–Huxley model even in computer simulation. Inherent noise in the membrane has a far larger scale than the fractal structure in the Hodgkin–Huxley model for its standard parameters. However, we believe that the phenomenon seen here on fine scales may well be present on larger scales in other neural systems or for different parameter values of the Hodgkin–Huxley model. The significance of our results is that they establish the subtlety of the concept of threshold: the excitability of a neural membrane to fire an action potential may be more complex than a smooth hypersurface that divides subthreshold and suprathreshold membrane potentials.

REFERENCES

[1] K. Aihara and G. Matsumoto, *Chaotic oscillations and bifurcations in squid giant axons*, in Chaos, A. Holden, ed., Manchester University Press, Manchester, UK, 1986, pp. 257–269.

[2] D. Campbell and H. Rose, eds., *Order in chaos*, Phys. D, 7 (1983), pp. 1–362.

[3] J. R. Clay, *Excitability of the squid giant axon revisited*, J. Neurophysiol., 80 (1998), pp. 903–913.

[4] E. Doedel, H. B. Keller, and J. P. Kernévez, *Numerical analysis and control of bifurcation problems* I, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 1 (1991), pp. 493–520.

[5] S. Doi and S. Kumagai, *Nonlinear dynamics of small-scale biophysical neural networks*, in Biophysical Neural Networks, R. R. Poznanski, ed., Mary Ann Liebert, Inc., Larchmont, NY, 2001, pp. 261–301.

[6] J.-P. Eckmann, *Roads to turbulence in dissipative dynamical systems*, Rev. Modern Phys., 53 (1981), pp. 643–654.

[7] R. Fitzhugh, *Impulses and physiological states in models of nerve membrane*, Biophys. J., 1 (1961), pp. 445–466.

[8] H. Fukai, T. Nomura, S. Doi, and S. Sato, *Hopf bifurcations in multiple-parameter space of the Hodgkin-Huxley equations*, I, II, Biol. Cybern., 82 (2000), pp. 215–222; 223–229.

[9] P.-L. Gong and J.-X. Xu, *Global dynamics and stochastic resonance of the forced Fitzhgh-Nagumo neuron model*, Phys. Rev. E. (3), 63 (2001), pp. 1–10.

[10] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.

[11] J. Guckenheimer and I. S. Labouriau, *Bifurcation of the Hodgkin–Huxley equations: A new twist*, Bull. Math. Biol., 55 (1993), pp. 937–952.

[12] J. Guckenheimer and B. Meloon, *Computing periodic orbits and their bifurcations with automatic differentiation*, SIAM J. Sci. Comput., 22 (2000), pp. 951–985.

[13] B. Hassard, *Bifurcation of periodic solutions of the Hodgkin-Huxley model for the squid giant axon*, J. Theoret. Biol., 71 (1978), pp. 401–420.

[14] B. Hassard and L.-J. Shiau, *A special point of Z_2-codimension three Hopf bifurcation in the Hodgkin-Huxley model*, Appl. Math. Lett., 9 (1996), pp. 31–34.

[15] A. L. Hodgkin and A. F. Huxley, *A quantitative description of membrane current and its applications to conduction and excitation in nerve*, J. Physiol. (Lond.), 116 (1952), pp. 500–544.

[16] I. S. Labouriau, *Degenerate Hopf bifurcation and nerve impulse* II, SIAM J. Math. Anal., 20 (1989), pp. 1–12.

[17] S. W. McDonald, C. Grebogi, E. Ott, and J. A. Yorke, *Fractal basin boundaries*, Phys. D, 17 (1985), pp. 125–153.

[18] J. Moser, *Stable and Random Motions in Dynamical Systems*, Princeton University Press, Princeton, NJ, 1973.

[19] J. Rinzel and R. Miller, *Numerical calculation of stable and unstable periodic solutions to the Hodgkin-Huxley equations*, Math. Biosci., 49 (1980), pp. 27–59.

[20] S. Smale, *Diffeomorphisms with many periodic points*, in Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), Princeton University Press, Princeton, NJ, 1965, pp. 63–80.

[21] I. N. Stewart, *Does God Play Dice?: The Mathematics of Chaos*, Basil Blackwell, Oxford, UK, 1989.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 1, pp. 115–145

Calculation of the Stability Index in Parameter-Dependent Calculus of Variations Problems: Buckling of a Twisted Elastic Strut*

Kathleen A. Hoffman[†], Robert S. Manning[‡], and Randy C. Paffenroth[§]

Abstract. We consider the problem of minimizing the energy of an inextensible elastic strut with length 1 subject to an imposed twist angle and force. In a standard calculus of variations approach, one first locates equilibria by solving the Euler–Lagrange ODE with boundary conditions at arclength values 0 and 1. Then one classifies each equilibrium by counting conjugate points, with local minima corresponding to equilibria with no conjugate points. These conjugate points are arclength values $\sigma \leq 1$ at which a second ODE (the Jacobi equation) has a solution vanishing at 0 and σ.

Finding conjugate points normally involves the numerical solution of a set of initial value problems for the Jacobi equation. For problems involving a parameter λ, such as the force or twist angle in the elastic strut, this computation must be repeated for every value of λ of interest.

Here we present an alternative approach that takes advantage of the presence of a parameter λ. Rather than search for conjugate points $\sigma \leq 1$ at a fixed value of λ, we search for a set of special parameter values λ_m (with corresponding Jacobi solution $\boldsymbol{\zeta}^m$) for which $\sigma = 1$ is a conjugate point. We show that, under appropriate assumptions, the index of an equilibrium at *any* λ equals the number of these $\boldsymbol{\zeta}^m$ for which $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle < 0$, where $\mathcal{S}$ is the Jacobi differential operator at λ. This computation is particularly simple when λ appears linearly in $\mathcal{S}$.

We apply this approach to the elastic strut, in which the force appears linearly in $\mathcal{S}$, and, as a result, we locate the conjugate points for any twisted unbuckled rod configuration without resorting to numerical solution of differential equations. In addition, we numerically compute two-dimensional sheets of buckled equilibria (as the two parameters of force and twist are varied) via a coordinated family of one-dimensional parameter continuation computations. Conjugate points for these buckled equilibria are determined by numerical solution of the Jacobi ODE.

Key words. elastic rods, anisotropy, stability index, conjugate points, buckling, parameter continuation, isoperimetric constraints

AMS subject classifications. 49K15, 34B08, 74K10, 74G60, 65P30, 65L10

PII. S1111111101396622

1. Introduction.

The classification of equilibria is a familiar idea from finite-dimensional optimization. Given an equilibrium (or critical point) of a function, one computes the eigenvalues of the Hessian matrix and determines the type of the equilibrium by computing the index, the number of negative eigenvalues. Equilibria with index 0 are local minima, and those

[*] Received by the editors October 17, 2001; accepted for publication (in revised form) by P. Holmes May 6, 2002; published electronically July 23, 2002.

http://www.siam.org/journals/siads/1-1/39662.html

[†] Department of Mathematics and Statistics, University of Maryland, Baltimore County, MD 21250 (khoffman@math.umbc.edu).

[‡] Mathematics Department, Haverford College, Haverford, PA 19041 (rmanning@haverford.edu). The research of this author was supported by NSF grant DMS-9973258.

[§] Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA 91125 (redrod@acm.org). The research of this author was supported by NSF grant KDI/NCC Molecular information and computer modeling in electrophysiology, SBR-9873173.

with index 1 are saddle points with one downward direction, etc. The goal is analogous in infinite dimensions, when the quantity J to be optimized is itself a function of a function $\mathbf{q}(s)$, say, with $0 \leq s \leq 1$.[1] Now the equilibria $\mathbf{q}_0(s)$ are found by solving the Euler–Lagrange ordinary differential equation (ODE). The role of the Hessian is played by the second-variation operator, and the notion of index, the number of negative eigenvalues in the spectrum of this operator, remains, with local minima again characterized as having index 0. In many applications, s is a spatial variable and J a potential energy,[2] and we think of the index as determining "stability," with the idea that, when time-dependence and a kinetic energy are added to this potential energy, the index-0 local minima will likely be dynamically stable. We emphasize, however, that, in this paper, "stability" is purely a classification of the type of equilibrium for an optimization problem.

In the calculus of variations, there is a powerful theory that allows a relatively simple determination of the index (see, e.g., [9]). Given an equilibrium, one consider the associated Jacobi equation, a linear second-order ODE. One then defines the notion of a conjugate point, which is a number σ, such that the Jacobi equation has a nonzero solution that vanishes at $s = 0$ and $s = \sigma$. Jacobi's strengthened condition says that the equilibrium is a local minimum if it has no conjugate point in $(0, 1]$. Morse [20] later generalized this theory to show that, assuming 1 is not a conjugate point, the number of conjugate points in $(0, 1)$ is the index of the equilibrium. This Euler–Jacobi–Morse index theory thus achieves a remarkable simplification: it reduces the problem of counting the number of negative eigenvalues of the second-variation operator of J in infinite-dimensional space to the problem of solving a second-order ODE and counting conjugate points. Thus the only numerical approximation arises in the discretization of ODEs, namely, the Euler–Lagrange and Jacobi equations.

Here we present a further simplification applicable to parameter-dependent problems and especially problems in which a parameter λ appears linearly in the second variation—for example, in problems containing a Lagrange multiplier. In this case, the counting of conjugate points at $\sigma \leq 1$ for a fixed value of λ can be reduced to the problem of analyzing conjugate points at the fixed value $\sigma = 1$ as λ is allowed to vary. Often, the determination of conjugate points at $\sigma = 1$ is easier than the general analysis of conjugate points for $\sigma \leq 1$, so this reduction can be a significant simplification. Of course, in cases where the relevant differential equations cannot be solved in closed form, this simplification is impractical, and one must revert to computing conjugate points numerically.

We further show that this entire theory can be carried through to the case of problems with isoperimetric constraints, where the only required change is an update to the definition of a conjugate point. This updated definition of a "constrained conjugate point" is the one taken, e.g., by Bolza [2] and rephrased in functional analytic language in Manning, Rogers,

[1] To rigorously define optimization in infinite dimensions, a normed space must be specified for the input functions $\mathbf{q}$. Following standard practice, we use the $C^1([0, 1])$ norm, in which case the resulting minima are traditionally referred to as "weak minima" [9].

[2] We note in passing an interesting exception. When J is a Lagrangian action functional and s represents time, the equilibria are classical mechanical trajectories, and the index plays a crucial role in the semiclassical approximation of these trajectories as quantum mechanical limits are approached [13]. The example in this paper involves a potential energy functional, but the general theory presented could equally well be applied to the determination of the semiclassical index for classical trajectories dependent on some parameter.

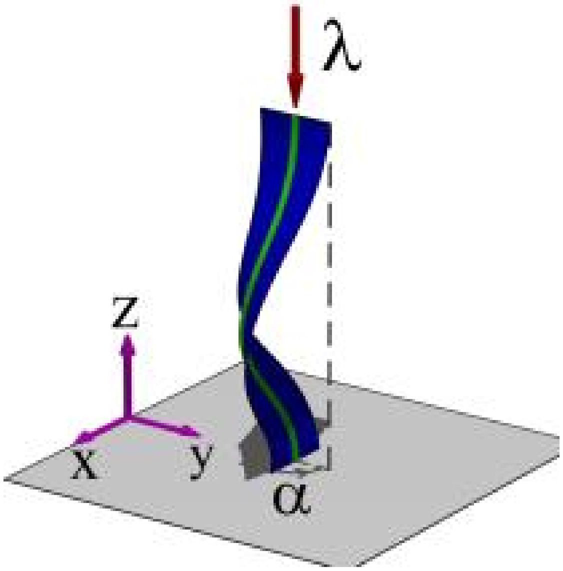


Figure 1.1. *An elastic strut clamped at each end, with a relative twist angle α, and with one end allowed to slide vertically in response to an imposed force λ. The cross-sections of the rod are elliptical, which we depict by showing the curve through the centers of the cross-sections as a green tube and tracking the major axes of the cross-sections with a blue ribbon.*

and Maddocks [19].

We apply this method for direct index computation to the example of a thin elastic strut with an elliptical (anisotropic) or circular (isotropic) cross-section, under the constraints shown in Figure 1.1. One end of the rod is clamped at the origin, with the major axis of the ellipse along the x-axis, the minor axis along the y-axis, and the tangent vector to the rod along the z-axis. The other end is constrained to lie along the z-axis, with its tangent vector also along the z-axis and its cross-section twisted by an angle α with respect to the x-y plane, but can slide up and down the z-axis in response to an applied vertical force λ.

Such a rod buckling problem has fed the curiosity of scientists since the time of Euler and Lagrange [17]. In 1883, Greenhill [11] considered the plane (over all values of α and λ) of unbuckled configurations for an isotropic strut and derived the condition for the index to be zero. More recently, Champneys and Thompson [3] and van der Heijden and Thompson [27] performed bifurcation analyses of an anisotropic elastic rod but did not focus on the question of stability. Goriely, Nizette, and Tabor [10] considered the dynamic stability of the unbuckled configurations for both the isotropic and anisotropic cases, and van der Heijden et al. [26] inferred stability for unbuckled and buckled configurations for the isotropic case from the shape of solution branches in a particular "distinguished" coordinate system. Neukirch and Henderson [21] performed an in-depth classification of buckled solutions for the isotropic problem, including the computation of two-dimensional sheets of equilibria.

Here we present stability results for this problem that both complement the results cited above and serve as an illustration for the general index theory we develop. Specifically, this theory allows a semianalytic determination of the index on the plane of unbuckled configurations for both the isotropic and anisotropic strut. (The determination is semianalytic in that it requires the numerical solution of an algebraic equation but no differential equations.) In addition, we compute sheets of buckled solutions using a family of one-dimensional branches generated by the parameter continuation package AUTO [5, 6]. We determine the index of configurations on these sheets via a numerical implementation of the conjugate point test developed in [19] since a closed-form solution of the differential equations is not feasible in this case.

We begin in section 2 by reviewing conjugate point theory for unconstrained and constrained problems. Section 3 presents our direct method for computing the index for unconstrained problems, which is then extended to problems with isoperimetric constraints in section 4. In section 5, we summarize the standard Kirchhoff theory of inextensible and unshearable elastic struts. The index on the plane of unbuckled equilibria for both the isotropic and anisotropic strut is determined in section 6, and the buckled configurations and their stability are presented in section 7.

2. A review of conjugate point theory.

2.1. Unconstrained conjugate points. In this section, we summarize conjugate point theory for unconstrained calculus of variations problems. This theory is an established part of the classic unconstrained calculus of variations literature and can be found in many standard texts, e.g., [7, 9, 14, 24]. However, index theory appears only in more modern treatments [12, 20].

We consider a functional of the form

$$J[\mathbf{q}] = \int_0^1 L(\mathbf{q}, \mathbf{q}', s)\, ds, \quad \mathbf{q}(s) \in \mathbb{R}^p, \qquad \text{subject to } \ \mathbf{q}(0) = \mathbf{h}_0, \quad \mathbf{q}(1) = \mathbf{h}_1. \tag{2.1}$$

Equilibria $\mathbf{q}_0(s)$ are solutions to the standard Euler–Lagrange equations for the functional J. Classification of these equilibria involves an analysis of the second variation of J at $\mathbf{q}_0$, namely,

$$\delta^2 J[\boldsymbol{\zeta}] = \frac{1}{2}\int_0^1 \left[(\boldsymbol{\zeta}')^T \mathbf{P}\boldsymbol{\zeta}' + (\boldsymbol{\zeta}')^T \mathbf{C}^T \boldsymbol{\zeta} + \boldsymbol{\zeta}^T \mathbf{C}\boldsymbol{\zeta}' + \boldsymbol{\zeta}^T \mathbf{Q}\boldsymbol{\zeta}\right]\, ds \tag{2.2}$$

for $\mathbf{P} \equiv L^0_{\mathbf{q}'\mathbf{q}'}$, $\mathbf{C} \equiv L^0_{\mathbf{q}\mathbf{q}'}$, and $\mathbf{Q} \equiv L^0_{\mathbf{q}\mathbf{q}}$, where, here and throughout, superscripting by T denotes the transpose, subscripting by $\mathbf{q}$ or $\mathbf{q}'$ denotes partial differentiation, and superscripting by 0 denotes evaluation at $\mathbf{q}_0(s)$. We assume that Legendre's strengthened condition holds,

$$\mathbf{P} > 0; \tag{2.3}$$

i.e., the symmetric matrix $\mathbf{P}$ is positive definite. Here $\boldsymbol{\zeta}$ is a variation in $\mathbf{q}$ that, due to the boundary conditions on $\mathbf{q}$, must lie in the set of admissible variations:

$$\mathcal{H}_d \equiv \left\{\boldsymbol{\zeta} \in H^2(\mathbb{R}^p, (0,1)) : \boldsymbol{\zeta}(0) = \boldsymbol{\zeta}(1) = \mathbf{0}\right\}.$$

Here we have chosen the Sobolev space H^2 of functions with integrable weak second derivatives because, after an integration by parts, the second variation will take the form

$$\delta^2 J[\boldsymbol{\zeta}] = \frac{1}{2}\langle \boldsymbol{\zeta}, \mathcal{S}\boldsymbol{\zeta}\rangle,$$

where $\mathcal{S}$ is the self-adjoint second-order vector differential operator

$$\mathcal{S}\boldsymbol{\zeta} \equiv -\frac{d}{ds}\left[\mathbf{P}\boldsymbol{\zeta}' + \mathbf{C}^T\boldsymbol{\zeta}\right] + \mathbf{C}\boldsymbol{\zeta}' + \mathbf{Q}\boldsymbol{\zeta}, \tag{2.4}$$

and $\langle\cdot,\cdot\rangle$ denotes the usual inner product in $L^2(\mathbb{R}^p, (0,1))$:

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^1 [\mathbf{f}(s)]^T \mathbf{g}(s)\, ds.$$

A sufficient condition for $\mathbf{q}_0$ to be a local minimum of J is a combination of Legendre's strengthened condition (2.3) with Jacobi's strengthened condition that $\mathbf{q}_0$ has no *conjugate point* in $(0,1]$, where a conjugate point is defined to be a value σ for which there is a nontrivial solution to

$$\mathcal{S}\boldsymbol{\zeta} = \mathbf{0}, \quad 0 < s < \sigma, \quad \boldsymbol{\zeta}(0) = \boldsymbol{\zeta}(\sigma) = \mathbf{0}. \tag{2.5}$$

Morse [20] extended Jacobi's condition to equate the number of conjugate points with the index that quantifies the dimension of the set on which $\delta^2 J$ is negative.

In some cases, (2.5) can be solved analytically, but generically a numerical procedure for computing conjugate points is required. For completeness, we briefly summarize a standard algorithm for computing unconstrained conjugate points [24, p. 152]. We numerically compute a basis of solutions $\{\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_p\}$ to the homogeneous second-order initial value problem

$$\mathcal{S}\boldsymbol{\zeta} = \mathbf{0}, \quad \boldsymbol{\zeta}(0) = \mathbf{0}. \tag{2.6}$$

A conjugate point occurs when a nontrivial linear combination of $\{\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_p\}$ vanishes, i.e., when there is a nontrivial solution to the $p \times p$ linear system

$$\begin{bmatrix} \boldsymbol{\zeta}_1 & \dots & \boldsymbol{\zeta}_p \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = \mathbf{0}. \tag{2.7}$$

Therefore, as we build up solutions $\boldsymbol{\zeta}_j(\sigma)$ to (2.6) as σ grows from 0 to 1, we track the determinant of the $p \times p$ matrix $\begin{bmatrix} \boldsymbol{\zeta}_1(\sigma) & \dots & \boldsymbol{\zeta}_p(\sigma) \end{bmatrix}$ and count as a conjugate point every time this determinant crosses zero.

2.2. Constrained conjugate points. In this section, we review a definition of a conjugate point appropriate for calculus of variations problems with isoperimetric constraints:

$$W[\mathbf{q}] = \int_0^1 L(\mathbf{q}, \mathbf{q}', s)\, ds, \quad \mathbf{q}(s) \in \mathbb{R}^p,$$

$$\text{subject to} \quad \int_0^1 g_i(\mathbf{q})\, ds = 0, \quad i = 1, \dots, n,$$

$$\mathbf{q}(0) = \mathbf{h}_0, \quad \mathbf{q}(1) = \mathbf{h}_1.$$

According to the usual multiplier rule, an associated functional

$$J[\mathbf{q}] = \int_0^1 \left(L + \mathbf{g}^T \boldsymbol{\nu}\right)\, ds$$

is constructed, and constrained equilibria $(\mathbf{q}_0(s), \boldsymbol{\nu}_0)$ of W are solutions of the standard unconstrained Euler–Lagrange equations for the functional J, with the multiplier $\boldsymbol{\nu}_0$ determined by the integral constraints. The second variation $\delta^2 J$ and its associated operator $\mathcal{S}$ take the unconstrained forms (2.2) and (2.4) but now with $\mathbf{Q} = [\mathbf{g}^0_{\mathbf{qq}}]^T \boldsymbol{\nu}_0 + L^0_{\mathbf{qq}}$. Admissible variations $\boldsymbol{\zeta}$ must satisfy the linearized constraints

$$\langle \boldsymbol{\zeta}, \mathbf{T}_i \rangle = 0, \quad i = 1, \dots , n,$$

where

$$\mathbf{T}_i \equiv (g_i)^0_{\mathbf{q}}.$$

We assume that the $\mathbf{T}_i(s)$ are linearly independent on $(0, \sigma)$ for each $\sigma \in (0, 1]$. So the admissible set of constrained variations is

$$\mathcal{H}^{cons}_d \equiv \{\boldsymbol{\zeta} \in \mathcal{H}_d : \langle \boldsymbol{\zeta}, \mathbf{T}_i \rangle = 0, \;\; i = 1, \dots , n\}\,.$$

Bolza [2], citing the work of Weierstrass and Kneser, defined the relevant notion of conjugate point for an isoperimetric problem, namely, that σ is called a conjugate point if the following system has a nontrivial solution:

$$\begin{aligned} &\mathcal{S}\boldsymbol{\zeta} = \sum_{i=1}^{n} \breve{c}_i \mathbf{T}_i, \quad 0 < s < \sigma, \quad \text{for some constants } \breve{c}_i, \\ &\boldsymbol{\zeta}(0) = \boldsymbol{\zeta}(\sigma) = \mathbf{0}, \quad \int_0^\sigma \boldsymbol{\zeta}^T \mathbf{T}_i \, ds = 0, \;\; i = 1, \dots, n. \end{aligned} \tag{2.8}$$

In [19], we introduced an orthogonal projection operator $\mathcal{Q}$ onto the L^2-orthogonal complement of span$(\mathbf{T}_1, \dots , \mathbf{T}_n)$ to rewrite Bolza's conjugate point condition (2.8) in the equivalent form

$$\begin{aligned} &(\mathcal{Q}\mathcal{S}\mathcal{Q})\boldsymbol{\zeta} = \mathbf{0}, \quad 0 < s < \sigma, \\ &\boldsymbol{\zeta}(0) = \boldsymbol{\zeta}(\sigma) = \mathbf{0}, \quad \int_0^\sigma \boldsymbol{\zeta}^T \mathbf{T}_i \, ds = 0, \;\; i = 1, \dots, n. \end{aligned} \tag{2.9}$$

Further, the arguments in [19], involving a proof of the monotonicity of the eigenvalues of $\mathcal{Q}\mathcal{S}\mathcal{Q}$ as a function of σ, demonstrate the analogue to Jacobi's strengthened condition: the lack of an isoperimetric conjugate point implies the existence of a local minimum. In addition, similar to Morse's theory in the unconstrained case, the number of isoperimetric conjugate points equals the maximal dimension of a subspace of $\mathcal{H}^{cons}_d$ on which the second variation is negative.

There is also a technique for the numerical determination of these conjugate points similar to the procedure described in section 2.1 for the unconstrained case. We summarize this technique here and refer the interested reader to [19] for complete details. In addition to numerically determining a basis of solutions to (2.6), we also compute solutions $\breve{\boldsymbol{\zeta}}_i$ to each of the n nonhomogeneous initial value problems

$$\mathcal{S}\boldsymbol{\zeta} = \mathbf{T}_i, \quad \boldsymbol{\zeta}(0) = \mathbf{0}.$$

A conjugate point occurs when there exists a nontrivial linear combination of $\{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_p, \breve{\boldsymbol{\zeta}}_1, \ldots \breve{\boldsymbol{\zeta}}_n\}$ that vanishes and also obeys the linearized constraints $\int_0^\sigma \boldsymbol{\zeta}^T \mathbf{T}_i ds = 0$. This condition has a $(p+n)$-by-$(p+n)$ matrix form analogous to the p-by-p unconstrained matrix equation (2.7). As in the unconstrained case, the matrix entries are built up by solving initial value problems as σ grows from 0 to 1, and the conjugate points are found by zero-crossings of the $(p+n)$-by-$(p+n)$ determinant.

3. Analytic determination of the index for unconstrained parameter-dependent problems. In this section, we describe an alternative approach to determining the index for equilibria of (2.1) when L, and hence the second-variation operator $\mathcal{S}$, depends on a parameter λ. This approach allows, in some cases, an analytic determination of the index. The key idea is to relate the index at a *specific* value of the parameter λ to the set of all parameter values λ that yield conjugate points at $\sigma = 1$, i.e., solutions of

$$\mathcal{S}(\lambda)\boldsymbol{\zeta} = \mathbf{0}, \quad \boldsymbol{\zeta}(0) = \boldsymbol{\zeta}(1) = \mathbf{0}. \tag{3.1}$$

We will denote solutions to (3.1) by $(\lambda_m, \boldsymbol{\zeta}^m)$ and refer to λ_m as *branch points*. This name is appropriate since, in all cases we consider, new branches of equilibria will arise at λ_m. In many applications, the analytic determination of $(\lambda_m, \boldsymbol{\zeta}^m)$ is possible even when the general conjugate point equation (2.5) cannot be solved in closed form.

We will assume that $\{\boldsymbol{\zeta}^m\}$ form a basis for $\mathcal{H}_d$ and that they are $\mathcal{S}$-orthogonal, i.e., that $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^n \rangle = 0$ for all $m \neq n$. These assumptions hold for many physically motivated problems, including cases in which (3.1) is a standard Sturm–Liouville eigenvalue problem (see, e.g., [23, p. 273]). This basis of solutions can then be used to diagonalize $\mathcal{S}$ and determine the index directly, as shown by the following lemma.

Lemma 3.1. *Assume that $\{\boldsymbol{\zeta}^m\}$ form an $\mathcal{S}$-orthogonal basis of $\mathcal{H}_d$. Then the index equals the number of $\boldsymbol{\zeta}^m$ for which $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle < 0$.*

Proof. Let $\mathcal{N}$ be the subspace of $\mathcal{H}_d$ spanned by those $\boldsymbol{\zeta}^m$ for which $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle < 0$. Any $\boldsymbol{\chi} \in \mathcal{H}_d$ that is L^2-orthogonal to $\mathcal{N}$ can be written as

$$\boldsymbol{\chi} = \sum_{\boldsymbol{\zeta}^m \notin \mathcal{N}} c_m \boldsymbol{\zeta}^m,$$

where the notation $\sum_{\boldsymbol{\zeta}^m \notin \mathcal{N}}$ denotes a sum over all m for which $\boldsymbol{\zeta}^m \notin \mathcal{N}$. Then, by $\mathcal{S}$-orthogonality,

$$\langle \boldsymbol{\chi}, \mathcal{S}\boldsymbol{\chi} \rangle = \sum_{\boldsymbol{\zeta}^m \notin \mathcal{N}} (c_m)^2 \langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle.$$

By definition, $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle \geq 0$ for all terms in the sum, so $\langle \boldsymbol{\chi}, \mathcal{S}\boldsymbol{\chi} \rangle \geq 0$.

Table 3.1
Index for the planar Euler buckling problem.

Range of λ	Functions $\boldsymbol{\zeta}^m$ for which $\langle \boldsymbol{\zeta}^m, \mathcal{S}(\lambda)\boldsymbol{\zeta}^m \rangle < 0$	Index
$0 < \lambda < \pi^2$	none	0
$\pi^2 < \lambda < 4\pi^2$	ζ^1	1
$4\pi^2 < \lambda < 9\pi^2$	ζ^1, ζ^2	2
$9\pi^2 < \lambda < 16\pi^2$	$\zeta^1, \zeta^2, \zeta^3$	3

Thus $\mathcal{N}$ is a maximal subspace in $\mathcal{H}_d$ on which the second variation is negative. The index is defined to be the dimension of $\mathcal{N}$, which, by definition of $\mathcal{N}$, is the number of basis functions $\boldsymbol{\zeta}^m$ for which $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle < 0$. ■

The above diagonalization is particularly useful when the parameter λ appears linearly in $\mathcal{S}$,

$$\mathcal{S}\boldsymbol{\zeta} = \mathcal{S}_1\boldsymbol{\zeta} + \lambda\mathcal{S}_2\boldsymbol{\zeta}, \tag{3.2}$$

because it is then routine to determine those basis functions $\boldsymbol{\zeta}^m$ for which $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle < 0$. In fact, the quantity $\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle$ is simply related to the λ-independent quantity $\langle \boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m \rangle$ by the following lemma.

Lemma 3.2. *Suppose $\mathcal{S}$ takes the form (3.2). Then, for any solution $(\lambda_m, \boldsymbol{\zeta}^m)$ of (3.1),*

$$\langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle = (\lambda - \lambda_m)\langle \boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m \rangle.$$

Proof. The key idea is to exploit the fact that $(\mathcal{S}_1 + \lambda_m\mathcal{S}_2)\boldsymbol{\zeta}^m = \mathbf{0}$:

$$\begin{aligned} \langle \boldsymbol{\zeta}^m, \mathcal{S}\boldsymbol{\zeta}^m \rangle &= \langle \boldsymbol{\zeta}^m, (\mathcal{S}_1 + \lambda\mathcal{S}_2)\boldsymbol{\zeta}^m \rangle \\ &= \langle \boldsymbol{\zeta}^m, (\mathcal{S}_1 + \lambda_m\mathcal{S}_2 - \lambda_m\mathcal{S}_2 + \lambda\mathcal{S}_2)\boldsymbol{\zeta}^m \rangle = (\lambda - \lambda_m)\langle \boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m \rangle. \end{aligned}$$ ■

Lemmas 3.1 and 3.2 lead immediately to the following corollary, the central result of this article.

Corollary 3.3. *If the solutions of (3.1) form an $\mathcal{S}$-orthogonal basis for $\mathcal{H}_d$ and if $\mathcal{S}$ has the form (3.2), then the index of an equilibrium at parameter value λ equals the number of branch points λ_m such that $\lambda > \lambda_m$ and $\langle \boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m \rangle < 0$, plus the number of branch points λ_m such that $\lambda < \lambda_m$ and $\langle \boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m \rangle > 0$.*

Example (planar buckling of a strut). We parametrize the strut by arclength s and choose a length scale so that $0 \leq s \leq 1$. We let $\theta(s)$ denote the angle that the strut makes with vertical at position s. We impose the boundary conditions that the strut be vertical at $s = 0$ and $s = 1$ and impose a vertical force λ. We then have the calculus of variations problem to minimize the total energy

$$\int_0^1 \left[\frac{1}{2}K(\theta'(s))^2 + \lambda\cos(\theta(s))\right] ds, \quad \theta(0) = \theta(1) = 0,$$

where $K > 0$ is a stiffness parameter.

Consider the unbuckled equilibrium $\theta(s) = 0$. The second-variation operator is $\mathcal{S}\zeta = -K\zeta'' - \lambda\zeta$, which is of the form (3.2), with $\mathcal{S}_2 = -1$. The nontrivial solutions to (3.1) are $\zeta^m(s) = A\sin(m\pi s)$, $m = 1, 2, 3, \ldots$, with $\lambda_m = m^2\pi^2$. These solutions are an $\mathcal{S}$-orthogonal basis for $\mathcal{H}_d$ since $\mathcal{S}$ is a Sturm–Liouville operator. We observe that $\langle\zeta^m, \mathcal{S}_2\zeta^m\rangle < 0$ for all m and then use Lemma 3.2 to conclude that the sign of $\langle\zeta^m, \mathcal{S}(\lambda)\zeta^m\rangle$ is the same as the sign of $m^2\pi^2 - \lambda$. Hence we conclude that the index for arbitrary λ is as given in Table 3.1.

4. Analytic determination of the index for isoperimetrically constrained parameter-dependent problems.

Next we combine the ideas of the previous two sections to produce a method for directly computing the index in a calculus of variations problem with isoperimetric constraints, when the second-variation operator $\mathcal{S}$ depends on a parameter λ. As in section 3, we assume that solutions $\boldsymbol{\zeta}^m$ of

$$\begin{aligned} &\mathcal{S}(\lambda)\boldsymbol{\zeta} = \sum_{i=1}^{n} \breve{c}_i \mathbf{T}_i, \quad 0 < s < 1, \quad \text{for some constants } \breve{c}_i, \\ &\boldsymbol{\zeta}(0) = \boldsymbol{\zeta}(1) = \mathbf{0}, \qquad \int_0^1 \boldsymbol{\zeta}^T \mathbf{T}_i \, ds = 0, \quad i = 1, \ldots, n, \end{aligned} \tag{4.1}$$

form an $\mathcal{S}$-orthogonal basis for the relevant function space $\mathcal{H}_d^{cons}$. Although (4.1) reflects Bolza's definition of conjugate point at $\sigma = 1$, we will exploit the equivalence of (2.8) and (2.9) to rewrite the differential equation $\mathcal{S}(\lambda)\boldsymbol{\zeta} = \sum_{i=1}^n \breve{c}_i\mathbf{T}_i$ as $(\mathcal{Q}\mathcal{S}(\lambda)\mathcal{Q})\boldsymbol{\zeta} = \mathbf{0}$. As before, the basis $\{\boldsymbol{\zeta}^m\}$ will be used to diagonalize the projected operator $\mathcal{QSQ}$ on $\mathcal{H}_d^{cons}$. In fact, Lemma 3.1 holds for the constrained case, provided $\mathcal{H}_d$ is replaced with the space $\mathcal{H}_d^{cons}$, noting that, since $\mathcal{Q}$ is self-adjoint, $\langle\boldsymbol{\zeta}, (\mathcal{QSQ})\boldsymbol{\zeta}\rangle = \langle\boldsymbol{\zeta}, \mathcal{S}\boldsymbol{\zeta}\rangle$ for $\boldsymbol{\zeta} \in \mathcal{H}_d^{cons}$.

If, in addition, $\mathcal{S}$ takes the form (3.2), then Lemma 3.2 can be extended to problems with isoperimetric constraints as follows.

Lemma 4.1. *Suppose $\mathcal{S}$ takes the form* (3.2). *Then, for any solution $\boldsymbol{\zeta}^m$ of* (4.1),

$$\langle\boldsymbol{\zeta}^m, (\mathcal{QSQ})\boldsymbol{\zeta}^m\rangle = (\lambda - \lambda_m)\langle\boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m\rangle.$$

Proof. As in Lemma 3.2, the key idea is to exploit the fact that $(\mathcal{Q}(\mathcal{S}_1 + \lambda_m\mathcal{S}_2)\mathcal{Q})\boldsymbol{\zeta}^m = \mathbf{0}$, combined with the fact that $\mathcal{Q}\boldsymbol{\zeta}^m = \boldsymbol{\zeta}^m$ for $\boldsymbol{\zeta}^m \in \mathcal{H}_d^{cons}$:

$$\begin{aligned} \langle\boldsymbol{\zeta}^m, (\mathcal{QSQ})\boldsymbol{\zeta}^m\rangle &= \langle\boldsymbol{\zeta}^m, (\mathcal{Q}(\mathcal{S}_1 + \lambda\mathcal{S}_2)\mathcal{Q})\boldsymbol{\zeta}^m\rangle \\ &= \langle\boldsymbol{\zeta}^m, (\mathcal{Q}(\mathcal{S}_1 + \lambda_m\mathcal{S}_2 - \lambda_m\mathcal{S}_2 + \lambda\mathcal{S}_2)\mathcal{Q})\boldsymbol{\zeta}^m\rangle \\ &= \langle\mathcal{Q}\boldsymbol{\zeta}^m, (-\lambda_m\mathcal{S}_2 + \lambda\mathcal{S}_2)\mathcal{Q}\boldsymbol{\zeta}^m\rangle \\ &= (\lambda - \lambda_m)\langle\boldsymbol{\zeta}^m, \mathcal{S}_2\boldsymbol{\zeta}^m\rangle. \qquad \blacksquare \end{aligned}$$

Thus the strategy for analytically determining the index in problems with isoperimetric constraints is the same as for unconstrained problems. This strategy will be illustrated in section 6 for the example of the buckling of an elastic strut under imposed force and twist. In preparation for this example, we next summarize the basic elastic strut equations.

5. The elastic strut.

5.1. Equilibrium equations.

In the Kirchhoff theory of inextensible and unshearable elastic struts [1, 4, 16, 17], the configuration of a strut is described by a centerline $\mathbf{r}(s)$ (written as a function of arclength s) and a set of directors $\{\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s)\}$ that form an orthonormal frame giving the orientation of the cross-section of the strut. For convenience, we choose a length scale so that $0 \leq s \leq 1$. Let the superscript $'$ denote a derivative with respect to s. The assumptions of inextensibility and unshearability of the strut are incorporated in the requirement that $\mathbf{d}_3(s)$, the director orthogonal to the strut cross-section, equals $\mathbf{r}'(s)$, the tangent vector to the centerline.

Orthonormality of the directors implies the existence of a (Darboux) vector $\mathbf{u}(s)$ defined by the kinematic relations

$$\mathbf{d}_i'(s) = \mathbf{u}(s) \times \mathbf{d}_i(s), \quad i = 1, 2, 3.$$

The components of $\mathbf{u}$ in the strut frame are denoted by $u_i(s) = \mathbf{u}(s) \cdot \mathbf{d}_i(s)$ and are called the *strains.*

It will be convenient to describe the directors via Euler parameters or quaternions $\mathbf{q} \in \mathbb{R}^4$. Quaternions provide an alternate formulation to Euler angles for parametrizing rotation matrices in $SO(3)$ (see, e.g., [25, p. 462]). The directors $\mathbf{d}_i$ are the columns of a rotation matrix and can be expressed by rational functions of the quaternions,

$$\mathbf{d}_1 = \frac{1}{|\mathbf{q}|^2} \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 \\ 2q_1q_2 + 2q_3q_4 \\ 2q_1q_3 - 2q_2q_4 \end{bmatrix}, \quad \mathbf{d}_2 = \frac{1}{|\mathbf{q}|^2} \begin{bmatrix} 2q_1q_2 - 2q_3q_4 \\ -q_1^2 + q_2^2 - q_3^2 + q_4^2 \\ 2q_2q_3 + 2q_1q_4 \end{bmatrix},$$

$$\mathbf{d}_3 = \frac{1}{|\mathbf{q}|^2} \begin{bmatrix} 2q_1q_3 + 2q_2q_4 \\ 2q_2q_3 - 2q_1q_4 \\ -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix},$$

and therefore the strains can be expressed as

$$u_1 = \frac{2}{|\mathbf{q}|^2} \left(q_1'q_4 + q_2'q_3 - q_3'q_2 - q_4'q_1 \right), \quad u_2 = \frac{2}{|\mathbf{q}|^2} \left(-q_1'q_3 + q_2'q_4 + q_3'q_1 - q_4'q_2 \right),$$

$$u_3 = \frac{2}{|\mathbf{q}|^2} \left(q_1'q_2 - q_2'q_1 + q_3'q_4 - q_4'q_3 \right).$$

For convenience, we define $d_{3i}(s)$ to be the ith component of the vector $\mathbf{d}_3$. We impose the following constraints on the elastic strut:

(5.1)
$$\int_0^1 d_{31}(\mathbf{q}(s))\, ds = \int_0^1 d_{32}(\mathbf{q}(s)) ds = 0, \quad \mathbf{q}(0) = (0, 0, 0, 1), \quad \mathbf{q}(1) = (0, 0, \sin(\alpha/2), \cos(\alpha/2)).$$

The boundary condition on $\mathbf{q}(0)$ forces the $s = 0$ end of the strut to be tangent to the z-axis. The pair of integral constraints imply, due to the inextensibility-unshearability condition $\mathbf{r}'(s) = \mathbf{d}_3(s)$, that $x(0) = x(1)$ and $y(0) = y(1)$, i.e., that the $s = 1$ end of the strut lies directly above the $s = 0$ end. Finally, the boundary condition on $\mathbf{q}(1)$ implies that the $s = 1$ end of the strut is tangent to the z-axis and is twisted by an angle α with respect to the $s = 0$ end. Figure 1.1 depicts a solution that satisfies (5.1).

The elastic energy of the strut is expressed in terms of the strains, and we assume here a commonly used quadratic energy

$$E[\mathbf{q}] \equiv \int_0^1 \left[\sum_{i=1}^{3} \frac{1}{2} K_i \left[u_i(\mathbf{q}(s), \mathbf{q}'(s)) \right]^2 + \lambda d_{33} \right] ds,$$

where K_i are the bending ($i = 1, 2$) and twisting ($i = 3$) stiffnesses of the strut, and λ is a force pushing downward on the strut. Unbuckled twisted equilibria are described by

$$\mathbf{q}_0(s) = \begin{bmatrix} 0 \\ 0 \\ \sin(\frac{\alpha s}{2}) \\ \cos(\frac{\alpha s}{2}) \end{bmatrix},$$

which correspond to directors

$$\mathbf{d}_1 = \begin{bmatrix} \cos(\alpha s) \\ \sin(\alpha s) \\ 0 \end{bmatrix}, \quad \mathbf{d}_2 = \begin{bmatrix} -\sin(\alpha s) \\ \cos(\alpha s) \\ 0 \end{bmatrix}, \quad \mathbf{d}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

These are readily verified to be solutions to the Euler–Lagrange equations of the functional E. The goal of section 6 is to use the theory of section 4 to assign to these configurations a stability index.

5.2. The second variation. In this section, we show a routine but technical computation of the second variation of E, finding in the end that it takes the form $\mathcal{S} = \mathcal{S}_1 + \lambda \mathcal{S}_2$ required for our results. The computation is somewhat involved due to our choice to parametrize $SO(3)$ by four-dimensional quaternions.

As outlined in section 2.2, for each equilibrium $\mathbf{q}_0$, we define an *allowed variation* to be any $\delta\mathbf{q}$ so that $\delta\mathbf{q}(0) = \delta\mathbf{q}(1) = \mathbf{0}$ and $\langle \delta\mathbf{q}, (d_{3i})^0_{\mathbf{q}} \rangle = 0$ for $i = 1, 2$. The second variation of E is

$$\delta^2 E[\delta\mathbf{q}] = \int_0^1 \left[(\delta\mathbf{q}')^T L^0_{\mathbf{q}'\mathbf{q}'} \delta\mathbf{q}' + (\delta\mathbf{q}')^T L^0_{\mathbf{q}'\mathbf{q}} \delta\mathbf{q} + (\delta\mathbf{q})^T L^0_{\mathbf{q}\mathbf{q}'} \delta\mathbf{q}' + (\delta\mathbf{q})^T L^0_{\mathbf{q}\mathbf{q}} \delta\mathbf{q} \right] ds, \tag{5.2}$$

where L is the integrand of E with the appropriate Lagrange multiplier terms added to it.

We note first a property of $\delta^2 E$ peculiar to the example at hand. The integrand L is invariant to a scaling of $\mathbf{q}$, i.e.,

$$L(c\mathbf{q}, c\mathbf{q}') = L(\mathbf{q}, \mathbf{q}')$$

for any $c \in \mathbb{R}$. This degeneracy arises from our use of four-dimensional quaternions to represent the three-dimensional space $SO(3)$ of directors. If we write $\delta\mathbf{q}(s)$ as $\beta(s)\mathbf{q}_0(s) + \mathbf{w}(s)$, where, at each s, $\mathbf{w}(s)$ is perpendicular to $\mathbf{q}_0(s)$, then

$$E[\mathbf{q}_0 + \epsilon\delta\mathbf{q}] = E[(1 + \epsilon\beta)\mathbf{q}_0 + \epsilon\mathbf{w}] = E\left[\mathbf{q}_0 + \frac{\epsilon}{1 + \epsilon\beta} \mathbf{w} \right].$$

Therefore, $E[\mathbf{q}_0+\epsilon\delta\mathbf{q}]<E[\mathbf{q}_0]$ for ϵ sufficiently small if and only if $E[\mathbf{q}_0+\epsilon\mathbf{w}]<E[\mathbf{q}_0]$ for ϵ sufficiently small. Thus it is sufficient to consider only those allowed variations $\delta\mathbf{q}$ that are orthogonal to $\mathbf{q}_0$ at each s. For example, if we define a projection matrix $\mathbf{\Pi}(s)\in\mathbb{R}^{4\times 3}$ whose columns span the orthogonal complement of $\mathbf{q}_0(s)$, then an arbitrary $\delta\mathbf{q}(s)$ can be written as $\mathbf{\Pi}(s)\boldsymbol{\zeta}(s)+\beta(s)\mathbf{q}_0(s)$ for some $\boldsymbol{\zeta}(s)\in\mathbb{R}^3$ and $\beta(s)\in\mathbb{R}$, and the second variation we will need to consider is $\delta^2E[\mathbf{\Pi}\boldsymbol{\zeta}]$. Thus we seek the maximal dimension of a subspace of functions $\boldsymbol{\zeta}(s)\in\mathbb{R}^3$ obeying $\delta^2E[\mathbf{\Pi}\boldsymbol{\zeta}]<0$, $\boldsymbol{\zeta}(0)=\boldsymbol{\zeta}(1)=\mathbf{0}$, and $\langle\mathbf{\Pi}\boldsymbol{\zeta},(d_{3i})^0_{\mathbf{q}}\rangle=0$ for $i=1,2$. For convenience, we observe that the constraints $\langle\mathbf{\Pi}\boldsymbol{\zeta},(d_{3i})^0_{\mathbf{q}}\rangle=0$ may be rewritten as $\langle\boldsymbol{\zeta},\mathbf{T}_i\rangle=0$ if we define $\mathbf{T}_i\equiv\mathbf{\Pi}^T(d_{3i})^0_{\mathbf{q}}$. The matrix $\mathbf{\Pi}$ will be defined differently for the anisotropic and the isotropic cases in sections 6.1 and 6.2, respectively.

Inserting $\delta\mathbf{q}=\mathbf{\Pi}\boldsymbol{\zeta}$ into (5.2) and integrating by parts terms starting with $(\boldsymbol{\zeta}')^T$, we find

$$\delta^2J[\mathbf{\Pi}\boldsymbol{\zeta}]=\langle\boldsymbol{\zeta},\mathcal{S}\boldsymbol{\zeta}\rangle,$$

where $\mathcal{S}$ is the operator

$$\mathcal{S}\boldsymbol{\zeta}=-\bar{\mathbf{P}}\boldsymbol{\zeta}''+\bar{\mathbf{C}}\boldsymbol{\zeta}'+\bar{\mathbf{Q}}\boldsymbol{\zeta}$$

with coefficient matrices

$$\begin{aligned}
&\bar{\mathbf{P}}=\mathbf{\Pi}^TL^0_{\mathbf{q}'\mathbf{q}'}\mathbf{\Pi},\\
&\bar{\mathbf{C}}=(\mathbf{\Pi}')^TL^0_{\mathbf{q}'\mathbf{q}'}\mathbf{\Pi}+\mathbf{\Pi}^TL^0_{\mathbf{q}\mathbf{q}'}\mathbf{\Pi}-\left(\mathbf{\Pi}^TL^0_{\mathbf{q}'\mathbf{q}'}\mathbf{\Pi}'+\mathbf{\Pi}^TL^0_{\mathbf{q}'\mathbf{q}}\mathbf{\Pi}\right)-\left(\mathbf{\Pi}^TL^0_{\mathbf{q}'\mathbf{q}'}\mathbf{\Pi}\right)',\\
&\bar{\mathbf{Q}}=(\mathbf{\Pi}')^TL^0_{\mathbf{q}'\mathbf{q}'}\mathbf{\Pi}'+\mathbf{\Pi}^TL^0_{\mathbf{q}\mathbf{q}}\mathbf{\Pi}+\mathbf{\Pi}^TL^0_{\mathbf{q}\mathbf{q}'}\mathbf{\Pi}'+(\mathbf{\Pi}')^TL^0_{\mathbf{q}'\mathbf{q}}\mathbf{\Pi}-\left(\mathbf{\Pi}^TL^0_{\mathbf{q}'\mathbf{q}'}\mathbf{\Pi}'+\mathbf{\Pi}^TL^0_{\mathbf{q}'\mathbf{q}}\mathbf{\Pi}\right)'.
\end{aligned}$$

Our choices of $\mathbf{\Pi}$ will cause the last term in each of $\bar{\mathbf{C}}$ and $\bar{\mathbf{Q}}$ to vanish and will yield simple s-independent expressions for $\bar{\mathbf{P}}$, $\bar{\mathbf{C}}$, and $\bar{\mathbf{Q}}$. The resulting expression for $\mathcal{S}$ will take the form required by Lemma 3.2:

$$\mathcal{S}=\mathcal{S}_1+\lambda\mathcal{S}_2.$$

Explicit forms for $\mathcal{S}_1$ and $\mathcal{S}_2$ depend on the choice of $\mathbf{\Pi}$ and so will be shown individually in sections 6.1 and 6.2.

6. Stability of unbuckled configurations.

6.1. The anisotropic strut.

For the anisotropic strut, we choose

$$\mathbf{\Pi}(s)\equiv\begin{bmatrix}\cos(\frac{\alpha s}{2}) & -\sin(\frac{\alpha s}{2}) & 0\\ \sin(\frac{\alpha s}{2}) & \cos(\frac{\alpha s}{2}) & 0\\ 0 & 0 & \cos(\frac{\alpha s}{2})\\ 0 & 0 & -\sin(\frac{\alpha s}{2})\end{bmatrix},$$

which gives a second-variation operator $\mathcal{S}=\mathcal{S}_1+\lambda\mathcal{S}_2$ with

$$\begin{aligned}\mathcal{S}_1\boldsymbol{\zeta}\equiv&\begin{bmatrix}-4K_1 & 0 & 0\\ 0 & -4K_2 & 0\\ 0 & 0 & -4K_3\end{bmatrix}\boldsymbol{\zeta}''+\begin{bmatrix}0 & 4\alpha(K_1+K_2-K_3) & 0\\ -4\alpha(K_1+K_2-K_3) & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\boldsymbol{\zeta}'\\
&+\begin{bmatrix}4\alpha^2(K_2-K_3) & 0 & 0\\ 0 & 4\alpha^2(K_1-K_3) & 0\\ 0 & 0 & 0\end{bmatrix}\boldsymbol{\zeta}\end{aligned}\tag{6.1}$$

and

$$\mathcal{S}_2\boldsymbol{\zeta} \equiv \begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\zeta}. \tag{6.2}$$

For this choice of $\mathbf{\Pi}$, the projected constraints take the explicit form

$$\mathbf{T}_1(s) = \begin{bmatrix} 2\sin(\alpha s) \\ 2\cos(\alpha s) \\ 0 \end{bmatrix}, \quad \mathbf{T}_2(s) = \begin{bmatrix} -2\cos(\alpha s) \\ 2\sin(\alpha s) \\ 0 \end{bmatrix}.$$

6.1.1. Verification of hypotheses from section 4. Before applying the results of section 4, we remove the third component of $\boldsymbol{\zeta}$, which plays a trivial role. Any variation $\boldsymbol{\zeta}$ can be written as $\boldsymbol{\zeta} = \bar{\boldsymbol{\zeta}} + \boldsymbol{\zeta}^*$, where $\bar{\boldsymbol{\zeta}}$ contains zeros in the first two slots, and $\boldsymbol{\zeta}^*$ contains a zero in the third slot. Using the explicit form for $\mathcal{S}$, and integrating by parts,

$$\langle \bar{\boldsymbol{\zeta}}, \mathcal{S}\bar{\boldsymbol{\zeta}} \rangle = 4K_3 \langle \bar{\boldsymbol{\zeta}}', \bar{\boldsymbol{\zeta}}' \rangle \geq 0. \tag{6.3}$$

By (6.1) and (6.2), $\mathcal{S}\bar{\boldsymbol{\zeta}}$ contains zeros in the first two slots, and $\mathcal{S}\boldsymbol{\zeta}^*$ contains a zero in the third slot, and, therefore,

$$\langle \boldsymbol{\zeta}, \mathcal{S}\boldsymbol{\zeta} \rangle = \langle \bar{\boldsymbol{\zeta}} + \boldsymbol{\zeta}^*, \mathcal{S}(\bar{\boldsymbol{\zeta}} + \boldsymbol{\zeta}^*) \rangle = \langle \bar{\boldsymbol{\zeta}}, \mathcal{S}\bar{\boldsymbol{\zeta}} \rangle + \langle \boldsymbol{\zeta}^*, \mathcal{S}\boldsymbol{\zeta}^* \rangle \geq \langle \boldsymbol{\zeta}^*, \mathcal{S}\boldsymbol{\zeta}^* \rangle. \tag{6.4}$$

So, given any basis $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n$ for a maximal subspace on which $\langle \boldsymbol{\zeta}, \mathcal{S}\boldsymbol{\zeta} \rangle < 0$, the vectors $\boldsymbol{\zeta}_1^*, \ldots, \boldsymbol{\zeta}_n^*$ must be linearly independent as functions of s since otherwise some linear combination of $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n$ would have zeros in the first two slots and hence a nonnegative second variation by (6.3). Then, by (6.4), the span of $\boldsymbol{\zeta}_1^*, \ldots, \boldsymbol{\zeta}_n^*$ is also an n-dimensional (hence maximal) subspace on which $\langle \boldsymbol{\zeta}, \mathcal{S}\boldsymbol{\zeta} \rangle < 0$. Thus we may restrict our attention to variations $\boldsymbol{\zeta}$ with a zero in the third slot.

Thus we define

$$\boldsymbol{\zeta}^* = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}, \quad \mathbf{T}_1^* = \begin{bmatrix} 2\sin(\alpha s) \\ 2\cos(\alpha s) \end{bmatrix}, \quad \mathbf{T}_2^* = \begin{bmatrix} -2\cos(\alpha s) \\ 2\sin(\alpha s) \end{bmatrix},$$

and

$$\mathcal{H}_d^{cons,*} = \{\boldsymbol{\zeta}^* \in H^2(\mathbb{R}^2, (0,1)) : \boldsymbol{\zeta}^*(0) = \boldsymbol{\zeta}^*(1) = \mathbf{0}, \;\; \langle \boldsymbol{\zeta}^*, \mathbf{T}_1^* \rangle = \langle \boldsymbol{\zeta}^*, \mathbf{T}_2^* \rangle = 0\}.$$

Plugging $\boldsymbol{\zeta}^*$ into (4.1), we find

$$(\mathcal{S}_3 + \lambda \mathcal{S}_4)\boldsymbol{\zeta}^* = \breve{c}_1 \mathbf{T}_1^* + \breve{c}_2 \mathbf{T}_2^*, \qquad \boldsymbol{\zeta}^* \in \mathcal{H}_d^{cons,*}, \tag{6.5}$$

where

$$\begin{aligned} \mathcal{S}_3\boldsymbol{\zeta}^* \equiv \begin{bmatrix} -4K_1 & 0 \\ 0 & -4K_2 \end{bmatrix} (\boldsymbol{\zeta}^*)'' \quad &+ \begin{bmatrix} 0 & 4\alpha(K_1 + K_2 - K_3) \\ -4\alpha(K_1 + K_2 - K_3) & 0 \end{bmatrix} (\boldsymbol{\zeta}^*)' \\ &+ \begin{bmatrix} 4\alpha^2(K_2 - K_3) & 0 \\ 0 & 4\alpha^2(K_1 - K_3) \end{bmatrix} \boldsymbol{\zeta}^*, \\ \mathcal{S}_4\boldsymbol{\zeta}^* \equiv \begin{bmatrix} -4 & 0 \\ 0 & -4 \end{bmatrix} &\boldsymbol{\zeta}^*. \end{aligned}$$

If we let $\mathcal{Q}$ denote the projection onto the space of functions L^2-orthogonal to both $\mathbf{T}_1^*$ and $\mathbf{T}_2^*$, then (6.5) may be rewritten as $(\mathcal{Q}\mathcal{S}_3\mathcal{Q})\boldsymbol{\zeta}^* = 4\lambda\boldsymbol{\zeta}^*$ for $\boldsymbol{\zeta}^* \in \mathcal{H}_d^{cons,*}$. We thus have an eigenvalue problem for the operator $\mathcal{Q}\mathcal{S}_3\mathcal{Q}$ on the space $\mathcal{H}_d^{cons,*}$. Relying on the fact that the spectrum of $\mathcal{S}$ in $\mathcal{H}_d$ consists purely of isolated eigenvalues, each with finite multiplicity, one can show that $\mathcal{S}_3$ has the same type of spectrum (see the argument on p. 3067 of [19]). It then follows from the spectral theorem for self-adjoint operators [8, p. 233] that the solutions of this equation form an orthogonal basis for $\mathcal{H}_d^{cons,*}$. Using the eigenvalue equation, we can see that this basis is also $(\mathcal{S}_3 + \lambda\mathcal{S}_4)$-orthogonal, as we need to apply Lemma 3.1.

6.1.2. Determination of the index. We observe that, for any continuous $\boldsymbol{\zeta}^* = (\zeta_1, \zeta_2)$ with ζ_1 and ζ_2 not both identically zero,

$$\langle \boldsymbol{\zeta}^*, \mathcal{S}_4\boldsymbol{\zeta}^* \rangle = -4\int_0^1 (\zeta_1(s)^2 + \zeta_2(s)^2)ds < 0.$$

Therefore, by Lemma 4.1, the index of the unbuckled equilibrium at (α, λ) is equal to the number of branch points λ_n below λ at that particular angle α. Thus, once we have determined the branch points, we will have determined the index at any (α, λ).

The differential equation appearing in (6.5) can be written out explicitly (after dividing through by $-4K_1$) as

$$\begin{aligned} \zeta_1'' - A\zeta_2' + B\zeta_1 &= -\frac{\sin(\alpha s)\breve{c}_1}{2K_1} + \frac{\cos(\alpha s)\breve{c}_2}{2K_1}, \\ \rho\zeta_2'' + A\zeta_1' + C\zeta_2 &= -\frac{\cos(\alpha s)\breve{c}_1}{2K_1} - \frac{\sin(\alpha s)\breve{c}_2}{2K_1}, \end{aligned}$$

where $\gamma = K_3/K_1$, $\rho = K_2/K_1$, $A = \alpha(1+\rho-\gamma)$, $B = \alpha^2(\gamma-\rho)+\frac{\lambda}{K_1}$, and $C = \alpha^2(\gamma-1)+\frac{\lambda}{K_1}$. The general solution can be found in closed form, and, applying the four boundary conditions $\zeta_{1,2}(0) = \zeta_{1,2}(1) = 0$ and two linearized constraints, we see that branch points occur when

$$\det \begin{bmatrix} \mathbf{M}_1(0) & \mathbf{M}_2(0) & \mathbf{M}_3(0) \\ \mathbf{M}_1(1) & \mathbf{M}_2(1) & \mathbf{M}_3(1) \\ \int_0^1 \mathbf{T}(s)^T\mathbf{M}_1(s)ds & \int_0^1 \mathbf{T}(s)^T\mathbf{M}_2(s)ds & \int_0^1 \mathbf{T}(s)^T\mathbf{M}_3(s)ds \end{bmatrix} = 0, \tag{6.6}$$

where $\mathbf{T}(s) \equiv \begin{bmatrix} \mathbf{T}_1^*(s) & \mathbf{T}_2^*(s) \end{bmatrix}$,

$$\mathbf{M}_{1,2}(s) = \begin{bmatrix} \sigma_{1,2}\cos(\sqrt{\omega_{1,2}}s) & -\sigma_{1,2}\sin(\sqrt{\omega_{1,2}}s) \\ \sin(\sqrt{\omega_{1,2}}s) & \cos(\sqrt{\omega_{1,2}}s) \end{bmatrix}, \quad \mathbf{M}_3(s) = -\frac{1}{2\lambda}\begin{bmatrix} \sin(\alpha s) & -\cos(\alpha s) \\ \cos(\alpha s) & \sin(\alpha s) \end{bmatrix}$$

for ω_1, ω_2, the (possibly complex) quantities

$$\omega_{1,2} = \frac{A^2 + B\rho + C \pm \sqrt{(A^2 + B\rho + C)^2 - 4BC\rho}}{2\rho},$$

and

$$\sigma_{1,2} = -\frac{A\sqrt{\omega_{1,2}}}{\omega_{1,2} - B}.$$

(We note that, for a few isolated parameter values, the characteristic equation of (6.5) has repeated roots, so the branch point equation does not take the above form.)

Further simplification of the determinant of this 6-by-6 matrix is difficult, so we instead determine its zeros numerically in the results in section 6.3 (taking care to handle separately the special cases where the characteristic equation has repeated roots). However, in the case when $K_1 = K_2$, this numerical search for zeros is difficult to implement since, due to symmetry, the above determinant does not cross zero transversely but rather intersects it tangentially. Thus we present in the next section a separate derivation of the index for an isotropic strut $K_1 = K_2$.

6.2. The isotropic strut.

When $K_1 = K_2$, we choose a slightly different projection matrix:

$$\mathbf{\Pi}(s) \equiv \begin{bmatrix} \cos(\frac{\alpha s}{2}) & \sin(\frac{\alpha s}{2}) & 0 \\ -\sin(\frac{\alpha s}{2}) & \cos(\frac{\alpha s}{2}) & 0 \\ 0 & 0 & \cos(\frac{\alpha s}{2}) \\ 0 & 0 & -\sin(\frac{\alpha s}{2}) \end{bmatrix}.$$

Then the second-variation operator is $\mathcal{S} = \mathcal{S}_1 + \lambda \mathcal{S}_2$, where

$$\mathcal{S}_1 \boldsymbol{\zeta} \equiv \begin{bmatrix} -4K_1 & 0 & 0 \\ 0 & -4K_1 & 0 \\ 0 & 0 & -4K_3 \end{bmatrix} \boldsymbol{\zeta}'' + \begin{bmatrix} 0 & -4\alpha K_3 & 0 \\ 4\alpha K_3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\zeta}'$$

and

$$\mathcal{S}_2 \boldsymbol{\zeta} \equiv \begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\zeta}.$$

The projected constraints take the explicit form

$$\mathbf{T}_1(s) = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{T}_2(s) = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}.$$

We may follow the same procedure as in section 6.1.1 to show that the set of solutions to (6.5) forms a basis for $\mathcal{H}_d^{cons,*}$, with the operator $\mathcal{S}_3$ now taking the form

$$\mathcal{S}_3 \boldsymbol{\zeta}^* \equiv \begin{bmatrix} -4K_1 & 0 \\ 0 & -4K_1 \end{bmatrix} (\boldsymbol{\zeta}^*)'' + \begin{bmatrix} 0 & -4\alpha K_3 \\ 4\alpha K_3 & 0 \end{bmatrix} (\boldsymbol{\zeta}^*)'.$$

For any continuous $\boldsymbol{\zeta}^* = (\zeta_1, \zeta_2)$ with ζ_1 and ζ_2 not both identically zero,

$$\langle \boldsymbol{\zeta}^*, \mathcal{S}_4 \boldsymbol{\zeta}^* \rangle = -4 \int_0^1 (\zeta_1(s)^2 + \zeta_2(s)^2) ds < 0,$$

and thus, by the theory of section 4, the index of the unbuckled equilibrium at (α, λ) is equal to the number of branch points λ_n below λ at that particular angle α.

The differential equation appearing in (6.5) can be written out explicitly (after dividing through by $-4K_1$) as

$$\begin{aligned}\zeta_1'' + A\zeta_2' + B\zeta_1 &= \frac{\breve{c}_2}{2K_1},\\ \zeta_2'' - A\zeta_1' + B\zeta_2 &= -\frac{\breve{c}_1}{2K_1},\end{aligned}$$

where $\gamma = K_3/K_1$, $A = \gamma\alpha$, and $B = \frac{\lambda}{K_1}$. The general solution can again be found in closed form, and, applying the boundary conditions and constraints, we see that (apart from the special cases $A = 0$, $B = 0$, and $A^2 + 4B = 0$) branch points occur when there is a nontrivial solution $(C_1, C_2, C_3, C_4, \breve{c}_1, \breve{c}_2)$ to

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & \frac{1}{2K_1B} \\ 0 & -1 & 0 & -1 & -\frac{1}{2K_1B} & 0 \\ \cos\omega_1 & \sin\omega_1 & \cos\omega_2 & \sin\omega_2 & 0 & \frac{1}{2K_1B} \\ \sin\omega_1 & -\cos\omega_1 & \sin\omega_2 & -\cos\omega_2 & -\frac{1}{2K_1B} & 0 \\ \frac{2(1-\cos\omega_1)}{\omega_1} & -\frac{2\sin\omega_1}{\omega_1} & \frac{2(1-\cos\omega_2)}{\omega_2} & -\frac{2\sin\omega_2}{\omega_2} & -\frac{1}{K_1B} & 0 \\ -\frac{2\sin\omega_1}{\omega_1} & -\frac{2(1-\cos\omega_1)}{\omega_1} & -\frac{2\sin\omega_2}{\omega_2} & -\frac{2(1-\cos\omega_2)}{\omega_2} & 0 & -\frac{1}{K_1B} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ \breve{c}_1 \\ \breve{c}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

where $\omega_{1,2}$ are the (possibly complex) quantities $\frac{A \pm \sqrt{A^2+4B}}{2}$.

Denote the upper-left 4-by-4 block of this 6-by-6 matrix as $\mathbf{M}_{11}$, the upper-right 4-by-2 block as $\mathbf{M}_{12}$, the lower-left 2-by-4 block as $\mathbf{M}_{21}$, and the lower-right 2-by-2 block as $\mathbf{M}_{22}$. Then $\det \mathbf{M}_{11} = 2(1 - \cos(\omega_1 - \omega_2))$. As long as $\omega_1 - \omega_2 \neq 2n\pi$, $n = 1, 2, 3, \ldots$, we then find that

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = -\mathbf{M}_{11}^{-1}\mathbf{M}_{12} \begin{bmatrix} \breve{c}_1 \\ \breve{c}_2 \end{bmatrix}. \tag{6.7}$$

Thus

$$-\mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12} \begin{bmatrix} \breve{c}_1 \\ \breve{c}_2 \end{bmatrix} + \mathbf{M}_{22} \begin{bmatrix} \breve{c}_1 \\ \breve{c}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Simplifying the above equation, we find

(6.8)
$$\frac{1}{K_1 B \omega_1 \omega_2 \sin\left(\frac{\omega_1-\omega_2}{2}\right)} \left(2(\omega_1 - \omega_2)\sin\frac{\omega_1}{2}\sin\frac{\omega_2}{2} - \omega_1\omega_2 \sin\left(\frac{\omega_1 - \omega_2}{2}\right)\right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \breve{c}_1 \\ \breve{c}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus, we have only the trivial solution $\breve{c}_1 = \breve{c}_2 = C_1 = C_2 = C_3 = C_4 = 0$ unless

$$2(\omega_1 - \omega_2)\sin\frac{\omega_1}{2}\sin\frac{\omega_2}{2} - \omega_1\omega_2\sin\left(\frac{\omega_1 - \omega_2}{2}\right) = 0,$$

or, in terms of A and B:

$$2\sqrt{A^2+4B}\sin\frac{A+\sqrt{A^2+4B}}{4}\sin\frac{A-\sqrt{A^2+4B}}{4}+B\sin\left(\frac{\sqrt{A^2+4B}}{2}\right)=0. \tag{6.9}$$

If this condition does hold, then any $(\breve{c}_1, \breve{c}_2)$ will satisfy (6.8), with (C_1, C_2, C_3, C_4) then determined by (6.7). Thus we have a branch point when (6.9) holds, and the corresponding conjugate point has multiplicity two.

A check of the special cases $\omega_1 - \omega_2 = 2n\pi$, $n = 1, 2, 3, \ldots$, reveals that the conjugate point condition again takes the form (6.9), again with each conjugate point having multiplicity two.

In the special cases $A = 0$, $B = 0$, and $A^2 + 4B = 0$, the general solution used above is not valid, but we may directly analyze these cases to find the following:

- The case $B = 0, A \neq 0$ yields the branch point condition

  $$\frac{A}{2}\cos\left(\frac{A}{2}\right)=\sin\left(\frac{A}{2}\right), \tag{6.10}$$

  which matches the dominant term in a Taylor series expansion of (6.9) in the limit that $B \to 0$. Again, the conjugate points in this case have multiplicity 2.
- The case $A = 0, B \neq 0$ yields the branch condition

  $$\sqrt{B}\sin(\sqrt{B})-4\sin^2\left(\frac{\sqrt{B}}{2}\right)=0,$$

  which matches (6.9) when $A = 0$. Again, the conjugate points in this case have multiplicity 2.
- The case $A^2 + 4B = 0$ can exhibit no branch points.

In summary, branch points are the solutions to (6.9) when $B \neq 0$ and $A^2+4B \neq 0$. When $A^2 + 4B = 0$, there are no branch points, while, when $B = 0$, they are the solutions to (6.10).

6.3. Results. In Figure 6.1, we show the index of the unbuckled configuration (color-coded by the scheme given in Figure 6.2) for various values of the twist angle α, loading force λ, bending stiffness ratio ρ, and twisting-to-bending stiffness ratio γ. In each frame of the figure, the black lines indicate the locations of the branch points as a function of α and λ/K_1 for ρ and γ held fixed. The top row of frames is for the isotropic problem $\rho = 1$, so the lines were determined by (6.9), while, for the remaining frames, they were determined using (6.6). The coloring of the diagrams is a computation of the index via the numerical conjugate point test described in section 2 and verifies the result proven in this section, namely, that the index at a point $(\alpha, \lambda/K_1)$ is equal to the number of lines that lie below it. (In the isotropic case $\rho = 1$, each line is counted with multiplicity two, as shown in section 6.2.)

The stability results in the isotropic case $\rho = 1$ match the standard intuition about buckling. For zero twist, the unbuckled configuration is stable up to a maximal loading force and becomes more and more unstable thereafter. As twist (either negative or positive) is added, the same behavior is observed, except that the threshold for instability is lowered since the twist destabilizes the unbuckled configuration.

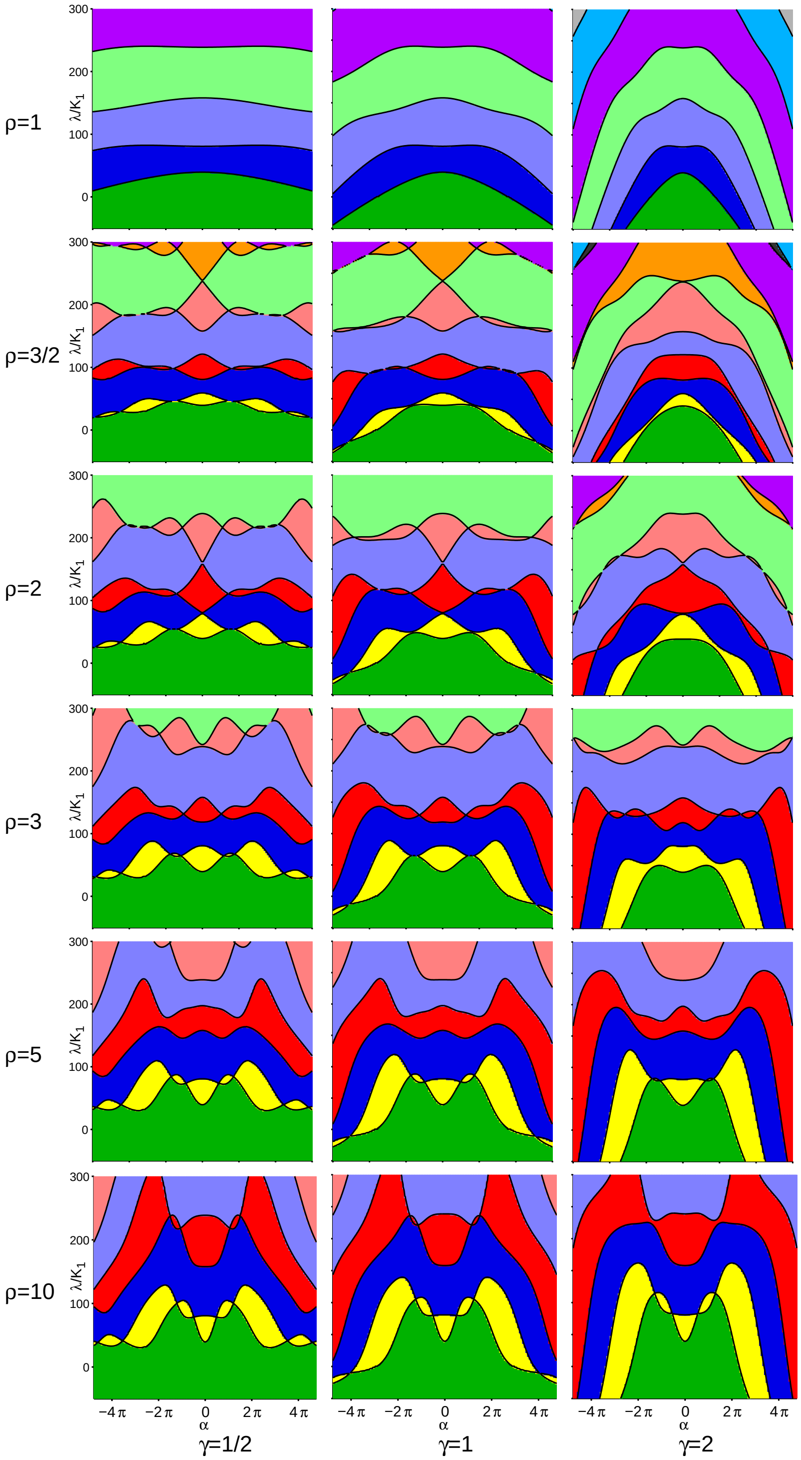


Figure 6.1. *The index on the plane of unbuckled equilibria for various values of ρ, the ratio of bending stiffnesses, and of γ, the ratio of twisting to bending stiffness. In each graph, the index of the unbuckled equilibrium with twist angle α and endloading λ is indicated by color, following the scheme in Figure* 6.2*. Black curves indicate the set of branch points as determined by the analysis in sections* 6.1 *and* 6.2*.*

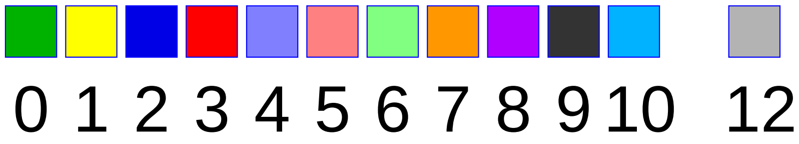


Figure 6.2. *Color scheme used to represent the index (index* 11 *is omitted since it does not appear in any of our figures).*

Comparing the $\rho = 1$ row to the $\rho = 3/2$ row, we can see the splitting generated by the anisotropy, as each curve where the index jumps by two when $\rho = 1$ splits into a pair of interlaced curves, where the index jumps by one when $\rho = 3/2$. As ρ grows, we also note a prominent trend whereby a twist angle α can cause a stabilization of the first buckling mode, as seen by the emergence of two protuberances symmetrically about $\alpha = 0$, e.g., at $\rho = 3/2$ for $\gamma = 1/2$, at $\rho = 2$ for $\gamma = 1$, or at $\rho = 3$ for $\gamma = 2$. Physically, this corresponds to the fact that an imposed twist effectively transfers some of the enhanced bending stiffness K_2 to the direction governed by K_1 in the untwisted configuration. For α sufficiently large, this effect is eventually countered by the fact that the act of twisting itself imposes a certain degree of instability so that the green-yellow boundary curve eventually begins to decrease for values of α further from the line $\alpha = 0$. As ρ increases further, the protuberances become larger. The bands with higher index mimic the shape of the green (stable) region.

As γ decreases, we can see that the dominant trend is simply to stretch the diagram laterally away from $\alpha = 0$. In the isotropic case, this stretching transformation is obeyed exactly. (We can see in the theoretical derivation that the index depends on α only through the term $\alpha\gamma$.) In the anisotropic case, a change in γ causes some slight qualitative changes in the index diagram in addition to this lateral stretch.

We note that we have restricted ρ to be greater than 1 in Figure 6.1 since any $\rho < 1$ diagram may be deduced from an appropriate $\rho > 1$ diagram. Specifically, given $\rho < 1$ and any values for γ and K_1, we have a strut with weaker bending stiffness equal to K_2, stronger bending stiffness equal to $1/\rho$ times the weaker bending stiffness ($K_1 = (1/\rho)K_2$), and twisting stiffness equal to γ/ρ times the weaker bending stiffness ($K_3 = (\gamma/\rho)K_2 = \gamma K_1$). Thus, if we consider the diagram with $\rho^* = 1/\rho > 1$ and $\gamma^* = \gamma/\rho$ and relabel the y-axis as λ/K_2 rather than λ/K_1, we will have the correct index diagram for (K_1, K_2, K_3). (If one wants a diagram with λ/K_1 to match the other diagrams, one would simply scale vertically by $K_1/K_2 = 1/\rho$.)

7. Stability of buckled configurations.

7.1. "Slices" of the sheets of buckled equilibria.

In the previous section, for fixed values of ρ and γ, the index of each unbuckled configuration in the $\alpha - \lambda$ plane was determined. The curves on this plane where the index changes indicate where sheets of buckled equilibria intersect this plane. In this section, we discuss these sheets of buckled equilibria. They are fairly complicated due to folding and branching, so we compute only those portions of the sheets corresponding to some of the simplest buckled configurations.

To build up to displaying the sheets of buckled equilibria, we first consider "slices" in which the angle α is fixed, while the force λ is allowed to vary. For example, we consider $(\rho, \gamma) = (1, 1)$

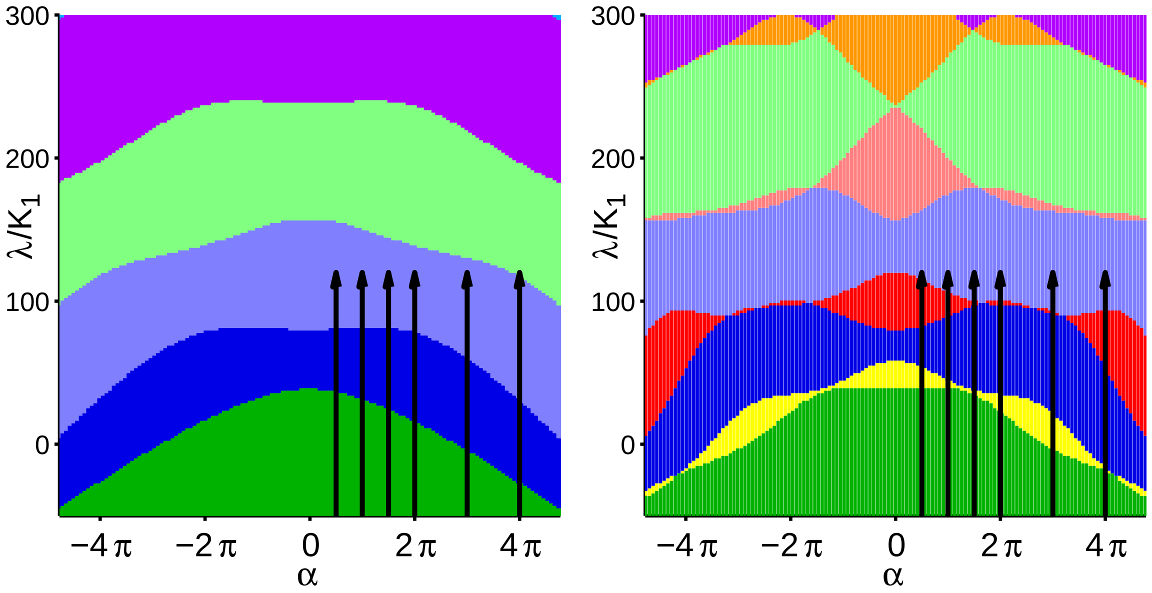


Figure 7.1. *A buckling problem at various values of the twist angle α. For bending stiffness ratios $\rho = 1$ (left) and $\rho = 3/2$ (right) and twisting-to-bending stiffness ratio $\gamma = 1$, we show the plane of unbuckled equilibria colored by stability from Figure 6.1. Now we fix α at several values ($\pi/2$, π, $3\pi/2$, 2π, 3π, and 4π) and vary the force λ. Branches of buckled configurations will bifurcate at each point where an arrow meets a change in color.*

and $(\frac{3}{2}, 1)$. In each case, we fix several values of α, namely, $\alpha = \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi, 3\pi, 4\pi$, and then we vary the force λ, as shown in Figure 7.1.

For each slice, we compute one-dimensional branches of buckled equilibria using AUTO [5, 6], which uses numerical parameter continuation to compute solutions to boundary value problems (BVPs) such as we have here with the Euler–Lagrange equations in $\mathbf{q}(s)$ subject to the constraints in (5.1). We present here only the briefest of introductions to continuation methods and refer the reader to [5, 6] and references therein for a full description.

As the name suggests, numerical parameter continuation is the computation of a family of solutions to a BVP as a parameter (in this case, λ) is varied. After discretization in s, the BVP becomes a finite-dimensional equation of the form

$$\mathbf{F}(\mathbf{X}) = \mathbf{0}, \qquad \mathbf{F} : \mathbb{R}^{n+1} \to \mathbb{R}^n. \tag{7.1}$$

This system has one more variable than it has equations; the extra variable is the continuation parameter λ. Therefore, given a solution $\mathbf{X}_0$, there generally exists a locally unique one-dimensional family of points, called a solution branch, that passes through $\mathbf{X}_0$. AUTO computes a numerical approximation to this solution branch using a well-known algorithm called the *pseudoarclength continuation method* [15]. As implemented in AUTO, branch points where new solution branches intersect the current branch can be detected and followed, and thus we may begin with a known solution $\mathbf{X}_0$ on the plane of unbuckled equilibria, follow solutions numerically on this plane, and then detect and follow the bifurcating branches which emerge at the color changes in Figure 7.1.

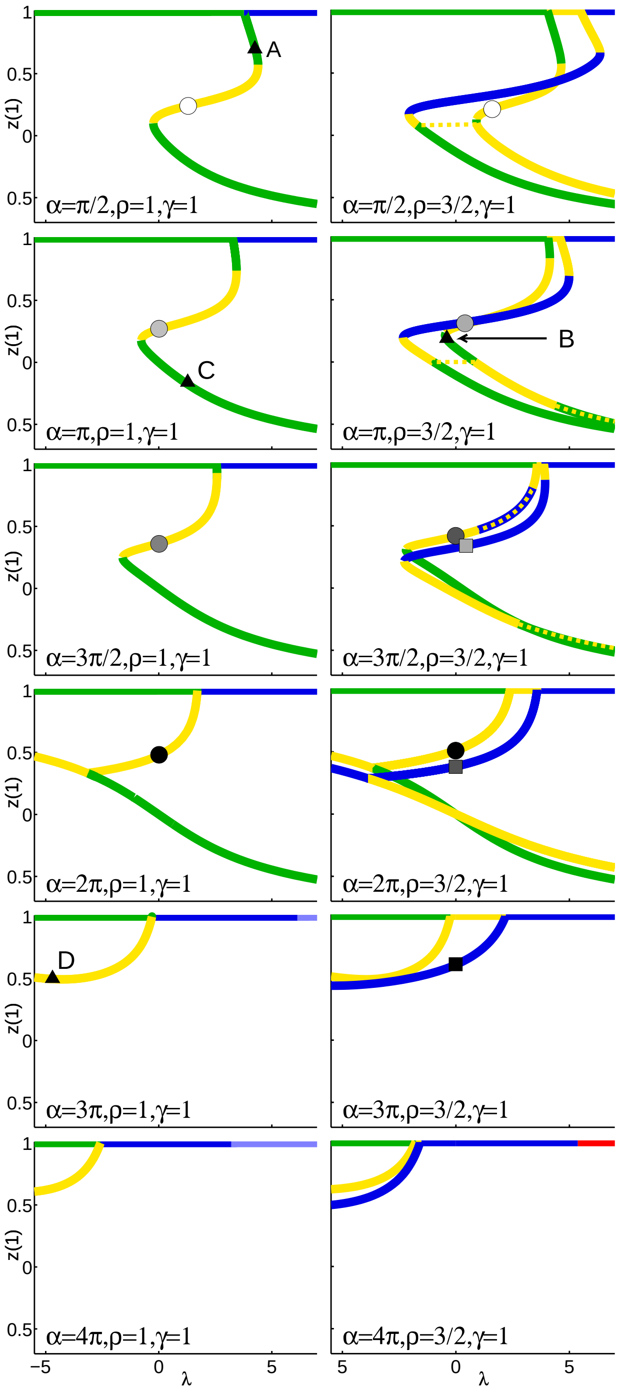


Figure 7.2. *Force-length diagrams for several values of the twist angle α and the bending stiffness ratio ρ: $\rho = 1$ in the left column, and $\rho = 3/2$ in the right column; α increases from $\pi/2$ to 4π vertically; $\gamma = 1, K_1 = 0.1$ throughout. In each graph, the length $z(1)$ is plotted against the force λ. Branches are colored by stability index, using the color scheme from Figure 6.2. Horizontal branches represent unbuckled configurations. For $\rho = 1$, we show the branch of buckled configurations arising from the first branch point, which splits into the two branches shown in the $\rho = 3/2$ diagrams. Some secondary bifurcating branches are shown as dashed lines. Circles and squares indicate points at the edges of the surfaces in Figures 7.9 and 7.11. Triangles with corresponding letters* A–D *indicate configurations depicted in Figure 7.3.*

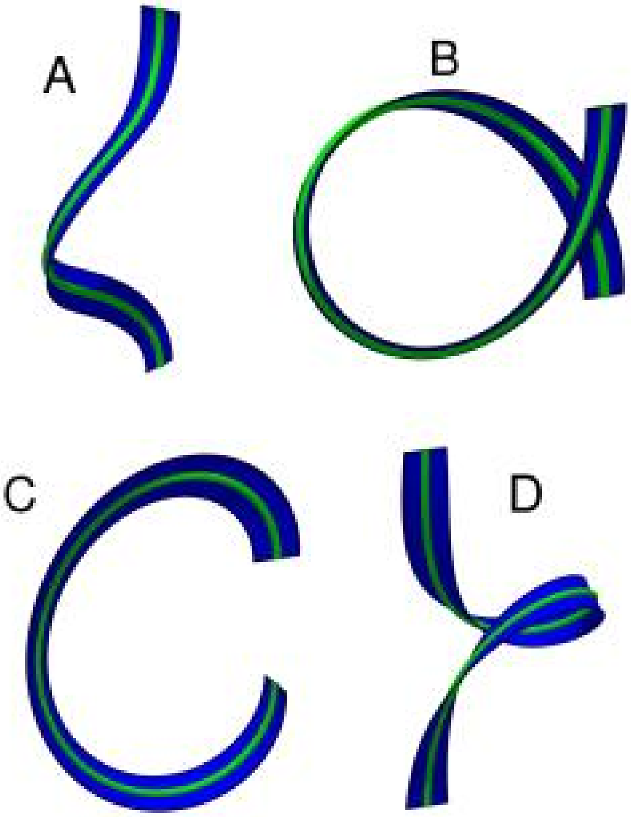


Figure 7.3. *Physical configurations appearing in the bifurcation diagrams in Figure* 7.2. *The centerline of the strut is shown as a green tube, while the director* $\mathbf{d}_1$ *is shown as a blue ribbon.*

To visualize these branches, we create force-length bifurcation diagrams, i.e., two-dimensional graphs that plot the force λ against the "length" $z(1)$, the distance between the two ends of the strut. (Thus negative lengths indicate that the endpoint of the strut corresponding to $s = 1$ has passed through the origin to the other side of the z-axis.) Figure 7.2 contains diagrams for the simplest buckling solutions. For the isotropic problem $\rho = 1$, these simple solutions arise from the first branch point, the green-to-blue transition on the plane of unbuckled equilibria. In the anisotropic problem $\rho = \frac{3}{2}$, these simple solutions perturb to two branches of solutions, the green-to-yellow and the yellow-to-blue transitions on the plane of unbuckled equilibria.

Some sample physical configurations appearing in these bifurcation diagrams are shown in Figure 7.3. Configurations close to the plane of unbuckled equilibria contain a small amount of buckling, as in configuration A, reminiscent of the first buckling mode of the classic untwisted problem. For some values of α, these are initially stable and then become unstable at a fold in λ, while, for sufficiently high α, they are always unstable. Further down the branch are more drastically buckled configurations, still with positive length, such as configuration B. Some of these are stable, as seen in the portions of the lowermost green branches that have $z(1) > 0$. There are also negative length configurations such as configuration C, and, again, some of

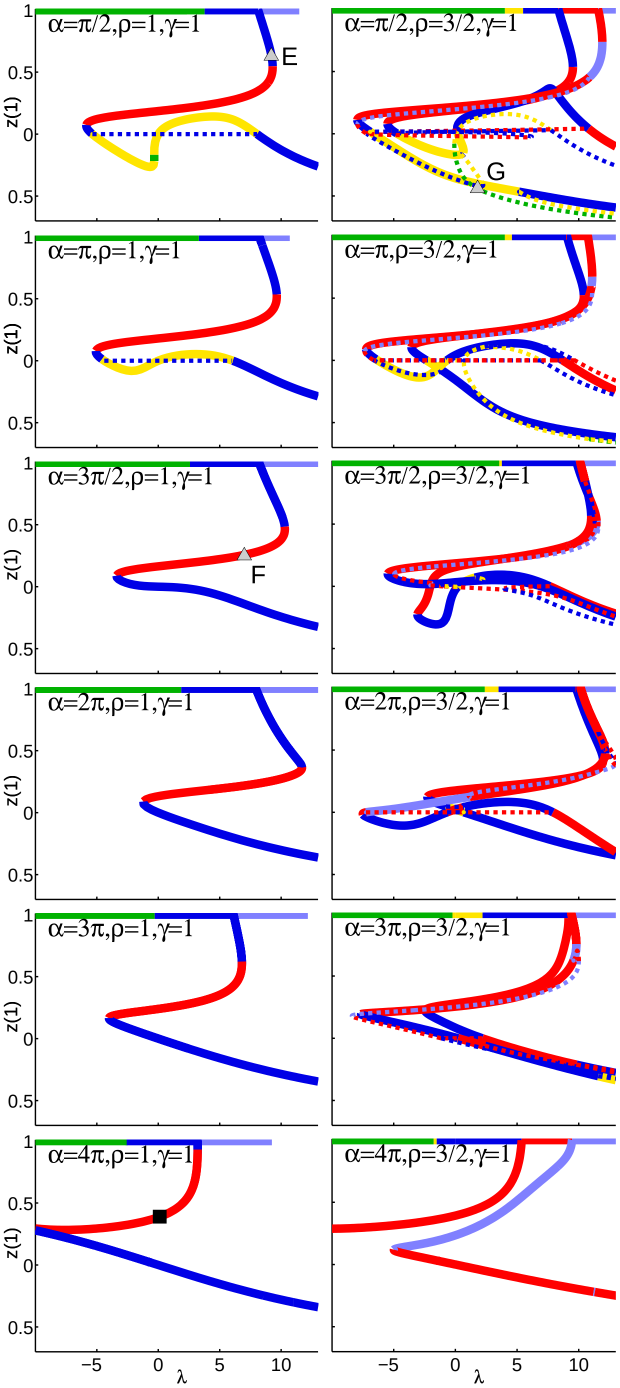


Figure 7.4. *Force-length diagrams for the next-higher buckling modes as compared to those shown in Figure* 7.2*. For $\rho = 1$, $\gamma = 1$, we plot the branch of buckled configurations arising from the second branch point, while, for $\rho = 3/2$, $\gamma = 1$, we plot the branches of buckled configurations arising from the third and fourth branch points, which are the splitting of the branch in the first column. The square in the lower left graph indicates a point at the edge of the surface in Figure* 7.9*. Triangles with corresponding letters* E–G *indicate configurations depicted in Figure* 7.5*.*

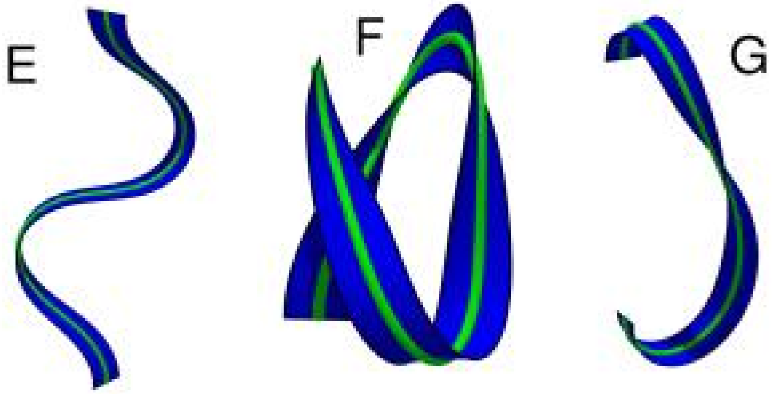


Figure 7.5. *Physical configurations appearing in the bifurcation diagrams in Figure 7.4. The centerline of the strut is shown as a green tube, while the director $\mathbf{d}_1$ is shown as a blue ribbon.*

these are stable. Finally, in some diagrams, there are branches leaving the graph at the left edge; the corresponding configurations look like configuration D, and are all unstable. The anisotropic diagrams are relatively simple splittings of the isotropic diagrams into two images, although there are a few secondary bifurcating branches that appear, sometimes connecting the two images to each other, and in one case ($\alpha = \frac{3\pi}{2}, \rho = \frac{3}{2}$) connecting a branch to itself.

Figure 7.4 illustrates the next-simplest buckling configurations. For the isotropic problem, these solutions arise from the second branch point, the blue-to-light-blue transition on the plane of unbuckled equilibria; for the anisotropic problem, these solutions perturb to two branches of solutions, the blue-to-red and the red-to-light-blue transitions. Some sample physical configurations appearing in these bifurcation diagrams are shown in Figure 7.5. The strut initially buckles to a two-mode shape like configuration E, as in the classic untwisted problem. Further on the branch, there are more drastically buckled configurations such as configuration F. Negative length configurations also exist, such as configuration G; in one case ($\alpha = \frac{\pi}{2}, \rho = \frac{3}{2}$), these are stable. Here, the perturbation in the diagram caused by anisotropy is quite complicated, leading in many cases to multiple secondary bifurcating branches with complicated connectivity. We expect that this complication would only worsen for higher buckling modes.

7.2. The sheets of buckled equilibria. Since AUTO is designed to compute one-dimensional branches of solutions as described in the previous section, it does not directly compute the two-dimensional sheets of buckled equilibria that we are interested in. However, by using a judicious itinerary of parameter switching, one may compute an approximation of the two-dimensional surface using a collection of one-dimensional curves [18, 22].

Consider first the computation of the plane of unbuckled equilibria. Of course, the configurations on this plane are known in closed form, so it is hardly necessary to determine them numerically, but nevertheless it serves as a useful first example. Furthermore, it is computationally expedient to compute this plane numerically since in so doing AUTO will detect branch points to be used as starting points for the computation of the sheets of buckled equilibria. We begin the computation of the plane of unbuckled equilibria at $(\lambda, \alpha) = (0, 0)$. We fix

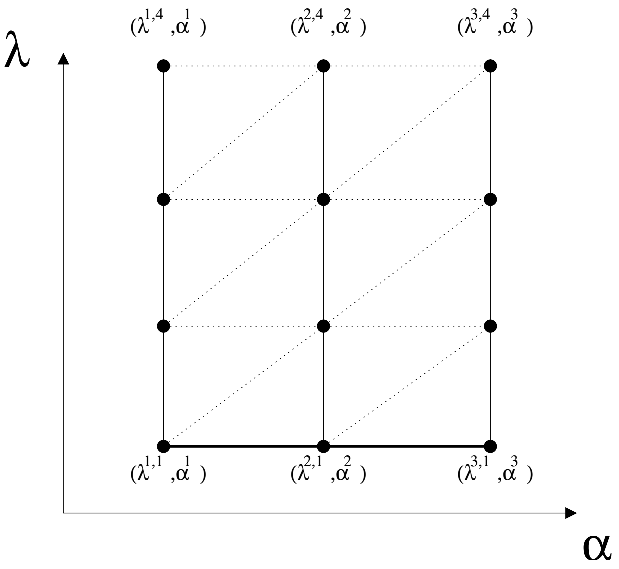


Figure 7.6. *Triangulation of plane of unbuckled equilibria: The solid horizontal line at the bottom of the grid represents the initial calculation fixing λ and allowing α to vary. Each of the solid vertical black lines represents a continuation in λ keeping α fixed. Each $(\lambda^{i,j}, \alpha^i)$ is marked with a black circle, and the triangulation is shown with the dotted lines.*

$\lambda = 0$ and allow α to vary, giving a system of the form (7.1), with α but not λ a component of $\mathbf{X}$. We then perform N steps of the continuation calculation described in section 7.1, giving a solution branch which we will denote by $(\lambda^{i,1}, \alpha^i), 1 \leq i \leq N$. (Note that $\lambda^{i,1} = 0$ for all i.) We now begin a new one-dimensional continuation calculation at each of the solutions $(\lambda^{i,1}, \alpha^i)$ by fixing $\alpha = \alpha^i$ and allowing λ to vary. This leads to a solution branch which we shall denote by $(\lambda^{i,j}, \alpha^i), 1 \leq j \leq M$. We choose M to be fixed for all i so that the result is a "grid" of points in the plane of unbuckled equilibria. Even though $\lambda^{i,1} = 0$ for $1 \leq i \leq N$, it is not necessarily the case that $\lambda^{i,j} = \lambda^{k,j}$ for $j > 1, k \neq i$, since AUTO automatically adapts the step size Δt during the calculation. We then triangulate the plane by choosing as the vertices of our triangles $(\lambda^{i,j}, \alpha^i), (\lambda^{i+1,j}, \alpha^{i+1}), (\lambda^{i+1,j+1}, \alpha^{i+1})$ and $(\lambda^{i,j}, \alpha^i), (\lambda^{i,j+1}, \alpha^i), (\lambda^{i+1,j+1}, \alpha^{i+1})$ for $1 \leq i \leq N-1, 1 \leq j \leq M-1$, as shown in Figure 7.6.

The sheets of buckled equilibria are calculated in a similar fashion. During each one-dimensional continuation in λ on the plane of unbuckled equilibria, AUTO is set to automatically report all branch points that it detects. Let $\lambda_k^{i,1}$ denote the value of λ at the kth branch point on the branch with $\alpha = \alpha^i$. We note that, in general, $\lambda_k^{i,1} \neq \lambda_k^{m,1}$ when $m \neq i$ since the position of the branch point is a function of α. At each such branch point, AUTO can be used to switch branches and compute a solution branch on the sheet of buckled equilibria emanating from the branch point. So, similar to the unbuckled case, we have an initial set of solutions $(\lambda_k^{i,1}, \alpha^i)$ from which we perform continuation calculations, keeping α fixed and allowing λ to vary. These calculations give rise to a set of buckled equilibria; we denote the values of λ and α for these equilibria by $(\lambda_k^{i,j}, \alpha^i)$ for $1 \leq i \leq N$ and $1 \leq j \leq M$. Finally, we triangulate the kth sheet by choosing the vertices of our triangles to be $(\lambda_k^{i,j}, \alpha^i), (\lambda_k^{i+1,j}, \alpha^{i+1}), (\lambda_k^{i+1,j+1}, \alpha^{i+1})$

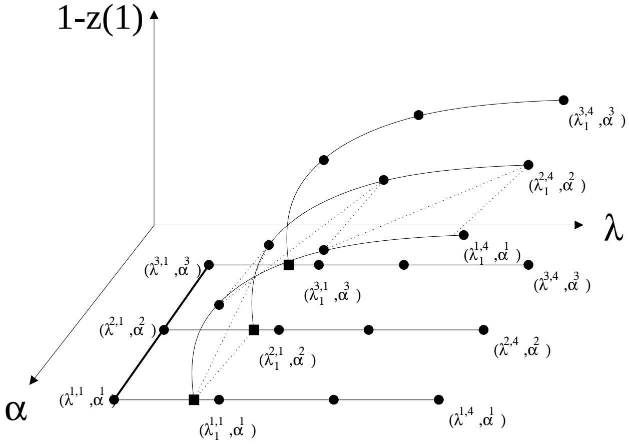

1-z(1)
$(\lambda_1^{3,4},\alpha^3)$
$(\lambda_1^{2,4},\alpha^2)$
λ
$(\lambda_1^{1,4},\alpha^1)$
$(\lambda^{3,1},\alpha^3)$
$(\lambda^{3,4},\alpha^3)$
$(\lambda_1^{3,1},\alpha^3)$
$(\lambda^{2,1},\alpha^2)$
$(\lambda_1^{2,1},\alpha^2)$
$(\lambda^{2,4},\alpha^2)$
α
$(\lambda^{1,1},\alpha^1)$
$(\lambda_1^{1,1},\alpha^1)$
$(\lambda^{1,4},\alpha^1)$

Figure 7.7. *Triangulation of a sheet of buckled equilibria: As in Figure 7.6, the thick black line in the (λ, α) plane represents the initial continuation in α keeping λ fixed, and the thin black lines in the plane represent the continuations in λ keeping α fixed. We now represent the branch points by black squares and show the branches of buckled equilibria emanating from each of them. The vertices of the triangulation of the sheet are shown as circles on the branches. This triangulation is indicated by the dotted lines, although, for clarity, we show only the triangulation between the first pair of branches.*

and $(\lambda_k^{i,j}, \alpha^i), (\lambda_k^{i,j+1}, \alpha^i), (\lambda_k^{i+1,j+1}, \alpha^{i+1})$ for $1 \leq i \leq N-1, 1 \leq j \leq M-1$, as shown in Figure 7.7.

This method for generating bifurcation surfaces is fairly intuitive, but it may not produce smooth results when the surfaces are complicated, such as in the case shown in the second column of Figure 7.4. Folds and sharp bends can be difficult to track smoothly, and, because the different one-dimensional branches are computed independently, the grid may become significantly more skewed than in the relatively regular cases shown in Figures 7.6 and 7.7. However, if the geometry of the surface is sufficiently simple, or if the discretization is made sufficiently fine, then the surface can be smoothly portrayed.

For example, we now give two examples of the computation of sheets of relatively simple buckled equilibria. First, we consider the region of the $\rho = 1$, $\gamma = 1$ plane of unbuckled equilibria highlighted in Figure 7.8, and then we show in Figure 7.9 two sheets of buckled equilibria branching from this region. Each sheet required approximately 6 hours of computation on a 700 MHz Pentium III Xeon and consisted of 40 MB of data. For each equilibrium, the index was calculated using a numerical implementation of the conjugate point technique discussed in section 2.2, and the surface is color coded by index with the color scheme from Figure 6.2. The first sheet contains portions of the slices appearing in column 1 of Figure 7.2, while the second sheet matches up with column 1 of Figure 7.4. On the first sheet, buckled solutions very close to the plane of unbuckled equilibria are stable (green) if α is sufficiently small but have index 1 (yellow) for larger values of α. Continuing further on the sheet, we see that the stable equilibria give way to index-1 equilibria as buckling continues. This pattern is repeated on the second sheet but with index-2 equilibria giving way to index-3 equilibria.

As a second example, we consider the region of the $\rho = \frac{3}{2}$, $\gamma = 1$ plane of unbuckled equilibria highlighted in Figure 7.10, and then we show in Figure 7.11 two sheets of buckled

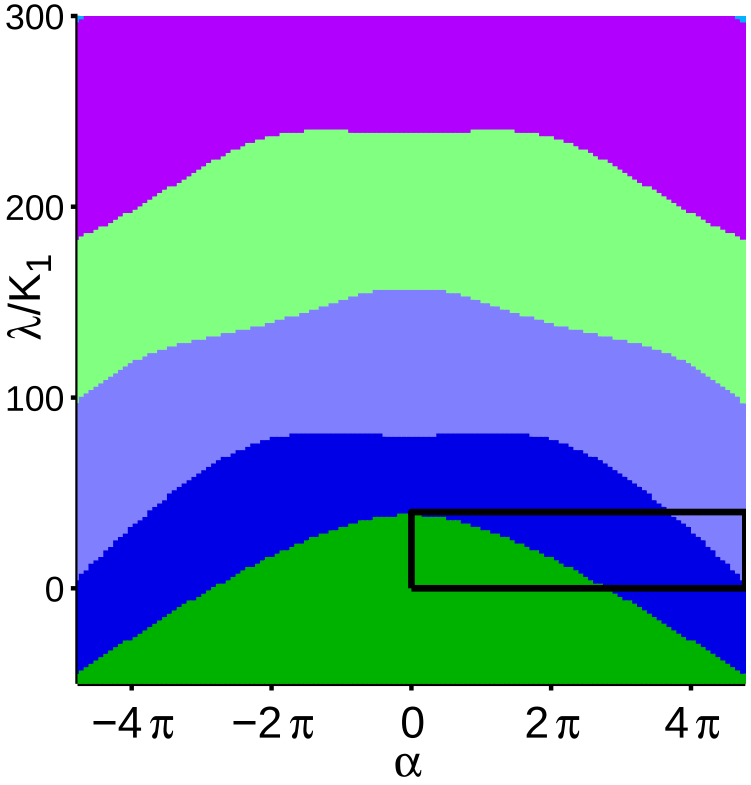


Figure 7.8. *A portion of the plane of unbuckled equilibria for an isotropic rod ($\rho = 1$) with twisting-to-bending stiffness ratio $\gamma = 1$. Within the region marked by the box, we will compute sheets bifurcating from the lines of color changes; these sheets are shown in Figure 7.9.*

equilibria branching from this region. In this case, the sheets of buckled equilibria match up with the slices appearing in column 2 of Figure 7.2, i.e., the splitting of the first sheet of buckled equilibria from the isotropic problem. The first sheet follows approximately the same pattern of index-0 equilibria giving way to index-1 equilibria that appeared in the isotropic case, while the second sheet involves a similar pattern of index-1 equilibria yielding to index-2 equilibria. This pattern is consistent with the usual expectation for stability of equilibria under symmetry-breaking perturbations: one of the perturbed images has the index of the symmetric case, and one has index one higher. In addition, we see a new feature in this case: the presence of a blue (index-2) patch on the first sheet. This blue patch exists because of a secondary bifurcating sheet, which is not shown in Figure 7.11 but can be clearly seen in the ($\alpha = \frac{3\pi}{2}, \rho = \frac{3}{2}$) slice in Figure 7.2 as a yellow dashed branch running on top of a solid blue branch. The solid blue branch is a slice of the blue patch in Figure 7.11, and thus we can infer that the blue patch is spanned by a nearly parallel yellow patch of which the yellow dashed branch is a slice. Looking through the slices in Figures 7.2 and 7.4, we can see that this is the simplest example of the type of surface branching and reconnection that proliferates through the bifurcation surface for more complicated buckled equilibria, not to mention the dependence of this topological structure on ρ and γ.

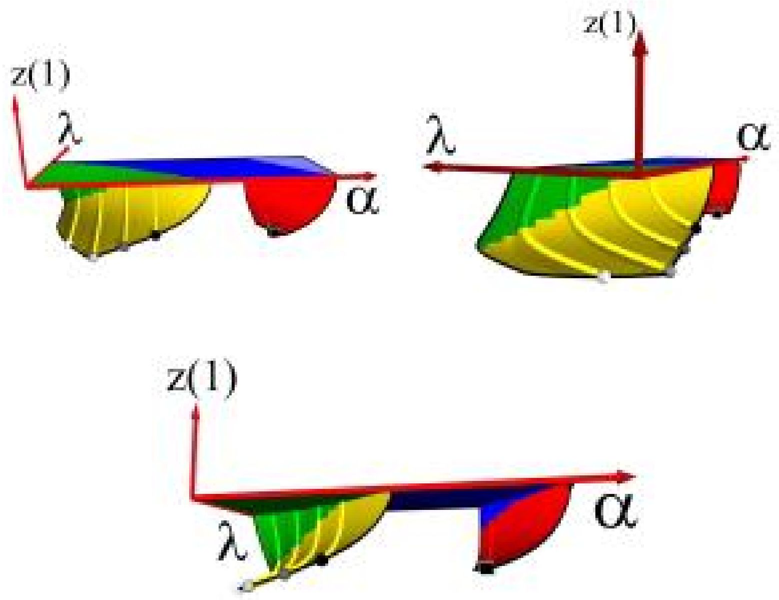


Figure 7.9. *Portion of the surface of buckled equilibria for an isotropic rod ($\rho = 1$) with twisting-to-bending stiffness ratio $\gamma = 1$, colored by the index of the equilibria according to the color scheme in Figure 6.2. For each buckled equilibrium, the value of the force λ, twist angle α, and length $z(1)$ are plotted, and three perspectives of the resulting surface are shown. The plane is the portion of the plane of unbuckled equilibria shown in Figure 7.8, and two sheets bifurcate from it. The origin of the red axes is at $(\alpha, \lambda, z(1)) = (0, 0, 1)$. The spheres and cube correspond to marked points in the left columns of Figures 7.2 and 7.4.*

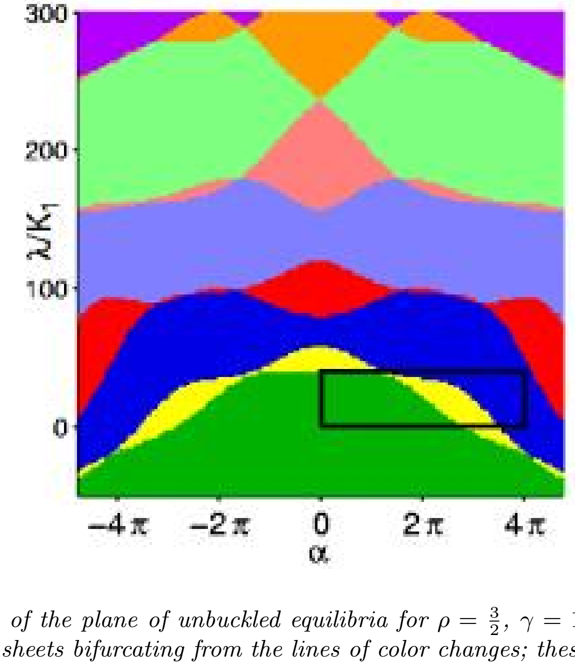


Figure 7.10. *A portion of the plane of unbuckled equilibria for $\rho = \frac{3}{2}$, $\gamma = 1$. Within the region marked by the box, we will compute sheets bifurcating from the lines of color changes; these sheets are shown in Figure 7.11.*

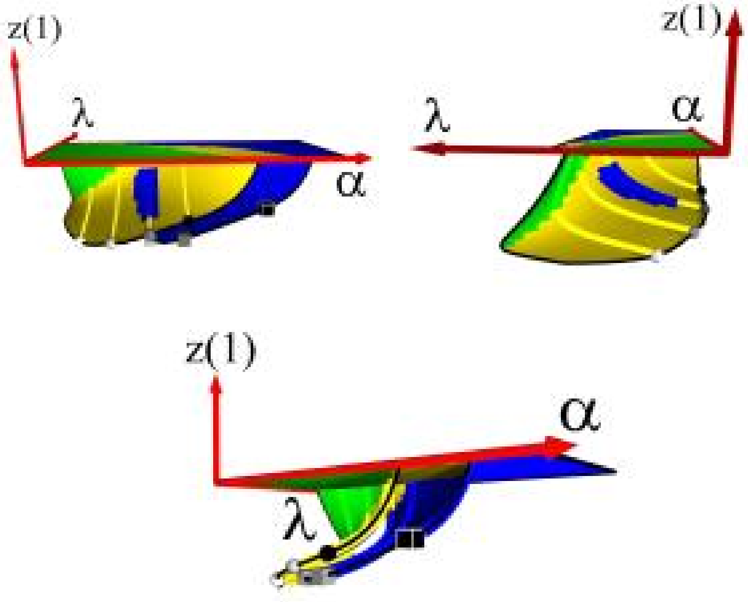


Figure 7.11. *Portion of the surface of buckled equilibria for $\rho = \frac{3}{2}$ and with twisting-to-bending stiffness ratio $\gamma = 1$, colored by the index of the equilibria according to the color scheme in Figure 6.2. As in Figure 7.9, three perspectives of the same surface are shown. The plane is the portion of the plane of unbuckled equilibria shown in Figure 7.10, and two sheets bifurcate from it. The origin of the red axes is at $(\alpha, \lambda, z(1)) = (0, 0, 1)$. The spheres and cubes correspond to marked points in the right column of Figure 7.2.*

8. Conclusions.

In this article, we developed a significant simplification of the standard technique for determining the index of equilibria in parameter-dependent calculus of variations optimization problems. One might argue that in practice only local minima are of interest, in which case it would seem more efficient to employ a minimization algorithm directly rather than pursue the apparently circuitous strategy of finding all equilibria and then performing a conjugate point computation to locate the local minima. For a single optimization problem, this argument may be correct, but for a problem with one or more parameters, it is less clear. As the parameters are varied, the set of equilibria tend to form connected sets that are readily tracked by parameter continuation, as we have seen in the one-dimensional and two-dimensional bifurcation diagrams in this paper. In contrast, the local minima may exist in several disconnected regions of parameter space and hence may be more difficult to locate and track using standard minimization techniques.

Such computational issues aside, the equilibrium approach offers major advantages in terms of qualitative understanding. The differential equations describing equilibrium may yield to qualitative analysis, at least for a subset of the solutions, that leads to insight into the overall structure of the problem. The central result of this paper, the corollary in section 3 and its extension to isoperimetrically constrained problems in section 4, is an example of this type of analysis. This result applied to the unbuckled configurations of the elastic strut led to a complete understanding of the stability index of these configurations, as shown in Figure 6.1. These unbuckled configurations were in turn the skeleton of the full set of equilibria, as seen in Figures 7.2, 7.4, 7.9, and 7.11, in that they provided the starting point for the numerical continuation necessary to compute the buckled configurations. As this pattern of complicated "nontrivial" solutions bifurcating from a "trivial" family of solutions is quite common, the combination of analysis and computation presented here could be applied to a variety of other parameter-dependent problems.

Acknowledgments. We thank John Maddocks for helpful discussions on some of these questions. Physical configurations and bifurcation surfaces were produced in POV-Ray.

REFERENCES

[1] S. S. Antman, *Nonlinear Problems of Elasticity*, Springer-Verlag, New York, 1995.

[2] O. Bolza, *Lectures on the Calculus of Variations*, 2nd ed., Chelsea, New York, 1961.

[3] A. R. Champneys and J. M. T. Thompson, *A multiplicity of localized buckling modes for twisted rod equations*, Proc. Roy. Soc. London A, 452 (1996), pp. 2467–2491.

[4] E. H. Dill, *Kirchhoff's theory of rods*, Arch. Hist. Exact Sci., 44 (1992), pp. 1–23.

[5] E. J. Doedel, H. B. Keller, and J. P. Kernévez, *Numerical analysis and control of bifurcation problems.* I. *Bifurcation in finite dimensions*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 1 (1991), pp. 493–520.

[6] E. J. Doedel, H. B. Keller, and J. P. Kernévez, *Numerical analysis and control of bifurcation problems.* II. *Bifurcation in infinite dimensions*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 1 (1991), pp. 745–772.

[7] G. M. Ewing, *Calculus of Variations with Applications*, Dover, New York, 1969.

[8] A. Friedman, *Foundations of Modern Analysis*, Holt, Rinehard, and Winston, New York, 1970.

[9] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice-Hall, Englewood Cliffs, NJ, 1963.

[10] A. Goriely, M. Nizette, and M. Tabor, *On the dynamics of elastic strips*, J. Nonlinear Sci., 11 (2001), pp. 3–45.

[11] A. G. Greenhill, Proc. Inst. Mech. Eng. (1883), p. 182.

[12] J. Gregory and C. Lin, *Constrained Optimization in the Calculus of Variations and Optimal Control Theory*, Van Nostrand Reinhold, New York, 1992.

[13] M. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, Springer-Verlag, New York, 1990.

[14] M. R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley and Sons, New York, 1966.

[15] H. B. Keller, *Numerical solution of bifurcation and nonlinear eigenvalue problems*, in Applications of Bifurcation Theory, P. H. Rabinowitz, ed., Academic Press, New York, 1977, pp. 359–384.

[16] G. Kirchhoff, *Über das Gleichgewicht und die Bewegung eines unendlich dünnen elastischen Stabes*, J. Reine Angew. Math., 56 (1859), pp. 285–313.

[17] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity,* 4th ed., Dover, New York, 1927.

[18] J. H. Maddocks, R. S. Manning, R. C. Paffenroth, K. A. Rogers, and J. A. Warner, *Interactive computation, parameter continuation, and visualization*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 7 (1997), pp. 1699–1715.

[19] R. S. Manning, K. A. Rogers, and J. H. Maddocks, *Isoperimetric conjugate points with application to the stability of DNA minicircles*, Proc. Roy. Soc. London A, 454 (1998), pp. 3047–3074.

[20] M. Morse, *Introduction to Analysis in the Large*, 2nd ed., Institute for Advanced Study, Princeton, NJ, 1951.

[21] S. Neukirch and M. E. Henderson, *Classification of the Spatial Clamped Elastica,* I *and* II, preprint, available online from http://lcvmsun9.epfl.ch/~neukirch/publi.html.

[22] R. C. Paffenroth, *Mathematical Visualization, Parameter Continuation, and Steered Computations*, Ph.D. thesis, University of Maryland, College Park, MD, 1999.

[23] M. Renardy and R. Rogers, *An Introduction to Partial Differential Equations*, Texts Appl. Math. 13, Springer-Verlag, New York, 1993.

[24] H. Sagan, *Introduction to the Calculus of Variations*, Dover, New York, 1969.

[25] M. D. Schuster, *A survey of attitude representations*, J. Astronaut. Sci., 41 (1994), pp. 439–518.

[26] G. H. M. van der Heijden, S. Neukirch, V. G. A. Goss, and J. M. T. Thompson, *Instability and Self-Contact Phenomena in the Writhing of Clamped Rods*, preprint, available online from http://lcvmsun9.epfl.ch/~neukirch/publi.html.

[27] G. H. M. van der Heijden and J. M. T. Thompson, *Lock-on to tape-like behaviour in the torsional buckling of anisotropic rods*, Phys. D, 112 (1998), pp. 201–224.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 1, pp. 146–174

Synchronized Activity and Loss of Synchrony Among Heterogeneous Conditional Oscillators[*]

Jonathan Rubin[†] and David Terman[‡]

Abstract. The inspiratory phase of the respiratory rhythm involves the synchronized bursting of a network of neurons in the brain stem. This paper considers activity patterns in a reduced model for this network, namely, a system of conductance-based ordinary differential equations with excitatory synaptic coupling, incorporating heterogeneities across cells. The model cells are relaxation oscillators; that is, no spikes are included. In the continuum limit, under assumptions based on the disparate time scales in the model, we derive consistency conditions sufficient to give tightly synchronized oscillations; when these hold, we solve a fixed point equation to find a unique synchronized periodic solution. This solution is stable within a certain solution class, and we provide a general sufficient condition for its stability. Allowing oscillations that are less cohesive but still synchronized, we derive an ordinary differential equation boundary value problem that we solve numerically to find a corresponding periodic solution. These results help explain how heterogeneities among synaptically coupled oscillators can enhance the tendency toward synchronization of their activity. Finally, we consider conditions for synchrony to break down.

Key words. oscillations, synchrony, synaptic coupling, heterogeneity, respiratory rhythm

AMS subject classifications. 34C15, 34C26, 37C25, 45J05, 92C20

PII. S111111110240323X

1. Introduction.

The inspiratory phase of the respiratory rhythm may originate in a network of interacting cells in a region of the brain stem called the pre-Bötzinger complex (pre-BötC) [17]. Within the pre-BötC, when coupling among cells is blocked, there are silent cells, cells that spike repeatedly, and intrinsically bursting cells that generate groups of spikes separated by pauses [12, 14, 17]. The burst frequencies vary among different bursting cells, depending on differences both in intrinsic cell properties and in inputs to the cells from other brain regions. Moreover, cells in all groups seem to be capable of bursting, if provided with appropriate inputs experimentally; hence they are referred to as conditional pacemakers.

Experiments in brain slices have shown that a coupled network of pre-BötC cells typically displays synchronized bursting oscillations, even though some uncoupled cells are silent or repeatedly spiking. Indeed, simulations suggest that with only about 10% of the cells in the network operating in the intrinsic bursting mode when uncoupled, a model pre-BötC network can still generate a synchronized population rhythm [3, 6]. Within the synchronized

[*] Received by the editors February 27, 2002; accepted for publication (in revised form) by T. Kaper July 23, 2002; published electronically August 2, 2002.

http://www.siam.org/journals/siads/1-1/40323.html

[†] Department of Mathematics and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA 15260 (rubin@math.pitt.edu). The work of this author was partially supported by NSF grants DMS-9804447 and DMS-0108857.

[‡] Department of Mathematics, Ohio State University, Columbus, OH 43210 (terman@math.ohio-state.edu). The work of this author was partially supported by NSF grant DMS-0103822.

cell population, details of the bursting pattern, such as precise onset and offset times, may differ slightly from cell to cell, presumably relating to the presence of intrinsic and input heterogeneities. Synchronization can also break down as it is replaced by a more complex bursting pattern such as a 2:1 or 4:1 rhythm, in which some subset of cells joins in only on 1 out of every 2 or 4 bursts of the rest of the cells [3], or perhaps even chaos [7].

In two papers, Butera and collaborators presented simulation results for models for individual cells in the pre-BötC [2] as well as for a network composed of these cells [3]. In the network, only excitatory coupling was included, as synchronized respiratory rhythms in pre-BötC persist under experimental blockage of inhibition but not under blockage of excitation [14]. For the most part, each cell was coupled to all other cells, since qualitatively similar results were found for sparse and full connectivities [3]. Cells also received tonic external excitatory input, with input strengths varying across the population.

The simulations in [2] captured the fact that individual neurons in the pre-BötC can be transformed from silent to bursting to repeated spiking states by varying certain parameters. When coupled, the simulated cells tended to engage in synchronized oscillations. Interestingly, while coupling among identical cells increased the range of external input levels over which synchronized oscillations occurred, relative to the oscillatory range for a single cell, networks of coupled cells with *heterogeneities* in certain parameter values displayed the broadest such dynamic range [3]. Thus heterogeneities in intrinsic cellular parameters and in external input levels are hypothesized to play key roles in enhancing the robustness of the respiratory rhythm and in shaping the details of cellular activity during these oscillations.

In this paper, we consider a synaptically coupled network of pre-BötC pacemaker cells, each featuring heterogeneities in certain parameters, with each cell governed by a reduced version of the conductance-based neuronal model presented in [2]; the model is introduced in section 2. We treat this system via both simulation and analysis. We perform simulations on a network of 20 heterogeneous cells with all-to-all excitatory synaptic coupling, with results presented in section 3. Our simulations, done with XPPAUT (developed by G. B. Ermentrout [8] and available at http://www.math.pitt.edu/~bard/xpp/xpp.html), provide a useful tool for phase space visualization of dynamics of nonidentical coupled oscillators. In particular, in animations of our results, we display nullclines, which dynamically evolve according to coupling levels present. For these nullclines, we also show the corresponding curves of knees, as defined in section 2, which evolve similarly. For oscillatory dynamics, the curves of knees are crucial in determining which cells are able to oscillate on each cycle of network activity, and the visualization that we provide gives an extremely clear way to view their role while network activity is in progress.

For our analytical treatment, we work in the continuum limit, in which the number of cells in the population is infinite. Our results thus yield a good approximation of network behavior with large numbers of cells, which is the biologically realistic scenario but for which direct simulations become difficult. Our analysis, in sections 4 and 5, provides conditions for the existence of stable synchronized relaxation oscillations in a reduced pre-BötC model. When these conditions are satisfied, all cells in the network will begin and terminate their active phases together, although they will do so from a distribution of voltage levels. Our analysis shows that, when it exists, this synchronized oscillatory solution is unique. It also yields a formula that pinpoints the location of cells at onset of activity, in an appropriate phase space. We prove that this solution is always stable within a certain solution class, and we provide a general sufficient condition for its stability. Alternatively, for the case of a more gradual onset of activity, observed in our simulations and in [3] for some parameter values, we derive in section 6 an ordinary differential equation boundary value problem that can be solved numerically for the location of the cells at activity onset (or termination). The same approach can be used to give conditions for synchrony with a gradual termination of activity. Finally, in section 7, we give a condition under which a cell configuration in phase space will not generate a synchronized oscillation, and we conclude with a discussion in section 8.

Synchronization of coupled oscillators has received significant attention in past works. In particular, synchrony is one of many activity patterns considered previously in modeling studies of networks of synaptically coupled relaxation oscillators (reviewed in [13, 16]). A novel feature of this paper is the inclusion of the effects of heterogeneity in such a network (see also [10, 19, 20, 5]). Heterogeneity has been found previously to compromise the robustness of synchronization in networks of *spiking* neurons with inhibitory coupling [4, 21]. Our analysis explains how heterogeneity can actually promote synchrony in a network of relaxation oscillators with excitatory synaptic coupling, even when some cells in the network would be unable to oscillate in isolation or with synaptic input only from cells identical to them. Similar mechanisms are likely to act to promote synchrony in heterogeneous populations of *bursting* neurons.

2. Model.

2.1. Single cell.

We model each cell by a system of ordinary differential equations of the form

$$\begin{aligned} v' &= f(v,h) + I_{app} + I_{syn}, \\ h' &= \epsilon g(v,h). \end{aligned} \tag{2.1}$$

Here $v(t)$ represents the membrane potential of the cell, and h is a channel state variable, as described below. The parameter I_{app} denotes an applied current, and $\epsilon > 0$ is assumed to be a small, singular perturbation parameter. The term I_{syn} encompasses coupling from other cells; it is described in detail below.

Suppose, for now, that $I_{syn} = 0$. We assume that the v-nullclines $\{f(v,h) + I_{app} = 0\}$ are cubic-shaped for all values of I_{app} of interest. Moreover, the h-nullcline $\{g(v,h) = 0\}$ is a monotone decreasing curve that intersects each of the v-nullclines at a single point, denoted by $p_0 = p_0(I_{app})$; see Figure 1. We further assume that $v' > 0$ (< 0) above (below) the v-nullcline and $h' > 0$ (< 0) below (above) the h-nullcline. It follows that, if p_0 lies on the middle branch of the v-nullcline, then (1) exhibits a stable limit cycle for all ϵ sufficiently small, while, if p_0 lies on either the left or right branch of the cubic nullcline, then p_0 is a globally stable fixed point for all ϵ sufficiently small.

In the simulations that follow, we consider a specific instance of (2.1), namely, the conductance-based model

$$\begin{aligned} C_m v' &= -g_{Na} m_\infty(v) h (v - v_{Na}) - g_L(v - v_L) + I_{syn} + I_{app}, \\ h' &= (h_\infty(v) - h)/\tau_h(v), \end{aligned} \tag{2.2}$$

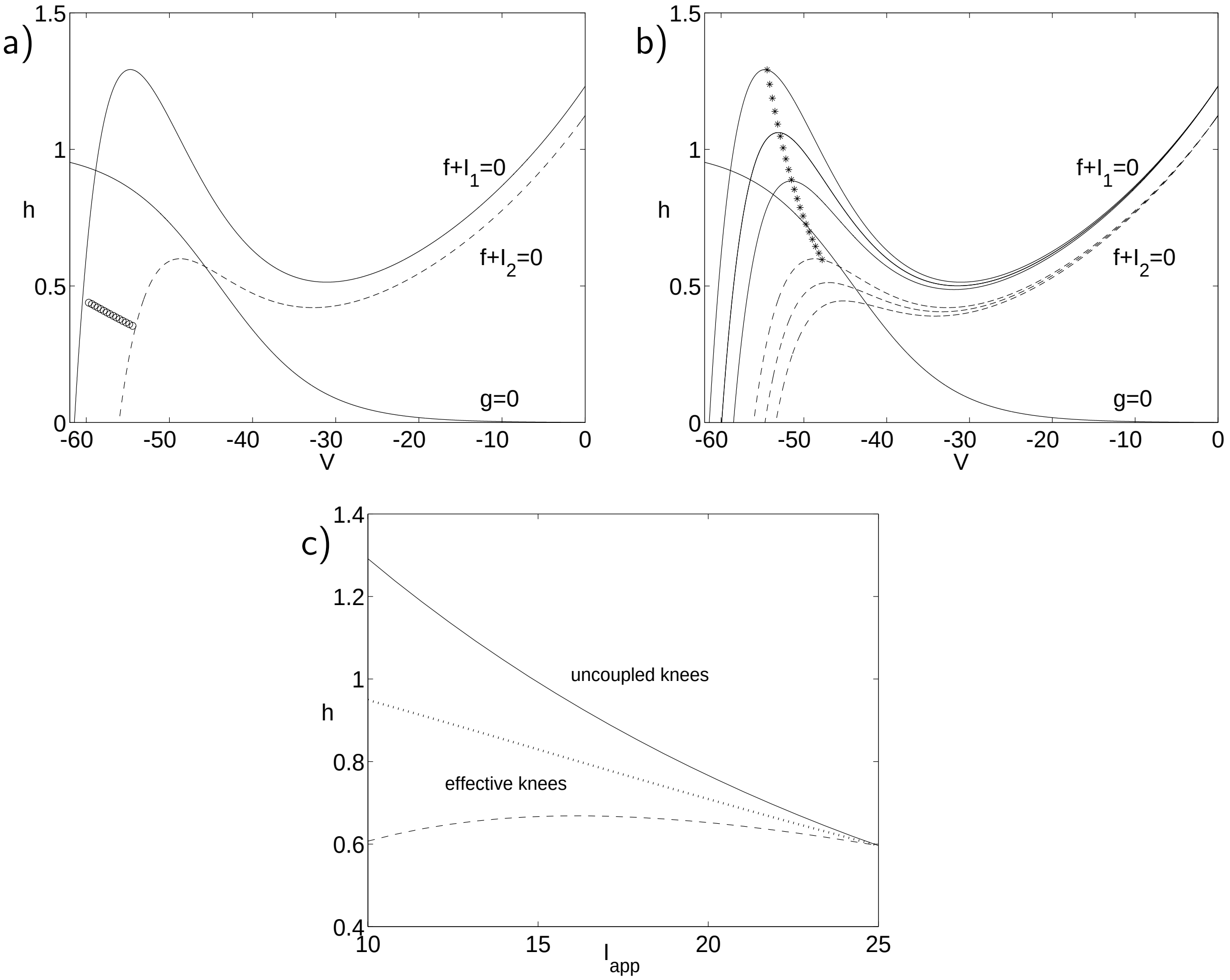


Figure 1. *Numerically generated nullclines for* (2.2) *in the* (v, h)*-phase plane, with parameters from the appendix.* (a) *Increasing excitatory input* $I_{app} > 0$ *(from* $I_1 = 10$ *to* $I_2 = 25$*) lowers the* v*-nullcline (here* $I_{syn} = 0$*). Cells with a range of* I_{app} *values can all be visualized in the same phase plane; an example representing the position of* 15 *cells is shown by the dark curve of circles.* (b) *Excitatory synaptic inputs further lower nullclines for* $v < v_{syn}$*. The dark curve denoted by asterisks gives a numerical approximation to the left curve of knees for* $I_{syn} = 0$ *and for* I_{app} *ranging from* I_1 *up to* I_2*.* (c) *Uncoupled* $(h_{LK}(I_{app}))$ *and effective left knees. The solid curve shows the left knee curve from part* (b)*. The other curves show effective knees computed numerically for* $g_{syn}/N = .005$ *(dotted) and* $g_{syn}/N = .01$ *(dashed) with* I_{syn} *computed by assuming that cells jump up to the active phase in order of decreasing* I_{app}*. Larger* g_{syn} *lowers the effective knees, promoting synchrony.*

where h_∞, m_∞ are monotone decreasing and increasing sigmoidal functions, respectively. The full functional forms and parameter values used are given in the appendix; these are based on models in [2, 3] but with sodium and potassium spiking currents blocked. The first equation in (2.2) describes the evolution of the voltage across a cell's membrane, with capacitance C_m, in terms of a persistent sodium current (I_{NaP} in [2, 3]), a leak current, and input currents. The second equation describes the slow inactivation of the persistent sodium current. For biophysically relevant parameter values, (2.2) can be considered as singularly perturbed, since h evolves much more slowly than v.

The v-nullclines of (2.2) for different values of I_{app}, with $I_{syn} = 0$, are shown in Figure 1a, along with the h-nullcline. On each nullcline, the left branch corresponds to the *silent phase*, where a neuron fires no spikes, and the right branch corresponds to the *active phase*, where a neuron is said to be spiking. For all values of I_{app} considered, the v-nullcline has two saddle-node points, or knees, where the branches coalesce; see Figure 1b–c. We refer to these as a left knee at a smaller value of v and a right knee at a larger value; we denote the left curve of knees by $(v_{LK}, h_{LK})(I_{app})$ and the right curve by $(v_{RK}, h_{RK})(I_{app})$. Let I_{max} denote the maximum of the values of I_{app} over the cell population, and set $h_{LK} = h_{LK}(I_{max})$.

Note from (2.2) that increasing I_{app} lowers the v-nullclines. If I_{app} is sufficiently small, then the fixed point p_0 lies on the left branch of the v-nullcline, and the system is said to be excitable. For larger values of I_{app}, the fixed point lies on the middle branch of the cubic nullcline, and the system is oscillatory, with a periodic solution that jumps up from the silent phase to the active phase and then down from the active phase to the silent phase. Thus this system captures the conditional pacemaker property of pre-BötC cells.

Simulations in [3] show that the biological effects of heterogeneities among cells in the pre-BötC are reproduced by introducing heterogeneities in the applied current I_{app} and in the intrinsic parameters v_L and g_{Na} in (2.2). The influences of heterogeneities in I_{app} and v_L in (2.2) can be combined by defining a new parameter $\tilde{I}_{app} = g_L v_L + I_{app}$. For notational convenience, we will instead, without loss of generality, fix $g_L v_L$ and restrict variations to I_{app}. In this paper, we will focus on heterogeneities in I_{app}, although we mention specific influences of heterogeneities in g_{Na} in the discussion.

2.2. Synaptic coupling. We now describe I_{syn}, the coupling between cells. First, consider a population of N discrete cells, and let the coupling to cell j be given by

$$I_{syn} = \frac{g_{syn}}{N}\left(\sum_{k=1}^{N} s_\infty(v_k)\right)(v_{syn} - v_j), \tag{2.3}$$

where

$$s_\infty(v) = 1/(1+\exp((v-\theta_s)/\sigma_s)).$$

Here we are assuming that the coupling is all-to-all and homogeneous. (That is, the form of s_∞ does not depend on the index k.) Note that, if σ_s is very small, then $s_\infty(v) \approx H(v-\theta_s)$, where H is the Heaviside step function; that is, $s_\infty(v) \approx 0$ if $v < \theta_s$, and $s_\infty(v) \approx 1$ if $v > \theta_s$. In the analysis, we assume that $s_\infty(v) = H(v - \theta_s)$. The value θ_s is such that a cell's voltage increases through θ_s as it jumps up to the active phase.

In (2.3), $g_{syn} > 0$ represents the maximal synaptic conductance, and v_{syn} is the synaptic reversal potential. We will choose v_{syn} so that $v_k(t) < v_{syn}$ for each k along every solution of interest. This implies that I_{syn} is always positive, corresponding to excitatory coupling; see Figure 1b.

In the analysis, we will consider the continuum limit of infinitely many cells. We assume that each cell is parameterized by a point x in some domain D. One may view D as some subset of $\mathbb{R}^3$; however, this is not necessary. We then denote the dependent variables as

$v(x,t)$ and $h(x,t)$. As before, heterogeneities will be in the applied currents, which we denote as $I_{app}(x)$. We then let

$$I_{syn}(x,t) \;=\; (g_{syn}/Vol)\,(v_{syn} - v(x,t)) \int_D s_\infty(v(x,t))dx, \tag{2.4}$$

where s_∞ is defined as above and $Vol = \int_D dx$.

When $I_{syn} = 0$, the h-values $h_{LK}(I_{app}), h_{RK}(I_{app})$ for the knees of the v-nullcline are defined by solving the two equations $f + I_{app} = 0$ and $\partial f/\partial v = 0$. Suppose that all cells with the same value of I_{app} are synchronized, such that the parameterization by x is equivalent to a parameterization by I_{app}. For $I_{syn} > 0$ given by (2.4), the equations $f + I_{syn} + I_{app} = 0$ and $\partial(f + I_{syn})/\partial v = 0$ then define curves $h(I_{app})$ as well. When solutions exist, we refer to these as *effective knees*. The effective knees will be crucial for determining which cells jump up to the active phase in a network oscillation. An example appears in Figure 1c, based on the assumption that cells jump up to the active phase in order of decreasing I_{app}, such that I_{syn} becomes larger for cells with smaller I_{app}. The role of the effective knees is illustrated in the numerical simulations in the next section. Note that, in Figure 1c, the effective curve of knees may be nonmonotone due to the larger value of I_{syn} that occurs for smaller I_{app} when cells jump up to the active phase in order of decreasing I_{app}.

3. Numerical simulations of network activity.

In this section, we present numerical simulations of (2.2) that illustrate different population rhythms exhibited by the model network. We consider a population of 20 cells and denote the dependent variables corresponding to cell i, $1 \le i \le 20$, as $(v_i(t), h_i(t))$. We assume that the heterogeneity parameter I^i_{app} varies in a uniform linear fashion between $I_{min} = 10$ and $I_{max} = 25$. Note that, when $I_{app} = I_{min}$, system (2.2) with $I_{syn} = 0$ is excitable, while, if $I_{app} = I_{max}$, then (2.2) is oscillatory. We demonstrate how the population rhythm changes as we vary the parameter $\bar{g}_{syn} = g_{syn}/20$.

First suppose that $\bar{g}_{syn} = .012$. Then the cells' activities are fairly well synchronized. This is demonstrated in Figure 2a, where we show the evolution of each $v_i(t)$. Note that all of the cells jump up to and down from the active phase at approximately the same times. Further, ordered jumping up occurs, such that cells with larger values of I_{app} jump up before cells with smaller I_{app}, as seen in the simulations in [3].

Perhaps a more illuminating way to present this solution is presented in Figure 3. This shows the evolution of each $(v_i(t), h_i(t))$ in the same (v, h) phase plane, along with the cubic-shaped v-nullclines (in black for ilow $\equiv I_{min}$ and in red for ihigh $\equiv I_{max}$) and effective knees corresponding to different levels of applied current and synaptic coupling. The synaptic coupling level is displayed in the upper-right corner as s_{tot}, defined as $\sum_{k=1}^{20} s_\infty(v_k)$. Cells are color-coded according to their I_{app} values, with red for largest I_{app} and dark blue for smallest I_{app}, and the effective left and right knees for each cell share its coloring. As s_{tot} changes, the effective knee positions move correspondingly. From this animation, we can observe that the cell with largest I_{app} jumps up first. Moreover, since the excitatory coupling lowers the nullclines and corresponding knees for cells while any cells are in the active phase, we see that synaptic coupling here prolongs active phases and slows oscillation frequency relative to the uncoupled case (as a consequence of fast threshold modulation; see [18]). This is similar, but complementary, to the mechanism by which inhibition can speed up rebound burst frequency

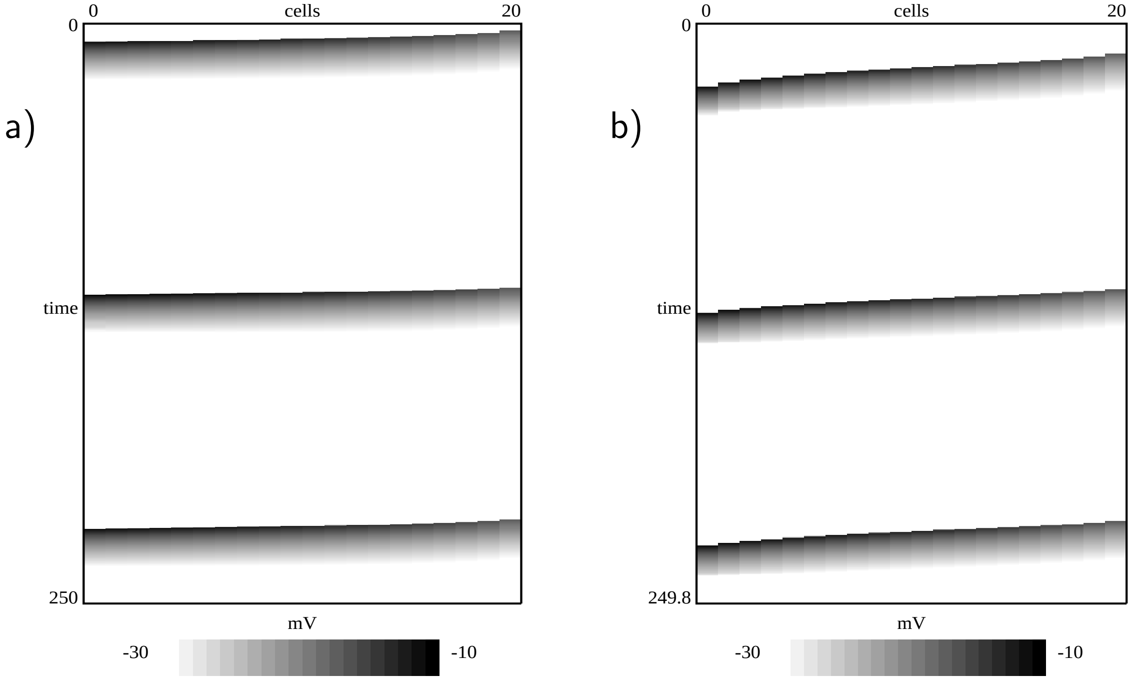


Figure 2. *Voltage versus time for synchronized oscillations of* 20 *heterogeneous pre-BötC cells. Moving horizontally across the figure corresponds to picking out different cells in the network. Time evolves downward. Voltage is coded in greyscale. The range of voltages encoded is limited to* $-30mV$ *up to* $-10mV$ *in order to focus sharply on the active phases of oscillations, which correspond to the dark bands in the figure. Cells with larger* I_{app} *values are labelled by higher numbers, appearing to the right in the figure. Note that cells with large* I_{app} *tend to jump up earliest in each cycle.* (a) $\bar{g}_{syn} = 0.012$ *gives a fairly unified jump-up.* (b) $\bar{g}_{syn} = 0.009$ *gives a more gradual jump-up.*

in networks [11], and it agrees with the observation in [3] that excitatory synaptic coupling slows oscillation frequency.

Figure 3 clearly demonstrates that each cell, while in the silent (active) phase, lies along the left (right) branch of the cubic corresponding to the level of applied current and synaptic input it is receiving. The positions of all of the cells at each fixed time approximates a curve that evolves in the (v, h) phase plane. We refer to this curve as a *snake of synchrony*. The primary goals of this paper are to derive conditions for when such a synchronous solution exists and to derive an analytic expression for the corresponding evolving snake curve.

Remark 3.1. Generally, a synchronous solution can be defined as one in which all cells jump up to the active phase on each cycle, and no active cell jumps down until all cells have jumped up. In our analysis, we will consider different types of synchronous solutions. In one, cells will jump up at the same moment in time, in an appropriate sense. In another, we allow cells to jump up at different times, but we require that no active cell jumps down until all of the other cells are active. In both cases, we assume that cells with $I_{app} = I_{max}$ jump up first and that cells jump up in order of decreasing I_{app}, as observed in simulations.

As we gradually decrease g_{syn}, and thus $\bar{g}_{syn}$, different cells may jump down first, and then the jump-up may become less unified (Figures 2b and 4). Note from Figure 2b and

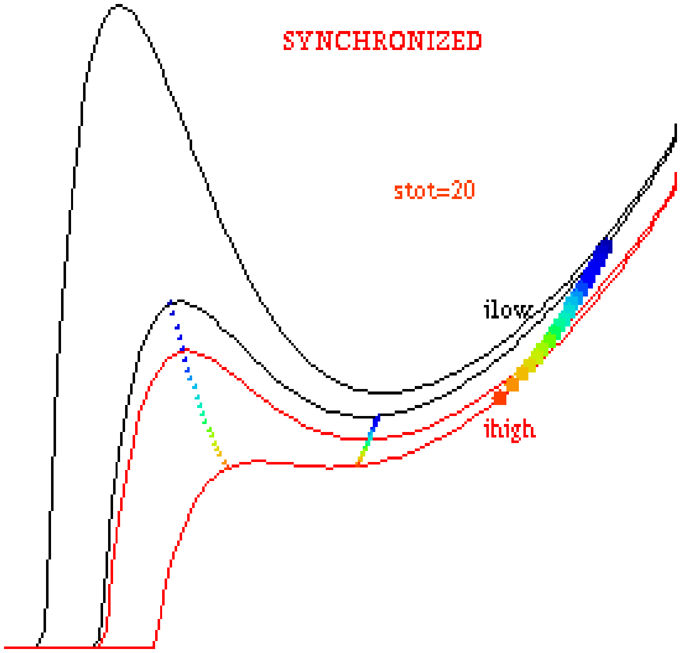


Figure 3. *Animation of a simulated synchronized oscillation with unified jumps ($\bar{g}_{syn} = .012$). Since all cells jump up at similar times, they end up fairly close together in the active phase, as shown in this still frame. Since all cells are in the active phase in the still frame, $s_{tot} = \sum_{k=1}^{20} s_\infty(v_k) = 20$.*

Figure 4 that the same cell (with $I = I_{max}$) jumps up and down first; as g_{syn} is weakened and the jump-up becomes more gradual, cells become unable to catch up to the lead cell with $I = I_{max}$ in the active phase. Finally, for still smaller g_{syn}, synchrony is lost, and more exotic population behaviors arise. For example, suppose that $\bar{g}_{syn} = .00825$, and consider the solution shown in Figures 5a and 6. Note that the entire population breaks up into two groups. Cells within each group are fairly well synchronized; however, one of the groups jumps up to the active phase only during every second cycle of the other group. As we decrease g_{syn}, solutions become increasingly more complicated. Figures 5b and 7, for instance, show that, when $\bar{g}_{syn} = .0035$, the solution appears to be quite irregular and possibly chaotic. Moreover, cells with $I_{app} < I_{max}$ may now jump up first on certain cycles, as seen in Figure 7. This can happen because, when a cell fails to jump up on one cycle, it can end up quite close to its uncoupled knee in the silent phase after active cells jump down on that cycle. Finally, when g_{syn} is sufficiently small, cells behave essentially as if they are uncoupled.

4. Analysis of snakes—preliminaries.

4.1. Introduction. Here and in section 5, we derive an analytic expression for a periodic solution to (2.1), consisting of a snake of synchrony for which all cells become active on each cycle of network activity, and conditions for when such a solution exists. In order to derive the analytic formulas, it will be necessary to make several simplifying assumptions on the nonlinear functions in (2.1). These assumptions are based on the numerical simulations of

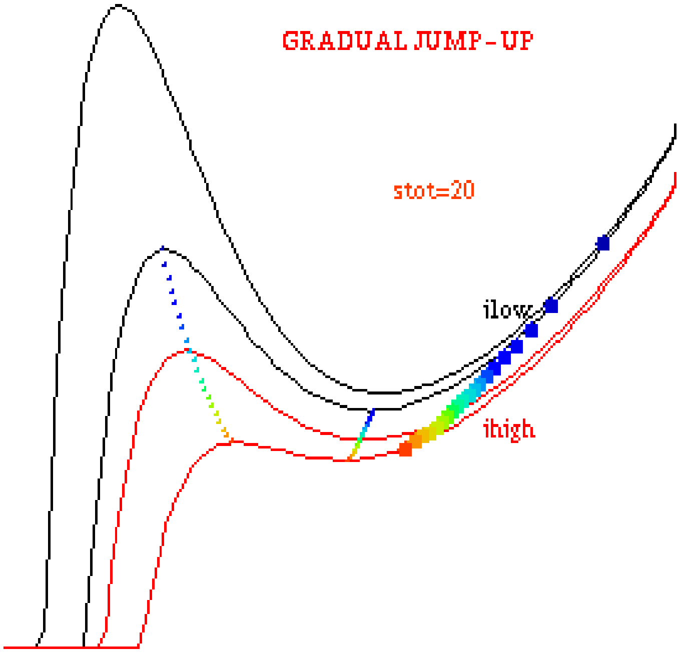


Figure 4. *Animation of a simulated synchronized oscillation with gradual jump-up ($\bar{g}_{syn}$ = .009). The gradual jump-up causes cells to be quite spread out in the active phase, as shown in this still frame.*

(2.2) discussed above as well as the forms of the nonlinearities in (2.2), as discussed below and in the appendix. In particular, to derive explicit formulas, we assume that all cells jump up and down together, in a sense to be made precise below. This is quite accurate for large g_{syn}. In section 6, we allow for more gradual jumping. This does not lead to an explicit snake formula but rather to a single ordinary differential (with respect to I_{app}) equation boundary value problem that can be solved numerically for the periodic solution, expressed as a curve parameterized by I_{app}.

Recall that a snake of synchrony is a curve in the (v, h) phase plane, parameterized by the position $x \in D$, that evolves in time. For our analysis, it will be more convenient to parameterize the snakes by the heterogeneity parameter I_{app}, which we usually write as simply I. As noted earlier, this is justified if all of the cells with the same input I_{app} are completely synchronized; that is, if $I_{app}(x_1) = I_{app}(x_2)$, then $(v(x_1, t), h(x_1, t)) = (v(x_2, t), h(x_2, t))$ for all t. Under this assumption, we denote the snake as $(v(I, t), h(I, t))$ for $I_{min} \le I \le I_{max}$, where I_{min} and I_{max} are the minimum and maximum values of I_{app} for $x \in D$, respectively. (It is assumed that each of I_{min} and I_{max} is attained for some x in D, justifying the notation, and that both are finite.) We shall refer to either the cells with input I or the position in phase space of these cells as *cell(I)*.

We assume that, initially, the snake is in the silent phase with one of the cells at a left knee ready to jump up. This will turn out to be the cell (or cells) with the maximum applied current I_{max}, as discussed in the previous section and Remark 3.1. We then follow the snake around in phase space until it completes one cycle. This cycle consists of four pieces: the

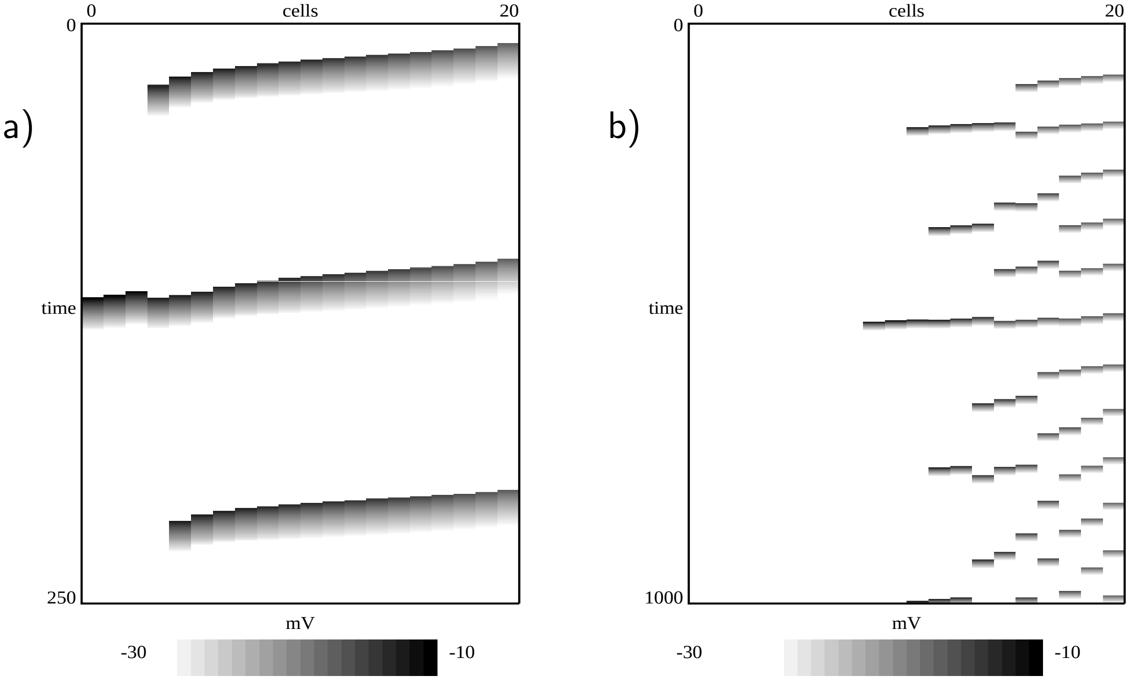


Figure 5. *Voltage versus time for asynchronous rhythms of* 20 *heterogeneous pre-BötC cells.* (a) *With* $\bar{g}_{syn} = .00825$*, three oscillation cycles are shown; in the first and third, several of the cells with small* I_{app} *fail to become active.* (b) *With* $\bar{g}_{syn} = .0035$*, participation in the oscillations is irregular, especially for cells on the edge of the participating and nonparticipating regions of the population.*

jump-up, the active phase, the jump-down, and the silent phase. We analyze the evolution of the snake over each of these pieces in the subsections below. The analytic formula for the snake is then obtained by assuming that the snake returns after one complete cycle to precisely the position from which it started. We derive the formula for the position of the snake at jump-up, although this could be done similarly for the snake position at other stages in a cycle.

4.2. The silent and active phases. We now derive equations for the evolution of the cells during the silent and active phases. The first step is to introduce the slow time scale $\tau = \epsilon t$ in (2.1). We then set $\epsilon = 0$ and $I_{tot}(x,t) = I_{app} + I_{syn}(x,t)$ to obtain the reduced equations

$$\begin{aligned} 0 &= f(v,h) + I_{tot}, \\ h' &= g(v,h), \end{aligned} \tag{4.1}$$

where differentiation is now with respect to τ. The first equation states that the cells lie along either the left or right branch of the cubic v-nullcline determined by I_{tot}; we refer to these nullclines simply as cubics below. We denote these branches as

$$v = v_L(h, I_{tot}) \quad \text{and} \quad v = v_R(h, I_{tot}),$$

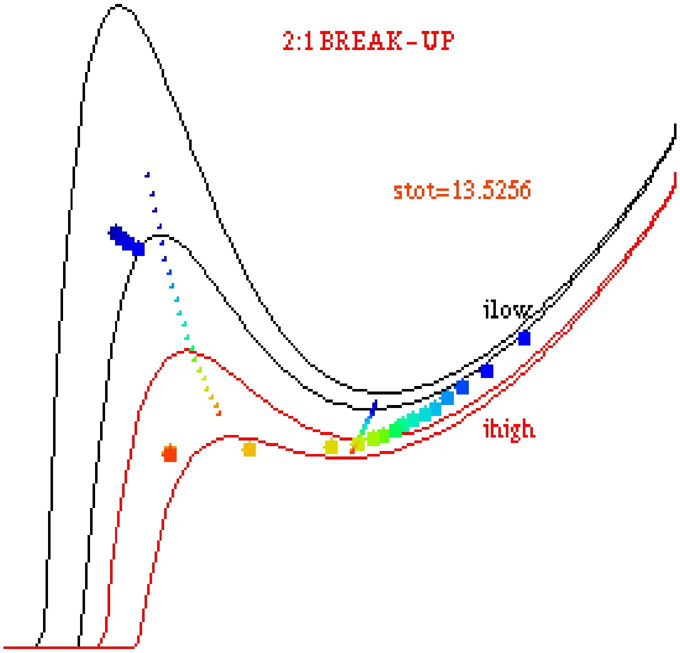


Figure 6. *Animation of a simulated solution that breaks up ($\bar{g}_{syn}$ = .00825). This still frame shows the red and orange cells, with largest I_{app}, beginning to jump down, while the darkest blue cells, with smallest I_{app}, have failed to reach their effective knees for jump-up.*

respectively. Substituting this into the second equation of (4.1), we find that the slow variables h satisfy scalar equations of the form

$$h' = g(v_\alpha(h, I_{tot}), h) \;\equiv\; G_\alpha(h, I_{tot}), \tag{4.2}$$

where $\alpha = L$ or R depending on whether the cell lies in the silent or active phase, respectively.

We next make some simplifying assumptions on the nonlinear functions. These will allow us to solve the scalar equations (4.2) explicitly. We first consider the active phases, during which the cells lie along the right branches of certain cubics. For these values of (v, h), we assume that

$$h' = G_R(h, I_{tot}) = -\rho h \tag{4.3}$$

for some positive constant ρ. In order to justify this assumption, we consider the biophysical model (2.2), in which

$$g(v, h) = (h_\infty(v) - h)/\tau_h(v). \tag{4.4}$$

For the parameter values given in the appendix, one finds that $h_\infty(v)$ is extremely small and $\tau(v)$ is nearly constant while the cells are in their active phases. This then leads to the approximation (4.3).

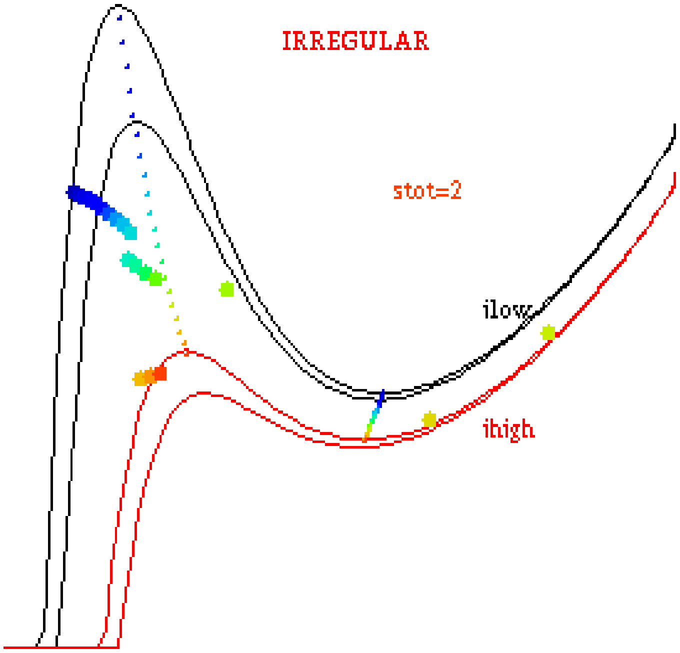


Figure 7. *Animation of a simulated irregular solution ($\bar{g}_{syn} = .0035$). In this still frame, note that cells with $I_{app} < I_{max}$ have jumped up first in this oscillation cycle. Since only two cells are active in the still frame, $s_{tot} = 2$.*

Now consider the silent phases, during which cells lie along the left branches of the appropriate cubics. We will assume that $G_L(h, I_{tot})$ is linear. More precisely, assume that there exist positive constants a, b, and c such that

$$h' = G_L(h, I_{tot}) = -ah - bI_{tot} + c. \tag{4.5}$$

This will be the case if $g(v, h)$ is given by (4.4), $h_\infty(v)$ is linear, $\tau_h(v)$ is constant, and each of the left branches is linear. Of course, none of these conditions are precisely satisfied. However, we demonstrate later that these assumptions lead to a very good approximation of the synchronized snake.

Remark 4.1. Recall that $h_{LK} = h_{LK}(I_{max})$, the h-value of the left knee of the v-nullcline for $I_{app} = I_{max}$. Since we assume that there is no fixed point on the left branch of the v-nullcline for $I_{app} = I_{max}$, we have

$$-ah_{LK} - bI_{max} + c > 0. \tag{4.6}$$

4.3. Jumping up. In section 5, we will consider synchronous snakes with the property that all of the cells jump up together, with respect to the slow time scale. In section 6, we consider snakes that jump up more gradually. The reason why simultaneous jump-up is possible is that, when one cell jumps up, the synaptic inputs to all of the other cells increase. This lowers the cubics associated with the other cells. If one of the other cells lies above the left knee of its lowered cubic (i.e., above its effective left knee), then that cell will also jump up to the active phase, leading to the further lowering of the cubics. Since the jump-up takes place on the fast time scale and the Heaviside synaptic variables $s_\infty(v)$ respond instantaneously, it is possible for all of the cells to jump up together with respect to the slow time variable τ.

We now derive an expression for when the cells do jump up together as described above. Cells with $I_{app} = I_{max}$ jump up when they reach their left knee. Motivated by Remark 3.1, we assume that the cells that jump up do so in order of decreasing I_{app}. Fix $I \in [I_{min}, I_{max})$, and assume that each of the cells with $I < I_{app} < I_{max}$ jumps up at the same moment as $cell(I_{max})$. We wish to determine whether $cell(I)$ must also jump up at that moment, that is, whether it lies above the left knee of the appropriate cubic.

Let $(v(I), h(I))$ denote the position of $cell(I)$ in (v, h) phase space, where $v(I) = v_L(h(I), I)$, at the moment when the cells with $I_{app} > I$ jump up. From (2.4), it follows that the synaptic input to $cell(I)$ is

$$I^u_{syn}(I) \equiv (g_{syn}/Vol)(v_{syn} - v(I))\mu\{x : I_{app}(x) > I\}, \tag{4.7}$$

where μ is the usual Lebesgue measure and Vol is the same normalization factor given in (2.4). Essentially, $\mu\{x : I_{app}(x) > I\}$ gives the volume of the subset of D on which $I_{app}(x) > I$. We assume henceforth that $I^u_{syn}(I)$ is a continuously differentiable function of I on (I_{min}, I_{max}). Following the notation introduced in the preceding section, we find that $cell(I)$ will jump up if

$$h(I) > h_{LK}(I + I^u_{syn}(I)). \tag{4.8}$$

Hence, if this last condition is satisfied for all $I \in [I_{min}, I_{max}]$, then all of the cells will jump up together. In subsection 5.2, we demonstrate how this leads to an explicit condition on parameters and nonlinear functions in (2.2) for the existence of a snake of synchrony.

4.4. Jumping down. For the solutions under consideration in sections 5 and 6, all cells jump down at the same time with respect to τ, from right knees $h_{RK}(I)$ which depend on I. This is possible through a mechanism analogous to that described above for synchronous jump-up. Jump-down is initiated when $cell(I_d)$ reaches its right knee for some $I_d \in [I_{min}, I_{max}]$. Once this occurs, then, for each $I \neq I_d$, the loss of synaptic input to $cell(I)$ from the jumping down of other cells raises the effective right knee of $cell(I)$ sufficiently high that $h(I) < h_{RK}(I + I_{syn}(I))$ for the appropriate value of $I_{syn}(I)$, and $cell(I)$ jumps down. Since any $cell(I_d)$ may initiate the jump-down, writing an expression analogous to (4.7) for $I_{syn}(I)$ at jump-down becomes complicated (although, for special cases, we can derive a jump-down condition analogous to (4.8)). Instead, we simply assume that all cells jump down together. This assumption is based on numerics showing unified jump-down for all synchronous solutions. Moreover, while no cells have critical points on their right branches, all cells' right knees come close to $h_\infty(v)$ in the active phase for the parameter values in the appendix. Thus, based on (4.4), all cells become compressed toward their right knees in the active phase, promoting unified jump-down.

5. Linear snakes: Cells jump up and jump down together.

5.1. Snake formula. We assume throughout this section that all of the cells jump up and jump down together, with respect to the slow time scale. Although this condition may not hold, in general, we will demonstrate that, for strong coupling, it does lead to a very good approximation for the snake. We shall derive an explicit formula for the initial (jump-up) position $h(I) \equiv h(I,0)$ of a periodic snake of synchrony. It will follow that $h(I)$ must be linear in the case of synchronized jumps.

Suppose that the first cell to jump down is *cell*(I_d) and this cell jumps down at the right knee, whose position we denote by h^d_{RK}. (Note that it is possible that $I_d \neq I_{max}$.) Then the time that cells spend in the active phase after they jump up is the time for *cell*(I_d) to evolve from $h_d \equiv h(I_d)$ to h^d_{RK} under (4.3), namely, $T_A = \frac{1}{\rho}\ln(h_d/h^d_{RK})$. At this jump-down time, each cell has a position given by $h(I,T_A) = h(I)h^d_{RK}/h_d$.

We next consider when the cells are in the silent phase after they jump down. During this time, $I_{syn} = 0$. Hence each cell evolves according to (4.5) with $I_{tot} = I$ and initial position $h(I,T_A)$. It follows that, while in the silent phase,

$$h(I,\tau) = h(I)\frac{h^d_{RK}}{h_d}e^{a(T_A-\tau)} + \Lambda_I(1-e^{a(T_A-\tau)}), \tag{5.1}$$

where we set

$$\Lambda_I = (c-bI)/a, \tag{5.2}$$

which is the value of the critical point of (4.5) for $I_{syn} = 0$. One cycle is completed when *cell*(I_{max}) returns to its left knee, $h_{LK} \equiv h_{LK}(I_{max})$. If this is at time T, then, setting $T_S = T - T_A$, (5.1) yields

$$h(I,T) = h(I)\frac{h^d_{RK}}{h_d}e^{-aT_S} + \Lambda_I(1-e^{-aT_S}) \equiv M(h(I)). \tag{5.3}$$

Now a periodic snake of synchrony corresponds to a fixed point of the operator $M(h(I))$.

In particular, for $I = I_{max}$, setting $\Lambda_I = \Lambda_M \equiv \Lambda_{I_{max}}$ in (5.3) gives

$$h_{LK} \;=\; h_{LK}\frac{h^d_{RK}}{h_d}e^{-aT_S} + \Lambda_M(1-e^{-aT_S}).$$

Multiplying through by $h(I)/h_{LK}$ gives

$$h(I) = h(I)\frac{h^d_{RK}}{h_d}e^{-aT_S} + \frac{\Lambda_M h(I)}{h_{LK}}(1-e^{-aT_S}). \tag{5.4}$$

We now have expressions for $M(h(I))$, from the right-hand side of (5.3), and $h(I)$, from (5.4). The corresponding fixed point equation $M(h(I)) = h(I)$ has a unique solution, given by the analytic expression

$$h(I) = h_{LK}\left(\frac{\Lambda_I}{\Lambda_M}\right) \;=\; h_{LK}\left(\frac{c-bI}{c-bI_{max}}\right). \tag{5.5}$$

Note that the value of $h(I)$ given in (5.5) is attainable by the evolution of (4.5), since it lies below the critical point of (4.5) for each I. This holds since $h_{LK} < \Lambda_M$ by (4.6), while the critical point of the h-equation for fixed I is Λ_I. In Figure 8, the fixed point snake position at jump-up predicted by (5.5) is compared to the snake position at jump-up from numerical simulations of the pre-BötC network (2.2), with 20 cells, for $g_{syn}/20 = .012$, which leads to relatively simultaneous jumps (see Figures 2a and 3).

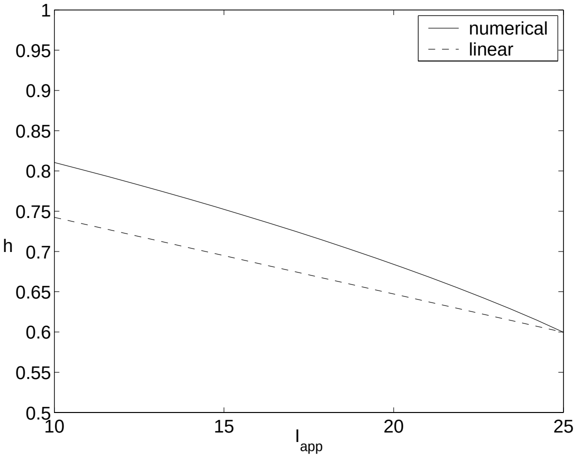


Figure 8. *Fixed point snake positions at jump-up. The dashed curve shows the estimate from formula* (5.5)*, and the solid curve shows the result from numerical simulations of* (2.2)*, with* 20 *cells, both for* $g_{syn}/20 = .012$.

Remark 5.1. Interestingly, formula (5.5) for the snake does not depend on any parameter associated with the active phase or the synaptic coupling between cells. Such parameters may affect whether or not cells all jump up together, but if this does happen, then (5.5) gives the position of the snake of synchrony at jump-up.

5.2. Jump-up condition. When deriving the snake formula, we assumed that all of the cells jumped up together. Here we use (4.8) to derive conditions on the parameters for when this must be the case. For simultaneous jump-up, the left-hand side of (4.8) is given by the snake formula given in (5.5). It remains, therefore, to estimate terms on the right-hand side of (4.8).

Choose λ_1 so that, if (v, h) lies in the silent phase (with $I_{syn} = 0$), then

$$v_{syn} - v > \lambda_1. \tag{5.6}$$

We assume that there exists $\lambda_2 > 0$ such that

$$\mu\{x : I_{app}(x) > I\} \geq \lambda_2(I_{max} - I) \tag{5.7}$$

for all $I \in [I_{min}, I_{max}]$. The existence of a strictly positive λ_2 such that (5.7) holds requires that the distribution of I values does not have an exponentially decaying “tail” near I_{max}. Such a tail would lead to a gradual jump-up, which is discussed in section 6.

Now let $I^u_{syn}(I)$ be as in (4.7), and set $\lambda_0 = \lambda_1\lambda_2$. It then follows from (4.7), (5.6), and (5.7) that

$$I^u_{syn}(I) \;>\; g_{syn}\lambda_0(I_{max} - I). \tag{5.8}$$

Finally, we assume that we can bound $h_{LK}(I)$ between two linear functions of I; that is, there exist positive constants m_1 and m_2 such that

$$m_1(I_{max} - I) < h_{LK}(I) - h_{LK} \;<\; m_2(I_{max} - I) \tag{5.9}$$

for all $I \in [I_{min}, I_{max}]$. (Recall that $h_{LK} \equiv h_{LK}(I_{max})$.) Together with (5.8), this implies that

$$h_{LK}(I + I^u_{syn}) < h_{LK} + m_2(I_{max} - I)(1 - g_{syn}\lambda_0). \tag{5.10}$$

A straightforward calculation, combining (4.8), (5.5), and (5.10), then demonstrates that (4.8) is satisfied if

$$g_{syn} > \frac{1}{\lambda_0}\left(1 - \frac{bh_{LK}}{am_2\Lambda_M}\right), \tag{5.11}$$

that is, if the synaptic coupling is sufficiently large.

When deriving the snake formula, we also assumed that *cell*(I_{max}) is the first to jump up. This will be the case if

$$h(I) \;=\; h_{LK}\frac{\Lambda_I}{\Lambda_M} \;<\; h_{LK}(I)$$

for all $I < I_{max}$. Using the left-hand side of (5.9), the definition of Λ_I from (5.2), and the notation $\Lambda_M := \Lambda_{I_{max}}$, we find that this holds if

$$m_1 > \frac{bh_{LK}}{a\Lambda_M}, \tag{5.12}$$

that is, if the curve of left knees is sufficiently steep.

Remark 5.2. In the simulations throughout this paper, the right-hand side of (5.12) is considerably smaller than m_1. This fits nicely with the fact that *cell*(I_{max}) jumps up first in all of the synchronized solutions that we observe.

Remark 5.3. In a similar manner to the above, we can derive conditions for which a cell hits its right knee first and for whether all cells jump down together when this occurs. Note, however, that the jump-down is in general observed to be well synchronized in our simulations, in full model simulations [3], and in experiments [12]. In the appendix, we also discuss how the parameters in (5.11) and (5.12) can be easily approximated from system (2.2), giving a means to predict whether system (2.2) can be expected to support synchronized oscillations with unified jump-up for particular parameter values. In the next section, we derive a more general formula for the snake in which we do not assume that all of the cells jump up at the same time on the fast time scale.

5.3. Stability of the fixed point snake of synchrony. In this subsection, we consider stability of the fixed point (5.5) representing the snake of synchrony, from two perspectives. The map given in (5.3) is nonlocal since the term $h(I_d)$ appears explicitly in it for all I, and this complicates stability analysis. We start by considering the class of linear snakes. That is, we restrict ourselves to snakes that satisfy initial conditions of the form

$$h(I,0) \;=\; h(I) \;=\; h_{LK} + \alpha(I_{max} - I) \tag{5.13}$$

at jump-up for an arbitrary parameter α. It is easy to check that a solution of (4.3) and (4.5), with simultaneous jumps between phases, remains linear if it satisfies (5.13) at jump-up. We derive a map on slopes α defined for linear snakes, with a fixed point corresponding to (5.5), and use this to prove that the snake of synchrony is stable within the class of linear snakes. After this, we derive a sufficient condition for this snake of synchrony to be nonlinearly stable, without restriction of solution class.

Because solutions with initial conditions that are linear functions of I remain linear for all time, the initial condition (5.13) evolves after one cycle into another linear function of I, possibly with a different slope. We write this as

$$h(I,T) \;=\; h_{LK} + \pi(\alpha)(I_{max} - I).$$

Here T is the cycle duration. This naturally gives rise to a real map $\alpha \to \pi(\alpha)$. We wish to derive a formula for this map, find the fixed point corresponding to (5.5), and determine its stability.

Recall that $h' = -\rho h$ in the active phase. Consistent with our earlier notation, we fix I_d such that $cell(I_d)$ jumps down first. As previously, let $h_d = h(I_d, 0)$, let h^d_{RK} denote the value of h at the right knee for $I = I_d$, and recall that T_A denotes the time spent by all cells in the active phase. Then (4.5) yields

$$h(I,T_A) = (h_{LK} + \alpha(I_{max} - I))\, e^{-\rho T_A}, \tag{5.14}$$

where

$$e^{-\rho T_A} = \frac{h^d_{RK}}{h_d}. \tag{5.15}$$

As previously, in the silent phase, $h' = -ah - bI + c$, and $\Lambda_I = (c - bI)/a$. Solving this silent phase equation with initial condition (5.14) yields that, for $T_A < \tau < T$,

$$h(I,\tau) = h(I,T_A)e^{a(T_A-\tau)} + \Lambda_I(1 - e^{a(T_A-\tau)})$$

or, letting $T_S = T - T_A$,

$$h(I,T) = (h_{LK} + \alpha(I_{max} - I))\, e^{-\rho T_A} e^{-aT_S} + \Lambda_I(1 - e^{-aT_S}). \tag{5.16}$$

Note that $h(I_{max}, T) = h_{LK}$. Hence, for $\Lambda_M := \Lambda_{I_{max}}$, as previously,

$$h_{LK} = h_{LK}e^{-\rho T_A} e^{-aT_S} + \Lambda_M(1 - e^{-aT_S}). \tag{5.17}$$

We use this last equation along with (5.16) to conclude that

$$h(I,T) \;=\; h_{LK} + \alpha(I_{max} - I)e^{-\rho T_A}e^{-aT_S} + (\Lambda_I - \Lambda_M)(1 - e^{-aT_S})$$

or

$$h(I,T) \;=\; h_{LK} + (I_{max} - I)\left(\alpha e^{-\rho T_A}e^{-aT_S} + \frac{b}{a}(1 - e^{-aT_S})\right).$$

It follows that

$$\pi(\alpha) = \alpha e^{-\rho T_A}e^{-aT_S} + \frac{b}{a}(1 - e^{-aT_S}). \tag{5.18}$$

To rewrite (5.18), note that, from (5.17),

$$e^{-\rho T_A}e^{-aT_S} \;=\; 1 - \frac{\Lambda_M}{h_{LK}}(1 - e^{-aT_S}), \tag{5.19}$$

which implies, after some rearrangement, that

$$\pi(\alpha) = \alpha + \left(\frac{b}{a} - \alpha\frac{\Lambda_M}{h_{LK}}\right)(1 - e^{-aT_S}). \tag{5.20}$$

Finally, solving $\pi(\alpha_0) = \alpha_0$ yields an expression for the fixed point α_0, namely,

$$\frac{b}{a} - \alpha_0\frac{\Lambda_M}{h_{LK}} = 0 \quad \text{or} \quad \alpha_0 = \frac{b}{a}\frac{h_{LK}}{\Lambda_M}. \tag{5.21}$$

Remark 5.4. Note that the slope specified in (5.21) is exactly that of the fixed point snake (5.5), and so our two calculations are consistent. However, the calculation here is not as general as the earlier one in subsection 5.1, which, in theory, allowed for the possibility of nonlinear fixed points.

Next we consider the stability of the fixed point. We differentiate (5.20) to find that

$$\pi'(\alpha) \;=\; 1 - \frac{\Lambda_M}{h_{LK}}(1 - e^{-aT_S}) + \left(\frac{b}{a} - \alpha\frac{\Lambda_M}{h_{LK}}\right)ae^{-aT_S}\frac{dT_S}{d\alpha}. \tag{5.22}$$

The last term is zero for $\alpha = \alpha_0$ because of (5.21). Hence

$$\pi'(\alpha_0) = 1 - \frac{\Lambda_M}{h_{LK}}(1 - e^{-aT_S}). \tag{5.23}$$

In order to make sense of this, we substitute (5.15) into (5.17) to find that

$$h_{LK} \;=\; h_* e^{-aT_S} + \Lambda_M(1 - e^{-aT_S}),$$

where

$$h_* \equiv h_{LK}\frac{h_{RK}^d}{h_d}. \tag{5.24}$$

Hence

$$e^{-aT_S} = \frac{\Lambda_M - h_{LK}}{\Lambda_M - h_*}. \tag{5.25}$$

Apply this in (5.23), and use (5.24) to find

$$\begin{aligned} \pi'(\alpha_0) &= 1 - \tfrac{\Lambda_M}{h_{LK}} \tfrac{(h_{LK} - h_*)}{(\Lambda_M - h_*)} \\ &= 1 - \Lambda_M \left(\frac{1 - \frac{h^d_{RK}}{h_d}}{\Lambda_M - h_{LK}\frac{h^d_{RK}}{h_d}} \right). \end{aligned} \tag{5.26}$$

The last term in (5.26) is positive because $h^d_{RK} < h_d$ and $h_{LK} < \Lambda_M$ from (4.6). Therefore, we wish to prove that it is less than 2. We will, in fact, show that this last term is less than 1, and, therefore, $0 < \pi'(\alpha_0) < 1$. This follows if

$$\Lambda_M - \Lambda_M \frac{h^d_{RK}}{h_d} < \Lambda_M - h_{LK}\frac{h^d_{RK}}{h_d},$$

which is true because $h_{LK} < \Lambda_M$. Thus the snake of synchrony is stable within the class of linear snakes.

Remark 5.5. When $I_d = I_{max}$, the map in (5.3) no longer depends on an unknown $h_d \equiv h(I_d, 0)$ since, by construction, $h(I_{max}, 0) = h_{LK}$. Thus (5.3) becomes linear in h and can be differentiated directly with respect to h. The derivative of the map is exactly the expression given in (5.26). Thus, for $I_d = I_{max}$, the fixed point snake of synchrony is always stable. However, this calculation is not possible for $I_d \neq I_{max}$.

To consider stability without restriction to the class of linear snakes or to $I_d = I_{max}$, we consider perturbations sufficiently small such that they do not change which cells jump down first (i.e., the value of I_d). Clearly such perturbations exist; as an example, recall from section 3 that there is a range of g_{syn} values that gives a snake of synchrony for which cells with $I = I_{max}$ jump down first. For such a snake, a perturbation that retards the other cells slightly and is sufficiently small to preserve synchrony conserves I_d.

Let $h(I)$ denote the snake of synchrony, and let $p(I) = h(I) + \epsilon(I)$ denote a perturbation of $h(I)$ at jump-up. We will measure distance under the supremum norm $||f(I)|| = \sup\{|f(I)| : I \in [I_{min}, I_{max}]\}$. We will derive a condition under which, for $M(h(I))$ defined in (5.3), we have $||M(p) - M(h)|| \leq L||p - h||$ for a constant $0 < L < 1$. Equation (5.3) gives

$$||M(p) - M(h)|| = ||h(I)h^d_{RK}e^{-aT_S}/h_d - p(I)h^d_{RK}e^{-a\tilde{T}_S}/p_d - \Lambda_I(e^{-aT_S} - e^{-a\tilde{T}_S})||,$$

where e^{-aT_S} is given by (5.25), $p_d = p(I_d)$, and $\tilde{T}_S$ is the value of T_S obtained by substituting p_d for h_d in $h_* = h_{LK}h^d_{RK}/h_d$. Some algebraic manipulation yields, for $\epsilon_d = \epsilon(I_d)$,

$$||M(p) - M(h)|| \leq (\Lambda_M - h_{LK}) \left|\left| \frac{p(I)h^d_{RK} - \Lambda_I(h_d + \epsilon_d)}{K - \Lambda_M \epsilon_d} - \frac{h(I)h^d_{RK} - \Lambda_I h_d}{K} \right|\right|, \tag{5.27}$$

where $K = h_{LK}h^d_{RK} - \Lambda_M h_d$ is independent of I. Note that $K < 0$ since $\Lambda_M > h_{LK}$ and $h_d > h^d_{RK}$.

Expand the right-hand side of (5.27) in ϵ_d, and note that $|\epsilon_d| \leq ||\epsilon||$ to obtain

$$||M(p) - M(h)|| \leq \left(\frac{(\Lambda_M - h_{LK})||\epsilon||}{|K|}\right)\left(h^d_{RK} + \left|\left|\frac{\Lambda_M(p\, h^d_{RK} - \Lambda_I h_d) - \Lambda_I K}{K}\right|\right|\right) + O(\epsilon^2).$$

For ϵ sufficiently small, the higher order terms can be neglected, and application of the definition of K yields

$$(5.28) \qquad ||M(p) - M(h)|| \leq \left(\frac{(\Lambda_M - h_{LK})h^d_{RK}||\epsilon||}{|K|}\right)\left(1 + \left|\left|\frac{p\,\Lambda_M - \Lambda_I h_{LK}}{K}\right|\right|\right).$$

Let S denote $\sup_I(p(I)\Lambda_M - \Lambda_I h_{LK})$, the difference of two positive terms. It remains to estimate S. Suppose that $S > 0$. Then the two relations $p(I) \leq \Lambda_I$ and $\Lambda_{I_{min}} = \max_I(\Lambda_I)$, which follow from (5.2), and $\Lambda_M > h_{LK}$ from (4.6) give $S \leq \Lambda_{I_{min}}(\Lambda_M - h_{LK})$. Suppose instead that $S < 0$. For concreteness, we assume that $p(I) \geq h_{LK}$ for all I, which holds for sufficiently small perturbations. Then $|S| \leq h_{LK}(\Lambda_{I_{min}} - \Lambda_M)$. Thus $|S| \leq \max\{\Lambda_{I_{min}}(\Lambda_M - h_{LK}), h_{LK}(\Lambda_{I_{min}} - \Lambda_M)\} \equiv \bar{S}$.

For this $\bar{S}$, (5.28) implies that, if the condition

$$\frac{(\Lambda_M - h_{LK})h^d_{RK}}{|K|}\left(1 + \frac{\bar{S}}{|K|}\right) < 1$$

holds, then the snake of synchrony is stable with respect to sufficiently small perturbations. This condition holds for a variety of numerical examples that we have considered. Note that it certainly holds when h_{LK} lies near the fixed point Λ_M, since K is bounded away from zero as $h_{LK} \to \Lambda_M$, as long as the h-values on the curves of left and right knees remain separate.

6. Nonlinear snakes.

In the previous section, we assumed that the cells jump up and jump down together on the slow time scale. This will be the case if g_{syn}, the synaptic strength, is sufficiently strong. Numerical simulations demonstrate that, for smaller g_{syn}, the jump-up and jump-down processes are more gradual, as shown in Figures 2b and 4. Each cell jumps up or down when it reaches the left or right knee of its effective cubic. In the analysis here, we allow the cells to jump up at different times on the slow time scale, but we still assume that the cells jump down at the same time. We shall derive a nonlinear boundary value problem for the periodic snake of synchrony. An analogous derivation leads to a similar formula if there is a gradual jump-down.

We denote the position of the snake as $h(I, \tau)$. It will be convenient to choose the translation now so that $\tau = 0$ corresponds to the moment when all of the cells jump down. In addition to assuming that all of the cells jump down together, we will further assume that the jump-down process begins when $cell(I_{max})$ reaches its right knee at $h = h_{RK}$, consistent with numerical simulations of the gradual jump-up case (section 3). Let $h_0(I) \equiv h(I, 0)$ denote the corresponding initial position of the snake.

We assume that the first cells to jump up are $cell(I_{max})$. Moreover, the cells jump up in order of decreasing I. As before, let T_S be the time for $cell(I_{max})$ to evolve under (4.5) from h_{RK} up to h_{LK}. We let $h_\mu(I) \equiv h(I, T_S)$ denote the position of the snake when $cell(I_{max})$ jumps up. We then let $\Delta(I)$ denote the delay in the jump-up of $cell(I)$ relative to $cell(I_{max})$.

That is, $\Delta(I_{max}) = 0$, and $cell(I)$ jumps up at $\tau = T_S + \Delta(I)$. We shall derive three equations for the three unknown functions $h_0(I), h_\mu(I)$, and $\Delta(I)$.

First, we consider the cells while all are in the silent phase. The cells then evolve according to (4.5) from their initial position $h_0(I)$ to $h_\mu(I)$ at time T_S. Solving (4.5), we find that

$$h_0(I) = (h_\mu(I) - \Lambda_I)e^{aT_S} + \Lambda_I.$$

Substitution of the formula for e^{aT_S} (see (5.25)) yields the first equation,

$$h_0(I) = (h_\mu(I) - \Lambda_I)\left(\frac{h_{RK} - \Lambda_M}{h_{LK} - \Lambda_M}\right) + \Lambda_I. \tag{6.1}$$

We next follow the cells forward after they jump up. Denote the position of the left knee at which $cell(I)$ jumps up as

$$(v_{LK}^{eff}, h_{LK}^{eff})(I) \;\equiv\; (v_{LK}, h_{LK})(I + I_{syn}^u(I)).$$

Recall that this is found by solving $f + I + I_{syn}^u(I) = 0$, where $I_{syn}^u(I)$ is given by (4.7) and implicitly depends on v, for $v = v_L(h, I)$, and then solving $\partial(f + I_{syn}^u(I))/\partial v = 0$ for $(v_{LK}^{eff}(I), h_{LK}^{eff}(I))$ on this curve. Since cells jump up at different times but jump down together, they spend different amounts of time in the active phase. We again use $h' = -\rho h$ for evolution in the active phase, as indicated in (4.3). We saw earlier that, according to this equation, $cell(I_{max})$ spends time $T_A = (1/\rho)\ln(h_{LK}/h_{RK})$ in the active phase. Each cell with $I < I_{max}$ spends time $T_A - \Delta(I)$ in the active phase, starting from initial condition $h_{LK}^{eff}(I)$ and ending at its jump-down position $h_0(I)$. Using this to solve (4.3) yields the second equation,

$$h_0(I) = h_{LK}^{eff}(I)\left(\frac{h_{RK}}{h_{LK}}\right)e^{\rho\Delta(I)}. \tag{6.2}$$

To obtain the third equation, we follow each $cell(I)$ in the silent phase from $\tau = T_S$ until $cell(I)$ jumps up. First, we need to introduce some notation. Let $I_{syn}(\tau)$ denote the synaptic input at time τ. We earlier let $I_{syn}^u(I)$ be the synaptic input when $cell(I)$ jumps up. Since $cell(I)$ jumps up when $\tau = T_S + \Delta(I)$, it follows that $I_{syn}^u(I) = I_{syn}(T_S + \Delta(I))$, where $I_{syn}^u(I)$ is given by (4.7) with $v(I) = v_{LK}^{eff}(I)$.

Now each $cell(I)$ satisfies (4.5) with $I_{tot} = I + I_{syn}(\tau)$. Moreover, $h(I, T_S) = h_\mu(I)$, and $cell(I)$ jumps up at $h_{LK}^{eff}(I)$, the left knee of its effective cubic, when $\tau = T_S + \Delta(I)$. Solving (4.5) with these boundary conditions, we find that

$$h_{LK}^{eff}(I) = (h_\mu(I) - \Lambda_I)e^{-a\Delta(I)} + \Lambda_I - be^{-aT_S}e^{-a\Delta(I)}\int_{T_S}^{T_S+\Delta(I)} I_{syn}(\tau)e^{a\tau}\,d\tau. \tag{6.3}$$

Equations (6.1), (6.2), and (6.3) constitute a system of three equations in the three unknowns $h_0(I), h_\mu(I)$, and $\Delta(I)$. From (6.2), $\Delta(I)$ can be expressed as a function of $h_0(I)$. Equations (6.1) and (6.3) are easily converted into formulas for $h_\mu(I)$. Equating these formulas gives, after minor rearrangement,

$$h_0(I) = \Lambda_I + \Gamma\left((h_{LK}^{eff}(I) - \Lambda_I)e^{a\Delta(I)} + be^{-aT_S}\int_{T_S}^{T_S+\Delta(I)} I_{syn}(\tau)e^{a\tau}\,d\tau\right), \tag{6.4}$$

where $\Delta(I)$ is now a function of $h_0(I)$ and where $\Gamma = (h_{RK} - \Lambda_M)/(h_{LK} - \Lambda_M)$.

To solve (6.4) for $h_0(I)$, we first differentiate with respect to I to obtain

$$\begin{aligned} h_0'(I) = \Lambda_I' &+ \Gamma((h_{LK}^{eff})'(I) - \Lambda_I')e^{a\Delta(I)} \\ &+ a\Gamma(h_{LK}^{eff}(I) - \Lambda_I)\Delta'(I)e^{a\Delta(I)} \\ &+ b\Gamma I_{syn}(T_S + \Delta(I))\Delta'(I)e^{a\Delta(I)}. \end{aligned} \tag{6.5}$$

Recall that $I_{syn}(T_S + \Delta(I)) = I_{syn}^u(I)$, where $I_{syn}^u(I)$ is given by (4.7) with $v(I) = v_{LK}^{eff}(I)$. Moreover, we can use (6.2) to solve for $\Delta(I)$ and $\Delta'(I)$. In fact, if we let

$$\Phi(I) \;\equiv\; e^{\rho\Delta(I)} \;=\; \left(\frac{h_{LK}}{h_{RK}}\right)\frac{h_0(I)}{h_{LK}^{eff}(I)}, \tag{6.6}$$

then a straightforward calculation starting from (6.2) demonstrates that

$$\Delta'(I)e^{a\Delta(I)} \;=\; \frac{1}{\rho}\Phi'(I)\Phi^{(a-\rho)/\rho}.$$

Substituting this into (6.5), we find that

$$\begin{aligned} h_0'(I) = \Lambda_I' &+ \Gamma((h_{LK}^{eff})'(I) - \Lambda_I')\Phi^{a/\rho}(I) \\ &+ \frac{a}{\rho}\Gamma(h_{LK}^{eff}(I) - \Lambda_I)\Phi'(I)\Phi^{(a-\rho)/\rho}(I) \\ &+ \frac{b}{\rho}\Gamma I_{syn}^u(I)\Phi'(I)\Phi^{(a-\rho)/\rho}. \end{aligned} \tag{6.7}$$

We can use (6.6) to write $\Phi(I)$ and $\Phi'(I)$ in terms of $h_0(I)$ and $h_0'(I)$. Equation (6.7) then gives an ordinary differential equation for $h_0(I)$, the initial (jump-down) position of the periodic snake. Note that the solution must satisfy the boundary condition $h(I_{max}) = h_{RK}$. Solving the resulting boundary value problem numerically with XPPAUT gives an estimate of $h_0(I)$. We can compute $h_\mu(I)$, the position of the snake when $cell(I_{max})$ is ready to jump up, directly from this, using (6.1). Figure 9 compares the resulting $h_\mu(I)$ (dotted curve) with the snake position from a full numerical simulation of the pre-BötC model network (solid curve) and with the linear snake formula (5.5), all for $g_{syn}/20 = .009$, which corresponds to gradual jump-up (see Figures 2b and 4). For the nonlinear and linear curves, parameters were estimated from the pre-BötC model without simulations, using the methods discussed below in the appendix. Because it takes gradual jump-up into account, the nonlinear result gives a better approximation of the snake position than does the linear formula, in the gradual jump-up case.

Remark 6.1. We can perform analogous calculations to derive an ordinary differential equation boundary value problem for the jump-up snake of synchrony $h_\mu(I)$ in the case of gradual jump-down and simultaneous jump-up. The resulting ordinary differential equation analogous to (6.7) is simpler because there is one fewer unknown: $h_\mu(I)$ and $h_{LK}^{eff}(I)$ collapse to the same curve. The ordinary differential equation in that case is explicitly nonlocal, however, in that Γ depends on $h_\mu(I_{min})$, such that the right-hand side depends on I_{min} for all I; nonetheless, it can be solved numerically with XPPAUT without a problem.

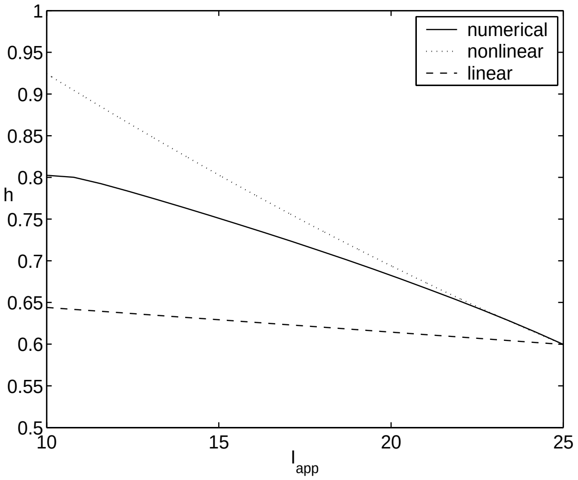


Figure 9. *Snake of synchrony at the moment when leading cells jump up. The plot shows curves for* $g_{syn}/20 = .009$*, corresponding to gradual jump-up.*

7. Loss of synchrony.

7.1. Break-up conditions. The analysis in section 5 shows that there is only one possible fixed point snake for the case in which all cells jump up and jump down together. Suppose that condition (4.8) fails for some I values, so that this solution does not exist and not all cells jump up together. In section 6, we consider another form of synchronized oscillation in which cells jump up more gradually until all are in the active phase. In this subsection, we find a sufficient condition for break-up of the synchronized solution by computing a condition under which not all cells reach jump-up before the leading cells, with $I = I_{max}$, jump down from the active phase. This computation can also lead to an estimate for the I value at which the snake will break.

First, we assume that cells with $I = I_{max}$ jump down first on every oscillation cycle. For fixed intrinsic cellular parameters, this corresponds to taking g_{syn} small enough. This is quite natural for the consideration of loss of synchrony since a large synaptic coupling strength promotes synchronization.

We start at time $\tau = 0$ with cells at $h(I, 0)$, a position for which cells with $I = I_{max}$ are about to jump up. To derive a break-up condition, we now compute a condition under which not all cells can evolve in the silent phase from $h(I, 0)$ to their effective knees before cells with $I = I_{max}$ jump down. Since we aim for a sufficient condition for break-up, we assume the fastest possible silent phase evolution corresponding to $I_{syn} = 0$. Specifically, we solve

$$\begin{aligned} h' &= -ah - bI + c, \\ h(0) &= h(I, 0) \end{aligned} \tag{7.1}$$

for $h(I, \tau)$, up until time $T_A = \frac{1}{\rho} \ln \frac{h_{LK}}{h_{RK}}$, when the active cells jump down; this easily can be done analytically. The break-up condition is that, for some $I_b \in (I_{min}, I_{max})$,

$$h(I_b, T_A) = h_{LK}(I_b + I^u_{syn}(I_b)), \tag{7.2}$$

where $h_{LK}(I_b+I^u_{syn}(I_b))$ denotes the effective knee for $I = I_b$. If we restrict ourselves to snake configurations for which cells jump up in order of decreasing I, then all cells with $I < I_b$ fail to jump up at all during the oscillation; see Figure 10.

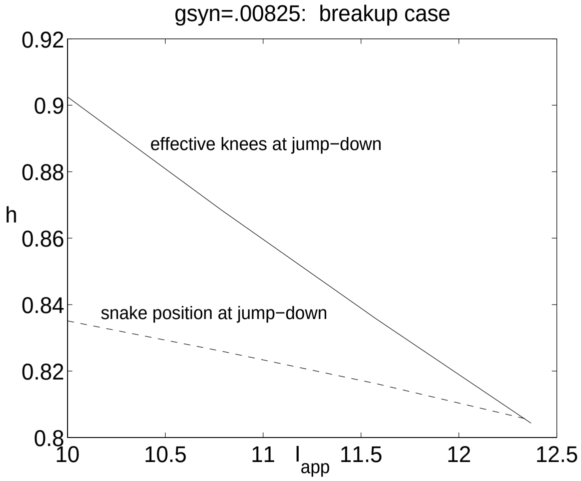


Figure 10. *Numerical illustration of the break-up of a snake. The network was simulated with* $g_{syn}/20 = .00825$*, as in Figures 5a and 6 in section 3. The cells gradually jump up, until cell(*$I_{max} = 25$*) jump down, after which no other cells jump up. The plot shows the silent phase positions of the cells that fail to jump up, at the moment when cell(*I_{max}*) jumps down, as a function of* I_{app}*. Note that the nonjumping cells lie below their numerically computed effective left knees, with cell position intersecting the effective knee curve at the break point of the snake.*

Equation (7.2) can allow us to solve for the I_b at which a break occurs (if such an I value exists). If we wish just to check whether or not a break occurs, again assuming jump-up in order of decreasing I, we simply need to compare $h(I_{min}, T_A)$ to $h_{LK}(I_{min}+I^u_{syn}(I_{min}))$. That is, break-up occurs for some $I \geq I_{min}$ if

$$h(I_{min}, T_A) < h_{LK}(I_{min} + I^u_{syn}(I_{min})). \tag{7.3}$$

As an example, we can compute $h(I_{min}, T_A)$ for the fixed point snake configuration derived earlier by solving (7.1), with $h(0) = h_{LK}(c-bI_{min})/(c-bI_{max})$ from (5.5), for time T_A, which yields

$$h(I_{min}, T_A) = \Lambda_{I_{min}} \left[1 + \left(\frac{h_{RK}}{h_{LK}}\right)^{a/\rho} \left(\frac{h_{LK}}{\Lambda_M} - 1\right)\right] < \Lambda_{I_{min}}.$$

Since I_{min} is the minimum value of I_{app}, for most I_{app} distributions, (4.7) gives $I^u_{syn}(I_{min}) \approx g_{syn}(v_{syn}-v(I_{min}))$. From the appendix, we thus have $I^u_{syn}(I_{min}) \approx g_{syn}(v_{syn}-(I_{min}/g_L+v_L))$. Hence all of the parameters needed to check inequality (7.3) can be estimated, as discussed in the appendix.

8. Discussion. We have considered a reduced model for a network of conditional pacemaker cells in the pre-BötC of the brain stem. In this model, each cell is represented by a pair of ordinary differential equations, with excitatory synaptic coupling between the cells. The network is heterogeneous in that cells take values of a parameter I_{app} from a distribution.

Heterogeneities in I_{app} represent different levels of applied current (including sustained excitatory inputs from other brain regions) and different leak reversal potentials among different cells.

We analyze this system in the continuum limit for which the coupling terms become integrals. We consider oscillatory solutions for which each cell undergoes sustained silent and active phases, with occasional rapid jumps between the two, but we do not consider individual spikes within the active phases. This leads to a natural further reduction of the model, such that each cell is governed by the scalar equations (4.5) and (4.3) for the evolution of the slow variable h in the silent and active phases, respectively.

With these simplifications, we treat two forms of periodic, synchronized solutions, or "snakes": solutions for which all cells jump up and down simultaneously (on a slow time scale), and solutions which feature gradual jumps. Indeed, a key point here is that heterogeneous networks of synaptically coupled oscillating cells can support periodic, synchronized solutions with a continuum of phase shifts between jump times of different cells. This allows for a richer range of dynamics than has been observed for a homogeneous network of relaxation oscillators [16] or a network of oscillators with dynamics and coupling based on phases [22, 15, 9, 1].

We provide natural geometric conditions for the case of simultaneous jumps to occur. When these hold, we prove the existence of a unique periodic snake, characterized by a simple formula. This formula does not depend on any of the parameters associated with the active phase or synaptic coupling; however, these do appear in the conditions for simultaneous jumping. We also show that this snake is linearly stable to certain small perturbations, and we provide a general nonlinear stability condition. We note that our analysis of simultaneous jump solutions generalizes immediately to any finite population of oscillators; the continuum limit is not required here. In particular, with a finite number of oscillators, the effective left knee for cells with fixed I is determined by computing the synaptic input that results when all cells with larger I jump up, now based on the discrete synaptic current (2.3) rather than the continuum current (2.4).

When the synaptic coupling strength in the network is weakened, the conditions for simultaneous jumping may fail. In the case of gradual jumps up from the silent phase to the active phase, we derive a single nonlinear ordinary differential equation boundary value problem for the position of the periodic snake in phase space at a certain stage in its oscillatory cycle. This generalizes naturally to gradual jumps down or gradual jumps in both directions.

For our analysis, although we require that each cell have a cubic v-nullcline when uncoupled, we do not require that all cells be intrinsic oscillators when uncoupled, which is consistent with experimental observations. Thus our analysis illustrates how heterogeneities in a network can lead to robust, stable oscillations by allowing intrinsically active cells to recruit silent cells via synaptic coupling. We emphasize that such solutions can be periodic and synchronized in the sense that all cells begin and end their active phases together in time despite significant heterogeneities. In the synchronized solutions that we consider, cells may jump up to the active phase by reaching a curve of knees, or saddle-node bifurcation points of a fast subsystem. Alternately, synaptic coupling may cause cells to suddenly lie above this curve and therefore to jump up immediately. Based on other studies of bursting (see [16]) and our own simulations, we expect that the inclusion of spiking currents will not qualitatively affect the relevant bifurcations and the corresponding knees for jump-up. Since the dynamics of the jump-up appears to be the key determinant of how well the network synchronizes, we therefore posit our results as an explanation for how heterogeneities enhance the tendency for a network of bursting cells, such as the pre-BötC, to fire synchronized bursts.

For all of our analysis, we assume all-to-all coupling. In the all-to-all case, all cells receive the same amount of coupling, depending only on the proportion of cells in the population that are active. Simulations in [3] yielded qualitatively similar network behavior for full and sparse network connectivities. Other coupling architectures would complicate analysis, however, especially if nonlocal or random connections were included.

In addition to analysis, we present numerical simulations to illustrate our results. These include comparisons of snake positions from full network simulations, performed with a discrete population of 20 cells with a uniform distribution of I_{app}, to snake positions computed from our analysis. To generate the latter, we estimate a variety of parameters, in particular those appearing in (5.5), (6.6), and (6.7), directly from the network equations (2.2); this can be done quite easily, as discussed in the appendix. Perhaps the most useful numerical results presented are animations of full network simulations. These show cells' positions in phase space, along with relevant nullclines and curves of knees. The nullclines and knees move as synaptic coupling strength changes over the course of a simulation. This allows for very clear visualization of the jumping behavior of individual cells, highlighting its dependence on the heterogeneity parameter I_{app}. We anticipate that such animations will be useful for a wide variety of studies of systems of coupled oscillators.

As g_{syn} drops still farther from the gradual jumping case, synchrony may in fact be lost. We provide a sufficient geometric condition for synchrony to fail. When synchrony breaks down, interesting periodic or possibly chaotic solutions can arise; we have illustrated two of these numerically. In general, the population's activity pattern for fixed parameter values can be classified according to the population of cells that become active on each cycle. Following [1], we can distinguish between locking, in which all cells fire on each cycle, partial locking, in which all cells fire but some cells skip some cycles, partial death, in which some cells never fire, and death, in which no cells fire. In simulations, we find that, for uniform distributions of I_{app}, with fixed mean I_{app} but different distribution widths γ, certain general trends emerge. For fixed γ, as coupling strength g_{syn} increases, the tendency to lock increases. Correspondingly, the population undergoes transitions from partial death, to partial locking, to locking as g_{syn} increases. Larger g_{syn} is required for more unified activity with larger γ, corresponding to a broader distribution of I_{app}; thus the partial locking and partial death regions form positively sloped bands in (γ, g_{syn}) parameter space. As the system switches from partial death to partial locking, there is a decrease in the variance across cells, in terms of the proportion of cycles during which each cell is active, until finally no variance remains in the locked state. Nonuniform distributions of I_{app} are expected to yield qualitatively similar trends. Work to analyze asynchronous solutions is in progress.

The heterogeneities that we consider are restricted to the parameter I_{app}, which includes heterogeneities both in the leak current reversal potential v_L and in applied current. Earlier studies have suggested that the biological effects of heterogeneities in pre-BötC cells are captured by heterogeneities in I_{app} (defined to include v_L) and g_{Na} [2]. The former, which we have considered here, is perhaps more relevant to ongoing biological experimentation because the leak reversal potential can be controlled by manipulation of potassium ion concentration in an experimental preparation [6, 7]. Nonetheless, the role of g_{Na} should be explored to gain a full understanding of pre-BötC behavior.

9. Appendix.

9.1. Model equations and parameter values.

In system (2.2), we have, for $x = m$ or h,

$$x_\infty(v) = 1/(1 + \exp((v - \theta_x)/\sigma_x)) \text{ and } \tau_h(v) = (\epsilon \cosh((v - \theta_h)/2\sigma_h))^{-1}.$$

When we simulate a discrete population of 20 cells, we take

$$I_{syn} = g_{syn} \left(\sum_{k=1}^{20} s_\infty(v_k) \right) (v_{syn} - v_j)$$

as the synaptic input to cell j (that is, g_{syn} is already scaled to take into account the population size, since we always use 20 cells); $v_{syn} > v_j$ for solutions v_j, for the parameter values used. We further take $s_\infty(v_k) = 1/(1 + \exp((v_k - \theta_s)/\sigma_s))$.

Parameter values used in simulations are given in Table 1, with units omitted. The parameter g_{syn} is varied, as indicated in figure captions. We use 20 cells, and I ranges over 20 equally spaced values, starting with $I_{min} = 10$ and ending with $I_{max} = 25$. For these parameters, there is a transition from excitable to oscillatory at around $I = 12.5$, such that cells with $I < 12.5$ will converge to a rest point in the silent phase without synaptic input. Thus we have cells that intrinsically are silent and cells that intrinsically are oscillators represented in our simulations. Note that the value of C_m used here is smaller than that in [2, 3, 6]. This accentuates the relaxation aspect of the oscillations that we study.

Table 1

Basic set of parameter values for the reduced pre-BötC cell model.

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
g_{Na}	2.8	v_{Na}	50	θ_m	-37	σ_m	-6
				θ_h	-44	σ_h	6
g_L	2.8	v_L	-65				
		v_{syn}	0	θ_s	-43	σ_s	-0.1
C_m	0.21	ϵ	0.01				

9.2. Estimation of parameters for numerics.

To generate the snake position numerically, given (5.5), we need only to estimate the parameters h_{LK}, b, c. Let the point (v_{LK}, h_{LK}) denote the solution to the two equations $F(v, h) = 0$ and $F_v(v, h) = 0$, where $F(v, h)$ denotes the right-hand side of the v-equation in (2.2) for $I_{syn} = 0$ and $I_{app} = I_{max}$. These easily can be solved numerically. In particular, they can be solved dynamically through the following procedure, which was used to generate the knee positions for all I_{app} in our animations. First, note that $F(v, h) = 0$ can be solved algebraically for $h(v)$. Next, let $y' = \phi F_v(v, h(v))$ for a large constant ϕ designed to speed up the convergence of y to v_{LK}. Once y converges sufficiently close to a steady value, then we denote this by v_{LK}, and we read out $h_{LK} = h(v_{LK})$.

For fixed I_{app}, we can estimate the value of v for a solution of (2.2) in the silent phase when $I_{syn} = 0$. To do this, we note that $m_\infty(v) \approx 0$ in the silent phase, such that $C_m v' \approx$

$-g_L(v - v_L) + I_{app}$. But the silent phase is defined by $v' = 0$. This yields

$$v \approx v_{sil}(I_{app}) := I_{app}/g_L + v_L, \tag{9.1}$$

which we use below. Now $\tau_h(v)$ tends to an asymptotic value τ_h^- as $v \to -\infty$, and $v_{sil}(I_{app})$ is sufficiently negative for parameters considered such that we can take $\tau_h(v) \approx \tau_h^-$ in the silent phase. Based on (2.4), (4.3), and (4.4), we thus approximate $a \approx 1/\tau_h^-$.

To estimate b and c, note that the function $h_\infty(v)$ is sigmoidal, with horizontal asymptotes at 1 as $v \to -\infty$ and at 0 as $v \to \infty$. Over the transitional region between these asymptotes, $h_\infty(v)$ is approximately linear. Further, since cells with $I = I_{max}$ are oscillatory, it is likely that the v values of cells in the silent phase lie in this transitional region. Thus we derive a least squares linear estimate of $h_\infty(v)$ over this region. Using $v \approx I/g_L + v_L$ in the silent phase from (9.1) converts the approximation of $h_\infty(v)$ into a linear function of I, namely, $BI + C$, which we substitute into (4.4). With this substitution, a comparison of (4.4) and (4.5) yields $b = aB$ and $c = aC$.

For simulation of the nonlinear boundary value problem (6.7), with $h(I_{max}) = h_{RK}$, several additional parameters are needed. We can approximate ρ, which appears in (4.3), as $\rho \approx 1/\tau_h^+$, where τ_h^+ denotes the positive asymptotic value of $\tau_h(v)$. Additionally, we require expressions for $h_{RK}, h_{LK}^{eff}(I)$, and $I_{syn}^u(I)$. The former two can be solved for dynamically at discrete I values (with $I = I_{max}$ for h_{RK}), analogously to the estimation of h_{LK} described above, and the results for $h_{LK}^{eff}(I)$ can be interpolated. A specific form must be assumed for $I_{syn}^u(I)$. Suppose that the I values in the network are distributed uniformly over $[I_{min}, I_{max}]$ and that cells jump down in order of decreasing I, as was the case in our simulations. Then, for $cell(I)$ in the silent phase, $I_{syn}^u(I) \approx g_{syn}(v_{syn} - v_{sil}(I))(I_{max} - I)/(I_{max} - I_{min})$.

Acknowledgments. The authors thank Rob Butera for introducing them to this problem and for input on simulations, and they thank John Rinzel for many helpful discussions.

REFERENCES

[1] J. T. Ariaratnam and S. H. Strogatz, *Phase diagram for the Winfree model of coupled nonlinear oscillators*, Phys. Rev. Lett., 86 (2001), pp. 4278–4281.

[2] R. Butera, J. Rinzel, and J. Smith, *Models of respiratory rhythm generation in the pre-Bötzinger complex.* I. *Bursting pacemaker neurons*, J. Neurophysiology, 81 (1999), pp. 382–397.

[3] R. Butera, J. Rinzel, and J. Smith, *Models of respiratory rhythm generation in the pre-Bötzinger complex.* II. *Populations of coupled pacemaker neurons*, J. Neurophysiology, 81 (1999), pp. 398–415.

[4] C. C. Chow, *Phaselocking in weakly heterogeneous neuronal networks*, Phys. D, 118 (1998), pp. 343–370.

[5] G. de Vries and A. Sherman, *From spikers to bursters via coupling: Help from heterogeneity*, Bull. Math. Biol., 63 (2002), pp. 371–391.

[6] C. DelNegro, S. Johnson, R. Butera, and J. Smith, *Models of respiratory rhythm generation in the pre-Bötzinger complex.* III. *Experimental tests of model predictions*, J. Neurophysiology, 86 (2001), pp. 59–74.

[7] C. DelNegro, C. Wilson, R. Butera, H. Rigatto, and J. Smith, *Periodicity, mixed-mode oscillations, and quasiperiodicity in a rhythm-generating neural network*, Biophys. J., 82 (2002), pp. 206–214.

[8] G. B. Ermentrout, *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*, Software Environ. Tools 14, SIAM, Philadelphia, 2002.

[9] G. B. Ermentrout and N. Kopell, *Frequency plateaus in a chain of weakly coupled oscillators*, SIAM J. Math. Anal., 15 (1984), pp. 215–237.

[10] D. Golomb and J. Rinzel, *Dynamics of globally coupled inhibitory neurons with heterogeneity*, Phys. Rev. E (3), 48 (1993), pp. 4810–4814.

[11] D. Golomb, X.-J. Wang, and J. Rinzel, *Synchronization properties of spindle oscillations in a thalamic reticular nucleus model*, J. Neurophysiology, 72 (1994), pp. 1109–1126.

[12] S. Johnson, J. Smith, G. Funk, and J. Feldman, *Pacemaker behavior of respiratory neurons in medullary slices from neonatal rat*, J. Neurophysiology, 72 (1994), pp. 2598–2608.

[13] N. Kopell and G. B. Ermentrout, *Mechanisms of phase-locking and frequency control in pairs of coupled neural oscillators*, in Handbook of Dynamical Systems, Vol. 2: Towards Applications, B. Fiedler ed., North–Holland, Amsterdam, 2002, pp. 3–54.

[14] N. Koshiya and J. Smith, *Neuronal pacemaker for breathing visualized in vitro*, Nature, 400 (1999), pp. 360–363.

[15] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer-Verlag, Berlin, 1984.

[16] J. Rubin and D. Terman, *Geometric singular perturbation analysis of neuronal dynamics*, in Handbook of Dynamical Systems, Vol. 2: Towards Applications, B. Fiedler ed., North–Holland, Amsterdam, 2002, pp. 93–146.

[17] J. Smith, H. Ellengerger, K. Ballanyi, D. Richter, and J. Feldman, *Pre-Bötzinger complex: A brainstem region that may generate respiratory rhythm in mammals*, Science, 254 (1991), pp. 726–729.

[18] D. Somers and N. Kopell, *Rapid synchronization through fast threshold modulation*, Biol. Cybern., 68 (1993), pp. 393–407.

[19] D. Somers and N. Kopell, *Waves and synchrony in networks of oscillators of relaxation and non-relaxation type*, Phys. D, 89 (1995), pp. 169–183.

[20] D. Terman and D. L. Wang, *Global competition and local cooperation in a network of neural oscillators*, Phys. D, 81 (1995), pp. 148–176.

[21] J. White, C. C. Chow, J. Ritt, C. Soto, and N. Kopell, *Synchronization and oscillatory dynamics in heterogeneous, mutually inhibited neurons*, J. Comput. Neurosci., 5 (1998), pp. 5–16.

[22] A. Winfree, *Biological rhythms and the behavior of coupled oscillators*, J. Theoret. Biol., 16 (1967), pp. 15–42.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 2, pp. 175–189

Chaotic Synchronization in Coupled Map Lattices with Periodic Boundary Conditions[*]

Wen-Wei Lin[†] and Yi-Qian Wang[‡]

Abstract. In this paper, we consider a lattice of the coupled logistic map with periodic boundary conditions. We prove that synchronization occurs in the one-dimensional lattice with lattice size $n = 4$ for any γ in the chaotic regime $[\gamma_\infty \approx 3.57, 4]$. It is worthwhile to emphasize that, despite of the fact that there is a rigorous proof for synchronization in many systems with continuous time, almost nothing is rigorously proved for the systems with discrete time.

Key words. synchronization, coupled map lattices, chaos, Liapunov method

AMS subject classifications. 37N35, 39A10, 39A11

PII. S1111111101395410

1. Introduction.

Synchronization is a fundamental phenomenon in physical systems with dissipation. Experimental observations show that subsystems manifest similar behavior in time, provided they are coupled with a dissipative coupling. If the behavior is periodic, then synchronization means matching frequencies and/or phases of signals generated by interacting oscillatory subsystems. Synchronization of periodic oscillations has been well studied and has many practical applications. However, if individual oscillations are chaotic, then the mathematical observations of "synchronized behavior" should be treated differently. Thus the problem of coming up with a rigorous description of the synchronized chaotic behavior of coupled subsystems appears to be attractive and important from both theoretical and application points of view; see, for example, Heagy, Carroll, and Pecora [5], Pecora et al. [8], [9], Vohra et al. [10], Cuomo and Oppenheim [2], [3], Wu and Chua [11], and the references therein. All results mentioned above considered synchronized chaos in continuous systems.

In 1999, Lin, Peng, and Wang [6] first considered synchronized chaos in discrete systems. More precisely, they studied the synchronized chaotic behavior of the popular model in coupled map lattices (CMLs) defined as follows: (1) $1D$ lattice for $1 \le i \le n$,

$$x_i(k+1) = f(x_i(k)) + c(f(x_{i-1}(k)) + f(x_{i+1}(k)) - 2f(x_i(k))) \tag{1.1}$$

with periodic boundary conditions $f(x_0(k)) = f(x_n(k))$ and $f(x_{n+1}(k)) = f(x_1(k))$, and (2)

[*] Received by the editors September 19, 2001; accepted for publication (in revised form) by A. Hagberg June 18, 2002; published electronically September 11, 2002.

http://www.siam.org/journals/siads/1-2/39541.html

[†] Department of Mathematics, National Tsing Hua University, Hsinchu, 300, Taiwan (wwlin@am.nthu.edu.tw).

[‡] Institute of Mathematics, Academia Sinica, Beijing, 100080, People's Republic of China. Current address: National Center for Theoretical Sciences, Mathematics Division, National Tsing Hua University, Hsinchu, 300, Taiwan (yqwang@math.cts.nthu.edu.tw).

$2D$ lattice for $i = (i_1, i_2)$ with $1 \le i_1, i_2 \le n$,

$$\begin{aligned} x_i(k+1) = f(x_i(k)) + c(&f(x_{i_1+1,i_2}(k)) + f(x_{i_1-1,i_2}(k)) + f(x_{i_1,i_2+1}(k)) \\ &+ f(x_{i_1,i_2-1}(k)) - 4f(x_i(k))) \end{aligned} \tag{1.2}$$

with $f(x_{0,i_2}(k)) = f(x_{n,i_2}(k)), f(x_{n+1,i_2}(k)) = f(x_{1,i_2}(k))$, $f(x_{i_1,0}(k)) = f(x_{i_1,n}(k))$ and $f(x_{i_1,n+1}(k)) = f(x_{i_1,1}(k))$, where f is a one-dimensional logistic map $x(k+1) = f(x(k)) = \gamma x(k)(1 - x(k))$ with $f : (0,1) \to (0,1)$ and $\gamma \in [\gamma_\infty \approx 3.57, 4]$. It is well known (see Gleick [4], Campbell [1], etc.) that the map f becomes chaotic whenever γ increases from 3.57 to 4, except that γ is at a very narrow interval of periodic windows near $3.63, 3.73$, or 3.83.

The simplest type of synchronization of CMLs in (1.1) or (1.2) occurs in stable spatially homogeneous regimes corresponding to the existence of attractive spatially homogeneous solutions. In other words, in such cases, there is a large (open) set of initial conditions such that a solution starting from an initial condition in the set becomes spatially homogeneous as discrete time k becomes very large; i.e., the coordinates of the individual maps become almost equal to each other (and are equal as $k \to \infty$). In established regimes, individual maps become indistinguishable, and we observe exact perfect synchronization. Thus it may occur that suitable coupling strength permits the existence of a spatially homogeneous solution provided all individual maps are identical.

In [6], they provided a complete numerical description of synchronization of $1D$ and $2D$ lattices with various lattice sizes. In other words, they found the region for γ and the corresponding coupling strengths c such that the synchronization occurs in corresponding CMLs. Moreover, for the first time, they gave a rigorous proof for synchronization in the case of $1D$ CMLs with lattice size $n = 2, 3$ for $\gamma \in [\gamma_\infty, 4]$ and with lattice size $n = 4$ for $\gamma \in [\gamma_\infty, 3.82]$ in the chaotic regime $[\gamma_\infty, 4]$.

However, it seems difficult to use the method of [6] to prove the same results for $\gamma \in [3.82, 4]$ in $1D$ CMLs with size $n = 4$, which is the gap between theoretical proof and numerical result.

In this paper, we will prove rigorously by the Liapunov method that synchronization occurs in the case of $1D$ CMLs with size $n = 4$ for every γ in the chaotic regime $[\gamma_\infty, 4]$. More precisely, we will prove the following theorem.

Theorem 1.1. *For any initial values $x_i(0) \in (0,1), i = 1,2,3,4,$ and every parameter $\gamma \in [\gamma_\infty, 4]$ of the logistic map in the CML of (1.1)$(n = 4)$, there exists $\delta = \delta(\gamma) > 0$, which is independent of initial values $x_i(0)(i = 1,2,3,4)$, such that*

$$\lim_{k\to\infty} |x_i(k) - x_j(k)| = 0$$

for $c \in (\frac{1}{3} - \delta(\gamma), \frac{1}{3} + \delta(\gamma))$ and $i, j = 1,2,3,4$.

Remark 1.2. *From [6], we know that all $x_i(k), i = 1,2,3,4,$ of (1.1) will lie in $[\frac{4-\gamma}{4}, \frac{\gamma}{4}]$ if k is large enough. So, without loss of generality, from now on we will always assume that $x_i(k)$ lie in this interval, $i = 1,2,3,4, k \ge 0$.*

Remark 1.3. *Although, for technical reasons, it is not easy to generalize the present results to $1D$ lattices with a larger size $(n \ge 5)$ as well as $2D$ lattices, our idea of the Liapunov method still provides a probable direction, and we will dwell on this in a future paper.*

Remark 1.4. *By the result of [6], we need only prove Theorem 1.1 for $\gamma \in [3.82, 4]$.*

This paper is organized into four sections. In section 2, we will describe our ideas using a simple case so that the mechanism of the Liapunov method can be easily understood. Then we will prove some important lemmas in section 3. The proof of Theorem 1.1 can be found in the last section.

For the sake of notation, we rewrite the iteration in the lattice of (1.1) for $n = 4$ by replacing $x_i(k+1)$ and $x_i(k)$ by $\bar{x}_i$ and x_i, respectively:

$$\begin{aligned}\bar{x}_1 &= f(x_1) + c(f(x_2) + f(x_4) - 2f(x_1)),\\ \bar{x}_2 &= f(x_2) + c(f(x_1) + f(x_3) - 2f(x_2)),\\ \bar{x}_3 &= f(x_3) + c(f(x_2) + f(x_4) - 2f(x_3)),\\ \bar{x}_4 &= f(x_4) + c(f(x_1) + f(x_3) - 2f(x_4)),\end{aligned} \tag{1.3}$$

where $f(x) = \gamma x(1-x)$. As mentioned in Remark 1.4, the parameter here is considered in the interesting range $[3.82, 4]$.

2. Proof for a simple case.

In this section, we shall prove that Theorem 1.1 holds using the basic Liapunov method for the simple case $x_1(0) = x_3(0)$ and $x_2(0) = x_4(0)$. Moreover, the proof for the general case can be reduced to the proof for the simple case by some technical lemmas which will be given in the next two sections. Under the conditions $x_1(0) = x_3(0), x_2(0) = x_4(0)$, one can easily check that $x_1(k) = x_3(k), x_2(k) = x_4(k)$ for all $k \geq 0$. Thus (1.1) becomes

$$\bar{x}_1 = f(x_1) + 2c(f(x_2) - f(x_1)), \quad \bar{x}_2 = f(x_2) + 2c(f(x_1) - f(x_2)). \tag{2.1}$$

Consequently, we have

$$\begin{aligned}\bar{x}_1 + \bar{x}_2 &= \quad f(x_1) + f(x_2),\\ \bar{x}_1 - \bar{x}_2 &= (1-4c)(f(x_1) - f(x_2)) = (1-4c)\gamma(1 - x_1 - x_2)(x_1 - x_2).\end{aligned} \tag{2.2}$$

We now define the Liapunov function for (2.1) as

$$L(x_1, x_2) = \frac{(x_1 - x_2)^2}{(x_1 + x_2)(2 - (x_1 + x_2))}, \tag{2.3}$$

where $x_1, x_2 \in [\frac{4-\gamma}{4}, \frac{\gamma}{4}]$ for the reason mentioned in Remark 1.2.

In the remainder of this paper, we will denote $f(x_i)$ by $f_i, i = 1, 2, 3, 4$, and we will denote $x_i + x_j, \bar{x}_i + \bar{x}_j, f_i + f_j, \bar{f}_i + \bar{f}_j$ by $x_{ij}, \bar{x}_{ij}, f_{ij}, \bar{f}_{ij}$ for $i, j = 1, 2, 3, 4$, respectively.

From (2.2), we obtain

$$\begin{aligned}L(\bar{x}_1, \bar{x}_2) &= \frac{(\bar{x}_1 - \bar{x}_2)^2}{\bar{x}_{12}(2 - \bar{x}_{12})} = \frac{(1-4c)^2(f(x_1) - f(x_2))^2}{f_{12}(2 - f_{12})}\\ &= \frac{(1-4c)^2\gamma^2(1 - x_{12})^2(x_1 - x_2)^2}{\gamma[x_{12} - (x_1^2 + x_2^2)](2 - \gamma[x_{12} - (x_1^2 + x_2^2)])}.\end{aligned} \tag{2.4}$$

Lemma 2.1. *For each $\gamma \in [3.82,\ 4]$ in (2.1), there exist numbers $\delta(\gamma), \lambda(\gamma) \in (0,1)$ independent of x_1 and x_2 such that, for $c \in (\frac{1}{3}-\delta(\gamma), \frac{1}{3}+\delta(\gamma))$, it holds that*

$$L(\bar{x}_1, \bar{x}_2) \leq \lambda(\gamma) L(x_1, x_2). \tag{2.5}$$

Remark 2.2. *For the suitable γ and c given in Lemma 2.1, we have*

$$L(x_1(k), x_2(k)) \leq \lambda(\gamma)^k \cdot L(x_1(0), x_2(0)),$$

which implies $L(x_1(k), x_2(k)) \to 0$, i.e., $|x_1(k) - x_2(k)| \to 0$ as $k \to \infty$.

Remark 2.3. *The width $\delta(\gamma)$ of the suitable interval for the coupling strength c can be carefully estimated as large as possible.*

Proof. From (2.4), proving (2.5) is equivalent to proving that

$$\frac{(1-4c)^2\gamma(1-x_{12})^2 x_{12}(2-x_{12})}{[x_{12}-(x_1^2+x_2^2)](2-\gamma[x_{12}-(x_1^2+x_2^2)])} \leq \lambda. \tag{2.6}$$

Let

$$F(x_1,x_2) = (1-x_{12})^2 x_{12}(2-x_{12}), \quad G(x_1,x_2) = [x_{12}-(x_1^2+x_2^2)](2-\gamma[x_{12}-(x_1^2+x_2^2)]),$$

and let

$$H(x_1,x_2) = \frac{F(x_1,x_2)}{G(x_1,x_2)}.$$

We want to find a suitable $\delta(\gamma)$ such that the maximal value of the function $H(x_1,x_2)$ is strictly less than $\frac{1}{\gamma(1-4c)^2}$ in the domain $[\frac{4-\gamma}{4}, \frac{\gamma}{4}] \times [\frac{4-\gamma}{4}, \frac{\gamma}{4}]$.

Then

$$\frac{\partial H}{\partial x_1} = \frac{F_{x_1}G - FG_{x_1}}{G^2}, \quad \frac{\partial H}{\partial x_2} = \frac{F_{x_2}G - FG_{x_2}}{G^2}.$$

Here F_{x_i} and G_{x_i} denote the partial derivatives of F and G with respect to x_i, respectively, for $i = 1, 2$.

Letting $\frac{\partial H}{\partial x_1} = \frac{\partial H}{\partial x_2} = 0$, we have

$$F_{x_1}G - FG_{x_1} = 0, \quad F_{x_2}G - FG_{x_2} = 0. \tag{2.7}$$

By direct computation, we obtain

$$\begin{aligned} F_{x_1} &= 2(1-x_{12})(2x_{12}^2 - 4x_{12} + 1), \\ F_{x_2} &= 2(1-x_{12})(2x_{12}^2 - 4x_{12} + 1), \\ G_{x_1} &= 2(1-\gamma[x_{12}-(x_1^2+x_2^2)])(1-2x_1), \\ G_{x_2} &= 2(1-\gamma[x_{12}-(x_1^2+x_2^2)])(1-2x_2). \end{aligned} \tag{2.8}$$

Since $F_{x_1} = F_{x_2}$ by (2.8), from (2.7) it follows that $FG_{x_1} = FG_{x_2}$. Thus the set of points at which the function H attains its local extremal value must satisfy the inclusion

$$\left\{\frac{\partial H}{\partial x_1} = \frac{\partial H}{\partial x_2} = 0\right\} \subset \{F \neq 0, \quad G_{x_1} = G_{x_2}\} \cup \{F = 0, \quad F_{x_1} = F_{x_2} = 0\} \equiv S_1 \cup S_2. \tag{2.9}$$

Case S_1. From (2.8) it follows that

$$1-\gamma[x_{12}-(x_1^2+x_2^2)]=0 \quad \text{or} \quad x_1=x_2.$$

If $1-\gamma[x_{12}-(x_1^2+x_2^2)]=0$, then we have $G(x_1,x_2)=\frac{1}{\gamma}$. On the other hand,

$$\begin{aligned} F(x_1,x_2) &= (1-x_{12})^2x_{12}(2-x_{12}) \\ &= (1-2x_{12}+x_{12}^2)x_{12}(2-x_{12}) \\ &= [1-(2x_{12}-x_{12}^2)](2x_{12}-x_{12}^2) \le \frac{1}{4} \end{aligned}$$

for $0 \le x_{12} \le 2$. Consequently, $H(x_1,x_2) \le \frac{\gamma}{4}$. Hence, for $c \approx \frac{1}{3}$ and $\gamma \in [3.82,4]$, we have

$$H(x_1,x_2) < \frac{1}{(1-4c)^2\gamma}. \tag{2.10}$$

If $x_1=x_2$, then $F(x_1,x_2)$ and $G(x_1,x_2)$ have simpler forms:

$$\begin{aligned} F(x_1,x_2=x_1) &= 4(1-2x_1)^2x_1(1-x_1), \\ G(x_1,x_2=x_1) &= 4(x_1-x_1^2)[1-\gamma(x_1-x_1^2)]. \end{aligned}$$

Thus

$$H(x_1,x_2=x_1)=\frac{1-4(x_1-x_1^2)}{1-\gamma(x_1-x_1^2)} \le 1.$$

Therefore, for $c \approx \frac{1}{3}$ and $\gamma \in [3.82,4]$, we have

$$H(x_1,x_2) < \frac{1}{(1-4c)^2\gamma}. \tag{2.11}$$

Case S_2. It is easily seen that $H(x_1,x_2)=0$.

It remains to check the extremal values of $H(x_1,x_2)$ on the boundary points. Let $\xi=\frac{4-\gamma}{4}$. Without loss of generality, we consider only two cases:

$$(H1) \quad x_2=\frac{4-\gamma}{4} \qquad \text{and} \qquad (H2) \quad x_2=\frac{\gamma}{4}.$$

Case $(H1)$. Since

$$f(x_1)+f(\xi) \ge \gamma x_1(1-x_1)+\gamma\xi(1-x_1)=\gamma(x_1+\xi)(1-x_1),$$

from $f(x_i)=\gamma x_i(1-x_i)(i=1,2)$ we have

$$\begin{aligned} \frac{1}{\gamma}H(x_1,\xi) &= \frac{(1-(x_1+\xi))^2(x_1+\xi)(2-(x_1+\xi))}{(f(x_1)+f(\xi))[2-(f(x_1)+f(\xi))]} \\ &\le \frac{(1-(x_1+\xi))(2-(x_1+\xi))}{\gamma(2-\gamma(x_1+\xi)(1-x_1))} \\ &\le \frac{1}{\gamma}\max_{0<y<1}\frac{(1-y)(2-y)}{2-\gamma y(1+\xi-y)}. \end{aligned} \tag{2.12}$$

From the relation $\gamma = 4(1-\xi)$, we obtain for $0 < y < 1$ that

$$\begin{aligned}\frac{(1-y)(2-y)}{2-\gamma y(1+\xi-y)} &= \frac{(1-y)(2-y)}{2-4(1-\xi^2)y+4(1-\xi)y^2}\\ &\le \frac{(1-y)(2-y)}{2-4(y-y^2)-4\xi y^2}\\ &\le \frac{(1-y)(2-y)}{2-4y+3.82y^2} \le 1.15 < \frac{1}{(1-4c)^2\gamma}.\end{aligned}$$

For simplicity, in the last two inequalities above, we use $\gamma \ge 3.82$ to obtain estimates. (However, this condition is not necessary.) Consequently, for $c \approx \frac{1}{3}$ and $\gamma \in [3.82, 4]$, we have

$$H(x_1,\xi) < \frac{1}{(1-4c)^2\gamma}. \tag{2.13}$$

Case $(H2)$. For $x_2 = \frac{\gamma}{4}$, since $x_1(1-x_1)+\xi(1-\xi) \ge (1+\xi-x_1)x_1$, we have

$$\begin{aligned}\frac{1}{\gamma}H(x_1,x_2) &= \frac{(1-(x_1+\frac{\gamma}{4}))^2(x_1+\frac{\gamma}{4})(2-(x_1+\frac{\gamma}{4}))}{(f(x_1)+f(\frac{\gamma}{4}))(2-(f(x_1)+f(\frac{\gamma}{4})))}\\ &= \frac{(\xi-x_1)^2(x_1+\frac{\gamma}{4})(1+\xi-x_1)}{(\gamma x_1(1-x_1)+\gamma\xi(1-\xi))(2-\gamma(x_1(1-x_1)+\xi(1-\xi)))}\\ &\le \frac{1}{\gamma}\frac{(x_1-\xi)(x_1+1-\xi)}{2-\gamma(1+\xi-x_1)x_1}.\end{aligned}$$

The last inequality follows from the facts that

$$(\xi-x_1)^2 \le x_1(x_1-\xi) \quad \text{and} \quad \left(x_1+\frac{\gamma}{4}\right) = (x_1+1-\xi).$$

Using the relation $(x_1-\xi)(x_1+1-\xi) = x_1(x_1+1)-((2x_1+1)-\xi)\xi \le x_1(x_1+1)$, we have

$$\frac{1}{\gamma}H\left(x_1,\frac{\gamma}{4}\right) \le \frac{x_1(x_1+1)}{\gamma(2-\gamma(1+\xi-x_1)x_1)} \le \frac{x_1(x_1+1)}{\gamma(2-4x_1+3.82x_1^2)}.$$

It can be easily shown that

$$\max_{0<y<1}\frac{y(y+1)}{2-4y+3.82y^2} \le 1.3.$$

Thus, for $c \approx \frac{1}{3}$ and $\gamma \in [3.82, 4]$, we get

$$H\left(x_1,\frac{\gamma}{4}\right) < \frac{1}{(1-4c)^2\gamma}. \tag{2.14}$$

From (2.10), (2.11), (2.13), and (2.14), we conclude that $H(x_1,x_2) < \frac{1}{(1-4c)^2\gamma}$ whenever c is near $\frac{1}{3}$ and thus that there is a $\lambda(\gamma) \in (0,1)$ such that (2.6) holds. ■

3. Some lemmas.

In section 2, we proved that synchronization occurs for a special case. In order to prove synchronization for the general case, we need the following lemmas.

Lemma 3.1. *If $f_1 = f_3, f_2 = f_4$, then*

$$\frac{1-2c}{2c} \le \frac{\bar f_{13}}{\bar f_{24}} \le \frac{2c}{1-2c}. \tag{3.1}$$

Proof. From the definition of the coupled map (1.1), we obtain

$$\begin{aligned} &\bar x_{13} = (1-2c)f_{13} + 2cf_{24}, \quad \bar x_{24} = (1-2c)f_{24} + 2cf_{13},\\ &\bar f_{13} = \gamma(\bar x_{13} - (\bar x_1^2 + \bar x_3^2)), \quad \bar f_{24} = \gamma(\bar x_{24} - (\bar x_2^2 + \bar x_4^2)). \end{aligned} \tag{3.2}$$

If $f_1 = f_3, f_2 = f_4$, then we have

$$\begin{aligned} &\bar x_1 = \bar x_3 = (1-2c)f_1 + 2cf_2,\\ &\bar x_2 = \bar x_4 = (1-2c)f_2 + 2cf_1,\\ &\bar f_{13} = 2\gamma((1-2c)f_1 + 2cf_2)(1 - ((1-2c)f_1 + 2cf_2)),\\ &\bar f_{24} = 2\gamma((1-2c)f_2 + 2cf_1)(1 - ((1-2c)f_2 + 2cf_1)). \end{aligned} \tag{3.3}$$

So

$$\frac{\bar f_{13}}{\bar f_{24}} = \frac{[(1-2c)f_1 + 2cf_2][1 - ((1-2c)f_1 + 2cf_2)]}{[(1-2c)f_2 + 2cf_1][1 - ((1-2c)f_2 + 2cf_1)]}. \tag{3.4}$$

Without loss of generality, we assume that $f_1 \le f_2$ and $4c > 1$. Then we have

$$(1-2c)f_1 + 2cf_2 \ge (1-2c)f_2 + 2cf_1. \tag{3.5}$$

To prove Lemma 3.1, we need to consider the following four cases:

$$\begin{aligned} &\text{(I)} \quad (1-2c)f_1 + 2cf_2 \le \frac{1}{2} \quad \text{and} \quad (1-2c)f_2 + 2cf_1 \le \frac{1}{2};\\ &\text{(II)} \quad (1-2c)f_1 + 2cf_2 \ge \frac{1}{2} \ge (1-2c)f_2 + 2cf_1 \quad \text{and}\\ &\qquad (1-2c)f_1 + 2cf_2 - \frac{1}{2} \le \frac{1}{2} - ((1-2c)f_2 + 2cf_1);\\ &\text{(III)} \quad (1-2c)f_1 + 2cf_2 \ge \frac{1}{2} \ge (1-2c)f_2 + 2cf_1 \quad \text{and}\\ &\qquad (1-2c)f_1 + 2cf_2 - \frac{1}{2} \ge \frac{1}{2} - ((1-2c)f_2 + 2cf_1);\\ &\text{(IV)} \quad (1-2c)f_1 + 2cf_2 \ge \frac{1}{2} \quad \text{and} \quad (1-2c)f_2 + 2cf_1 \ge \frac{1}{2}. \end{aligned} \tag{3.6}$$

Case (I). Obviously,

$$1 - [(1-2c)f_1 + 2cf_2] \le 1 - [(1-2c)f_2 + 2cf_1]. \tag{3.7}$$

From the assumption that $4c > 1$, it follows that

$$\frac{\bar{f}_{13}}{\bar{f}_{24}} \le \frac{(1-2c)f_1 + 2cf_2}{(1-2c)f_2 + 2cf_1} \le \frac{2c}{1-2c}. \tag{3.8}$$

Since $g(x) = x(1-x)$ is monotone increasing in the interval $(0, \frac{1}{2})$, from (3.3) and (3.5) we get $\bar{f}_{13} \ge \bar{f}_{24}$, i.e., $\frac{\bar{f}_{13}}{\bar{f}_{24}} \ge 1$. Combining the last inequality with (3.8), the assertion (3.1) holds.

Case (II). For this case, it is easily seen that $\bar{f}_{13} \ge \bar{f}_{24}$ by the graph of $g(x) = x(1-x)$. On the other hand, from (3.7), we also have

$$1 \le \frac{\bar{f}_{13}}{\bar{f}_{24}} \le \frac{(1-2c)f_1 + 2cf_2}{(1-2c)f_2 + 2cf_1} \le \frac{2c}{1-2c}. \tag{3.9}$$

The assertion (3.1) holds by (3.9).

Case (III). Let $A = 1 - [(1-2c)f_1 + 2cf_2]$ and $B = 1 - [(1-2c)f_2 + 2cf_1]$. Then we have

$$\bar{f}_{13} = A(1-A), \quad \bar{f}_{24} = B(1-B). \tag{3.10}$$

Moreover, A and B satisfy $A \le \frac{1}{2} \le B$, and $B - \frac{1}{2} < \frac{1}{2} - A$, which implies that $\bar{f}_{13} < \bar{f}_{24}$. On the other hand, from (3.5) and the assumption that $f_1 \le f_2$, we have

$$\frac{\bar{f}_{13}}{\bar{f}_{24}} = \frac{A(1-A)}{B(1-B)} \ge \frac{A}{B} \ge \frac{1-2c}{2c}.$$

Thus (3.1) can be concluded similarly.

Case (IV). Let $\tilde{f}_1 = 1 - f_1$ and $\tilde{f}_2 = 1 - f_2$. Then, from the assumption that $4c > 1$, the following inequality holds:

$$\tilde{f}_1 \le \tilde{f}_2, \qquad (1-2c)\tilde{f}_2 + 2c\tilde{f}_1 \le (1-2c)\tilde{f}_1 + 2c\tilde{f}_2 \le \frac{1}{2}.$$

Moreover, we have

$$\begin{aligned} \bar{f}_{13} &= (1 - [(1-2c)f_1 + 2cf_2])[(1-2c)f_1 + 2cf_2] \\ &= [(1-2c)\tilde{f}_1 + 2c\tilde{f}_2][1 - ((1-2c)\tilde{f}_1 + 2c\tilde{f}_2)], \\ \bar{f}_{24} &= (1 - [(1-2c)f_2 + 2cf_1])[(1-2c)f_2 + 2cf_1] \\ &= [(1-2c)\tilde{f}_2 + 2c\tilde{f}_1][1 - ((1-2c)\tilde{f}_2 + 2c\tilde{f}_1)]. \end{aligned}$$

Hence Case (IV) is reduced to Case (I) if f_2 and f_1 are replaced by $\tilde{f}_1$ and $\tilde{f}_2$, respectively. ■

In the following, we shall prove that the inequality (3.1) in Lemma 3.1 holds without the restrictions $f_1 = f_3$ and $f_2 = f_4$.

Lemma 3.2. *For $4c > 1$, it holds that*

$$\frac{1-2c}{2c} \le \frac{\bar{f}_{13}}{\bar{f}_{24}} \le \frac{2c}{1-2c}. \tag{3.11}$$

Proof. By direct computation, we have

$$\frac{\bar{f}_{13}}{\bar{f}_{24}} = \frac{(1-2c)f_{13} + 2cf_{24} - [((1-2c)f_1 + cf_{24})^2 + ((1-2c)f_3 + cf_{24})^2]}{(1-2c)f_{24} + 2cf_{13} - [((1-2c)f_2 + cf_{13})^2 + ((1-2c)f_4 + cf_{13})^2]}.$$

Denote

$$\begin{aligned} F &= (1-2c)f_{13} + 2cf_{24} - [((1-2c)f_1 + cf_{24})^2 + ((1-2c)f_3 + cf_{24})^2], \\ G &= (1-2c)f_{24} + 2cf_{13} - [((1-2c)f_2 + cf_{13})^2 + ((1-2c)f_4 + cf_{13})^2]. \end{aligned}$$

It is easy to verify that

$$G_{f_1} = G_{f_3}, \quad F_{f_2} = F_{f_4}. \tag{3.12}$$

The equations

$$\frac{\partial}{\partial f_1}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_2}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_3}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_4}\left(\frac{F}{G}\right) = 0 \tag{3.13}$$

are equivalent to

$$F_{f_i}G - FG_{f_i} = 0 \quad \text{for} \quad i = 1,2,3,4. \tag{3.14}$$

From (3.12) and (3.14), we obtain

$$F_{f_1} = F_{f_3}, \quad G_{f_2} = G_{f_4},$$

which implies

$$f_1 = f_3, \quad f_2 = f_4.$$

Thus the set of possible local extremal values satisfies the inclusion

$$\left\{\frac{\partial}{\partial f_1}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_2}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_3}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_4}\left(\frac{F}{G}\right) = 0\right\} \subset \{f_1 = f_3, \quad f_2 = f_4\}.$$

Hence (3.11) is obtained for the case of possible local extremal values by Lemma 3.1.

It remains to estimate the maximal value of $\frac{F}{G}$ on the boundary. Because of symmetry, we need only consider two cases: (i) $f_3 = 1 - \xi$ and (ii)$f_3 = \xi$, where $\xi = \frac{4-\gamma}{4}$.

Case (i). Let

$$\left\{\frac{\partial}{\partial f_1}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_2}\left(\frac{F}{G}\right) = \frac{\partial}{\partial f_4}\left(\frac{F}{G}\right)\right\} = 0,$$

which implies $f_2 = f_4$. Therefore, we have

$$\frac{\bar{f}_{13}}{\bar{f}_{24}} = \frac{[(1-2c)(f_1 + 1 - \xi) + 4cf_2] - [((1-2c)f_1 + 2cf_2)^2 + ((1-2c)(1-\xi) + 2cf_2)^2]}{2[(1-2c)f_2 + c(f_1 + 1 - \xi)] - 2[(1-2c)f_2 + c(f_1 + 1 - \xi)]^2}.$$

For $c = \frac{1}{3}$, we want to show that

$$\frac{2c}{1-2c} = 2 > \frac{\bar{f}_{13}}{\bar{f}_{24}} > \frac{1-2c}{2c} = \frac{1}{2},$$

i.e.,

$$2\bar{f}_{24} - \bar{f}_{13} > 0 \quad \text{and} \quad 2\bar{f}_{13} - \bar{f}_{24} > 0.$$

For $c = \frac{1}{3}$, we get by direct computation that

$$\begin{aligned}
\frac{1}{\gamma}\bar{f}_{13} &= \frac{1}{3}(f_1 + 1 - \xi) + \frac{4}{3}f_2 - \left[\left(\frac{1}{3}f_1 + \frac{2}{3}f_2\right)^2 + \left(\frac{1}{3}(1-\xi) + \frac{2}{3}f_2\right)^2\right],\\
\frac{1}{\gamma}\bar{f}_{24} &= \frac{2}{3}(f_1 + f_2 + 1 - \xi) - \frac{2}{9}(f_1 + f_2 + 1 - \xi)^2.
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{1}{\gamma}(2\bar{f}_{24} - \bar{f}_{13}) &= \frac{4}{3}(f_{12} + 1 - \xi)\\
&-\frac{4}{9}(f_{12} + 1 - \xi)^2 - \frac{1}{3}(f_1 + 1 - \xi) - \frac{4}{3}f_2\\
&+\left[\left(\frac{1}{3}f_1 + \frac{2}{3}f_2\right)^2 + \left(\frac{1}{3}(1-\xi) + \frac{2}{3}f_2\right)^2\right]\\
&= f_1 + 1 - \xi - \frac{1}{3}f_1^2 - \frac{1}{3}(1-\xi)^2 - \frac{8}{9}(1-\xi)f_1 + \left[-\frac{4}{9}f_1f_2 + \frac{4}{9}f_2^2 - \frac{4}{9}(1-\xi)f_2\right]\\
&= f_1 + 1 - \xi - \frac{1}{3}f_1^2 - \frac{1}{3}(1-\xi)^2 - \frac{8}{9}(1-\xi)f_1 + \left(\frac{2}{3}f_2 - \frac{1}{3}f_1 - \frac{1}{3}(1-\xi)\right)^2\\
&-\frac{1}{9}f_1^2 - \frac{1}{9}(1-\xi)^2 - \frac{2}{9}f_1(1-\xi)\\
&> 0.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\frac{1}{\gamma}(2\bar{f}_{13} - \bar{f}_{24}) &= \frac{2}{3}(f_1 + 1 - \xi) + \frac{8}{3}f_2 - 2\left[\left(\frac{1}{3}f_1 + \frac{2}{3}f_2\right)^2 + \left(\frac{1}{3}(1-\xi) + \frac{2}{3}f_2\right)^2\right]\\
&-\frac{2}{3}(f_{12} + 1 - \xi) + \frac{2}{9}(f_{12} + 1 - \xi)^2\\
&= 2f_2 + \frac{4}{9}(1-\xi)f_1 - \frac{4}{9}(1-\xi)f_2 - \frac{4}{9}f_1f_2 - \frac{14}{9}f_2^2\\
&= \frac{4}{9}f_1[(1-f_2) - \xi] + \frac{1}{9}f_2[14(1-f_2) + 4\xi] > 0.
\end{aligned}$$

Conclusively, we prove our assertion at $c = \frac{1}{3}$ for Case (i). Note that all of the inequalities above hold strictly about $c = \frac{1}{3}$. Consequently, there exists a suitable $\delta(\gamma)$ such that, for each $c \in (\frac{1}{3} - \delta(\gamma), \frac{1}{3} + \delta(\gamma))$, the inequality (3.11) holds.

The other extreme values on the boundary points can be estimated similarly to that above. We omit the estimations here.

Case (ii). The proof is similar to that of Case (i).

Thus we finish the proof of Lemma 3.2. ■

4. Proof of Theorem 1.1.

4.1. Definition of the generalized Liapunov function. In order to deal with the general case, we need to generalize the definition of the Liapunov function.

Define the function $\phi : [2\xi, 2 - 2\xi] \to [2\xi, 1]$ with $\xi = \frac{4-\gamma}{4}$ by

$$\phi(z) = \begin{cases} z(2-z) & \text{if} \quad |z| \le a \quad \text{or} \quad |2-z| \le a, \\ a(2-a) & \text{otherwise,} \end{cases} \tag{4.1}$$

where $a = \frac{1}{4} + \epsilon_0$ with ϵ_0 a small positive constant. Let $g(x, y) = \phi(x+y) : [\xi, 1-\xi]^2 \to [2\xi, 1]$. Now we define the new Liapunov function by

$$\mathcal{L}(x, y) = \frac{(x-y)^2}{g(x, y)}. \tag{4.2}$$

One can see that the definition of $\mathcal{L}$ in (4.2) is a generalization of (2.3). In the following, we shall prove a conclusion similar to that in Lemma 2.1 for the new Liapunov function $\mathcal{L}$ defined in (4.2).

Lemma 4.1. *For each $\gamma \in [3.82, 4]$ in (1.1), there exist numbers $\delta(\gamma)$ and $\lambda \in (0, 1)$ independent of x_1 and x_3 such that, for $c \in (\frac{1}{3} - \delta(\gamma), \frac{1}{3} + \delta(\gamma))$, the following inequality holds:*

$$\mathcal{L}(\bar{x}_1, \bar{x}_3) \le \lambda(\gamma)\mathcal{L}(x_1, x_3). \tag{4.3}$$

Proof. By direct computation, we have

$$\frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} = \frac{c^2\gamma^2(1 - x_{13})^2\phi(x_{13})}{\phi(\frac{1}{3}f_{13} + \frac{2}{3}f_{24})}. \tag{4.4}$$

We first consider the case in which $c = \frac{1}{3}$. There are nine different cases to be considered.

Case A. $|x_{13}| \le a$, and $|\frac{1}{3}f_{13} + \frac{2}{3}f_{24}| \le a$.

It follows that

$$\phi(x_{13}) = x_{13}(2 - x_{13}) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right) = \left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right)\left[2 - \left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right)\right].$$

If $f_{13} \leq f_{24}$, then $\frac{1}{3}f_{13} + \frac{2}{3}f_{24} \geq f_{13}$. Hence, from (4.4), we obtain

$$\begin{aligned}\frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} &\leq \frac{\gamma^2(1-x_{13})^2 x_{13}(2-x_{13})}{9f_{13}(2-(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}))} \\ &= \frac{2-f_{13}}{2-(\frac{1}{3}f_{13}+\frac{2}{3}f_{24})} \frac{\gamma^2(1-x_{13})^2 x_{13}(2-x_{13})}{9f_{13}(2-f_{13})} \\ &\leq \frac{2}{2-a}K,\end{aligned}$$

where $K = \frac{\gamma^2(1-x_{13})^2 x_{13}(2-x_{13})}{9f_{13}(2-f_{13})}$. The last inequality follows from the assumption that $|\frac{1}{3}f_{13} + \frac{2}{3}f_{24}| \leq a$.

If $f_{13} > f_{24}$, then, from Lemma 3.2, we have $f_{24} \geq \frac{1}{2}f_{13}$ and $2 - (\frac{1}{3}f_{13} + \frac{2}{3}f_{24}) \geq 2 - f_{13}$. Hence we get

$$\frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} \leq \frac{3}{2}\frac{\gamma^2(1-x_{13})^2 x_{13}(2-x_{13})}{9f_{13}(2-f_{13})} = \frac{3}{2}K.$$

Observe that K is similar to the left-hand side of (2.6). From (2.14), with the proof of Lemma 2.1, we have $K \leq \frac{1.3\gamma}{9}$. Hence, for Case A, we have

$$\frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} \leq \frac{1.3\gamma}{6}. \tag{4.5}$$

Case B. $|x_{13}| \leq a$, and $a \leq |\frac{1}{3}f_{13} + \frac{2}{3}f_{24}| \leq 2 - a$.

It follows that

$$\phi(x_{13}) = x_{13}(2-x_{13}) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right) = a(2-a).$$

On the other hand, since $x_{13} \leq a$, we have

$$\phi(x_{13}) \leq a(2-a) < \frac{1}{2}.$$

Therefore,

$$(1-x_{13})^2 x_{13}(2-x_{13}) \leq (1-a(2-a))a(2-a). \tag{4.6}$$

Then we have

$$\frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} \leq \frac{\gamma^2(1-a(2-a))a(2-a)}{9a(2-a)} = \left(\frac{1}{3}\gamma(1-a)\right)^2. \tag{4.7}$$

Next, we shall prove that Case C defined as below cannot occur.

Case C. $|x_{13}| \leq a$, and $|\frac{1}{3}f_{13} + \frac{2}{3}f_{24}| \geq 2 - a$.

It follows that $\phi(x_{13}) = x_{13}(2-x_{13})$. From $x_{13} \leq a$, we have $f_{13} = \gamma(x_{13} - (x_1^2 + x_3^2)) \leq \gamma(a - \frac{a^2}{2}) \leq 1$. Hence, combining this with the fact that $a = \frac{1}{4} + \epsilon_0$ with ϵ_0 small enough, we have $\frac{1}{3}f_{13} + \frac{2}{3}f_{24} \leq \frac{5}{3} < 2 - a$, which contradicts the assumption that $|\frac{1}{3}f_{13} + \frac{2}{3}f_{24}| \geq 2 - a$.

Now we have finished the proof of the first three cases.

Case D. $|x_{13}| \geq 2-a$, and $|\frac{1}{3}f_{13}+\frac{2}{3}f_{24}| \leq a$.

It follows that

$$\phi(x_{13}) = x_{13}(2-x_{13}) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right) = \left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right)\left[2-\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right)\right].$$

By the symmetry of $\phi(z)$ with respect to $z=1$, we can reduce Case D to Case A.

Case E. $|x_{13}| \geq 2-a$, and $a \leq |\frac{1}{3}f_{13}+\frac{2}{3}f_{24}| \leq 2-a$.

It follows that

$$\phi(x_{13}) = x_{13}(2-x_{13}) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right) = a(2-a).$$

Hence

$$\frac{\mathcal{L}(\bar{x}_1,\bar{x}_3)}{\mathcal{L}(x_1,x_3)} = \frac{\gamma^2(1-x_{13})^2 x_{13}(2-x_{13})}{9a(2-a)}.$$

By (4.6) and the symmetry of $\phi(z)$ with respect to $z=1$, we can reduce this case to Case B.

Case F. $|x_{13}| \geq 2-a$, and $|\frac{1}{3}f_{13}+\frac{2}{3}f_{24}| \geq 2-a$.

If follows that

$$\phi(x_{13}) = x_{13}(2-x_{13}) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right) = \left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right)\left[2-\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right)\right].$$

Hence we have

$$\frac{\mathcal{L}(\bar{x}_1,\bar{x}_3)}{\mathcal{L}(x_1,x_3)} = \frac{\gamma^2(1-x_{13})^2 x_{13}(2-x_{13})}{9(\frac{1}{3}f_{13}+\frac{2}{3}f_{24})[2-(\frac{1}{3}f_{13}+\frac{2}{3}f_{24})]}.$$

By the symmetry of $\phi(z)$ with respect to $z=1$, we can reduce this case to Case C.

It remains to check the last three cases.

Case G. $a \leq |x_{13}| \leq 2-a$, and $|\frac{1}{3}f_{13}+\frac{2}{3}f_{24}| \leq a$.

Then $\phi(x_{13}) = a(2-a)$. Moreover, we have

$$\phi\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right) = \left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right)\left[2-\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right)\right]$$

and

$$2-\left(\frac{1}{3}f_{13}+\frac{2}{3}f_{24}\right) \geq 2-a.$$

On the other hand, since $a \leq |x_{13}| \leq 2-a$, one can verify that $f_{13} \geq \gamma a(1-a)$, which implies that $\frac{1}{3}f_{13}+\frac{2}{3}f_{24} \geq \frac{2}{3}\gamma a(1-a)$ by Lemma 3.2. Hence we have

$$\frac{\mathcal{L}(\bar{x}_1,\bar{x}_3)}{\mathcal{L}(x_1,x_3)} \leq \frac{\gamma^2(1-x_{13})^2 a(2-a)}{6\gamma a(1-a)(2-a)} \leq \frac{\gamma(1-x_{13})^2}{6(1-a)} \leq \frac{\gamma(1-a)}{6}. \tag{4.8}$$

Case H. $a \leq |x_{13}| \leq 2-a$, and $|\frac{1}{3}f_{13}+\frac{2}{3}f_{24}| \geq 2-a$.

It follows that

$$\phi(x_{13}) = a(2-a) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right) = \left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right)\left[2 - \left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right)\right].$$

Hence we have

$$\begin{aligned} \frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} &= \frac{\gamma^2(1-x_{13})^2 a(2-a)}{(3f_{13}+6f_{24})[2-(\frac{1}{3}f_{13}+\frac{2}{3}f_{24})]} \\ &\leq \frac{\gamma^2(1-x_{13})^2 a(2-a)}{9(2-a)[2-(\frac{1}{3}f_{13}+\frac{2}{3}f_{24})]} \\ &\leq \frac{\gamma^2(1-x_{13})^2 a}{3(2-f_{13})} \\ &= \frac{\gamma^2(1-x_{13})^2 a}{3(2-\gamma(x_{13}-(x_1^2+x_3^2)))} \\ &\leq \frac{\gamma^2 a}{3} \cdot \frac{(1-x_{13})^2}{2-\gamma x_{13}+\frac{\gamma}{2}x_{13}^2} \\ &\leq \frac{\gamma^2 a}{6}. \end{aligned} \tag{4.9}$$

Case I. $a \leq |x_{13}| \leq 2-a$, and $a \leq |\frac{1}{3}f_{13} + \frac{2}{3}f_{24}| \leq 2-a$.
It follows that

$$\phi(x_{13}) = a(2-a) \quad \text{and} \quad \phi\left(\frac{1}{3}f_{13} + \frac{2}{3}f_{24}\right) = a(2-a).$$

Hence we have

$$\frac{\mathcal{L}(\bar{x}_1, \bar{x}_3)}{\mathcal{L}(x_1, x_3)} = \frac{\gamma^2(1-x_{13})^2 a(2-a)}{9a(2-a)} \leq \frac{\gamma^2(1-a)^2}{9}. \tag{4.10}$$

From (4.5), (4.7), (4.8), (4.9), and (4.10), we obtain that

$$\max\left\{\frac{1.3\gamma}{6}, \left(\frac{1}{3}\gamma(1-a)\right)^2, \frac{\gamma(1-a)}{6}, \frac{\gamma^2 a}{6}\right\} < 1, \tag{4.11}$$

which proves Lemma 4.1 for $c = \frac{1}{3}$. Moreover, since the inequality in (4.11) holds strictly, it is easily seen that there is an interval $(\frac{1}{3} - \delta(\gamma), \frac{1}{3} + \delta(\gamma))$ such that (4.3) holds for c in this interval. ■

4.2. Proof of Theorem 1.1. By using Lemma 4.1, we have proved that synchronization occurs between x_1 and x_3. Similarly, we can prove $|x_2(k) - x_4(k)| \to 0$ as $k \to \infty$. Subsequently, using Lemma 2.4 in [6] stated below and Lemma 2.1, we can prove that synchronization occurs between x_1 and x_2.

Lemma 2.4 (in [6]). *For any one-dimensional map*

$$u(k+1) = a(k)u(k) + b(k),$$

if $|b(k)|$ *is finite with* $|b(k)| \to 0$ *as* $k \to \infty$ *and*

$$\sup_{k \leq 0} |a(k)| = \lambda < 1,$$

then $|u(k)|$ *converges to zero.*

Acknowledgments. Yi-Qian Wang is currently a postdoc at the National Center for Theoretical Sciences in Taiwan. He would like to thank the hospitality of the institute and Prof. Song-Sun Lin of the National Chiao Tung University in Taiwan.

REFERENCES

[1] D. Campbell, *An introduction to nonlinear phenomena*, in Lectures in the Sciences of Complexity, D. L. Stein, ed., Addison–Wesley, Redwood City, CA, 1989, pp. 3–105.

[2] K. M. Cuomo and A. V. Oppenheim, *Synchronized Chaotic Circuits and Systems for Communications*, Electr. TR. MIT Res. Lab., 575, MIT, Cambridge, MA, 1992.

[3] K. M. Cuomo and A. V. Oppenheim, *Circuit implementation of synchronized chaos, with applications to communications*, Phys. Rev. Lett., 71 (1993), p. 65.

[4] J. Gleick, *Chaos: Making a New Science*, Viking, New York, 1987.

[5] J. F. Heagy, T. L. Carroll, and L. M. Pecora, *Synchronization with application to communication*, Phys. Rev. Lett., 74 (1995), pp. 5028–5031.

[6] W. W. Lin, C. C. Peng, and C. S. Wang, *Synchronization in coupled map lattices with periodic boundary condition*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 9 (1999), pp. 1635–1652.

[7] J. Maynard Smith, *Mathematical Ideas in Biology*, Cambridge University Press, Cambridge, UK, 1968.

[8] L. M. Pecora and T. L. Carroll, *Synchronization in chaotic systems*, Phys. Rev. Lett., 64 (1990), pp. 821–824.

[9] L. M. Pecora, T. L. Carroll, G. A. Johnson, D. J. Mar, and J. F. Heagy, *Fundamentals of synchronization in chaotic systems, concept and applications*, Chaos, 6 (1997), pp. 262–276.

[10] S. Vohra, M. Spano, M. Shlesinger, L. Pecora, and W. Ditto, Eds., *Proceedings First Experimental Chaos Conference*, World Scientific, River Edge, NJ, 1992.

[11] C. W. Wu and L. O. Chua, *A unified framework for synchronizations and control of dynamical systems*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 4 (1994), pp. 979–988.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 1, No. 2, pp. 190–214

A Multiparameter, Numerical Stability Analysis of a Standing Cantilever Conveying Fluid[*]

Nawaf M. Bou-Rabee[†], Louis A. Romero[‡], and Andrew G. Salinger[‡]

Abstract. In this paper, we numerically examine the stability of a standing cantilever conveying fluid in a multiparameter space. Based on nonlinear beam theory, our mathematical model turns out to be replete with exciting behavior, some of which was totally unexpected and novel, and some of which confirm our intuition as well as the work of others. The numerical bifurcation results obtained from applying the Library of Continuation Algorithms (LOCA) reveal a plethora of one, two, and higher codimension bifurcations. For a vertical or standing cantilever beam, bifurcations to buckled solutions (via symmetry breaking) and oscillating solutions are detected as a function of gravity and the fluid-structure interaction. The unfolding of these results as a function of the orientation of the beam compared to gravity is also revealed.

Key words. numerical bifurcation analysis, aero-elasticity, beam flutter, buckling

AMS subject classifications. 74F10, 37M20, 65P30, 74H10

PII. S1111111102400753

1. Introduction.

The dynamics of a standing cantilever conveying fluid has been thoroughly studied for the past century [18]. A major contribution to this problem dates back to 1961 with the work of Benjamin on articulated pipes [4]. Benjamin explained the physical mechanism behind flutter using a theoretical double pendulum model [4]. In a later work, Benjamin shows the destabilizing effect of damping [5].

Païdoussis' experimental and computational work in 1970 confirmed and developed Benjamin's results [15]. Païdoussis' model, based on conservation laws, took into account the flexural restoring force, the fluid inertia force, the gravity force, and the tube inertia force. He also added viscous effects due to external damping.

Using this model, Païdoussis showed that damping does destabilize the beam and that the fluid flow rate can prevent the beam from buckling [15]. In fact, dissipation-induced instabilities have been extensively studied [7], [8] in a wide variety of dynamical systems such

[*] Received by the editors January 10, 2002; accepted for publication (in revised form) by M. Dellnitz August 5, 2002; published electronically September 12, 2002. This research was supported by Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000. This work was supported by the US DOE through the MICS and ASCI programs. 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. Copyright is owned by SIAM to the extent not limited by these rights.

http://www.siam.org/journals/siads/1-2/40075.html

[†] California Institute of Technology Applied and Computational Mathematics, 1200 E. California Blvd., MC 217-50, Pasadena, CA 91125 (nawaf@acm.caltech.edu).

[‡] Sandia National Laboratories, P.O. Box 5800, MS 1110, Albuquerque, NM 87185-1110 (lromero@sandia.gov, agsalin@sandia.gov).

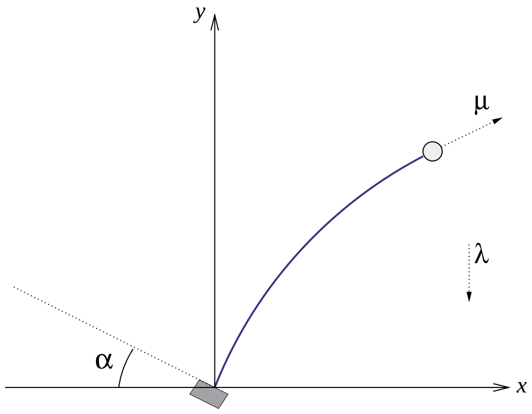


Figure 1.1. *The orientation of the cantilever with respect to gravity (α), the dimension follower force (μ), and the dimensionless gravity (λ) are all parameters in our equations.*

as the double spherical pendulum [13] and rotating systems with gyroscopic terms. Païdoussis also demonstrated using linear theory that there exist a curve of pitchfork bifurcation points and a curve of Hopf bifurcation points in parameter space. The works of Bajaj, Sethna, and Lundgren [1] and Bajaj and Sethna [2], [3] are also noteworthy references on this subject.

In this work, we use a simple model of a standing cantilever conveying fluid, namely, a cantilever beam clamped at one end and free at the other end as shown in Figure 1.1. We assume the fluid-structure interaction induces a concentrated force tangent to the free end of the cantilever and a point source of damping normal to the free end of the cantilever. Our model is simple because we neglect all other fluid structure interaction, including the fluid-induced coriolis forces, unlike the more complete model of Païdoussis [15]. The model is a good representation of a long, thin conduit with a nozzle at its free end. The nozzle maintains the necessary momentum flux, which induces the concentrated force that remains tangent to the free end of the beam. The ratio of the *follower force* to the restoring force due to structural rigidity is denoted as μ, and we will hereafter refer to this quantity as the dimensionless follower force. As we increase μ, the cantilever eventually experiences an oscillatory instability (a Hopf bifurcation). Likewise, the ratio of the gravitational body force to the restoring force due to structural rigidity will be denoted as λ, and we will hereafter refer to this quantity as the dimensionless gravity parameter. Because our model incorporates dimensionless gravity and the beam's orientation to gravity (α), when the standing cantilever is heavy enough, the beam can experience buckling (a pitchfork bifurcation).

This paper extends the results of previous work in several ways. First, we have been able to identify a quartic bifurcation point which more completely maps out the behavior of the beam in parameter space [25]. Second, we have shown analytically, using the perturbation theory of eigenvalues, that damping destabilizes the beam and that, for a point source of damping, the magnitude of damping has no effect on the Hopf bifurcation points. Finally, since we included the inclination of the beam at the clamped end as a parameter in our problem (α), we were able to numerically continue in α. With this capability, we demonstrate how the high codimensional pitchfork bifurcations unfold in α. More precisely, the paper makes the following assertions supported by numerical evidence and theory:

- When $\alpha = 0.0$, the parameter space is divided by four bifurcation curves: a curve of pitchfork bifurcation points, a curve of Hopf bifurcation points, a curve of turning points, and a curve of saddle-loop bifurcations. We tracked the first three curves presented in section 4.1 using the tracking algorithms in the Library of Continuation Algorithms (LOCA). The theory of the double zero eigenvalue predicts the existence of the saddle-loop bifurcation [12], but we were unable to obtain it using our current capabilities. The saddle-loop bifurcation occurs when a periodic orbit bifurcates to a solution with infinite period.
- When the angle of inclination at the clamped end of the beam is zero ($\alpha = 0.0$), there is a symmetry-breaking Takens–Bogdanov point, where the curve of Hopf bifurcations terminates on a path of pitchfork bifurcations. When we unfold this point in α, it becomes a double zero eigenvalue of codimension two, where the Hopf bifurcation curve terminates at a curve of turning points.
- Again, when $\alpha = 0.0$, there is another extremely degenerate bifurcation point, where the curve of turning points intersects the curve of pitchfork bifurcation points. At this quartic bifurcation point, a symmetry-breaking pitchfork bifurcation changes from supercritical to subcritical [10]. When we unfold this point in α, part of what we get is a codimension two bifurcation, where a curve of turning points terminates at a cusp. We use the theory of normal forms to predict the existence of a curve of turning points in a neighborhood of the quartic bifurcation point and unfold the quartic in section 4.2.
- As expected, a point source of damping had no influence on the stationary bifurcations. Unexpectedly, however, a point source of damping had no influence on the position of the Hopf bifurcation point. Using the perturbation theory of eigenvalues, we will show in section 4.3 why, for a point source of damping, the Hopf bifurcation point is completely independent of the magnitude of damping. This discussion culminates in an explanation of why the beam experiences flutter.

There are also existing techniques for the detection and tracking of bifurcation points. For example, techniques for computing higher codimension Takens–Bogdanov points can be found in [6], [26], and [19]. There is also a large literature on the unfolding of a symmetry-breaking Takens–Bogdanov point [23], [24].

2. Formulation.

2.1. Derivation of equations.

The beam has the following properties: length L, mass per unit length ρ, and flexural rigidity EI. We assume that the flexural rigidity about the x-axis is large enough to confine the cantilever beam to move in the x-y plane. The position of the beam is fully described by the vector $\mathbf{X} = (X(s,t), Y(s,t), 0)$, where s is the arclength $s \in [0, L]$ and t is time. The velocity of the beam is given by the vector $\mathbf{V} = (\dot{X}(s,t), \dot{Y}(s,t), 0) = (U(s,t), V(s,t), 0)$. The problem is planar, and we assume that moments are restricted to exist in the z-direction only. Therefore, the moment and force vectors take the following forms: $\mathbf{M} = (0, 0, M_z(s,t))$ and $\mathbf{F} = (F_x(s,t), F_y(s,t), 0)$.

Consider a differential element of the beam of length ds. The forces in the x- and y-directions and bending moment on this differential element are shown in Figure 2.1. A force

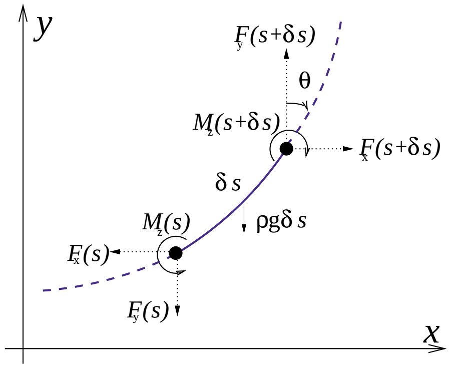


Figure 2.1. *A differential element of the beam.*

balance leads to

$$\rho \frac{\partial^2 X}{\partial t^2} = \frac{\partial F_x}{\partial s},$$
$$\rho \frac{\partial^2 Y}{\partial t^2} = \frac{\partial F_y}{\partial s} - \rho g,$$

and a moment balance results in

$$\left(M_z + \frac{\partial M_z}{\partial s}\, ds \right) - M_z = \mathbf{F} \cdot \mathbf{n}\, ds,$$

where $\mathbf{n}$ is the normal vector defined as

$$\mathbf{n} = (\cos(\theta), -\sin(\theta), 0).$$

The moment is related to the curvature in the following manner:

$$M_z = -EI \frac{\partial \theta}{\partial s}.$$

The negative sign in the moment-curvature relation is due to the definition of θ as being positive moving in the clockwise direction. Using this moment-curvature relation, the moment balance reduces to

$$-EI \frac{\partial^2 \theta}{\partial s^2} = \mathbf{F} \cdot \mathbf{n}.$$

2.2. Governing equations of motion. After including the equations relating the derivatives of X and Y to θ, we obtain a set of five coupled partial differential equations on the domain $s \in [0, L]$. The dependent variables are F_x, F_y, X, Y, and θ.

$$\begin{aligned}
\rho\frac{\partial^2 X}{\partial t^2} &= \frac{\partial F_x}{\partial s}, \\
\rho\frac{\partial^2 Y}{\partial t^2} &= \frac{\partial F_y}{\partial s} - \rho g, \\
0 &= EI\frac{\partial^2 \theta}{\partial s^2} + \mathbf{F}\cdot\mathbf{n}, \\
0 &= \frac{\partial X}{\partial s} - \sin\theta, \\
0 &= \frac{\partial Y}{\partial s} - \cos\theta.
\end{aligned} \tag{2.1}$$

The conveying fluid exerts a force on the free end of the beam. We make the reasonable assumption that the follower force remains tangent to the free end of the beam even as the beam moves. Furthermore, the force normal to this follower force contributes to a point source of damping in our model.

$$\mathbf{F}(1,t)\cdot\mathbf{t} = f,$$
$$-\mathbf{F}(1,t)\cdot\mathbf{n} = C\mathbf{V}(1,t)\cdot\mathbf{n},$$

where C is the damping parameter, f is the follower force, and $\mathbf{t}$ is the tangent vector defined as

$$\mathbf{t} = (\sin(\theta), \cos(\theta), 0).$$

We also assume that the couple at the free end of the beam vanishes. Since the couple is a multiple of the curvature of the beam, this condition reduces to

$$\frac{\partial \theta}{\partial s}(L,t) = 0.$$

On the left side, the beam is clamped, and, at an angle α with respect to gravity,

$$\theta(0,t) = \alpha, \quad X(0,t) = Y(0,t) = 0.$$

Since we will only examine the steady behavior of our governing equations and the linearized behavior about the steady-state, no initial conditions are needed. Together the above equations and boundary conditions are the governing equations of our system.

2.3. Dimensionless equations.

We will introduce the dimensionless variables

$$\xi = \tfrac{s}{L}, \quad x = \tfrac{X}{L}, \quad y = \tfrac{Y}{L}, \quad f_x = \tfrac{L^2 F_x}{EI}, \quad f_y = \tfrac{L^2 F_y}{EI}, \quad \tau = \tfrac{\sqrt{EI}}{\sqrt{L^4 \rho}}t.$$

Our dimensionless force and velocity vector take the following forms: $\mathbf{f} = (f_x(\xi,\tau), f_y(\xi,\tau), 0)$ and $\mathbf{v} = (u(\xi,\tau), v(\xi,\tau), 0)$. Substituting these dimensionless variables into (2.1), we have the equations

$$
\begin{aligned}
\ddot{x} &= f_x', \\
\ddot{y} &= f_y' - \lambda, \\
0 &= \theta'' + \mathbf{f} \cdot \mathbf{n}, \\
0 &= x' - \sin\theta, \\
0 &= y' - \cos\theta,
\end{aligned}
\tag{2.2}
$$

where the overdot and prime denote differentiation with respect to dimensionless time (τ) and dimensionless arclength (ξ), respectively. In terms of these dimensionless variables, the associated boundary conditions become

$$
\begin{gathered}
\theta(0,\tau) = \alpha, \qquad x(0,\tau) = y(0,\tau) = 0, \\
-\mathbf{f}(1,\tau) \cdot \mathbf{n} = \gamma \mathbf{v}(1,\tau) \cdot \mathbf{n}, \\
\theta'(1,\tau) = 0, \qquad \mathbf{f}(1,\tau) \cdot \mathbf{t} = \mu,
\end{gathered}
\tag{2.3}
$$

where μ, γ, and λ are the dimensionless follower force, dissipation, and gravity parameters, respectively, and are defined as

$$\mu = \tfrac{fL^2}{EI}, \quad \gamma = \tfrac{\sqrt{EI}}{\sqrt{L^4 \rho}} C, \quad \lambda = \tfrac{\rho g L^2}{EI}.$$

We can convert the second-order derivatives in dimensionless time (τ) in (2.2) into first-order derivatives by introducing two additional equations:

$$u = \dot{x}, \quad v = \dot{y}.$$

Then (2.2) can be put into the form

$$\mathbf{M}\dot{\mathbf{z}} = \mathbf{R}(\mathbf{z}), \quad \mathbf{z} \in \Re^n, \tag{2.4}$$

where the vector of unknowns $\mathbf{z}$, the mass matrix $\mathbf{M}$, and the function $\mathbf{R}(\mathbf{z})$ are

$$
\mathbf{z} = \begin{bmatrix} x \\ y \\ \theta \\ f_x \\ f_y \\ u \\ v \end{bmatrix}, \quad
\mathbf{M} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}, \quad
\mathbf{R}(\mathbf{z}) = \begin{bmatrix} f_x' \\ f_y' - \lambda \\ \theta'' + \mathbf{f} \cdot \mathbf{n} \\ x' - \sin\theta \\ y' - \cos\theta \\ u \\ v \end{bmatrix}.
$$

In this paper, we never analyze the transient dynamics of (2.4). Instead, we look for equilibrium solutions $\bar{\mathbf{z}}$ of (2.4) which satisfy

$$\mathbf{R}(\bar{\mathbf{z}}) = \mathbf{0}.$$

The stability of this solution can be determined by linearizing (2.4) about $\bar{\mathbf{z}}$. After substituting an expansion of $\bar{\mathbf{z}}$ ($\bar{\mathbf{z}} + \bar{\mathbf{z}}_\mathbf{1}$) into (2.4), we obtain the linear system

$$\mathbf{M}\dot{\bar{\mathbf{z}}}_\mathbf{1} = \mathbf{R}_\mathbf{z}(\bar{\mathbf{z}})\bar{\mathbf{z}}_\mathbf{1},$$

where $\mathbf{R_z}(\bar{\mathbf{z}})$ is the Fréchet derivative.

Substituting $\bar{\mathbf{z}}_\mathbf{1} = e^{\sigma t}\phi$ into the linearized system, we obtain

$$\sigma \mathbf{M}\phi = \mathbf{R_z}(\bar{\mathbf{z}})\phi. \tag{2.5}$$

Equation (2.5) is a generalized eigenvalue problem for the continuous problem with eigenfunction ϕ and eigenvalue σ. If all of the eigenvalues of our eigenvalue problem have negative real parts, the equilibrium solution $\bar{\mathbf{z}}$ is stable.

2.4. Linearization about standing cantilever. Equation (2.4) may not be familiar to readers with a background in beam theory. Linearizing, however, about the standing cantilever solution would result in a set of equations that more closely resembles the set of equations from beam theory. We simply evaluate the Fréchet derivative at the standing cantilever fixed point ($f_x = 0,\ f_y = -\mu + \lambda(s-1),\ x = 0,\ y = s$) in (2.5). After manipulating the resulting set of equations, we obtain a fourth-order differential equation in x. The stability of the standing cantilever fixed point is determined from the eigenvalues of the resulting linear operator

$$L(\phi) = \phi'''' - (((s-1)\lambda - \mu)\phi')' = -\sigma^2\phi \tag{2.6}$$

with the associated boundary conditions

$$\phi(0) = \phi'(0) = 0, \quad \phi''(1) = 0, \quad \phi'''(1) = \gamma\sigma\phi. \tag{2.7}$$

This linear operator has several interesting properties. First, L is not self-adjoint because of the boundary condition at the free end: $\phi'''(1) = \gamma\sigma\phi$. If we made a tiny change in this boundary condition,

$$\phi'''(1) + \mu\phi'(1) = 0,$$

the linear operator would be self-adjoint. This boundary condition change would result in a linear operator which represents the classical equations describing the buckling of a beam.

If $\gamma = 0$, the linear operator L exhibits time-reversal symmetry since, if σ is an eigenvalue of this linear operator, then so is $-\sigma$. The linear operator is also non-self-adjoint when $\gamma = 0$. Because of time-reversal symmetry, the eigenvalues of the linear operator for the undamped case indicate only neutral stability or instability. Numerically, we have determined (with $\alpha = 0$) that the beam loses neutral stability when $\mu = 20$. When a small amount of damping γ is introduced, however, the beam loses stability when $\mu = 16$. This apparent paradox was noted by Païdoussis [15] and will be explained in section 4.3.

We will need the adjoint eigenvalue problem when we apply the perturbation theory of eigenvalues in section 4.3. For the purpose of computing this adjoint, we will define the following inner product:

$$\langle f, g\rangle = \int_0^1 f^* g(s)ds,$$

where f^* is the complex conjugate of f. Since we will deal exclusively with real eigenfunctions in the neutrally stable regime, we will drop the complex conjugate notation. The adjoint of our linear operator L^* satisfies the following inner product:

$$\langle \psi, L(\phi)\rangle = \langle L^*\psi, \phi\rangle.$$

Taking the inner product of an arbitrary function ψ with the linear operator in (2.6), we obtain

$$\langle \psi, L(\phi)\rangle = \int_0^1 \psi(\phi'''' - ((\lambda(s-1)-\mu)\phi')')ds.$$

We can obtain the adjoint linear operator by repeatedly integrating this equation by parts to obtain

$$\begin{aligned}\langle \psi, L(\phi)\rangle = \int_0^1 \phi(\psi'''' - ((\lambda(s-1)-\mu)\psi')')ds + \psi\phi'''\big|_0^1 - \psi'\phi''\big|_0^1 \\ + \ \psi''\phi'\big|_0^1 - \psi'''\phi\big|_0^1 + \psi((\lambda(s-1)-\mu)\phi')'\big|_0^1 - \psi'(\lambda(s-1)-\mu)\phi'\big|_0^1.\end{aligned}$$

Several of these terms evaluated at the boundaries vanish due to the boundary conditions on ϕ except

$$\begin{aligned}\mu(\psi(1)\phi'(1) - \psi'(1)\phi(1)) - \psi'''(1)\phi(1) \\ + \psi''(1)\phi'(1) + \psi''(1)\phi'(1) + \psi'(0)\phi''(0) + \psi(0)\phi'''(0).\end{aligned}$$

We will specify the boundary conditions of the adjoint eigenvalue problem so that these terms vanish as well:

$$\begin{aligned}\psi(0) = \psi'(0) &= 0,\\ \psi'''(1) + \mu\psi'(1) &= 0,\\ \psi''(1) + \mu\psi(1) &= 0.\end{aligned}$$

The adjoint eigenvalue problem follows:

$$L^*(\psi) = \psi'''' - (((s-1)\lambda - \mu)\psi')' \tag{2.8}$$

with the boundary conditions

$$\begin{aligned}\psi(0) = \psi'(0) &= 0,\\ \psi'''(1) + \mu\psi'(1) &= 0,\\ \psi''(1) + \mu\psi(1) &= 0.\end{aligned} \tag{2.9}$$

This adjoint linear operator will be useful when we examine the effects of damping on stability in section 4.3.

3. Numerical technique.

We approximated the derivatives that appear in the steady form of (2.4) using a Chebyshev collocation method. The approximating functions employed by this spectral method are Chebyshev polynomials which are infinitely differentiable global functions. When evaluated at the Gauss–Labotto points, this spectral method produces highly accurate approximations to the derivative [9].

We computed the steady-state solutions of our discrete approximation to (2.4) using a Newton–Raphson iteration,

$$\mathbf{R_z}(\mathbf{z_o})\delta\mathbf{z} = -\mathbf{R}(\mathbf{z_o}),$$

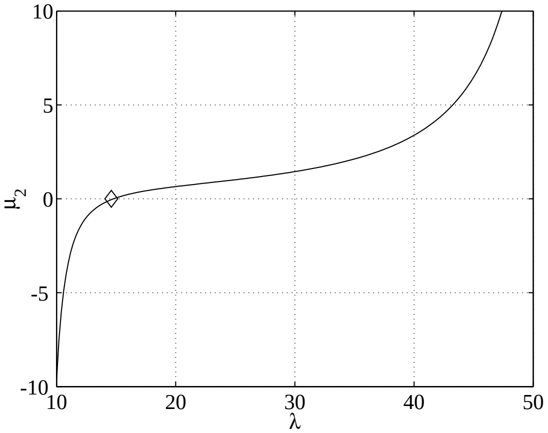


Figure 3.1. *This plot of μ_2 as a function of λ shows that μ_2 changes sign at approximately $\mu = 7.3447$ and $\lambda = 14$. The variation of the solution as a function of λ at this transition point (marked with a diamond) is shown in bifurcation diagram* (b) *in Figure* 4.4*. This transition point is called a quartic bifurcation point and is invariant under a change in the parameterization.*

where $\mathbf{z_o}$ is the solution at the previous iteration, $\delta\mathbf{z}$ is the update to $\mathbf{z_o}$ for this iterate, and $\mathbf{R_z}(\mathbf{z_o})$ is the Fréchet derivative. We also approximated the continuous eigenvalue problem in (2.4) with a discrete approximation to the Fréchet derivative, the vector of unknowns $\bar{\mathbf{z}}$, and the eigenfunction ϕ.

We used the arclength continuation, Hopf, pitchfork, and turning point tracking algorithms in LOCA to obtain the numerical bifurcation results in this paper. For more information on these algorithms, consult [14], [20], [21]. Branch switching was accomplished using an algorithm which perturbs the symmetric, unstable solution in the direction of the null vector, ϕ:

$$\mathbf{z}_{stable} = \mathbf{z}_{unstable} + \frac{\phi}{||\phi||}.$$

The transition from a supercritical to a subcritical pitchfork bifurcation can be determined using bifurcation theory. Werner and Spence discuss an analogous approach to detect whether a pitchfork bifurcation is supercritical or subcritical in [25]. We first transform our governing equations so that $\mathbf{z} = \mathbf{0}$ is the standing cantilever solution. We then introduce a regular perturbation expansion about $\mathbf{z_0} = \mathbf{0}$ to third order in ϵ. Let us choose as our bifurcation parameter the dimensionless follower force μ. We also expand this bifurcation parameter as follows:

$$\mu = \mu_0 + \epsilon\mu_1 + \epsilon^2\mu_2 + \epsilon^3\mu_3.$$

We substitute these expansions into our equations and retain terms up to third order in ϵ. If we were to collect terms of first order in ϵ, we obtain the eigenvalue problem at the bifurcation point:

$$\mathbf{J}(0, \mu_0)\mathbf{z_1} = 0.$$

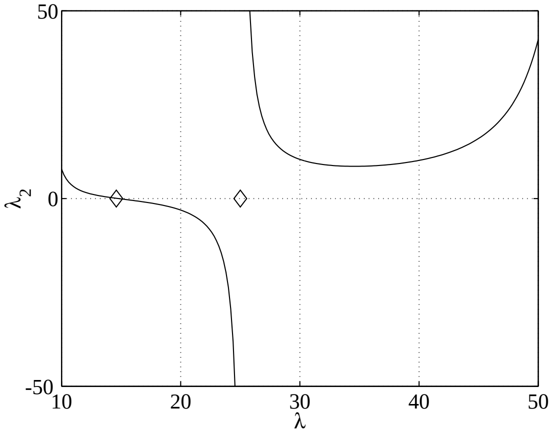


Figure 3.2. *This plot of λ_2 as a function of λ shows that there are two points (marked with diamonds) where λ_2 changes sign. The point that is also predicted in Figure 3.1 is special because it is invariant under a change in the parameterization, whereas the second point predicted when λ is chosen as the bifurcation parameter is an artifact of the parameterization we take but still accurately predicts a change in the criticality of the pitchfork bifurcation.*

Collecting terms of second order in ϵ, we obtain an equation of the form

$$\mathbf{J}(\mathbf{0},\mu_0)\mathbf{z_2}+\mu_1\beta=\mathbf{Q}(\mathbf{z_1}).$$

Finally, collecting terms of third order in ϵ, we obtain an equation of the form

$$\mathbf{J}(\mathbf{0},\mu_0)\mathbf{z_3}+\mu_2\beta=\mathbf{C}(\mathbf{z_1},\mathbf{z_2}).$$

We also have the normalization condition, which provides a nontrivial solution to these equations:

$$\phi\cdot\mathbf{z_k}=0,\quad k=1,2.$$

By solving the following system of equations simultaneously for $\mu_{k-1}, k=1,2$, we can obtain the sign of μ_2:

$$\begin{bmatrix}\mathbf{J}(0,\mu_0) & \beta\\ \phi & 0\end{bmatrix}\begin{bmatrix}\mathbf{z_k}\\ \mu_{k-1}\end{bmatrix}=\begin{bmatrix}\mathbf{R}(\mathbf{z_{k-1}})\\ 0\end{bmatrix}.$$

The pitchfork bifurcation is supercritical or subcritical based on the sign of μ_2. We solved these equations numerically and obtained the transition point (for $\alpha=0$) at $\mu=7.3447$, as seen in bifurcation diagram (b) in Figure 4.4. The variation of μ_2 as a function of λ is shown in Figure 3.1. If we were to choose the dimensionless gravity λ as our bifurcation parameter, we obtain two points where λ_2 changes sign once through 0 and once through ∞. The first transition point is the quartic bifurcation point, which is invariant under a change in the parameterization. The second transition point is an artifact of the parameterization we chose but is nonetheless a meaningful indicator of the change in criticality. A plot of λ_2 as a function of λ is shown in Figure 3.2.

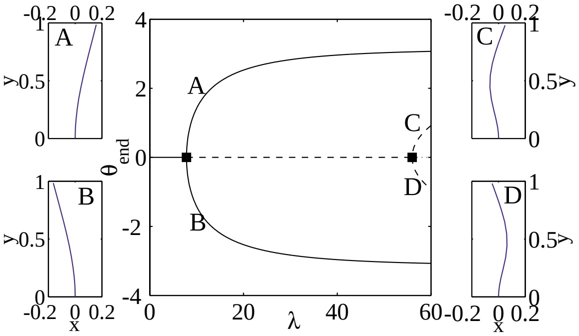


Figure 4.1. *The full and dashed lines represent stable and unstable solutions, respectively, and the square marker indicates a pitchfork bifurcation point. The symmetric solution is unstable past the pitchfork bifurcation point and tends toward the two stable branches labeled* A *and* B *in the diagram. Sample bifurcated solutions are shown.* A *and* B *correspond to the first unstable mode, and* C *and* D *correspond to the second unstable mode. The standing cantilever fixed point, not shown, corresponds to the horizontal* $\theta = 0$ *branch.*

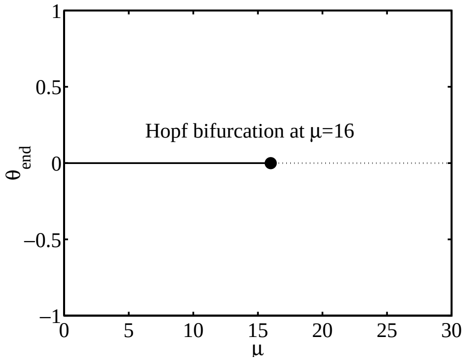
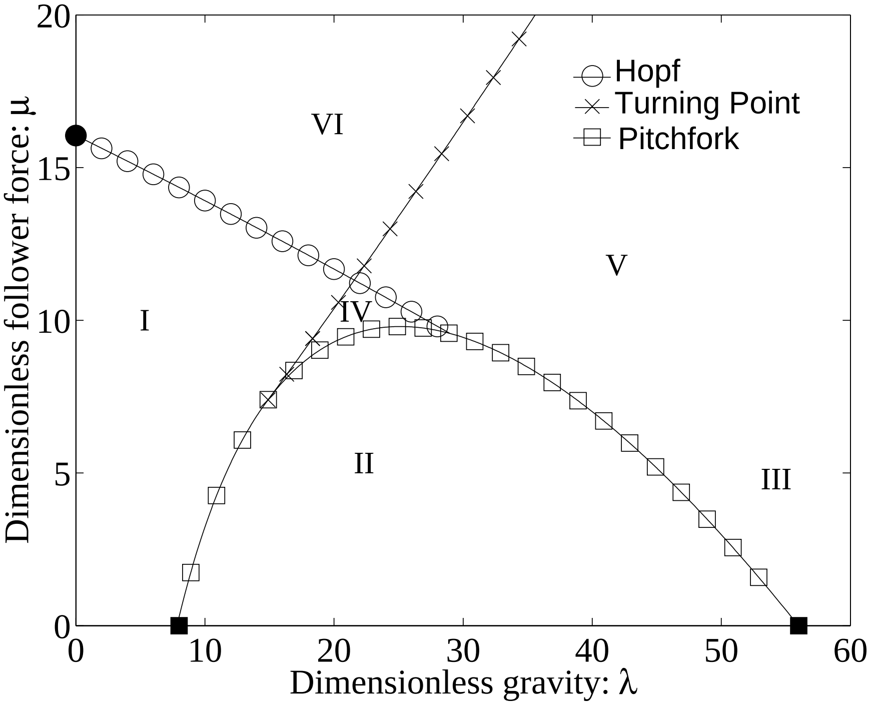


Figure 4.2. *The full and dotted lines represent stable and doubly unstable solutions, respectively. The value of* μ *at the onset of oscillatory behavior is shown in this figure.*

4. Results and analysis.

4.1. Numerical bifurcation results. Figure 1.1 illustrates the beam's orientation as we vary α. When $\alpha = 0$, the standing cantilever is a solution for all parameter values, but it may not be stable to small perturbations. In this section, we will explore the standing cantilever's stability in the parameter plane defined by λ and μ when the underlying equations of the beam exhibit reflectional symmetry $\alpha = 0.0$.

Consider the case when there is no follower force or $\mu = 0.0$. When gravity is pointing toward the clamped end of the beam, the standing cantilever will buckle under its own weight at a critical value of λ. As in the Euler beam problem, the standing cantilever will also have a second mode of instability at another critical value of λ. These points of instability are shown in Figure 4.1 with sample buckled solutions. This stationary bifurcation point is characterized as being supercritical since the branched solutions are stable and occur after the symmetric solution loses stability. Now consider the case when $\lambda = 0.0$. The beam experiences flutter at a critical value of the dimensionless follower force μ. This dynamic instability corresponds to



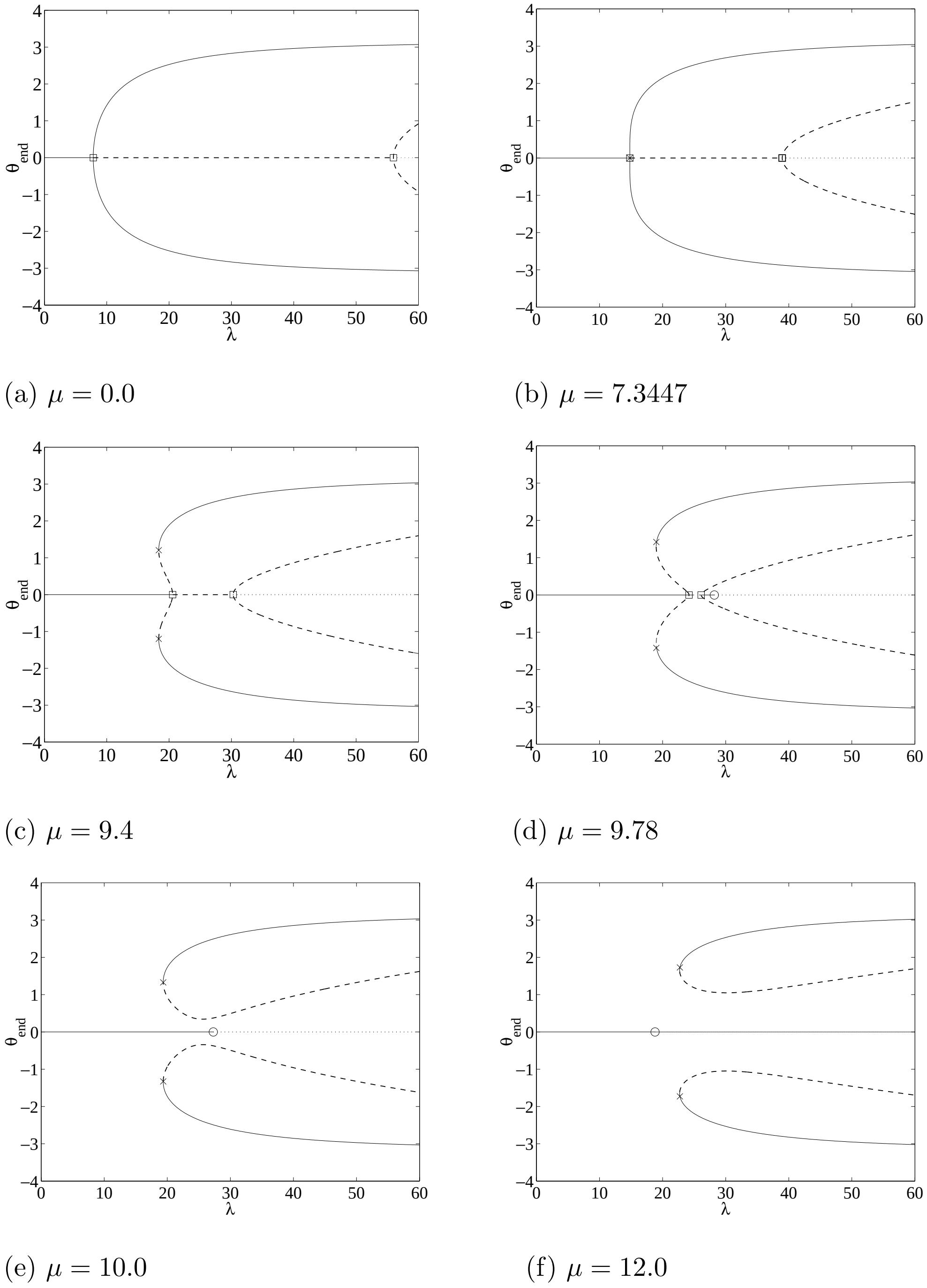
I. one symmetric solution
II. two buckled solutions
III. two buckled solutions
IV. one symmetric and two buckled solutions
V. one oscillatory solution
VI. oscillatory and buckled solutions

Figure 4.3. *For $\alpha = 0.0$ and $\gamma = 1.0$. Each region indicated with a Roman numeral and demarcated by the curves of bifurcation points has different stable solutions defined above. This two-parameter plot shows all solutions in this parameter-space. Please view Figure 4.4 for a better understanding of how the solutions vary as a function of μ and λ. Note that this figure contains two different bifurcation boundaries: one from the trivial state (the curve of pitchfork and Hopf bifurcations) and the other from an already buckled state (the curve of turning points).*

a Hopf bifurcation point and is shown in Figure 4.2. The dotted line in the figure represents an unstable solution which tends to an oscillatory solution. Since we know the bifurcations that the beam experiences for the trivial cases when $\mu = 0.0$ and $\lambda = 0.0$, we can use the tracking capabilities in LOCA to obtain the curve of pitchfork and Hopf bifurcation points.

As seen in Figure 4.3, the curve of pitchfork bifurcation points and the curve of Hopf bifurcation points intersect at a point which is a high codimension bifurcation. This point appears to be accounted for by Païdoussis in his stability map of the boundaries of buckling and oscillatory instabilities [15]. His stability map, however, misses the quartic bifurcation point, which we were able to obtain using an algorithm discussed in section 3 that detects whether a pitchfork bifurcation point is supercritical or subcritical.

By using this additional capability, we can obtain the quartic bifurcation point shown in Figure 4.3. This point is another a high codimension bifurcation point. Using the theory of normal forms, we show in section 4.2 that another curve of turning points should come tangent to this curve. This conclusion based on theory inspired us to search for the curves of turning



(a) $\mu = 0.0$

(b) $\mu = 7.3447$

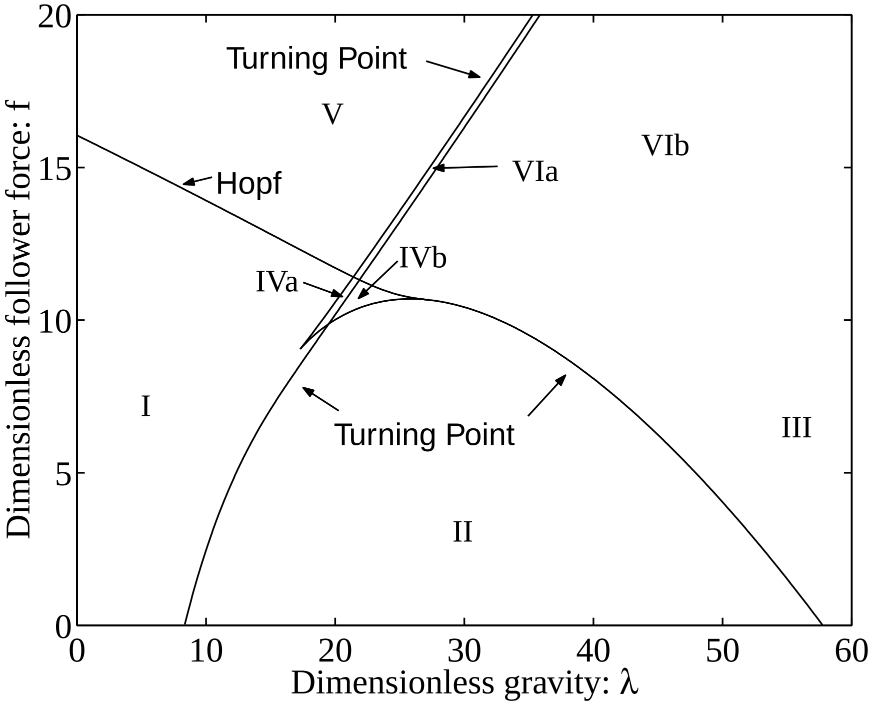
(c) $\mu = 9.4$

(d) $\mu = 9.78$

(e) $\mu = 10.0$

(f) $\mu = 12.0$

Figure 4.4. *For $\alpha = 0.0$ and $\gamma = 1.0$. A value characteristic of our entire solution θ_{end} is plotted as a function of λ for a variety of f. The full, dashed, and dotted lines represent stable solutions, solutions with one unstable mode, and doubly unstable solutions, respectively. Square, circle, and X markers are used to denote a pitchfork bifurcation, Hopf bifurcation, and turning point, respectively. Each bifurcation diagram was selected to represent a significant section of Figure 4.3. As the dimensionless follower force increases, the two pitchfork bifurcation points tend toward each other until they coalesce, leaving a Hopf bifurcation and two turning points in their wake.*

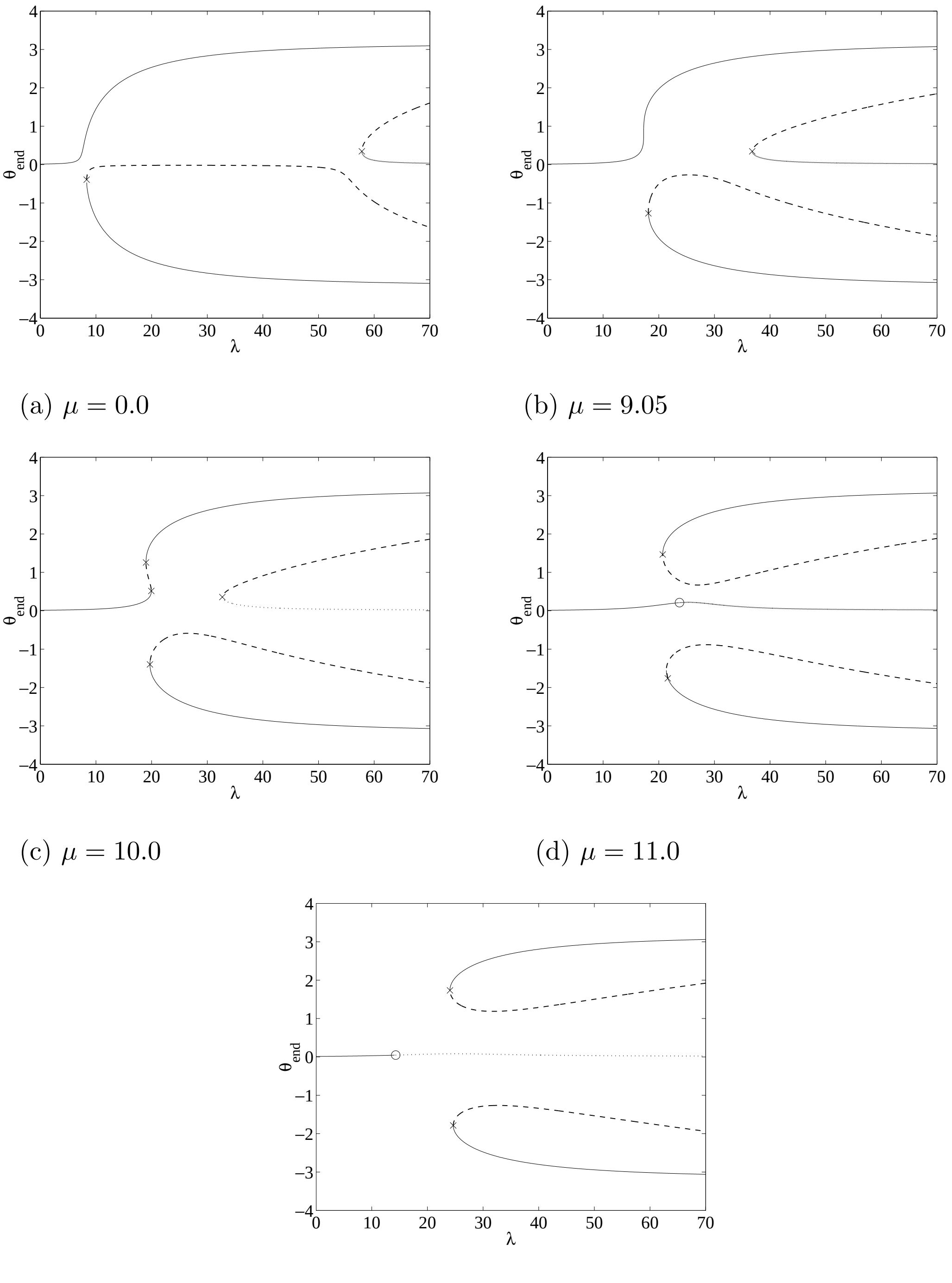


I. one near-symmetric/buckled solution
II. two buckled solutions
III. two buckled solutions
IVa. one near-symmetric and one buckled solution
IVb. one near-symmetric and two buckled solutions
V. one oscillatory solution
VIa. one oscillatory and one buckled solution
VIb. one oscillatory and two buckled solutions

Figure 4.5. *For $\alpha = 0.0125$ and $\gamma = 1.0$. When we break the symmetry by the introduction of a deflection $\alpha = 0.0125$, we obtain the following two-parameter plot. Notice that the curve of pitchfork bifurcations splits into two curves of turning points. The region of stability* (I) *does not appear to have increased dramatically. The near-symmetric/buckled modifier implies a solution which continuously transitions from a stable, almost symmetric solution to a buckled solution.*

points which appear in Figure 4.3. It should be noted that there are two curves of turning points shown in Figure 4.3 which happen to lie on the same curve in parameter space. Before we unfold these two higher codimension bifurcation points in α, we will highlight the effect the dimensionless follower force, gravity, and damping have on the boundaries of instability.

As can be seen in Figure 4.3, gravity makes the beam more likely to flutter and hence destabilizes the beam. This result agrees with our physical intuition that gravity is a force tending to make the beam oscillate. Contrary to physical intuition, the follower force μ makes the beam less likely to buckle and hence stabilizes the beam. This apparent paradox can be explained in the following way: if the beam is perturbed in the direction of its buckled state,



(a) $\mu = 0.0$

(b) $\mu = 9.05$

(c) $\mu = 10.0$

(d) $\mu = 11.0$

(e) $\mu = 13.0$

Figure 4.6. *For $\alpha = 0.0125$ and $\gamma = 1.0$. A value characteristic of our entire solution θ_{end} is plotted as a function of λ for a variety of f. The full, dashed, and dotted lines represent stable solutions, solutions with one unstable mode, and doubly unstable solutions, respectively. A circle and X symbols are used to denote a Hopf bifurcation and turning point, respectively.*

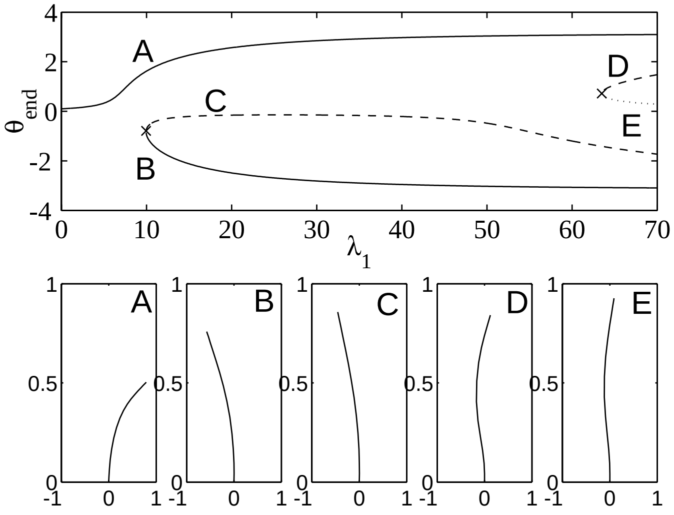


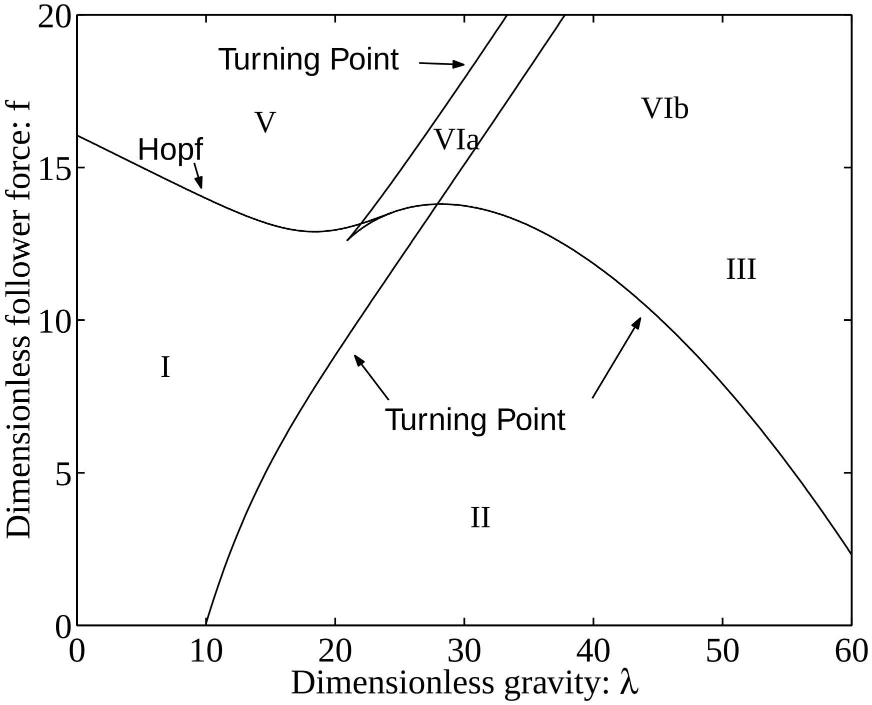
Figure 4.7. *For $\alpha = 0.1$ and $\gamma = 1.0$. The full and dashed lines represent stable and unstable solutions, respectively, and the X marker indicates a turning point. Clearly the solution corresponding to* A *is favored over solutions on branch* B*. The same pattern holds for the rest of the bifurcation diagram which corresponded to the second pitchfork bifurcation point in the symmetric case shown in Figure* 4.1*.*

then the follower force tends to push the beam back to the standing cantilever fixed point. The effect of the follower force agrees with Païdoussis' result that standing cantilevers which would ordinarily buckle without flow can actually become more stable with flow for a certain range of flow rates [15].

The follower force, however, also appears to excite the second pitchfork bifurcation mode by decreasing the critical value of λ, marking the onset of the second pitchfork instability. This phenomenon is apparent in bifurcation diagrams (a) and (b) in Figure 4.4. These bifurcation diagrams plot a characteristic value of our solution, the angle at the end of the beam, as a function of λ for fixed μ.

Bifurcation diagram (b) in Figure 4.4 marks the quartic bifurcation point—the transition between a supercritical and a subcritical pitchfork bifurcation. This point corresponds to the intersection of the turning point curves and the pitchfork bifurcation curve in Figure 4.3. The pitchfork bifurcation becomes subcritical because the branched solutions are unstable and occur when the symmetric solution is stable.

As the value of μ continues to increase, we notice the emergence of the Hopf bifurcation point in bifurcation diagram (d) in Figure 4.4. The second mode of instability becomes so excited by the increased follower force and the first mode so subdued by the stabilizing effect of the follower force that the two points coalesce at the higher codimension bifurcation point, where the Hopf bifurcation curve intersects the curve of pitchfork bifurcation points. As can be seen in bifurcation diagram (e), a Hopf bifurcation and two turning points remain in their wake. At a sufficiently high follower force, the Hopf bifurcation emerges as the first instability the symmetric system experiences as λ is increased. Bifurcation diagram (f) in Figure 4.4 shows that two turning points still remain, but the buckled solutions no longer coexist with a stable symmetric solution.

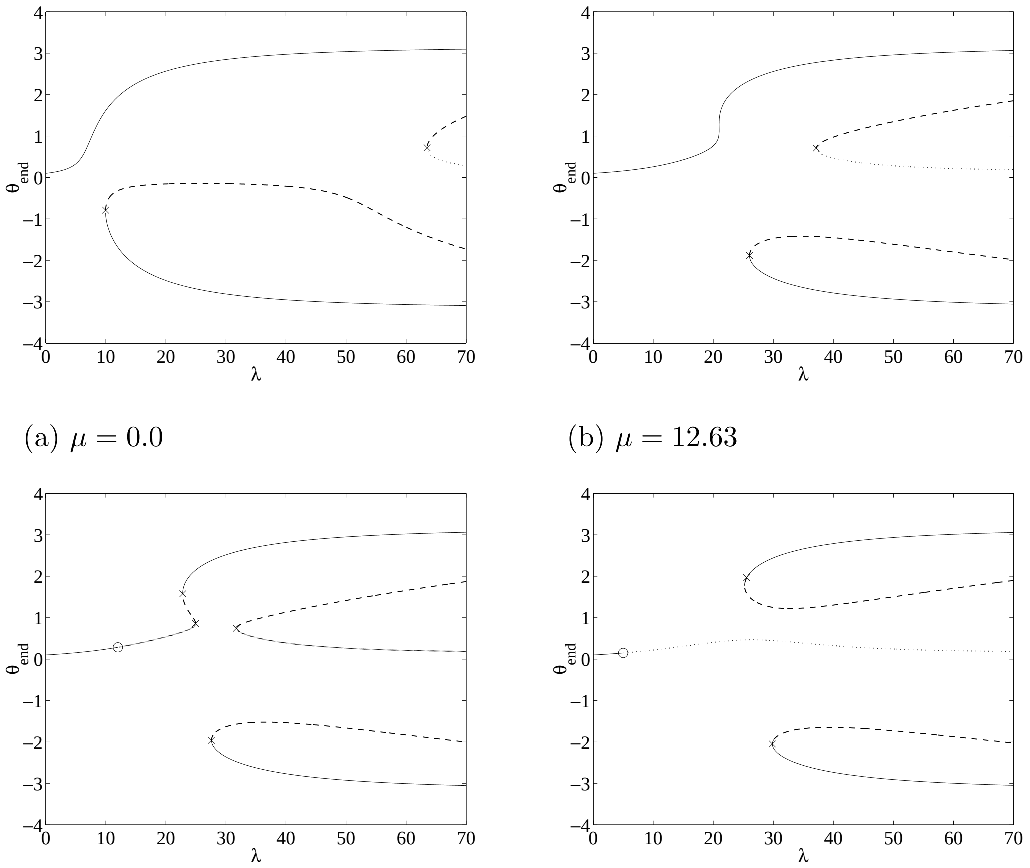


I. one near-symmetric/buckled solution
II. two buckled solutions
III. two buckled solutions
IV. one near-symmetric/buckled and two buckled solutions
V. oscillatory solution
VIa. one oscillatory and one buckled solution
VIb. one oscillatory and two buckled solutions

Figure 4.8. *For $\alpha = 0.1$ and $\gamma = 1.0$. Each region indicated with a Roman numeral and demarcated by the curves of bifurcation points has different stable solutions outlined above. The near-symmetric solution continuously transitions from a slightly deflected solution to a buckled solution in region* I *but never experiences a turning point or bifurcation.*

For $\lambda = 0$, without damping, the standing cantilever was neutrally stable up until $\mu = 20.0$, where the beam experienced flutter. With a small amount of damping, γ, the beam experienced flutter at a smaller value of μ, leaving us puzzled by the prospect that damping had a destabilizing effect on the beam. This important finding was also noted by others in the literature. However, a more puzzling consequence of our implementation of damping is that the Hopf bifurcation point is absolutely independent of γ. These puzzling results are explained in section 4.3.

The series of bifurcation diagrams in Figure 4.4 are used to define the regions in Figure 4.3 by the type of stable solutions which exist. Region I indicates that all solutions tend toward the standing cantilever fixed point. In Region II, the symmetric fixed point loses stability to one of the buckled branches depending on the direction of the perturbation. This scenario



(a) $\mu = 0.0$

(b) $\mu = 12.63$

(c) $\mu = 13.6$

(d) $\mu = 15.0$

Figure 4.9. $\alpha = 0.1$. *A value characteristic of our entire solution θ_{end} is plotted as a function of λ for a variety of f. The full, dashed, and dotted lines represent stable solutions, solutions with one unstable mode, and doubly unstable solutions, respectively. A circle and X symbols are used to denote a Hopf bifurcation and turning point, respectively.*

remains true even after the second pitchfork bifurcation point is passed in Region III. The symmetric and buckled solutions are all stable in Region IV. It should be noted that the two-parameter plot can be misleading because it contains bifurcations of different solutions on the same parameter space. In order to correctly interpret Figure 4.3, please view the one-parameter plots in Figure 4.4. It should also be noted that Regions VI and V do not have well-defined boundaries in Figure 4.3. Region VI has only oscillatory solutions, while, in Region V, both steady and oscillatory solutions are stable. LOCA cannot predict the end of the oscillatory solutions that occurs when the limit cycle has a period which reaches ∞ and ceases to exist. Time-integration of the governing equations of the motion of the beam is one way to obtain this boundary.

What happens to the two higher codimension bifurcation points as we unfold them in α? The authors were unable to locate any work that discussed the unfolding of these higher codimension bifurcation points in α. When we deflect the beam slightly ($\alpha = 0.0125$), we obtain the two-parameter plot shown in Figure 4.5. We notice that the curve of pitchfork bifurcations has given rise to two curves of turning points. The quartic bifurcation point has unfolded into a codimension two bifurcation, where a curve of turning points terminates at a cusp, and a codimension one bifurcation corresponding to the remaining curve of turning points. The intersection of the Hopf bifurcation and the pitchfork bifurcation curves has unfolded into a double zero eigenvalue which marks where the Hopf bifurcation curve intersects the curve of turning points in Figure 4.5. The effects of unfolding become more pronounced as we continue to increase the deflection as seen in Figure 4.8. We will briefly highlight some features of the deflected beam.

With a small deflection, rather than either branched solution being equally possible, stability appears to be biased toward the direction of the deflection. In fact, the solution in the direction of the deflection is stable for the deflected case. This bias becomes especially evident in the bifurcation diagrams shown in Figure 4.6.

Bifurcation diagram Figure 4.6 (a) clearly shows that one stable solution branch connects the standing cantilever and one of the buckled branches. Buckling in this case is a continuous transition and is not characterized by a bifurcation. The near-vertical standing cantilever equilibria become disconnected from the unstable near-standing solution, which is typical of how a pitchfork bifurcation diagram looks after the symmetry is broken. The second pitchfork bifurcation mode behaves as expected as well when the symmetry is broken. As we increase the dimensionless follower force, we arrive at the cusp shown in Figure 4.5. Similar to the symmetric case, the second turning point moves closer to the stable branches, until the stable deflected branch merges with the second turning point, leaving a Hopf bifurcation point. By the time $\mu = 13$ in bifurcation diagram (e), we notice that the Hopf bifurcation point becomes the first instability the near-symmetric solution experiences.

For larger deflections ($\alpha = 0.1$), Figure 4.7 shows how the beam actually looks as we move on the various stable, unstable, and doubly unstable solution branches. Notice that the solution in the direction of the deflection is stable and clearly favored over the solution on the turning point branch. The two-parameter plot for the $\alpha = 0.1$ case is shown in Figure 4.8. The bifurcation diagrams shown in Figure 4.9 are again revealing. Bifurcation diagram (a) shows that the solution corresponding to the direction opposing the direction of the deflection is unfavored and disconnected from the near-symmetric solution. The near-symmetric solution is initially stable and remains stable even as it becomes more deflected and loses more symmetry. This stability is due to the continuous nature of the transition from the near-symmetric solution to what we would consider a buckled solution. As the dimensionless follower force increases, we notice that the second turning point moves closer to the stable solution branch. By the time $\mu = 13.6$, the beam can flutter. The beam, however, restabilizes and continues on the deflected solution branch. The second pair of turning points then merges, leaving a Hopf bifurcation in its wake.

4.2. The quartic bifurcation point. By applying the theory of normal forms, we can obtain a one-dimensional equation which can validate the topological behavior of the two-parameter plots we obtained in the neighborhood of the quartic bifurcation point we noticed in Figure 4.3. At this point, the beam exhibits symmetry, and, because we are at a transition between a subcritical and a supercritical pitchfork bifurcation, we can ignore terms of order greater than five in x (since $F_{xxx} = 0$). We will introduce parameters (a, b) such that, at the quartic bifurcation point, $a = b = 0$. The one-dimensional equation which represents the basic topological behavior in the vicinity of a quartic bifurcation point follows:

$$g(x) = \pm x^5 + ax^3 + bx = 0, \tag{4.1}$$

where a and b are parameters which we can vary in this equation. This normal form is a standard example in the work of Golubitsky and Schaeffer [11]. We will choose the sign in front of the quintic term in (4.1) to be negative because the resulting equation better resembles the topological behavior in Figure 4.3. The solution, $x = 0$, corresponds to the symmetric solution. Furthermore, for any given value of a, $b = 0$ corresponds to the locus of pitchfork bifurcation points, where $\frac{\partial g}{\partial x}(0) = 0$. Also notice that, for $a < 0$, the pitchfork bifurcation point is supercritical and, for $a > 0$, the pitchfork bifurcation point is subcritical.

We are interested in seeing if there are any curves of turning points in the neighborhood of this quartic bifurcation point. By definition, at a turning point, the following is true:

$$\frac{dg}{dx}(x) = -5x^4 + 3ax^2 + b = 0. \tag{4.2}$$

Neglecting the trivial solution which corresponds to the locus of pitchfork bifurcation points, we can solve (4.1) and (4.2) to obtain an expression for the locus of turning points near the quartic bifurcation:

$$\frac{a^2}{4} + b = 0, \qquad x^2 = \frac{a}{2}.$$

A schematic of b as a function of a is given in Figure 4.10. Notice the locus of turning points comes tangent to the locus of pitchforks and terminates at the quartic bifurcation point. Also notice that the curve of turning points is actually two curves of turning points which happen to lie on the same curve in parameter space and correspond to the two different solutions: $x = \pm\frac{\sqrt{a}}{\sqrt{2}}$. This topological behavior is seen in Figure 4.3.

Figure 4.5 shows how the quartic bifurcation point unfolds when we break the symmetry in the beam equations. There are two notable features in the unfolding: a curve of turning points which terminates at a cusp and another curve of turning points nearby. Let us now show that, when we break the symmetry in the normal form (4.1), we also obtain these features. The asymmetric normal form follows:

$$g(x) = -x^5 + ax^3 + dx^2 + bx + c = 0. \tag{4.3}$$

We begin by looking for a cusp. When $d = 0$, the conditions for a cusp can be written as

$$\begin{aligned} g(x) &= -x^5 + ax^3 + bx + c = 0, \\ \frac{dg}{dx}(x) &= -5x^4 + 3ax^2 + b = 0, \\ \frac{d^2g}{dx^2}(x) &= -20x^3 + 6ax = 0. \end{aligned}$$

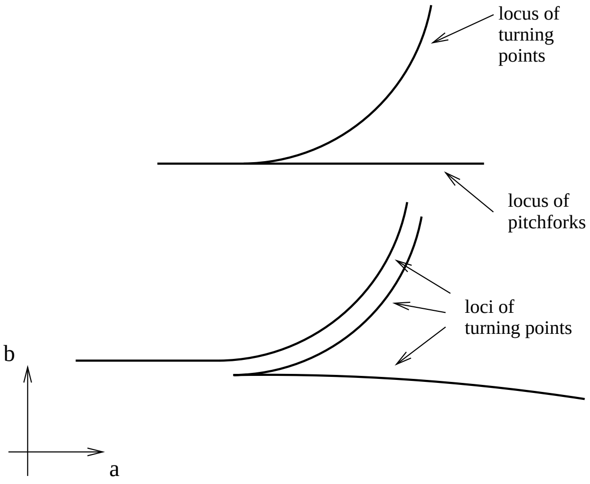


Figure 4.10. *From the top: A schematic of the locus of turning points near the quartic bifurcation point and a schematic of the locus of turning points after the symmetry is broken. The top schematic agrees with the topological behavior in the neighborhood of the quartic bifurcation point in Figure* 4.3*. The bottom schematic agrees with Figure* 4.5*, which shows the unfolding of the quartic bifurcation point.*

Solving these equations we obtain $a = \frac{5|c|^{2/5}}{2^{1/5}3^{3/5}}$, $b = -\frac{3^{7/5}|c|^{2/5}}{4 \cdot 2^{1/5}}$, and $x = \frac{3^{2/5}|c|^{2/5}}{2^{6/5}}$. Thus, for each value of c with $d = 0$, there is one and only one cusp. For $d \neq 0$, we can invoke the implicit function theorem to conclude that, for small d, this conclusion still holds.

In the symmetric normal form, we noticed that two curves of turning points happened to lie on the same curve in parameter space. When we break the symmetry, these curves of turning points split apart in parameter space, as shown in Figure 4.10. One of these curves terminates at a cusp, and the other is a locus of turning points that continues on and merges into what used to be the supercritical pitchforks.

4.3. Effect of damping.

When $\gamma = 0.0$, we noted in section 2.4 that the eigenvalue problem obtained from linearizing about the standing cantilever fixed point could only predict neutral stability or instability. In section 4.1, we found that, at $\mu = 20.0$, the beam experiences flutter. For $\mu < 20.0$, the neutrally stable modes do not necessarily tell us anything about the stability of the standing cantilever. One would assume that, when we add damping, those neutrally stable modes would become stable. We noticed numerically, however, that damping made some of those neutrally stable modes unstable. In addition, we analyzed the effect damping had on this point of instability by introducing extreme values of damping ($\gamma = 10^{-5}$ and $\gamma = 10^{5}$) and observed that damping had no effect on this point of instability. These deeply puzzling results are analytically clarified in this section.

By applying the perturbation theory of eigenvalues, we can determine the effect a small amount of damping would have on the sign of the real part of the perturbation in the eigenvalue. If the real part of the perturbation is positive, then the damped solution becomes unstable, and, if the real part of the perturbation is negative, then we know the solution will become stable. Assuming γ is small, let us introduce the following expansions for this purpose:

$$\begin{aligned} \phi &= \phi_0 + \gamma\phi_1, \\ \sigma &= \sigma_0 + \gamma\sigma_1. \end{aligned}$$

Substituting these expansions into (2.6) and collecting terms up to first-order in γ, we obtain

$$\begin{aligned} L(\phi_1) + \sigma_0^2\phi_1 &= -2\sigma_0\sigma_1\phi_0, \\ \phi_1(0) = \phi_1'(0) &= 0, \\ \phi_1''(0) &= 0, \\ \phi_1'''(1) &= \sigma_0\phi_0(1). \end{aligned} \tag{4.4}$$

Using Fredholm's alternative, we can take the inner product of both sides of the first equation in (4.4) with ψ_0, which satisfies the linear operator in (2.8) and the adjoint boundary conditions in (2.9). Repeatedly integrating this resulting equation by parts, we can then apply the boundary conditions in ϕ_0 and use the fact that $L\psi_0 + \sigma_0^2\psi_0 = 0$ to obtain the following equation:

$$\psi_0(1)\sigma_0\phi_0(1) = -2\int_0^1 \psi_0\sigma_0\sigma_1\phi_0 ds.$$

Finally, solving for σ_1, we obtain

$$\sigma_1 = -\frac{\psi_0(1)\phi_0(1)}{2\int_0^1 \psi_0\phi_0 ds}. \tag{4.5}$$

When there is no follower force ($\mu = 0$ and hence $\phi_0 = \psi_0$), σ_1 is negative. Therefore, a little bit of damping stabilizes the standing cantilever when there is no follower force. For small μ, we would expect σ_1 to remain negative since $\phi_0 \approx \psi_0$. As we increase μ, we can solve the purely inviscid problem and keep track of the undamped right and left eigenfunctions to obtain the point where a small amount of damping makes the standing cantilever unstable. σ_1 changes sign when $\psi_0(1)$, $\phi_0(1)$, or the denominator changes sign. Numerically, we have identified that $\psi_0(1)$ changes sign when $\mu = 16.0$. This important result proves that a point source of damping destabilizes the beam.

We have shown that an instability occurs for small values of γ at $\mu = 16.0$ and $\psi_0(1) = 0.0$. However, why does this point of instability hold for all values of γ? How can we explain this special property of the point source of damping in our model of the beam? Since $\psi_0(1) = 0$ at the point of instability, one of the adjoint boundary conditions in (2.9) becomes

$$\psi_0''(1) = 0.$$

Therefore, both ϕ_0 and ψ_0 satisfy all of the boundary conditions for the damped problem at the point of instability except for $\phi'''(1) = \sigma_0(1)\phi_0(1)$. We can thus obtain an eigenfunction which satisfies the damped linear operator and boundary conditions using a linear combination of the undamped left and right eigenfunctions ϕ_0 and ψ_0. Since we can determine an eigenfunction $\phi^d(\gamma)$ for the damped linear operator in (2.6) that ensures that the eigenvalue remains unchanged as we vary γ, we have proven that the point of instability is independent of damping. It turns out that this particular linear combination is

$$\phi = \phi_0 + k\gamma\sigma_0\psi_0,$$

where $k = -\frac{\phi_0(1)}{\mu\psi_0'(1)}$.

We have shown that, as long as a point source of damping is nonzero, damping has no effect on the Hopf bifurcation point. It should be emphasized that this is an interesting property of what is most probably a degenerate model and does not hold for distributed damping or even two point sources of damping for that matter. However, just as we learned a tremendous amount from the extremely degenerate high codimension bifurcations, we hypothesize that this degenerate point foretells that distributed damping would have a nominal effect on the point of instability. Our hypothesis is strengthened by the work of Païdoussis, who showed in Table 2 of his paper that the effect external distributed damping had on the critical flow velocity for the standing cantilever is small [15]. A more physical explanation of why damping actually destabilizes the beam follows, culminating in a discussion that restates conclusions made by Benjamin about the physical mechanism behind flutter [4].

4.4. Physical mechanism behind flutter.

Let us assume that $\sigma = i\omega$. By construction, a fixed point of the beam takes the form

$$\mathbf{x} = \Re(\phi_1 e^{i\omega t}),$$

where ϕ_1 is the eigenfunction of the linear operator appearing in (2.6). The work performed by the follower force is approximately

$$W = \mu \int_0^t u(1)\theta(1)dt.$$

We stated in section 2.4 that, when we linearize about the standing cantilever fixed point, we obtain a fourth-order differential equation in x, and we used ϕ_1 to denote the eigenfunction. For small deflections, θ is simply equal to ϕ_1' since, in (2.4), $x' = \sin(\theta)$. Putting this all together, $u(1)$ and $\theta(1)$ have the following form:

$$u(1) = \Re(i\omega\phi_1(1)e^{i\omega t}),$$
$$\theta(1) = \Re(\phi_1'(1)e^{i\omega t}).$$

Without loss of generality, we can normalize the eigenfunction ϕ_1 such that

$$\phi_1(1) = 1,$$
$$\phi_1'(1) = Ae^{i\beta},$$

where A and β are real constants obtained after normalizing the eigenfunction so that $\phi_1(1) = 1$. The u velocity and θ in terms of this normalized eigenfunction are

$$u(1) = -\omega\sin(\omega t),$$
$$\theta(1) = A\cos(\omega t + \beta),$$

so that the work performed by the follower force can be expressed as

$$W = -\mu \int_0^t \omega A \sin(\omega t) \cos(\omega t + \beta) dt.$$

This integral evaluates to zero over a full period, unless β is nonzero. When there is no damping $\gamma = 0$, then ϕ_1 and ϕ_1' are real and β is zero. Essentially, there is no energy transfer from the follower force to the beam when there is no damping. When there is damping, β becomes nonzero because ϕ_1 and ϕ_1' become complex. This energy due to the follower force is countered by the energy removed from the system by dissipation. At a certain point, the follower force outweighs the damping mechanism, causing the beam to become unstable. Thus the beam experiences an instability because the u velocity and θ become more in phase. These results are consistent with those found by Benjamin [4].

5. Conclusion. As a result of this research, we have outlined the development of a generic tool for detecting the criticality of a pitchfork bifurcation point. In the future, we will want to develop this tool for large-scale stability problems. Moreover, because a point source of damping had no influence on the Hopf bifurcation point, we predicted that distributed damping would have a nominal effect on stability. We would like to test the validity of this claim by implementing global damping in future work. In this study, we used α only to unfold our high codimensional pitchfork bifurcations. Future work may consider α as a continuation parameter and examine bifurcations in α. We may also wish to identify the saddle-loop bifurcation curve missing in our two-parameter plot by time-integrating our equations of motion.

Acknowledgments. We thank Jerry Marsden and Simon Tavener for their help with this paper. We also thank Danny Sorensen and Richard Lehoucq for making this research opportunity available to the authors.

REFERENCES

[1] A. K. Bajaj, P. R. Sethna, and T. S. Lundgren, *Hopf bifurcation phenomena in tubes carrying a fluid*, SIAM J. Appl. Math., 39 (1980), pp. 213–230.

[2] A. K. Bajaj and P. R. Sethna, *Flow-induced bifurcations to three-dimensional oscillatory motions in continuous tubes*, SIAM J. Appl. Math., 44 (1984), pp. 270–286.

[3] A. K. Bajaj and P. R. Sethna, *Effect of symmetry-breaking on flow-induced oscillations in tubes*, J. Fluids and Structures, 5 (1991), pp. 651–679.

[4] T. B. Benjamin, *Dynamics of a system of articulated pipes conveying fluid.* I. *Theory*, Proc. Roy. Soc. Ser. A, 261 (1961), pp. 457–486.

[5] T. B. Benjamin, *The threefold classification of disturbances in flexible surfaces bounding inviscid flows*, J. Fluid Mech., 16 (1963), pp. 436–450.

[6] W. J. Beyn, *Numerical analysis of homoclinic orbits emanating from Takens-Bogdanov point*, IMA J. Numer. Anal., 14 (1994), pp. 381–410.

[7] 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.

[8] A. M. Bloch, P. S. Krishnaprasad, J. E. Marsden, and T. S. Ratiu, *The Euler-Poincaré equations and double bracket dissipation*, Comm. Math. Phys., 175 (1996), pp. 1–42.

[9] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics*, Springer-Verlag, New York, 1988, pp. 7–9.

[10] K. A. Cliffe and A. Spence, *The calculation of high order singularities in the finite Taylor problem*, in Numerical Methods for Bifurcation Problems, Internat. Schriftenreihe Numer. Math. 70, Birkhäuser, Basel, 1984, pp. 129–144.

[11] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. I, Springer-Verlag, New York, 1985, p. 267.

[12] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983, pp. 290–295, 364–376.

[13] J. E. Marsden and J. Scheurle, *Lagrangian reduction and the double spherical pendulum*, Z. Angew. Math. Phys., 44 (1993), pp. 17–43.

[14] G. Moore and A. Spence, *The calculation of turning points of nonlinear equations*, SIAM J. Numer. Anal., 17 (1980), pp. 567–576.

[15] M. P. Païdoussis, *Dynamics of tubular cantilevers conveying fluid*, J. Mech. Engrg. Sci., 12 (1970), pp. 85–103.

[16] M. P. Païdoussis, *Fluid-Structure Interactions: Slender Structures and Axial Flow*, Academic Press, New York, 1998, pp. 111–132.

[17] M. P. Païdoussis and N. T. Issid, *Dynamics stability of pipes conveying fluid*, J. Sound Vibration, 33 (1974), pp. 267–294.

[18] M. P. Païdoussis and G. X. Li, *Pipes conveying fluid: A model dynamical problem*, J. Fluids and Structures, 7 (1993), pp. 137–204.

[19] G. Pfister, H. Schmidt, K. A. Cliffe, and T. Mullin, *Bifurcation phenomena in Taylor-Couette flow in a very short annulus*, J. Fluid Mech., 191 (1988), pp. 1–18.

[20] A. G. Salinger, N. M. Bou-Rabee, E. A. Burroughs, R. B. Lehoucq, R. P. Pawlowski, L. A. Romero, and E. D. Wilkes, *LOCA* 1.0: *Library of Continuation Algorithms, Theory and Implementation Manual*, Sandia Technical Report SAND2002-0396, Sandia National Laboratories, Albuquerque, NM, 2002, available online from http://www.cs.sandia.gov/LOCA.

[21] A. G. Salinger, R. B. Lehoucq, R. P. Pawlowski, and J. N. Shadid, *Computational bifurcation and stability studies of the* 8:1 *thermal cavity problem*, Internat. J. Numer. Methods Fluids, to appear.

[22] C. Semler, G. X. Li, and M. P. Païdoussis, *The non-linear equations of motion of pipes conveying fluid*, J. Sound Vibration, 169 (1994), pp. 577–599.

[23] S. W. Shaw and P. R. Sethna, *On the effects of asymmetries on a system near a codimension two point*, in Dynamical Systems Approaches to Nonlinear Problems in Systems and Circuits, F. M. A. Salam and M. Levi, eds., SIAM, Philadelphia, 1988, pp. 317–332.

[24] S. W. Shaw and P. R. Sethna, *On codimension-three bifurcations in the motion of articulated tubes conveying a fluid*, Phys. D, 24 (1987), pp. 305–327.

[25] B. Werner and A. Spence, *The computation of symmetry-breaking bifurcation points*, SIAM J. Numer. Anal., 21 (1984), pp. 388–399.

[26] W. Wu, A. Spence, and K. A. Cliffe, *Steady-state Hopf mode interaction at a symmetry-breaking Takens-Bogdanov point*, IMA J. Numer. Anal., 14 (1994), pp. 137–160.

Attracting Fixed Points for the Kuramoto–Sivashinsky Equation: A Computer Assisted Proof[*]

Piotr Zgliczynski[†]

Abstract. We present a computer assisted proof of the existence of several attracting fixed points for the Kuramoto–Sivashinsky equation

$$u_t = (u^2)_x - u_{xx} - \nu u_{xxxx}, \quad u(x,t) = u(x+2\pi,t), \quad u(x,t) = -u(-x,t),$$

where $\nu > 0$. The method is general and can be applied to other dissipative PDEs.

Key words. dissipative PDEs, fixed points, Galerkin projection, computer assisted proof

AMS subject classifications. 35B35, 35B45, 65G20, 65N30

PII. S111111110240176X

1. Introduction.

The goal of this paper is to extend the method of self-consistent a priori bounds developed in [ZM, Z] for a rigorous study of dynamics of dissipative PDEs. We present an approach which allows us to show that a given fixed point for a PDE is asymptotically stable. We apply the method to the Kuramoto–Sivashinsky (KS) equation subject to periodic and odd boundary conditions

$$u_t = (u^2)_x - u_{xx} - \nu u_{xxxx}, \quad u(x,t) = u(x+2\pi,t), \quad u(x,t) = -u(-x,t), \tag{1.1}$$

where $\nu > 0$. While the method will be explained in detail later, we would like to stress here its basic ingredients, which will also explain how this paper is dependent on [Z] and [ZM] and what is new here.

The approach starts as in [ZM]:

1. We have to find an approximate attracting fixed point x_0 for some Galerkin projection of (1.1). Then we construct a trapping region around x_0 (which is an example of self-consistent a priori bounds defined in [ZM]) using the algorithm presented in [ZM]. From this we conclude that there exists a fixed point $x^* \in R$, but we cannot claim its asymptotic stability.
2. In paper [Z], we obtained estimates for the Lipschitz constants for the flow induced by the Navier–Stokes equations on two-dimensional torus. Here we adopt this approach to construct a norm for which the induced flow is a contraction around x^*.

In the present work, for each attracting branch from a nonrigorous steady state bifurcation diagram presented in [JKT], we picked up a point on it, and we proved that it is attracting.

Below we include some of the attracting steady states we had proved rigorously to exist.

[*] Received by the editors February 1, 2002; accepted for publication (in revised form) by B. Sandstede August 19, 2002; published electronically October 17, 2002. This research was supported by Polish State Committee for Scientific Research grants 2 P03A 019 22 and 2P03A 011 18 and by NSF grant DMS-9706903.
http://www.siam.org/journals/siads/1-2/40176.html

[†] Institute of Mathematics, Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland (zgliczyn@im.uj.edu.pl).

- $\nu \in 0.75 + [-10^{-2}, 10^{-2}]$, two stable unimodal fixed points.
- $\nu \in 0.5 + [-10^{-4}, 10^{-4}]$, two stable unimodal fixed points.
- $\nu \in 0.3 + [-10^{-4}, 10^{-4}]$, two stable unimodal fixed points.
- $\nu \in 0.125 + [-10^{-4}, 10^{-4}]$, one stable bimodal fixed point.
- $\nu \in 0.1 + [-10^{-4}, 10^{-4}]$, one bimodal stable fixed point.
- $\nu \in 0.08 + [-10^{-6}, 10^{-6}]$, one bimodal stable fixed point. A pair of stable fixed points close to $R_3 t_2$ (see [JKT]).
- $\nu \in 0.062 + [-10^{-6}, 10^{-6}]$, two stable trimodal points and two stable points from giant branch.
- $\nu \in 0.045 + [-10^{-6}, 10^{-6}]$, $\nu \in 0.04 + [-10^{-7}, 10^{-7}]$, two stable points from giant branch.

In the above listing, when we write that, for $\nu \in 0.75 + [-10^{-2}, 10^{-2}]$, we have two stable fixed points, this means that, for all ν in this interval, these stable fixed points exist. In section 4, we present an example of a precise theorem about an existence of a fixed point obtained using our method.

In sections 2 and 3, we present the method in detail, and we prove Theorem 3.8, which is the main tool in our approach. In section 4, we present an example of a precise theorem, give an outline of the algorithm, and present numerical data from the proof. In section 5, we derive various estimates for the KS equation required in the rigorous check of assumptions of Theorem 3.8. In section 6, we discuss the directions in which this work can be extended further.

2. Uniform convergence of Galerkin projections on a trapping region.

We adopt here the notation used in sections 4 and 5 in [Z]. Let H be a real Hilbert space. Let $e_1, e_2, \dots$ form an orthonormal basis in H.

In what follows, we will quite often denote the elements of H by x, and we hope it will not be confused with the space variable in (1.1).

Let $A_n : H \to \mathbb{R}$ denote a projection onto a one-dimensional subspace $\langle e_n \rangle$; i.e., $x = \sum A_n(x) e_n$ for all $x \in H$. By X_n we will denote a space spanned by $\{e_1, \dots, e_n\}$. Let P_n denote the projection onto X_n, $Q_n = I - P_n$.

For $x \in \mathbb{R}^n$ or $x \in H$, we set $|x|$ to be a standard (Euclidean) norm, $|x|_\infty = \max_i |x_i|$ and $|x|_1 = \sum_i |x_i|$.

We investigate the Galerkin projections of the problem

$$x' = F(x) = L(x) + N(x), \tag{2.1}$$

where L is a linear operator and N is a nonlinear part of F. We assume that the basis $e_1, e_2, \dots$ of H is built from eigenvectors of L. We assume that the corresponding eigenvalues λ_k (i.e., $Le_k = \lambda_k e_k$) are ordered so that

$$\lambda_1 \geq \lambda_2 \geq \dots \quad \text{and} \quad \lim_{k \to \infty} \lambda_k = -\infty.$$

Hence L can have only a finite number of positive eigenvalues.

Definition 2.1. *Let $W \subset H$ and $F : dom(F) \to H$, W be closed. We say that W and F satisfy conditions* C1, C2, *and* C3 *if the following hold:*

C1. *There exists $M \geq 0$ such that $P_n(W) \subset W$ for $n \geq M$.*

C2. *Let $\hat{u_k} = \max_{x \in W} |A_k x|$. Then $\hat{u} = \sum \hat{u}_k e_k \in H$. In particular, $|\hat{u}| < \infty$.*

C3. *The function $x \mapsto F(x)$ is continuous on W, and $f = \sum_k f_k e_k$, given by $f_k = \max_{x \in W} |A_k F(x)|$, is in H. In particular, $|f| < \infty$.*

Observe that, if $W_m \subset X_m$ and $\{a_k^-, a_k^+\}$ form self-consistent a priori bounds (see [ZM, Def. 2.1]) for F, then $W = W_m \oplus \Pi_{k=m+1}^{\infty} [a_k^-, a_k^+]$ and F satisfy conditions C1, C2, and C3.

Definition 2.2. *We say that $W \subset H$ and $F = N + L$ satisfy condition* D *if, for any i, j, the function*

$$\frac{\partial N_i}{\partial x_j} : W \to \mathbb{R} \tag{2.2}$$

is continuous and the following condition holds:

D. *There exists $l \in \mathbb{R}$ such that, for all $k = 1, 2, \ldots,$*

$$1/2 \sum_{i=1}^{\infty} \left| \frac{\partial N_k}{\partial x_i} \right| (W) + 1/2 \sum_{i=1}^{\infty} \left| \frac{\partial N_i}{\partial x_k} \right| (W) + \lambda_k \leq l. \tag{2.3}$$

The main idea behind condition D is to ensure that Lipschitz constants of flows induced by Galerkin projections of (2.1) are uniformly bounded. (See the proof of Theorem 13 in [Z] for more details.)

Definition 2.3. *Consider an ODE*

$$x' = f(x), \tag{2.4}$$

where $x \in \mathbb{R}^n$. The compact set $W \subset \mathbb{R}^n$ is called a trapping region for (2.4) *if, for any solution $x(t)$ of (2.4), if $x(0) \in W$, then $x(t) \in W$ for all $t > 0$.*

The following easy lemma was used throughout this paper as a criterion for a set to be a trapping region.

Lemma 2.4. *Assume that W is a closure of an open set, with a piecewise smooth boundary. For any $x \in \partial W$, let $\nu(x)$ denote an outward normal vector to ∂W.*

If, for all $x \in \partial W$, we have $\nu(x) \cdot f(x) < 0$, then W is a trapping region for (2.4).

The following theorem was proved in [Z].

Theorem 2.5 (see [Z, Theorem 13]). *Assume that $R \subset H$ and F satisfy conditions* C1, C2, C3, *and* D *and that R is convex. Assume that $P_n(R)$ is a trapping region for the n-dimensional Galerkin projection of (2.1) for all $n > M_1$. Then the following hold.*

1. Uniform convergence and existence. *For a fixed $x_0 \in R$, let $x_n : [0, \infty] \to P_n(R)$ be a solution of $x' = P_n(F(x))$, $x(0) = P_n x_0$. Then x_n converges uniformly on compact intervals to a function $x^* : [0, \infty] \to R$, which is a solution of (2.1), and $x^*(0) = x_0$. The convergence of x_n on compact time intervals is uniform with respect to $x_0 \in R$.*
2. Uniqueness within R. *There exists only one solution of the initial value problem (2.1), $x(0) = x_0$ for any $x_0 \in R$, such that $x(t) \in R$ for $t > 0$.*
3. Lipschitz constant. *Let $x : [0, \infty] \to R$ and $y : [0, \infty] \to R$ be solutions of (2.1); then*

$$|y(t) - x(t)| \leq e^{lt} |x(0) - y(0)|.$$

4. Semidynamical system. *The map* $\varphi : \mathbb{R}_+ \times R \to R$, *where* $\varphi(\cdot, x_0)$ *is a unique solution of* (2.1), *such that* $\varphi(0, x_0) = x_0$, *defines a semidynamical system on* R*; namely,*
   - φ *is continuous,*
   - $\varphi(0, x) = x$, *and*
   - $\varphi(t, \varphi(s, x)) = \varphi(t + s, x)$.

In the context of this paper, the statement about the Lipschitz constant in Theorem 2.5 is of special importance. We can formulate it as

$$|\varphi(t, x) - \varphi(t, y)| \leq e^{lt}|x - y|, \quad t \geq 0, \tag{2.5}$$

where l is given in condition D and $x, y \in R$.

Assume that we have a trapping region, R, satisfying the assumptions of Theorem 2.5. The next step is to prove that the induced semiflow is contracting. This may be hard to achieve in the original norm, but, in section 3, we construct another norm (similar to the $|\cdot|_\infty$-norm) for which we are able to show that (2.5) holds with $l < 0$ for the steady states for the KS equation mentioned in the introduction.

3. Diagonalization and construction of a "contracting" norm.

As was mentioned in section 2, we would like to construct a "contracting" norm on trapping region R ($l < 0$ in (2.5)).

3.1. Block decomposition.

Our construction will be based on the approximate diagonalization of the matrix $\frac{\partial F}{\partial x}$. We want this matrix to be dominated by diagonal terms. This is achieved by an approximate diagonalization in case of real eigenvalues. The case of the complex eigenvalues forces us to consider blocks on the diagonal. We formalize this as follows.

Definition 3.1. *A decomposition of* H *into a sum of subspaces is called* a block decomposition of H *if the following conditions are satisfied.*

1. $H = \bigoplus_i H_i$.
2. *For every* i, $h_i = \dim H_i \leq h_{\max} < \infty$.
3. *For every* i, $H_i = \langle e_{i_1}, e_{i_2}, \dots, e_{i_{h_i}} \rangle$.
4. *If* $\dim H = \infty$, *then there exists* k *such that, for* $i > k$, $h_i = 1$.

For a block decomposition of H, we adopt the following notation, which makes a distinction between blocks and one-dimensional subspaces spanned by $\langle e_i \rangle$. For the blocks, we use $H_{(i)} = \langle e_{i_1}, \dots e_{i_k} \rangle$, where $(i) = (i_1, \dots i_k)$. The symbol H_i will always mean the subspace generated by e_i. For one-dimensional block (i), we adopt the following convention: the only element of (i) will be denoted by the same letter i.

For a given block decomposition of H and block (i), we set

$$\dim\ (i) = \dim\ H_{(i)}.$$

For any $x \in H$, by $x_{(i)}$ we will denote a projection of x onto $H_{(i)}$. For any a and $(i) = (i_1, \dots, i_k)$, we will say that $(i) \leq a$ if $i_s \leq a$ for all $s = 1, \dots, k$, and we say that $(i) > a$ if $i_s > a$ for all $s = 1, \dots, k$.

On each component $H_{(i)}$, we will use a norm induced from H. By $P_{(i)}$ we will denote an orthogonal projection onto $H_{(i)}$. By $\mathrm{Lin}(H_{(i)}, H_{(j)})$ we denote a set of all linear maps from $H_{(i)}$ to $H_{(j)}$ equipped with an operator norm $|A| = \max_{|v|=1, v \in H_{(i)}} |Av|$.

We have the following easy lemma.

Lemma 3.2. *Assume that we have a block decomposition of H, and let $W \subset H$. If, for any k, l, the function*

$$\frac{\partial F_k}{\partial x_l} : W \to \mathbb{R}$$

is continuous, then, for every (i) and (j), the map

$$\frac{\partial F_{(i)}}{\partial x_{(j)}} : W \to Lin(H_{(j)}, H_{(i)}) \approx \mathbb{R}^{\dim\ (j) \times \dim\ (i)}$$

is continuous.

For any square matrix $Q \in \mathbb{R}^{\dim H \times \dim H}$ (a linear map $Q : \operatorname{dom}(Q) \to H$) and for any blocks $(i), (j)$, we define a matrix $Q_{(i)(j)}$ as the matrix corresponding to an induced linear map $Q_{(i)(j)} : H_{(j)} \to H_{(i)}$ given by $Q_{(i)(j)}(x) = P_{(i)}(QP_{(j)}x)$.

3.2. A block-infinity norm for block decomposition. For a fixed block decomposition of H, we define the norm (*the block-infinity norm*) by

$$|x|_{b,\infty} = \max_{(i)} |P_{(i)}x|. \tag{3.1}$$

We have the following easy lemma.

Lemma 3.3. *Assume that $W \subset H$, W is closed and satisfies condition* C2. *Then on W the convergence in the norm $|\cdot|$ is equivalent to the convergence in the norm $|\cdot|_{b,\infty}$; namely, for any sequence $\{x_n\} \subset W$, $|x_n - x^*| \to 0$ if and only if $|x_n - x^*|_{b,\infty} \to 0$.*

Now we turn to the computation of the logarithmic norm for $|\cdot|_{b,\infty}$.

For any norm $\|\cdot\|$ on $\mathbb{R}^n$ following [HNW], we introduce the notion of the logarithmic norm of a matrix by the following definition.

Definition 3.4. *Let Q be a square matrix; then we call*

$$\mu(Q) = \limsup_{h>0, h\to 0} \frac{\|I + hQ\| - 1}{h} \tag{3.2}$$

the logarithmic norm *of Q.* Definition 3.4 differs slightly from Definition I.10.4 in [HNW] because, to avoid a question of the existence of the limit in (3.2), we use the lim sup.

By $\mu(Q)$ we denote the logarithmic norm induced by the Euclidean norm, and for all other norms we will use a subscript identifying it.

The following theorem was proved in [HNW].

Theorem 3.5 (see [HNW, Th. I.10.5]). *The logarithmic norm is obtained by the following formulas:*

$$\mu(Q) = \textit{the largest eigenvalue of} \quad 1/2(Q + Q^T), \tag{3.3}$$

$$\mu_\infty(Q) = \max_k \left(q_{kk} + \sum_{i, i\neq k} |q_{ki}| \right), \tag{3.4}$$

$$\mu_1(Q) = \max_i \left(q_{ii} + \sum_{k, k\neq i} |q_{ki}| \right). \tag{3.5}$$

The next lemma tells us how to compute the logarithmic norm induced by the block-infinity norm.

Lemma 3.6. *Assume that we have a block decomposition of* $\mathbb{R}^n$*; then*

$$\mu_{b,\infty}(Q) \leq \max_{(i)} \left(\mu(Q_{(i)(i)}) + \sum_{(k)\neq(i)} |Q_{(i)(k)}| \right). \tag{3.6}$$

Proof. Since, for any $h > 0$, (i), and $x \in \mathbb{R}^n$, we have

$$\begin{aligned} |P_{(i)}(I + hQ)x| &= \left| \left(I_{(i)(i)} + hQ_{(i)(i)} \right) x_{(i)} + h \sum_{(j),(j)\neq(i)} Q_{(i)(j)} x_{(j)} \right| \\ &\leq \left| \left(I_{(i)(i)} + hQ_{(i)(i)} \right) x_{(i)} \right| + h \sum_{(j),(j)\neq(i)} \left| Q_{(i)(j)} x_{(j)} \right| \\ &\leq \left(\left| I_{(i)(i)} + hQ_{(i)(i)} \right| + h \sum_{(j),(j)\neq(i)} \left| Q_{(i)(j)} \right| \right) |x|_{b,\infty}, \end{aligned}$$

then

$$|(I + hQ)|_{b,\infty} \leq \max_{(i)} \left(\left| I_{(i)(i)} + hQ_{(i)(i)} \right| + h \sum_{(j),(j)\neq(i)} \left| Q_{(i)(j)} \right| \right). \tag{3.7}$$

From the above equation and the definition of the logarithmic norm, one easily obtains the assertion of the theorem. ■

From Theorem 3.5, it follows that, when all blocks are one-dimensional, we have an equality in (3.6), but observe that this is not true in general as is shown by the following example.

Let $n = 4$. Consider the blocks $(e_1, e_2), (e_3)$, and (e_4) and the matrix

$$Q = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.8}$$

An easy computation shows that

$$\mu_{b,\infty}(Q) = 2 < \max_{(i)} \left(\mu(Q_{(i)(i)}) + \sum_{(k)\neq(i)} |Q_{(i)(k)}| \right) = 2\sqrt{2}.$$

3.3. Lipschitz constants in block-infinity norm and main theorem . The following theorem has exactly the same proof as Theorem 13 in [Z]. The only difference is that the standard norm in H is replaced by the block-infinity norm.

Theorem 3.7. *Assume that* $R \subset H$*,* R *is convex, and* F *satisfies conditions* C1, C2, *and* C3. *Assume that we have a block decomposition of* H *such that condition* Db *holds.*

Db. *There exists $l \in \mathbb{R}$ such that, for any (i) and $x \in R$,*

$$\mu\left(\frac{\partial F_{(i)}}{\partial x_{(i)}}(x)\right) + \sum_{(k),\,(k)\neq(i)} \left|\frac{\partial F_{(i)}}{\partial x_{(k)}}(x)\right| \leq l.$$

Assume that $P_n(R)$ is a trapping region for the n-dimensional Galerkin projection of (2.1) for all $n > M_1$. Then the following hold.

1. Uniform convergence and existence. *For a fixed $x_0 \in R$, let $x_n : [0,\infty] \to P_n(R)$ be a solution of $x' = P_n(F(x))$, $x(0) = P_n x_0$. Then x_n converges uniformly in a max-infinity norm on compact intervals to a function $x^* : [0,\infty] \to R$, which is a solution of (2.1) and $x^*(0) = x_0$. The convergence of x_n on compact time intervals is uniform with respect to $x_0 \in R$.*
2. Uniqueness within R. *There exists only one solution of the initial value problem (2.1), $x(0) = x_0$ for any $x_0 \in R$, such that $x(t) \in R$ for $t > 0$.*
3. Lipschitz constant. *Let $x : [0,\infty] \to R$ and $y : [0,\infty] \to R$ be solutions of (2.1); then*

$$|y(t) - x(t)|_{b,\infty} \leq e^{lt}|x(0) - y(0)|_{b,\infty}.$$

4. Semidynamical system. *The map $\varphi : \mathbb{R}_+ \times R \to R$, where $\varphi(\cdot, x_0)$ is a unique solution of (2.1), such that $\varphi(0, x_0) = x_0$, defines a semidynamical system on R; namely,*
   - *φ is continuous,*
   - *$\varphi(0, x) = x$,*
   - *$\varphi(t, \varphi(s, x)) = \varphi(t + s, x)$.*

The following theorem is the main tool in proving an existence of attracting fixed points.

Theorem 3.8. *We use the same assumptions on R,F and a block decomposition of H as in Theorem 3.7. Assume that $l < 0$.*

Then there exists a fixed point for (2.1), $x^* \in R$, unique in R, such that, for every $y \in R$,

$$|\varphi(t,y) - x^*|_{b,\infty} \leq e^{lt}|y - x^*|_{b,\infty} \quad \text{for } t \geq 0,$$
$$\lim_{t\to\infty} \varphi(t,y) = x^*.$$

Proof. It is enough to prove the existence of $x^* \in R$ such that $F(x^*) = 0$ because the uniqueness in R and all other assertions follow directly from the assumption $l < 0$ and the Lipschitz constant estimates given in Theorem 3.7.

It is easy to see that, for all $n > M_1$, there exists $x_n \in P_n R$, a fixed point for the nth Galerkin projection. Passing to the limit with n (by picking a subsequence eventually), we obtain x^* (see [ZM, Thm 2.16] for details). ■

4. An example of a theorem and a description of an algorithm.

In this section, we present an example of a theorem we prove using our method, give a description of an algorithm, and provide some numerical data from the proof.

Theorem 4.1. *Let*

$$u(x) = -2\left(0.711691 \sin(x) - 0.123059 \sin(2x) + 0.01011 \sin(3x)\right).$$

For any $\nu \in 0.75 + [-10^{-2}, 10^{-2}]$, there exists an equilibrium solution $u_\nu(x)$ to (1.1) such that

- u_ν *is attracting,*
- $\|u_\nu - u\|_{L^2} \leq 0.104$, $\|\partial_x u_\nu - \partial_x u\|_{L^2} \leq 0.132$, *and* $\|u_\nu - u\|_{C^0} \leq 0.084$.

The attracting fixed point from the above theorem had already been discovered (nonrigorously) by Jolly, Kevrekidis, and Titi in [JKT]. In the terminology used there, this is a unimodal fixed point.

The proof consists of two parts.

1. The first part is a construction of topologically self-consistent a priori bounds (see Definition 2.11 in [ZM]) for the KS equation, i.e., $W \subset X_m$ and $\{a_k^-, a_k^+\}_{k>m}$, an isolating block $N \subset W$ with empty exit set. We define a set R given by

$$R = N \oplus \Pi_{k=m+1}^{\infty}[a_k^-, a_k^+]. \tag{4.1}$$

From the construction, it follows that $P_n(R)$ is a trapping region for the nth Galerkin projection of the KS equation for $n \geq m$.

From the results in [ZM], it follows that there exists $u_\nu \in R$ such that $F(u_\nu) = 0$.

2. The second part is the computation of l. Now, if $l < 0$, then from Theorem 3.8 it follows that u_ν is attracting.

The method of construction of the tail in self-consistent bounds, i.e., the numbers $a_k^\pm$ for $k > m$, is described in section 3.3 in [ZM], but the construction of an isolating block N was not presented there; therefore, we present an outline of this algorithm here.

First, we introduce some notation and fix some terminology. We recall condition C4a from [ZM]:

C4a. Let $u \in W \oplus \Pi_{k=m+1}^{\infty}[a_k^-, a_k^+]$. Then, for $k > m$,

$$A_k u = a_k^+ \Rightarrow A_k F_\nu(u)) < 0, \tag{4.2}$$

$$A_k u = a_k^- \Rightarrow A_k F_\nu(u)) > 0. \tag{4.3}$$

Definition 4.2. *In the context of a block-decomposition of a finite-dimensional space,* $H = \bigoplus H_{(i)}$, *we consider a system of differential inclusions, which is a product of inclusions for each block of the form*

$$x_k' \in \lambda_k x_k + (b_k, B_k) \tag{4.4}$$

or a two-dimensional block $(k) = (k_1, k_2)$,

$$\begin{aligned} x_{k_1}' &\in \alpha_{(k)} x_{k_1} - \beta_{(k)} x_{k_2} + (b_{k_1}, B_{k_1}), \\ x_{k_1}' &\in -\beta_{(k)} x_{k_1} + \alpha_{(k)} x_{k_2} + (b_{k_2}, B_{k_2}). \end{aligned} \tag{4.5}$$

Let $N = \bigoplus N_{(i)}$, *where* $N_{(i)} \subset H_{(i)}$ *and*

$$\begin{aligned} N_{(i)} &= [n_{(i)}^-, n_{(i)}^+] \qquad \textit{if } \dim(i) = 1, \\ N_{(i)} &= \overline{B}(0, n_{(i)}) \qquad \textit{if } \dim(i) = 2. \end{aligned}$$

We will say that we have an isolation for the (k)-block on N *if the following hold:*

$$\lambda_k n_k^+ + (b_k, B_k) < 0, \qquad \lambda_k n_k^- + (b_k, B_k) > 0 \qquad \textit{if } \dim(k) = 1. \tag{4.6}$$

If $\dim(k) = 2$, *we require that*

$$\lambda_{(k)} n_{(k)} + \sqrt{(b_{k_1}, B_{k_1})^2 + (b_{k_2}, B_{k_2})^2} < 0. \tag{4.7}$$

The following easy lemma explains why we care for an isolation.

Lemma 4.3. *Let a block decomposition of H, set N, and differential inclusions be as in Definition 4.2. If we have an isolation for all blocks on N, then the set N is a trapping region. (In particular, it is an isolating block.)*

We perform our computations in an interval arithmetic [Mo]. We use the following notation and conventions.

By arabic letters we denote both single-valued objects like vectors, real numbers, and matrices and sets of these objects. Sometimes we will use square brackets, for example, $[r]$, to denote sets. Usually this will be some set constructed in an algorithm. In situations when we want to stress that we have a set in a formula involving both single-valued objects and sets together, we would rather use square brackets; hence we prefer to write $[S]$ instead of S to represent a set. From this point of view, $[S]$ and S are different symbols in the alphabet used to name variables, and, formally, there is no relation between the set represented by $[S]$ and the object represented by S. Sometimes both variables $[S]$ and S are used simultaneously; usually $S \in [S]$ in this situation, but this is always stated explicitly.

For a set $[S]$ by $[S]_I$, we denote an interval hull of $[S]$, i.e., the smallest product of intervals containing $[S]$. The symbol $\mathrm{hull}(x_1, \ldots, x_k)$ will denote an interval hull of intervals $x_1, \ldots, x_k$. For any interval set $[S] = [S]_I$, by $\mathrm{m}([S])$ we will denote a center point of $[S]_I$. For any interval $[a, b]$, we define a diameter by $\mathrm{diam}([a, b]) = b - a$. For an interval vector or an interval matrix $[S] = [S]_I$, by $\mathrm{diam}([S])$ we will denote the maximum of diameters of its components. For an interval $[x^-, x^+]$, we set $right([x^-, x^+]) = x^+$ and $left([x^-, x^+]) = x^-$.

4.1. A detailed outline of an algorithm.

Input data:

- m, M are dimensions describing self-consistent bounds.
- $[\nu] = \nu_0 + [-\delta\nu, \delta\nu]$ is a range of parameters.
- $x_0 \in \mathbb{R}^m$ such that $P_m F_{\nu_0}(x_0) \approx 0$; this is our candidate for a fixed point.
- Δ is a parameter defining an initial size of N.

1. *An approximate diagonalization of $dP_m F_{\nu_0}(x_0)$, a generation of new coordinates in $X_m = \mathbb{R}^m$, and a block decomposition of H.* From an approximate diagonalization of $dP_m F_{\nu_0}(x_0)$, we obtain new coordinates, which will be called *the block coordinates*. The coordinates a_k will be referred to as *the standard coordinates*. The block coordinates are obtained from standard coordinates through an affine transformation $T : \mathbb{R}^m \to \mathbb{R}^m$,

$$T(x) = T_l(x - x_0), \tag{4.8}$$

where $T_l \in \mathbb{R}^{m \times m}$.

We define a block decomposition of $H = \bigoplus_{(i)} H_i$ such that, for $(i) > m$, all blocks are given by $H_{(i)} = \langle e_i \rangle$; for $(i) < m$, each block $H_{(i)}$ is an eigenspace of $dP_m F_{\nu_0}(x_0)$. Complex eigenvalues give rise to two-dimensional blocks and real eigenvalues to one-dimensional blocks. Eigenvectors from one-dimensional blocks are normalized to a unit length; in a two-dimensional block, the length of a longer vector from the pair is normalized to one.

From now on, we will change the norm in H so that the blocks $H_{(i)}$ become orthogonal and, for two-dimensional blocks, the real and the imaginary parts of a complex eigenvector are orthogonal.

2. Preparation for the main loop:

- *An initialization of variables* $Iso[k]$, $1 \leq k \leq m$. We set $Iso[k] = 0$. It will later become true (i.e., equal to 1) if we will have an isolation for the block containing a kth variable).
- *An initialization of our initial guess for* $N = \bigoplus_{(i) \leq m} N_{(i)}$, where $N_{(i)} \subset H_{(i)}$. We set

$$N_{(i)} = [n^-_{(i)}, n^+_{(i)}] = [-\Delta, \Delta] \qquad \text{if } \dim(i) = 1,$$
$$N_{(i)} = \overline{B}(0, n_{(i)}) = \overline{B}(0, \Delta) \qquad \text{if } \dim(i) = 2.$$

3. Main loop:

- *An initialization of a local variable iso_change* $= 0$. The variable *iso_change* tells us if there is any new isolation for $1 \leq k \leq m$ or if our set N has changed, giving us the chance that repeating a loop once again will result in a better tail, which may produce new isolations.
- *A computation of* W. $W = [T^{-1}(N)]_I$. It is enough to define W as $W = T^{-1}(N)$, i.e., the set N in standard coordinates. However, if we evaluate this formula in interval arithmetics, we obtain the set $[T^{-1}N]_I$, which is larger than $T^{-1}(N)$ due to round-off errors and the wrapping effect [Mo].
- *A generation of self-consistent tail.* Using formulas derived in section 3 in [ZM], for the current values of W, m, M, we find $\{a^-_k, a^+_k\}$ such that conditions C1, C2, C3, and C4a are satisfied on $W \oplus \Pi^\infty_{k=m+1}[a^-_k, a^+_k]$.
  In principle, the procedure of generation $\{a^\pm_k\}$ may fail; in this case, we interrupt the algorithm and return *fail*.
- *A computation of an influence of the tail* $V = \Pi^\infty_{k=m+1}[a^-_k, a^+_k]$ *onto the* m*-dimensional Galerkin projection.* Using formulas from section 3 in [ZM], we find an interval vector $[\epsilon] \subset \mathbb{R}^n$ such that

$$P_m(F_\nu(x)) - P_m(F_\nu(P_m x)) \subset [\epsilon] \qquad \text{for } x \in W \oplus V. \tag{4.9}$$

  Our goal for the next step is to construct an isolating block N for an equation (in fact a differential inclusion)

$$x' \in P_m(F_\nu(x)) + [\epsilon], \qquad \text{where } x \in W. \tag{4.10}$$

- *A transformation of* (4.10) *to the block coordinates and "an interval diagonalization."* We transform (4.10) into the block coordinates; as a result, we obtain for $\nu \in \nu_0 + [-\delta\nu, \delta\nu]$, for one-dimensional blocks (i),

$$x'_i \in \lambda_i(\nu) x_i + f_i(x) + [\tilde{\epsilon}_i] \tag{4.11}$$

  and, for two-dimensional blocks $(i) = (i_1, i_2)$,

$$\begin{aligned} x'_{i_1} &\in \alpha_{(i)}(\nu) x_{i_1} + \beta_{(i)}(\nu) x_{i_2} + f_{i_1}(x) + [\tilde{\epsilon}_{i_1}], \\ x'_{i_2} &\in -\beta_{(i)}(\nu) x_{i_1} + \alpha_{(i)}(\nu) x_{i_2} + f_{i_2}(x) + [\tilde{\epsilon}_{i_2}]. \end{aligned} \tag{4.12}$$

Observe that, since we are using the block coordinates, which diagonalize dP_mFP_m, for a small N the values $f_i(x)$ for $x \in N$ will usually be very small.
For $1 \leq i \leq m$, we compute an interval (b_i, B_i) such that

$$f_i(W) + \tilde{\epsilon}_i \subset (b_i, B_i). \tag{4.13}$$

Instead of (4.11) and (4.12), we will now consider the equations

$$x_i' \in \lambda_i(\nu)x_i + (b_i, B_i) \tag{4.14}$$

and

$$\begin{aligned} x_{i_1}' &\in \alpha_{(i)}x_{i_1} + \beta_{(i)}x_{i_2} + (b_{i_1}, B_{i_1}),\\ x_{i_2}' &\in -\beta_{(i)}x_{i_1} + \alpha_{(i)}x_{i_2} + (b_{i_2}, B_{i_2}), \end{aligned} \tag{4.15}$$

respectively.
Let us stress that, in our computations, we have uniform bounds for $\lambda_{(i)}(\nu)$, $\alpha_{(i)}(\nu)$, and $\beta_{(i)}(\nu)$. Namely, we have intervals $[\lambda_{(i)}]$, $[\alpha_{(i)}]$, and $[\beta_{(i)}]$ such that, for all $\nu \in [\nu_0 - \delta\nu, \nu_0 + \delta\nu]$, the following hold:

$$\lambda_{(i)}(\nu) \in [\lambda_{(i)}], \quad \alpha_{(i)}(\nu) \in [\alpha_{(i)}], \quad \beta_{(i)}(\nu) \in [\beta_{(i)}].$$

- *An isolation for* $1 \leq i \leq m$. First observe that, since we are looking for an attracting fixed point, then $\lambda_i < 0$ and $\alpha_{(i)} < 0$ (provided x_0 is a good approximation).
For each block, we try to find an isolation as follows: for one-dimensional block (i), we set

$$d_i^+ = \text{right}\left(-\frac{B_i}{[\lambda_i]}\right), \qquad d_i^- = \text{left}\left(-\frac{b_i}{[\lambda_i]}\right). \tag{4.16}$$

Now if

$$[d_i^-, d_i^+] \subset [n_i^-, n_i^+], \tag{4.17}$$

then an easy computation shows that we have an isolation for the (i)th block.
For two-dimensional blocks $(i) = (i_1, i_2)$, we set

$$d_{(i)} = \text{right}\left(-\frac{B}{[\alpha_{(i)}]}\right), \tag{4.18}$$

where $B = \text{right}(\sqrt{(b_{i_1}, B_{i_1})^2 + (b_{i_2}, B_{i_2})^2})$. It is easy to see that we have an isolation for the (i)th block on N if

$$n_{(i)} \geq d_{(i)}. \tag{4.19}$$

If (4.17) holds for one-dimensional block (i), then we set

$$\begin{aligned} iso\_change &= 1,\\ n_i^+ = d_i^+, &\quad n_i^- = d_i^-,\\ Iso[i] &= 1. \end{aligned}$$

If (4.19) holds for two-dimensional block $(i) = (i_1, i_2)$, then we set

$$\begin{aligned} iso\_change &= 1, \\ n_{(i)} &= d_{(i)}, \\ Iso[i_1] = 1, &\qquad Iso[i_2] = 1. \end{aligned}$$

Observe that, with these updates, we achieve the following: if $iso\_change = 1$, then the current set N is a proper subset of the set N at the beginning of the loop. This guarantees that W and then also $[\epsilon]$ will be smaller in the next iterate. This implies also that, if in one loop we have an isolation for a block (i), then we will have an isolation for this block for all following iterations of the loop, and it creates a possibility for obtaining an isolation for other blocks in the next iterations.

- *A verification of an isolation for all blocks.* We check if, for all $1 \leq k \leq m$, $Iso[k] = 1$. If this is the case, then we leave the loop because N is the desired trapping region. Otherwise, we check if $iso\_change = 0$. If this is the case, then the algorithm failed. If $iso\_change = 1$, we repeat the loop again.

4. Computation of l: From previous steps, we have a block decomposition of H, a change of coordinates T, a trapping region N, and a tail $[a_k^-, a_k^+]_{k>m}$.

We compute l using formulas from section 5 on the set $W = [T^{-1}(N)]_I \oplus \Pi_{k>m}[a_k^-, a_k^+]$.

5. Output: We set $x_c = \mathrm{m}(N)$ (a center point of N). Let $x_* = T^{-1}x_c$. x_* is a center of a trapping region N expressed in the standard coordinates. We estimate $|y - x_*|$ in various norms for all $y \in T^{-1}(N) \oplus \Pi_{k>m}[a_k^-, a_k^+]$ (for example, L_2, H^1 etc.). If $l < 0$, then we can conclude that, close to x_*, there exists an attracting fixed point; otherwise, we can claim only the existence of a fixed point.

End of an algorithm.

4.2. Numerical data from the proof of Theorem 4.1. We have chosen $m = 3$ and $M = 10$. In fact, to complete the full algorithm successfully without checking that $l < 0$, i.e., to prove just the existence of a fixed point, it is enough to take $m = 2$ (see the proof of Theorem 4.1 in [ZM]), but, in this case, we were unable to verify that $l < 0$.

Other starting parameters for the algorithm were given by

$$\begin{aligned} x_0 &= (0.712361,\ -0.12324,\ 0.0101787), \\ \Delta &= 0.03125. \end{aligned}$$

x_0 was found by simply integrating forward a three-dimensional Galerkin projection of (1.1). We tried first $\Delta = 10^{-5}$ and then $\Delta = 5\Delta$ (we multiplied the current value of Δ by 5) until we were able to successfully complete the algorithm.

From the approximate diagonalization, we list approximate eigenvalues and several most significant digits of interval matrixes T_l and T_l^{-1}. Diameters of entries in T_l and T_l^{-1} were smaller than 10^{-15}.

$$\lambda_1 \approx -51.46617, \qquad \lambda_2 \approx -7.7545, \qquad \lambda_3 \approx -0.52575.$$

We see that, in our block decomposition, we have only one-dimensional blocks.

$$T_l = \begin{bmatrix} -0.0090621 & 0.09949 & 1.0087 \\ -0.39312 & -1.1041 & -0.069407 \\ -1.1408 & -0.21674 & -0.0066056 \end{bmatrix},$$

$$T_l^{-1} = \begin{bmatrix} 0.0065846 & 0.18519 & -0.94042 \\ -0.065071 & -0.97779 & 0.33745 \\ 0.99786 & 0.098106 & -0.041732 \end{bmatrix}.$$

We obtained an isolation for $1 \leq i \leq 3$ after four iterates of the main loop for the set N are given by

$$N = [-3.176878e-04, 2.272739e-04] \times [-3.618637e-03, 3.756375e-03] \\ \times[-2.566221e-02, 2.711549e-02].$$

For l, we have

$$l < -0.05716.$$

Below we list some other data obtained in the algorithm. The set $W = [T^{-1}(N)]_I$ is given by

$$\begin{aligned} W &= [a_1^-, a_1^+] \times [a_2^-, a_2^+] \times [a_3^-, a_3^+], \\ [a_1^-, a_1^+] &= 0.711691 + [-0.0255012, 0.0255012], \\ [a_2^-, a_2^+] &= -0.123059 + [-0.0125283, 0.0125283], \\ [a_3^-, a_3^+] &= 0.01011 + [-0.00173494, 0.00173494]. \end{aligned}$$

In Table 4.1, we list $a_k^\pm$ for $k > 4$.

Table 4.1

Estimates for the intervals $[a_k^-, a_k^+]$ representing self-consistent a priori bounds in the proof of Theorem 4.1.

k	$[a_k^-, a_k^+]$
4	$-6.77647 \cdot 10^{-4} + [-1.48185, 1.48185] \cdot 10^{-4}$
5	$3.95994 \cdot 10^{-5} + [-1.10021, 1.10021] \cdot 10^{-5}$
6	$-2.14113 \cdot 10^{-6} + [-7.10907, 7.10907] \cdot 10^{-7}$
7	$1.09725 \cdot 10^{-7} + [-4.21541, 4.21541] \cdot 10^{-8}$
8	$-5.41181 \cdot 10^{-9} + [-2.35893, 2.35893] \cdot 10^{-9}$
9	$2.60507 \cdot 10^{-10} + [-2.10033, 2.10033] \cdot 10^{-10}$
10	$-1.22454 \cdot 10^{-11} + [-3.57734, 3.57734] \cdot 10^{-10}$
> 10	$[-1, 1] \cdot 4176.07/k^{10}$

For $[\epsilon] = ([\epsilon_1], [\epsilon_2], [\epsilon_3])$, we obtained the following numbers:

$$\begin{aligned} [\epsilon_1] &= -1.42733 \cdot 10^{-5} + [-5.37438, 5.37438] \cdot 10^{-6}, \\ [\epsilon_2] &= 0.000342671 + [-0.000107623, 0.000107623], \\ [\epsilon_3] &= -0.00294653 + [-0.00074762, 0.00074762]. \end{aligned}$$

After passing to the block coordinates and an interval diagonalization, we obtained on $N \oplus \Pi_{k>m}[a_k^-, a_k^+]$

$$\begin{aligned} x_1' &\in [-0.01608764, 0.01150572] + [-52.28307, -50.65041]x_1, \\ x_2' &\in [-0.02740801, 0.02844858] + [-7.932687, -7.574724]x_2, \\ x_3' &\in [-0.01291743, 0.01364734] + [-0.5486830, -0.5033749]x_3. \end{aligned}$$

In Table 4.2, we list the computed upper bounds for l_i on $N \oplus \Pi_{k>m}[a_k^-, a_k^+]$. It is easy to see that $l = l_3 < -0.05716$.

Table 4.2

Estimates from above on l_i from the computation of l in the proof of Theorem 4.1.

k	l_k
1	-44.76
2	-6.06
3	-0.05716
4	-160.6
5	-425.8
6	-914.7
7	-1726.9
8	-2979.6
9	-4807.8
10	-7364.5
> 10	-10820.7

Observe that the value of $l = -0.05716$ is close to zero. This is essentially due to the fact that we wanted to extend as much as possible the parameter interval. If we just choose $\nu = 0.75$ with the same x_0, m, and M, then we obtain $l = -0.348938$. By increasing m and M, we can further decrease l up to λ_3. For example, for $m = 15$ and $M = 45$, we obtained $l = -0.52581$.

5. Details for the KS equation.

The goal of this section is to derive formulas from which, for a given block decomposition and trapping region, the constants $l_{(i)}$ for the KS equation may be computed, thus implementing step 4 of the algorithm of section 4.1.

5.1. The KS equation in Fourier representation.

For the KS equation in one space dimension with periodic and odd boundary conditions, we have the following infinite ladder of equations for the Fourier coefficients (see [ZM]):

$$\dot{a}_k = F_k(a) = \lambda_k a_k + N_k(a), \tag{5.1}$$

$$\lambda_k = k^2(1 - \nu k^2), \tag{5.2}$$

$$N_k(a) = -k \sum_{n=1}^{k-1} a_n a_{k-n} + 2k \sum_{n=1}^{\infty} a_n a_{n+k}. \tag{5.3}$$

Hence we obtain

$$\frac{\partial N_i}{\partial a_j} = 2i a_{i+j} \quad \text{for } i = j,$$

$$\begin{aligned}\frac{\partial N_i}{\partial a_j} &= -2ia_{i-j} + 2ia_{i+j} \quad \text{for } j < i,\\ \frac{\partial N_i}{\partial a_j} &= 2ia_{j-i} + 2ia_{i+j} \quad \text{for } j > i,\\ \frac{\partial F_i}{\partial a_j} &= i^2(1-\nu i^2)\delta_{ij} + \frac{\partial N_i}{\partial a_j}.\end{aligned}$$

5.2. Coordinate change and block-decomposition. The following lemma does not require any proof.

Lemma 5.1. *Let $A : H \to H$ be a linear coordinate change of the form*

$$\begin{aligned}A : X_m \oplus Y_m &\to X_m \oplus Y_m,\\ A(x \oplus y) &= Ax \oplus y.\end{aligned}$$

Let $\tilde{F} = A \circ F \circ A^{-1}$. ($\tilde{F}$ is F expressed in new coordinates.)

$$\begin{aligned}\frac{\partial \tilde{F}_i}{\partial x_j} &= \sum_{k,l=1}^{m} A_{ik}\frac{\partial F_k}{\partial x_l}A^{-1}_{lj} && \text{for } i \le m \text{ and } j \le m,\\ \frac{\partial \tilde{F}_i}{\partial x_j} &= \sum_{k\le m} A_{ik}\frac{\partial F_k}{\partial x_j} && \text{for } i \le m \text{ and } j > m,\\ \frac{\partial \tilde{F}_i}{\partial x_j} &= \sum_{l\le m} \frac{\partial F_i}{\partial x_l}A^{-1}_{lj} && \text{for } i > m \text{ and } j \le m,\\ \frac{\partial \tilde{F}_i}{\partial x_j} &= \frac{\partial F_i}{\partial x_j} && \text{for } i > m \text{ and } j > m.\end{aligned}$$

Consider now the KS equation and assume that $W = N \oplus \Pi_{k=m+1}^{\infty}[a_k^-, a_k^+]$ is a trapping region representing self-consistent bounds for a fixed point. Let the numbers $m < M$ be as in conditions C1, C2, and C3, and we assume that $a_k^{\pm} = \pm\frac{C}{k^s}$ for $k > M$ (as in [ZM]).

Let $A \in \mathbb{R}^{m\times m}$ be a coordinate change around an approximate fixed point in X_m for the m-dimensional Galerkin projection of (5.1). This matrix induces a coordinate change in H. For our purpose, it is optimal to choose A so that the m-dimensional Galerkin projection of F is very close to a diagonal matrix (or to a block diagonal matrix when complex eigenvalues are present).

From now on, we will use these new coordinates on H. We also change the norm so that the new coordinates become orthogonal. We define the splitting of P_mH into blocks which are either two-dimensional (complex eigenvalue) or one-dimensional (real eigenvalue). For the KS equation, there was no need to consider more complicated situations. For $(i) > m$, all blocks are one-dimensional. (These coordinates are not affected by our coordinate change.)

We would like to derive the formula for

$$l_{(i)} := \sup_{x\in W} \mu\left(\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}(x)\right) + \sum_{(j),(j)\neq(i)} \sup_{x\in W}\left|\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x)\right|. \tag{5.4}$$

We define $S(l)$ by

$$S(l) = \sum_{k \geq l} \sup_{a \in W} |a_k|.$$

We estimate $S(l)$ from above using the following lemma.

Lemma 5.2. *Assume that* $|a_k(W)| \leq \frac{C}{k^s}$ *for* $k > M$, $s > 1$*; then*

$$S(l) < \sum_{k=l}^{M} |a_k(W)| + \frac{C}{(s-1)M^{s-1}} \quad \textit{for } l \leq M,$$
$$S(l) < \frac{C}{(s-1)(l-1)^{s-1}} \quad \textit{for } l > M.$$

Proof. Observe that

$$\sum_{k=l}^{\infty} \frac{1}{k^s} < \int_{l-1}^{\infty} \frac{dx}{x^s} = \frac{1}{(s-1)(l-1)^{s-1}}. \qquad \blacksquare \tag{5.5}$$

We set

$$\overline{S}(l) = \sum_{k=l}^{M} |a_k(W)| + \frac{C}{(s-1)M^{s-1}} \quad \text{for } l \leq M, \tag{5.6}$$

$$\overline{S}(l) = \frac{C}{(s-1)(l-1)^{s-1}} \quad \text{for } l > M. \tag{5.7}$$

5.3. Formulas for one-dimensional blocks. Observe that, if $\dim(i) = 1$, then $\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}} \in \mathbb{R}$; hence

$$\mu\left(\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}\right) = \frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}. \tag{5.8}$$

Lemma 5.3. *Assume* $(i) \leq m$ *and* $\dim\,(i) = 1$.

$$\begin{aligned} &\sup_{x \in W} \mu\left(\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}(x)\right) + \sum_{(j) \neq (i)} \sup_{x \in W} \left|\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x)\right| \\ &\leq \bar{l}_i := \sup_{x \in W} \frac{\partial \tilde{F}_i}{\partial x_i}(x) + \sum_{j \neq i, j \leq m} \sup_{x \in W} \left|\frac{\partial \tilde{F}_i}{\partial x_j}(x)\right| \\ &+ 2 \sum_{k \leq m} k|A_{ik}|(\overline{S}(m-k+1) + \overline{S}(m+k+1)). \end{aligned}$$

Proof. Observe that, when $(j) = (j_1, j_2)$ is a two-dimensional block, then we need to compute the norm of $[\frac{\partial \tilde{F}_i}{\partial x_{j_1}}(x), \frac{\partial \tilde{F}_i}{\partial x_{j_2}}(x)]$. It is easy to see that this norm is less than or equal to

$$\left|\frac{\partial \tilde{F}_i}{\partial x_{j_1}}(x)\right| + \left|\frac{\partial \tilde{F}_i}{\partial x_{j_2}}(x)\right|.$$

This means that ignoring the block structure is safe (we have an inequality in the correct direction); hence we obtain

$$\sum_{(j)\neq(i)}\left|\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(W)\right| \leq \sum_{j\neq i}\left|\frac{\partial \tilde{F}_i}{\partial x_j}(W)\right|. \tag{5.9}$$

To finish the proof, it is enough to show that

$$\sum_{j>m}\left|\frac{\partial \tilde{F}_i}{\partial x_j}(W)\right| \leq 2\sum_{k\leq m} k|A_{ik}|(S(m-k+1)+S(m+k+1)). \tag{5.10}$$

Observe that, from Lemma 5.1, it follows that

$$\begin{aligned}\sum_{j>m}\left|\frac{\partial \tilde{F}_i}{\partial x_j}(W)\right| &\leq \sum_{j>m}\sum_{k\leq m}|A_{ik}|2k(|a_{j-k}(W)|+|a_{j+k}(W)|)\\ &= \sum_{k\leq m}|A_{ik}|2k\left(\sum_{j>m}|a_{j-k}(W)|+|a_{j+k}(W)|\right)\\ &= \sum_{k\leq m}|A_{ik}|2k\left(S(m-k+1)+S(m+k+1)\right). \qquad \blacksquare\end{aligned}$$

Observe that, from our assumptions about the decomposition of H, it follows that all blocks (i), such that $(i)>m$, are one-dimensional.

Lemma 5.4. *For $m<i\leq M$, we have*

$$\begin{aligned}&\sup_{x\in W}\mu\left(\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}(x)\right)+\sum_{(j),(j)\neq(i)}\sup_{x\in W}\left|\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x)\right|\\ &\leq \bar{l}_i = i^2(1-\nu i^2)+\sum_{j\leq M}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right|+2i(\overline{S}(M+1-i)+\overline{S}(i+M+1)).\end{aligned}$$

Proof. Just as in the proof of Lemma 5.3, we can ignore the block structure here. It is easy to see that

$$\sup_{x\in W}\frac{\partial \tilde{F}_i}{\partial x_i}(x)+\sum_{j\neq i}\sup_{x\in W}\left|\frac{\partial \tilde{F}_i}{\partial x_j}(x)\right| \leq i^2(1-\nu i^2)+\sum_{j=1}^{\infty}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right|. \tag{5.11}$$

Therefore, to finish the proof, it is enough to show that

$$\sum_{j=M+1}^{\infty}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right| < 2i(S(M+1-i)+S(i+M+1)). \tag{5.12}$$

We proceed as follows:

$$\begin{aligned}&\sum_{j=M+1}^{\infty}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right| = \sum_{j=M+1}^{\infty}\sup_{x\in W}\left|\frac{\partial N_i}{\partial x_j}(x)\right|\\ &\leq \sum_{j=M+1}^{\infty}2i(|a_{j-i}(W)|+|a_{i+j}(W)|)\leq 2i(S(M+1-i)+S(M+1+i)). \qquad \blacksquare\end{aligned}$$

Lemma 5.5. *For $(i) > m$, we have*

$$\begin{aligned}\sup_{x\in W} \mu\left(\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}(x)\right) + \sum_{(j)\neq(i)} \sup_{x\in W}\left|\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x)\right| \leq \bar{l}_i := i^2(1-\nu i^2)\\ +\ 2im(\overline{S}(i-m)+\overline{S}(i+1)) \max_{k,l=1,\dots,m}|A^{-1}_{kl}|\\ +\ 2i(\overline{S}(i+m+1)+2\overline{S}(1)).\end{aligned}$$

Proof. Just as in the proof of Lemma 5.3, we can ignore the block structure here. It is easy to see that

$$\sup_{x\in W}\frac{\partial \tilde{F}_i}{\partial x_i}(x) + \sum_{j\neq i}\sup_{x\in W}\left|\frac{\partial \tilde{F}_i}{\partial x_j}(x)\right| \leq i^2(1-\nu i^2) + \sum_{j=1}^{\infty}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right|. \tag{5.13}$$

Therefore, to finish the proof, it is enough to show that

$$\sum_{j=1}^{m}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right| \leq 2im(S(i-m)+S(i+1))\max_{k,l=1,\dots,m}|A^{-1}_{kl}|, \tag{5.14}$$

$$\sum_{j=m+1}^{\infty}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right| \leq 2i(S(i+m+1)+2S(1)). \tag{5.15}$$

To prove (5.14), observe that

$$\begin{aligned}\sum_{j=1}^{m}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right| &= \sum_{j=1}^{m}\sup_{x\in W}\left|\sum_{l=1}^{m}\frac{\partial N_i}{\partial x_l}(x)A^{-1}_{lj}\right|\\ &\leq \sum_{j=1}^{m}\sum_{l=1}^{m}2i(|a_{i-l}(W)|+|a_{i+l}(W)|)|A^{-1}_{lj}|\\ &\leq 2i\sum_{j=1}^{m}(S(i-m)+S(i+1))\max_{k,l=1,\dots,m}|A^{-1}_{kl}|\\ &= 2im(S(i-m)+S(i+1))\max_{k,l=1,\dots,m}|A^{-1}_{kl}|.\end{aligned}$$

To prove (5.15), we proceed as follows:

$$\begin{aligned}\sum_{j=m+1}^{\infty}\sup_{x\in W}\left|\frac{\partial \tilde{N}_i}{\partial x_j}(x)\right| &= \sum_{j=m+1}^{\infty}\sup_{x\in W}\left|\frac{\partial N_i}{\partial x_j}(x)\right|\\ &\leq \sum_{m<j<i}(2i(|a_{i-j}(W)|+|a_{i+j}(W)|))\\ &\quad + 2i|a_{2i}(W)| + \sum_{j>i}2i(|a_{j-i}(W)|+|a_{i+j}(W)|)\\ &\leq 2i\left(\sum_{j>m}|a_{i+j}(W)| + \sum_{m<j<i}|a_{i-j}(W)| + \sum_{j>i}|a_{j-i}(W)|\right)\\ &< 2i\left(S(i+m+1)+2S(1)\right). \qquad ■\end{aligned}$$

The following lemma shows how to handle the case of large i.

Lemma 5.6. *If, for some $n > m$, $\bar{l}_n < 0$, then*

$$0 > \bar{l}_i > \bar{l}_j \qquad \text{for} \qquad i < j, \quad i \geq n. \tag{5.16}$$

Proof. From Lemma 5.5, it follows that

$$\bar{l}_i = i((i - \nu i^3) + 2m(\overline{S}(i-m) + \overline{S}(i+1))a + 2(\overline{S}(i+m+1) + 2\overline{S}(1))),$$

where $a = \max_{k,l=1,\ldots,m} |A^{-1}_{kl}|$.

Hence

$$\bar{l}_i = i((i - \nu i^3) + f(i)), \tag{5.17}$$

where $f(i)$ is a positive decreasing function of i. Since $\bar{l}_n < 0$, then $(n - \nu n^3) < 0$ also, and it is easy to see that the function $i \mapsto (i - \nu i^3)$ is decreasing and negative for $i \geq n$. ■

5.4. Formulas for "complex" blocks. The purpose of this subsection is to derive a formula for $l_{(i)}$ in case of a two-dimensional block from the diagonalization corresponding to a complex eigenvalue. The main results are summarized in Lemmas 5.8 and 5.9.

Lemma 5.7. *Let $Q \in \mathbb{R}^{2\times 2}$; then (in the Euclidean norm)*

$$|Q| \leq \sqrt{Q_{11}^2 + Q_{12}^2} + \sqrt{Q_{21}^2 + Q_{22}^2}.$$

Proof. Let $v = (v_1, v_2)$; then

$$|Qv| \leq |Q_{11}v_1 + Q_{12}v_2| + |Q_{21}v_1 + Q_{22}v_2| \leq \sqrt{Q_{11}^2 + Q_{12}^2}|v| + \sqrt{Q_{21}^2 + Q_{22}^2}|v|. \quad ■$$

Lemma 5.8. *If $(i) = (i_1, i_2)$, then*

$$\sup_{x\in W} \mu\left(\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}}\right) \leq \max_{k=1,2}\left(\sup_{x\in W} \frac{\partial \tilde{F}_{i_k}}{\partial x_{i_k}}(x)\right) + \sup_{x\in W} 1/2\left|\frac{\partial \tilde{F}_{i_1}}{\partial x_{i_2}}(x) + \frac{\partial \tilde{F}_{i_2}}{\partial x_{i_1}}(x)\right|.$$

Proof. The proof is an immediate consequence of Theorem 3.5 and the Gershgorin theorem (see [QSS, Property 5.2]). ■

Observe that, from the diagonalization of a block corresponding to a complex eigenvalue, we obtain

$$\frac{\partial \tilde{F}_{(i)}}{\partial x_{(i)}} \approx \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix};$$

hence $\sup_{x\in W} 1/2|\frac{\partial \tilde{F}_{i_1}}{\partial x_{i_2}}(x) + \frac{\partial \tilde{F}_{i_2}}{\partial x_{i_1}}(x)|$ is usually very small.

The following lemma takes care of nondiagonal terms.

Lemma 5.9. *If $(i) = (i_1, i_2)$, $(i) \leq m$, then*

$$\sum_{(j),(j)\neq(i)} \sup_{x\in W}\left|\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x)\right| \leq \sum_{j\leq M, j\neq i_1,i_2} \sup_{x\in W}\sqrt{\left(\frac{\partial \tilde{F}_{i_1}}{\partial x_j}\right)^2 + \left(\frac{\partial \tilde{F}_{i_2}}{\partial x_j}\right)^2}$$
$$+ \sum_{l=1,2}\sum_{k\leq m} 2|A_{i_l,k}|k(\overline{S}(M+1-k) + \overline{S}(M+1+k)).$$

Proof. If dim $(j) = 1$, then

$$\frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x) = \left[\frac{\partial \tilde{F}_{i_1}}{\partial x_j}(x), \frac{\partial \tilde{F}_{i_2}}{\partial x_j}(x) \right]. \tag{5.18}$$

Therefore, we obtain

$$\left| \frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x) \right| = \sqrt{\left(\frac{\partial \tilde{F}_{i_1}}{\partial x_j}(x) \right)^2 + \left(\frac{\partial \tilde{F}_{i_2}}{\partial x_j}(x) \right)^2}. \tag{5.19}$$

From Lemma 5.7, it follows that we can ignore the block structure for all blocks different from (i) and use the above formula for all coordinates.

Therefore, we have

$$\sum_{(j)\neq(i)} \sup_{x\in W} \left| \frac{\partial \tilde{F}_{(i)}}{\partial x_{(j)}}(x) \right| \leq \sum_{j\neq i_1,i_2} \sup_{x\in W} \sqrt{\left(\frac{\partial \tilde{F}_{i_1}}{\partial x_j} \right)^2 + \left(\frac{\partial \tilde{F}_{i_2}}{\partial x_j} \right)^2}. \tag{5.20}$$

To finish the proof, it is enough to show that

$$\begin{aligned} &\sum_{j>M} \sup_{x\in W} \sqrt{\left(\frac{\partial \tilde{F}_{i_1}}{\partial x_j} \right)^2 + \left(\frac{\partial \tilde{F}_{i_2}}{\partial x_j} \right)^2} \\ &\leq \sum_{l=1,2} \sum_{k\leq m} 2|A_{i_l,k}| k (S(M+1-k) + S(M+1+k)). \end{aligned}$$

To make the notation less cumbersome, we will drop $\sup_{x\in W}$ from the computations below:

$$\sum_{j>M} \sqrt{\left(\frac{\partial \tilde{F}_{i_1}}{\partial x_j} \right)^2 + \left(\frac{\partial \tilde{F}_{i_2}}{\partial x_j} \right)^2} \leq \sum_{j>M} \left| \frac{\partial \tilde{F}_{i_1}}{\partial x_j} \right| + \sum_{j>M} \left| \frac{\partial \tilde{F}_{i_2}}{\partial x_j} \right|. \tag{5.21}$$

We have, for $l = 1, 2$ (observe that $i_l \leq m$),

$$\begin{aligned} \sum_{j>M} \left| \frac{\partial \tilde{F}_{i_l}}{\partial x_j} \right| &\leq \sum_{j>M} \sum_{k\leq m} |A_{i_l,k}| \left| \frac{\partial F_k}{\partial x_j} \right| \\ &= \sum_{k\leq m} |A_{i_l,k}| \sum_{j>M} \left| \frac{\partial F_k}{\partial x_j} \right| \leq \sum_{k\leq m} 2k |A_{i_l,k}| \sum_{j>M} (|a_{j-k}| + |a_{j+k}|) \\ &\leq \sum_{k\leq m} 2k |A_{i_l,k}| \left(S(M+1-k) + S(M+1+k) \right). \end{aligned}$$

This finishes the proof. ■

6. Conclusions and future work. We have shown that we can prove rigorously the existence of branches of attracting fixed points for the KS equation with odd periodic boundary conditions. Below we indicate some further possible developments and applications of the method:

- a rigorous steady state bifurcation diagram for the KS equation,
- applications of other dissipative PDEs, e.g., Navier–Stokes equations with periodic boundary conditions on the plane,
- an automatization of generation of formulas for tail (a content of section 5 in this paper and section 3 in [ZM]) for KS and other equations.

REFERENCES

[HNW] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations* I*, Nonstiff Problems*, Springer-Verlag, Berlin, Heidelberg, 1987.

[JKT] M. Jolly, I. Kevrekidis, and E. Titi, *Approximate inertial manifolds for the Kuramoto–Sivashinsky equation: Analysis and computations*, Phys. D, 44 (1990), pp. 38–60.

[Mo] R. E. Moore, *Methods and Applications of Interval Analysis*, SIAM Stud. Appl. Math. 2, SIAM, Philadelphia, 1979.

[QSS] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics*, Texts Appl. Math. 37, Springer-Verlag, New York, 2000.

[Z] P. Zgliczynski, *Trapping Regions and an ODE-Type Proof of an Existence and Uniqueness for Navier-Stokes Equations with Periodic Boundary Conditions on the Plane*, available online at http://arXiv.org/abs/math/0103053; see also http://www.im.uj.edu.pl/~zgliczyn.

[ZM] P. Zgliczynski and K. Mischaikow, *Rigorous numerics for partial differential equations: The Kuramoto-Sivashinsky equation*, Found. Comput. Math., 1 (2001), pp. 255–288.

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 1, No. 2, pp. 236–247

Development of Standing-Wave Labyrinthine Patterns[*]

Arik Yochelis[†], Aric Hagberg[‡], Ehud Meron[§], Anna L. Lin[¶], and Harry L. Swinney[¶]

Abstract. Experiments on a quasi-two-dimensional Belousov–Zhabotinsky (BZ) reaction-diffusion system, periodically forced at approximately twice its natural frequency, exhibit resonant labyrinthine patterns that develop through two distinct mechanisms. In both cases, large amplitude labyrinthine patterns form that consist of interpenetrating fingers of frequency-locked regions differing in phase by π. Analysis of a forced complex Ginzburg–Landau equation captures both mechanisms observed for the formation of the labyrinths in the BZ experiments: a transverse instability of front structures and a nucleation of stripes from unlocked oscillations. The labyrinths are found in the experiments and in the model at a similar location in the forcing amplitude and frequency parameter plane.

Key words. labyrinthine patterns, Belousov–Zhabotinsky reaction, complex Ginzburg–Landau equation

AMS subject classifications. 35, 37

PII. S1111111101397111

1. Introduction.

Labyrinthine patterns occur in a variety of equilibrium and nonequilibrium systems. Competition between two interacting phases in diblock copolymers [21], ferrofluids [25], and Langmuir films [25] results in labyrinthine domain patterns of the two phases at equilibrium. Labyrinths made of superconducting and normal phases are found in thin films of type-I superconductors [13]. Nonequilibrium labyrinthine patterns are observed in chemical reaction-diffusion systems with a Turing instability [22] and in bistable reaction-diffusion systems [14, 17]. Although periodically forced oscillatory systems have been studied in the context of traveling waves and spiral waves [2, 1, 26, 8], only a few studies have focused on labyrinthine standing-wave patterns. Labyrinthine patterns have been found in numerical simulations of the periodically forced Brusselator reaction-diffusion equations [18] and in numerical solutions of the normal form equation for the oscillation amplitude [1, 23].

Experiments on the periodically forced photosensitive Belousov–Zhabotinsky (BZ) reaction produce nonequilibrium labyrinthine patterns when the system is forced with time-

[*] Received by the editors October 30, 2001; accepted for publication (in revised form) by M. Golubitsky August 19, 2002; published electronically October 24, 2002. This research was supported by grant 9800129 from the US–Israel Binational Science Foundation.

http://www.siam.org/journals/siads/1-2/39711.html

[†] Department of Physics, Ben-Gurion University, Beer Sheva, 84105, Israel (yochelis@bgumail.bgu.ac.il).

[‡] Center for Nonlinear Studies and T-7, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 (hagberg@lanl.gov). The work of this author was supported by the U.S. Department of Energy under contract W-7405-ENG-36.

[§] Department of Solar Energy and Environmental Physics, Ben-Gurion University, Sede Boker Campus 84990, Israel (ehud@bgumail.bgu.ac.il).

[¶] Center for Nonlinear Dynamics and Department of Physics, The University of Texas at Austin, Austin, TX 78712 (alin@phy.duke.edu, swinney@chaos.ph.utexas.edu). The work of these authors was supported by the Engineering Research Program of the Office of Basic Energy Sciences of the U.S. Department of Energy and by the Robert A. Welch Foundation.

periodic pulses of light that are approximately twice the system's natural oscillation frequency [24, 18]. In this case, the two phases correspond to two phases of oscillation, each locked to the time-periodic forcing and shifted by π with respect to one another. We observe that the standing-wave labyrinthine patterns form in two distinct ways: from a transverse instability of planar fronts connecting the two phase-locked states or by nucleating stripes from unlocked oscillating domains. Even though the mechanisms are different, the resulting patterns in both cases are large amplitude labyrinths.

In this paper, we explain the experimental observations using a normal form equation for periodically forced oscillatory systems. We demonstrate the two mechanisms for labyrinthine pattern formation and give criteria for the parameter regions where they act. Since the mechanisms of labyrinthine pattern formation are found in a normal form equation, these mechanisms should also be observed in other periodically forced oscillatory systems such as electro-convection [15].

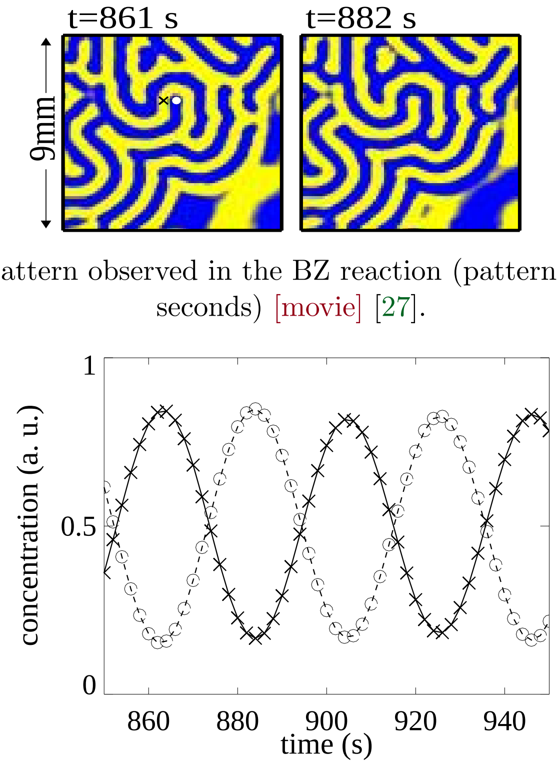
2. Experimental results.

We create chemical labyrinths in a porous membrane (0.4 mm thick, 22 mm diameter) fed by two reservoirs. Each reservoir contains a subset of the chemical reactants for the oscillatory photosensitive (ruthenium catalyzed) BZ reaction [18, 24], and the two reservoirs are in contact with opposite faces of the porous membrane. We force the reaction externally with spatially homogeneous time-periodic square waves of light. The patterns form in the membrane as variations in the concentration of the chemical catalyst Ru(III). The unforced pattern is a rotating spiral wave of Ru(III) concentration. Parametric forcing with light modulated in intensity at a frequency that is approximately twice that of the chemical oscillation frequency results in chemical patterns that oscillate once every two forcing cycles. These patterns, which we call 2:1 resonant patterns, consist of synchronous domains that oscillate with a relative phase difference of π.

Photobleaching experiments [19] reveal that the membrane supports spatially uniform oscillations over a range of chemical concentrations. We conduct our experiments within this range; thus our experimental conditions are far from the Hopf bifurcation. The labyrinthine patterns we describe here are 2:1 resonant with the natural frequency (the spatially homogeneous oscillation frequency), not the oscillation frequency of the spiral pattern [18].

Different resonant patterns are observed, depending on the forcing frequency ω_f and the forcing strength γ [18]. The region of 2:1 resonant dynamics forms a tongue in the $\omega_f - \gamma$ parameter space [12]. For most parameter values in the tongue, the patterns consist of irregularly shaped standing-wave domains differing in phase by π. Near the bottom of the tongue, there are rotating phase-locked spiral waves, and, on one side of the tongue, standing-wave labyrinthine patterns form [18], as shown in Figure 1. Outside the range of 2:1 resonant dynamics, patterns are either unlocked or locked at a different resonance.

A useful way to characterize the spatio-temporal patterns is in terms of the complex Fourier amplitude $a(x, y)$ of a particular mode in the time series of each pixel (x, y) in the pattern [18]. We look at $a(x, y)$ for the $\omega_f/2$ mode, the primary response mode of the pattern. For each pixel (x, y) in the labyrinthine pattern of Figure 1, $a(x, y)$ is plotted as a black dot in the Re-Im plane shown in Figure 2(a). The points located at the ends of the "S"-shaped curve correspond to pixels in one of the two phase-locked domains. The other points are from the interfaces between domains. The ends of the "S"-shaped curve are π out of phase, which



Labyrinthine pattern observed in the BZ reaction (pattern sampled every 2 seconds) [movie] [27].

Figure 1. *Resonant labyrinthine pattern in the* 2:1 *periodically forced BZ reaction. (top) The two images show the spatial* Ru(III) *concentration pattern in a 9 mm region of the chemical reactor. Yellow represents regions of high* Ru(III) *concentration, and blue represents regions of low concentration. The images are at two different times separated by one forcing period (*21 *s). (bottom) A time series from two locations in the top pattern (marked with "x" and "○") shows that the pattern is formed of regions of two different phases separated by* π*. The chemical conditions are given in Figure* 3*.*

shows that the phase-shift between the yellow and the blue domains pictured in Figure 1 is π. The distribution of points in Figure 2(a) as a function of phase angle $\theta = \arg(a(x, y))$ is shown in Figure 2(b). The histogram shows that most of the pattern is in one of the two phase-locked states, $\theta = 0$ and $\theta = \pi$, which correspond to the yellow and blue regions in Figure 1. From this representation of the data, we clearly see that the observed labyrinths are large amplitude patterns comprised of two phase-locked, π-shifted domains, as opposed to small amplitude modulations on one or both of the two phases.

The development of labyrinthine patterns by the two mechanisms is shown in Figure 3. A labyrinth can grow from a transverse instability of a front that separates two phase-locked domains, as shown in Figure 3(a). A labyrinth pattern can also develop, stripe by stripe, from an unlocked oscillatory state, as shown in Figure 3(b). The resulting labyrinthine pattern in Figure 3(b) is similar to the labyrinth shown in Figure 3(a); both patterns consist of standing-wave domains oscillating with a relative phase-shift of π.

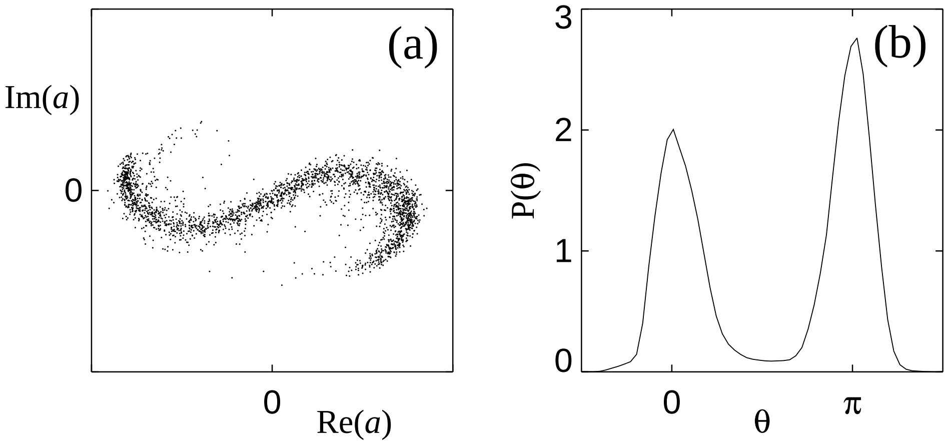


Figure 2. (a) *The complex Fourier amplitude $a(x, y)$ of the $\omega_f/2$ mode corresponding to the labyrinthine pattern shown in Figure 1, plotted in the* Re-Im *phase plane.* (b) *A histogram $P(\theta)$ of the phase angle $\theta = \arg(a)$. Most of the pattern is locked at one of two phase angles, $\theta = 0, \pi$.*

3. Theoretical analysis. To study the mechanisms by which these chemical labyrinths form, we model the oscillating BZ chemical reaction as an extended system with a Hopf bifurcation to uniform oscillations. Let $\mathbf{u}$ be a vector field of chemical concentrations responding at $\omega_f/2 \approx \omega_0$, where ω_0 is the frequency of the unforced system and ω_f is the forcing frequency. Near a supercritical Hopf bifurcation, the field can be written as

$$\mathbf{u} = \mathbf{u}_0 A e^{i\omega_f t/2} + \text{complex conjugate } (c.c.) + \cdots , \tag{3.1}$$

where $\mathbf{u_0}$ is a constant, A is a complex amplitude, and the ellipses denote higher order terms. The amplitude of oscillation $A(x, y, \tau)$ is slowly varying in space and time and is described by the complex Ginzburg–Landau equation [11, 1, 6, 7]

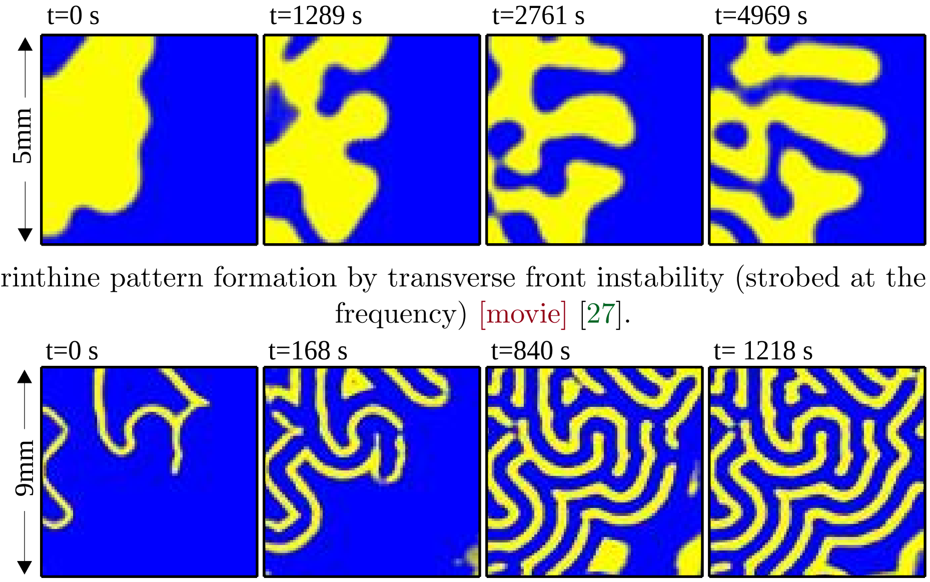
$$A_\tau = (\mu + i\nu)A + (1 + i\alpha)\nabla^2 A - (1 + i\beta)|A|^2 A + \gamma A^* \, . \tag{3.2}$$

The equation has been scaled to its reduced form, where μ is the distance from the Hopf bifurcation, ν is the detuning (the deviation of ω_f from $2\omega_0$), α is a dispersion parameter, and γ is the forcing amplitude. The term A^* is the complex conjugate of A and appears from the addition of 2:1 periodic forcing [11]. To simplify the following discussion, we set $\beta = 0$.

When $\mu > 0$, the spatially uniform solution $A = 0$ is unstable. In the unforced system ($\gamma = 0$), (3.2) has a continuous family of uniformly oscillating solutions

$$A = \sqrt{\mu} e^{i\nu\tau + i\phi_0} \, , \tag{3.3}$$

where ϕ_0 is an arbitrary constant phase. Equation (3.2) also has plane wave solutions, but we do not consider them here. The existence of a continuous family of solutions when $\gamma = 0$ is a consequence of the phase-shift invariance of (3.2), $A \to A\exp(i\phi_0)$, which follows from the time translation symmetry of the unforced oscillatory system. The forcing term γA^* restricts the phase-shift invariance to π phase-shifts. At $\gamma = |\nu|$, two stable uniform phase-locked



Labyrinthine pattern formation by transverse front instability (strobed at the pattern frequency) [movie] [27].

Labyrinthine pattern formation by nucleation of stripes (strobed at the pattern frequency) [movie] [27].

Figure 3. *Formation of labyrinthine patterns through (top) a transverse front instability and (bottom) a nucleation of stripes. The patterns are from a region of the BZ reactor, strobed at half the forcing frequency. Blue (yellow) represents regions of low (high)* Ru(III) *concentration.* (a) $\gamma = 600$ W/m^2, $\omega_f = 0.273$ *rad/s. Reservoir* I: 0.22 *M malonic acid,* 0.046 *M* $\mathrm{KBr0_3}$*,* 0.2 *M* KBr*, and* 0.8 *M* $\mathrm{H_2SO_4}$*; reservoir* II: 1.0 *mM* $\mathrm{Ru(2,2'-bipyridine)_3+2}$*,* 0.8 *M* $\mathrm{H_2SO_4}$*, and* 0.184 *M* $\mathrm{KBrO_3}$*. The flow rates through each reservoir are* 20 *ml/h and* 5 *ml/h, respectively. The reservoir volumes are each* 10 *ml.* (b) $\gamma = 228$ W/m^2, $\omega_f = 0.300$ *rad/s. Reservoir* I: 0.22 *M malonic acid,* 0.2 *M NaBr,* 0.264 *M* $\mathrm{KBrO_3}$*,* 0.8 *M* $\mathrm{H_2SO_4}$*; reservoir* II: 0.184 *M* $\mathrm{KBrO_3}$*,* 1×10^{-3} *M Tris(*2, 2*'-bipyridyl)dichlororuthenium*(II)*hexahydrate,* 0.8 *M* $\mathrm{H_2SO_4}$*. The volume of each reservoir was* 8.3 *ml. The flow rate through reservoir* I *was* 20 *ml/h; that through reservoir* II *was* 5 *ml/h.*

solutions are formed in a pair of saddle-node bifurcations on a circle of amplitude $|A|$. These solutions describe oscillations at precisely half the forcing frequency, $\omega_f/2$, even though for $\nu \neq 0$ the oscillation frequency of the unforced system differs from $\omega_f/2$. The phases of the two solutions differ by π.

To see how these solutions appear, we consider spatially uniform solutions of (3.2) and write the complex amplitude in a polar form, $A = |A|\exp(i\phi)$. Using this form in (3.2), we find

$$\phi_\tau = \nu - \gamma \sin(2\phi)\,. \tag{3.4}$$

In order for (3.1) to describe resonant dynamics, we must look for stationary solutions of (3.4) (so that $\mathbf{u}$ oscillates at $\omega_f/2$). Stationary solutions of this equation exist for $\gamma \geq |\nu|$:

$$\begin{aligned}\phi_S^- &= \frac{1}{2}\arcsin\left(\frac{\nu}{\gamma}\right), \qquad \phi_S^+ = \phi_S^- + \pi\,,\\ \phi_U^- &= \frac{\pi}{2} - \frac{1}{2}\arcsin\left(\frac{\nu}{\gamma}\right), \qquad \phi_U^+ = \phi_U^- + \pi\,,\end{aligned} \tag{3.5}$$

where the subscripts S and U refer to stable and unstable solutions, respectively. At $\gamma = |\nu|$, the two pairs of solutions, ϕ_S^-, ϕ_U^- and ϕ_S^+, ϕ_U^+ are born in saddle-node bifurcations, as shown in Figure 4(a). The condition $\gamma > |\nu|$ defines the 2:1 resonance tongue, where the system's frequency is locked to one half of the forcing frequency. The resonance tongue in the ν-γ plane is shown in Figure 4(b).

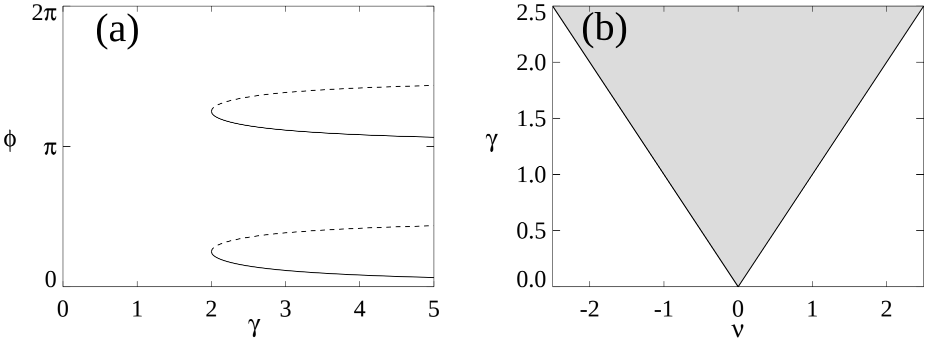


Figure 4. (a) *Bifurcation diagrams showing the formation of the four stationary phase solutions in* (3.5) *for fixed detuning* ($\nu = 2$) *and varying forcing strength* γ. *The solid (dashed) curves represent stable (unstable) solutions.* (b) *The existence range of the phase solutions in* (a), $\gamma > |\nu|$, *defines the resonance tongue in the* ν-γ *plane (shaded region) inside which the original system responds at half the forcing frequency.*

Inside the tongue, front structures form between the two phases ϕ_S^- and ϕ_S^+. At low forcing γ, the fronts travel, and the domains organize into two-phase spiral waves [1, 9]. As the forcing is increased, the system goes through a nonequilibrium Ising–Bloch (NIB) bifurcation [2, 8]. For $\nu = \alpha = 0$, the NIB bifurcation point is at $\gamma = \mu/3$. A numerically computed NIB bifurcation boundary for nonzero ν and α values is shown in Figure 5(a). Above the NIB bifurcation, only stationary Ising fronts exist. In the following, we confine ourselves to the Ising regime well beyond the NIB bifurcation.

When $\alpha \neq 0$, there is a range of parameters in the $\nu - \gamma$ parameter plane where the Ising fronts are unstable to transverse perturbations. The boundary of this linear transverse instability can be computed numerically and is shown as the line γ_T in Figure 5(b).

When $\gamma > \gamma_T$, the fronts are stable, and domains of the two phases persist for long periods. For $\nu < \gamma < \gamma_T$, stationary fronts are unstable. Perturbations along the front grow into fingers, which tip, split, and form labyrinthine patterns, as shown in Figure 6(a). The amplitude of the labyrinth approaches that of the uniform phase-locked states $|A| \sim (\mu + \sqrt{\gamma^2 - \nu^2})^{1/2}$ and is large because of the large distance μ from the Hopf bifurcation. Note the similarity to the experimental labyrinth formation shown in Figure 3(a).

Outside the tongue ($\gamma < \nu$), uniform phase-locked solutions do not exist, but resonant labyrinthine patterns still persist [23]. The labyrinthine patterns form in a range $\gamma_N < \gamma < \nu$ outside the tongue boundary and, similar to the labyrinths inside the tongue, are characterized by large amplitudes. This observation can be explained by a coupling of a finite-wavenumber (Turing) mode to a zero-wavenumber mode [5, 20, 3]. In the present case, the zero-wavenumber mode has uniform oscillations. In the following, we derive equations for the amplitudes of these modes and use them to obtain a criterion for the boundary γ_N.

Consider the dispersion relation associated with perturbations, $A \sim \exp(\sigma\tau - ikx)$, of the

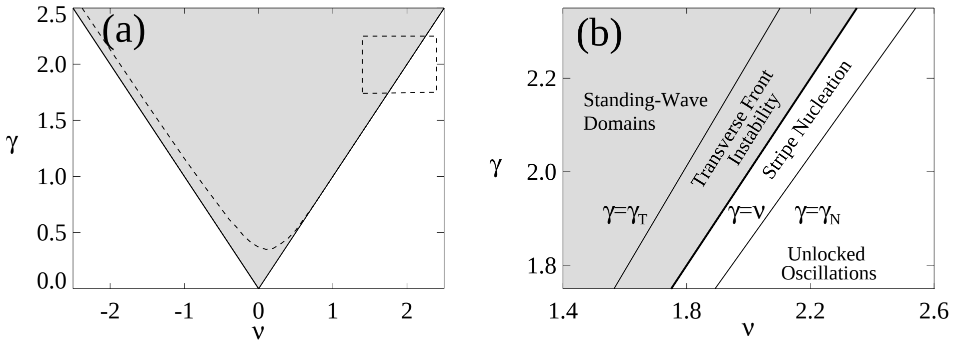


Figure 5. (a) *The NIB bifurcation boundary (dashed curve) inside the resonance tongue, obtained by numerical integration of* (3.2)*. Below the boundary, the fronts that connect the two stable phase-locked states, ϕ_S^- and ϕ_S^+, are traveling Bloch fronts. Above the boundary, the fronts are stationary Ising fronts.* (b) *A closeup of the rectangular region indicated in* (a) *showing the regions where a labyrinthine pattern forms. For $\nu < \gamma < \gamma_T$, a labyrinth forms by a transverse front instability, while, for $\gamma_N < \gamma < \nu$, it forms by stripe nucleation from the unlocked oscillating state. Outside these regions, the dominant pattern is unlocked oscillations ($\gamma < \gamma_N$) or irregularly shaped standing-wave domains with nearly stationary interfaces ($\gamma > \gamma_T$). The boundaries γ_N and γ_T are computed from direct numerical solution of* (3.2)*. Parameters: $\mu = 1$, $\beta = 0$, $\alpha = 0.5$.*

$A = 0$ state

$$\sigma(k) = \mu - k^2 + \sqrt{\gamma^2 - (\nu - \alpha k^2)^2}. \tag{3.6}$$

At the codimension 2 point, $\mu = 0$, $\gamma = \gamma_c$, where

$$\gamma_c = \frac{\nu}{\sqrt{1+\alpha^2}}\,,$$

two modes become marginally stable in a Turing–Hopf bifurcation [16, 4]. That is, the growth rates, shown in Figure 7, are zero for both a zero-k mode describing uniform oscillations, $k = 0$, $\omega = \omega_0$, and a finite-k mode describing a stationary pattern, $k = k_c$, $\omega = 0$, where $\omega = \mathrm{Im}(\sigma)$ and

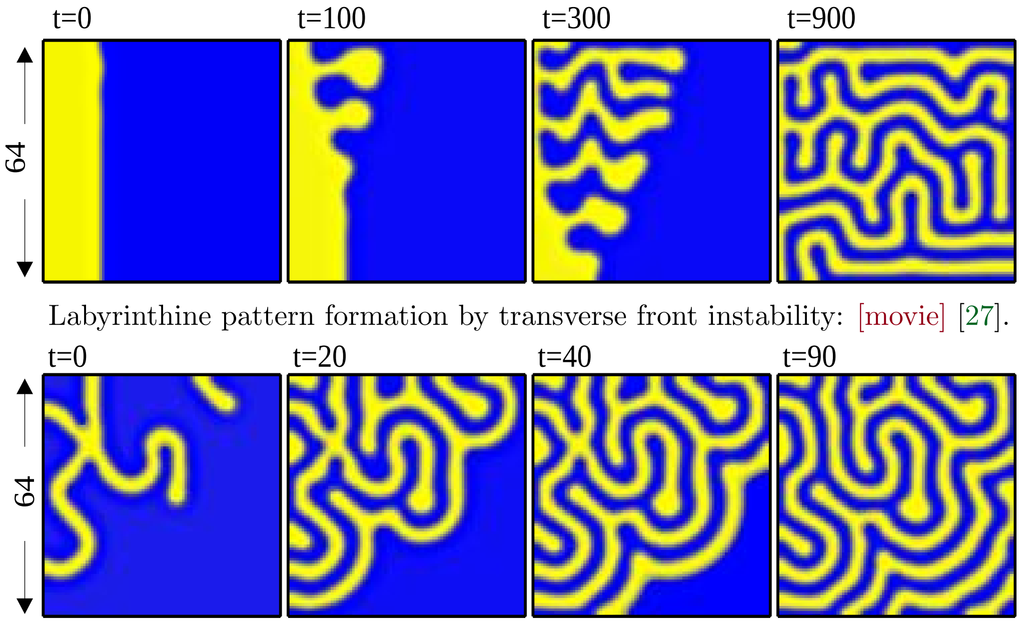
$$k_c^2 = \frac{\nu\alpha}{1+\alpha^2}\,,$$
$$\omega_0 = \frac{\nu\alpha}{\sqrt{1+\alpha^2}}\,.$$

To study the coupling of the two modes, we assume $|\mu| \sim |d| \ll 1$, where $d := \gamma - \gamma_c$, and we consider (3.2) in one space dimension. We then express A in terms of its real and imaginary parts, $A = U + iV$, and expand

$$\begin{pmatrix} U \\ V \end{pmatrix} = \sqrt{\mu}\begin{pmatrix} U_0 \\ V_0 \end{pmatrix} + \mu\begin{pmatrix} U_1 \\ V_1 \end{pmatrix} + \mu^{3/2}\begin{pmatrix} U_2 \\ V_2 \end{pmatrix} + \cdots\,, \tag{3.7}$$

where the ellipses denote higher order contributions and

$$\begin{pmatrix} U_0 \\ V_0 \end{pmatrix} = \mathbf{e_0} B_0\,(X,T)\,e^{i\omega_0\tau} + \mathbf{e_k} B_k\,(X,T)\,e^{ik_c x} + c.c. \tag{3.8}$$



Labyrinthine pattern formation by transverse front instability: [movie] [27].

Labyrinthine pattern formation by stripe nucleation: [movie] [27].

Figure 6. *Formation of labyrinthine patterns in* (3.2). *Blue and yellow regions are different phases separated by π. (top) When $\nu < \gamma < \gamma_T$ ($\gamma = 2.02$), the interface between the phase-locked domains is transversely unstable, and small perturbations grow and finger. (bottom) Outside the tongue, when $\gamma_N < \gamma < \nu$ ($\gamma = 1.98$), the pattern forms by nucleating stripes from the unlocked oscillating state. The stripes are unstable to the zigzag instability. Other parameters: $\mu = 1$, $\nu = 2.0$, $\alpha = 0.5$.*

Here $\mathbf{e_0}$ and $\mathbf{e_k}$ are the eigenvectors corresponding to the eigenvalues $\sigma(0)$ and $\sigma(k_c)$, respectively.

The amplitudes B_0 and B_k in (3.8) describe weak spatio-temporal modulations of the (relatively) fast oscillations associated with the zero-k mode and of the strong spatial variations associated with the finite-k mode. The weak dependence is expressed by the introduction of the slow variables $T = \mu\tau,\ X = \sqrt{\mu}x$.

Inserting the expansion (3.7) into (3.2), solving the linear equations at order μ, and evaluating the solvability condition at order $\mu^{3/2}$, we find the amplitude equations

$$\begin{aligned} \partial_T B_0 &= (\mu - i\zeta)B_0 - 4\left|B_0\right|^2 B_0 - (a - ib)\left|B_k\right|^2 B_0 + (1 + i\rho)\partial_X^2 B_0\,, \\ \partial_T B_k &= \eta B_k - c_1\left|B_k\right|^2 B_k - 8\left|B_0\right|^2 B_k + c_2\partial_X^2 B_k\,, \end{aligned} \tag{3.9}$$

with the coefficients

$$\begin{aligned} \zeta &= d/\alpha \quad (d = \gamma - \gamma_c)\,, \\ \eta &= \mu + \zeta\rho\,, \\ \rho &= \sqrt{1 + \alpha^2}\,, \\ a &= 8\rho(\rho + \alpha)\,, \\ b &= 4(\rho + \alpha)\,, \end{aligned}$$

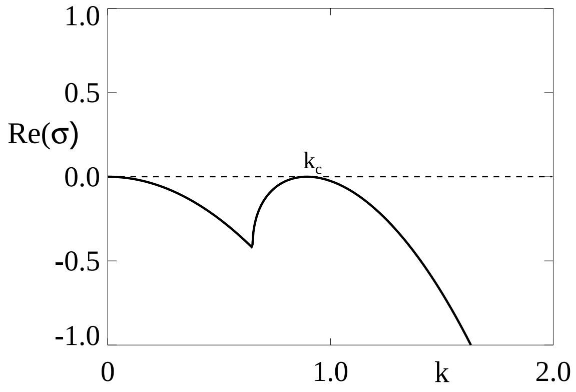


Figure 7. *The growth rate* $\mathrm{Re}(\sigma)$ *for perturbations of the* $A = 0$ *solution at the codimension* 2 *point:* $\mu = 0$, $\gamma = \gamma_c$. *Two modes become marginal at this point: a zero-k (Hopf) mode and a finite-k (Turing) mode. Parameters:* $\mu = 0$, $\nu = 2.0$, $\alpha = 0.5$, $\gamma = \gamma_c \approx 1.8$.

$$c_1 = 3a/4\,,$$
$$c_2 = 2\rho^2\,.$$

More details about the derivation of (3.9) will be presented elsewhere.

Equations (3.9) have two families of pure-mode uniform solutions,

$$B_0 = \frac{1}{2}\sqrt{\mu}e^{-i\zeta T + i\psi_1}\,, \qquad B_k = 0\,, \tag{3.10}$$

representing uniform oscillations, and

$$B_0 = 0\,, \qquad B_k = \sqrt{\eta/c_1}e^{i\psi_2}\,, \tag{3.11}$$

representing stationary periodic patterns, where ψ_1 and ψ_2 are arbitrary constants. Outside and close enough to the tongue boundary ($\gamma = \nu$), both families of solutions are linearly stable. The family of solutions representing stationary patterns, however, loses stability as γ is decreased past

$$\gamma_S = \gamma_c - \frac{\mu\alpha}{4\sqrt{1+\alpha^2}} = \frac{\nu - \mu\alpha/4}{\sqrt{1+\alpha^2}}\,. \tag{3.12}$$

Figure 8 shows the boundary $\gamma = \gamma_S$ as computed from (3.12) and compared with results from the direct numerical solution of (3.2) in one space dimension. The agreement is good despite the relatively large value of μ.

The amplitude equations (3.9) also have a mixed-mode family of uniform solutions ($B_0 \neq 0$, $B_k \neq 0$), but these solutions are unstable.

The existence boundary of resonant stripes $\gamma = \gamma_S$ is well below the boundary $\gamma = \gamma_N$, where stripes are observed to nucleate from the unlocked oscillation state. This observation can be understood by considering front solutions of (3.9) which are biasymptotic to the two states $(0, B_k)$ and $(B_0, 0)$ as $x \to \pm\infty$. Numerical studies of these equations in the range

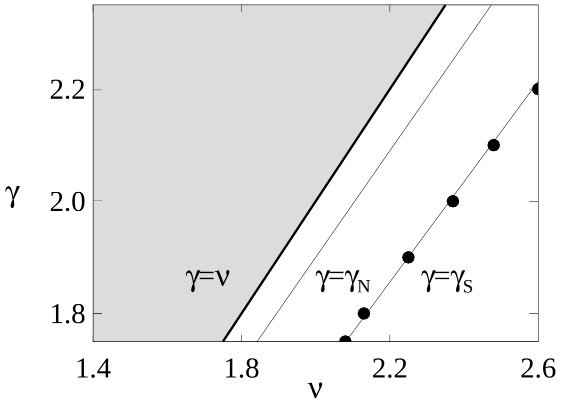


Figure 8. *The boundaries γ_S of stationary stripe patterns and γ_N of stripe nucleation outside of the resonance tongue. The line $\gamma = \gamma_S$ was computed using* (3.12)*, and the solid circles represent the numerical solution of the model equation* (3.2)*. The line $\gamma = \gamma_N$ was computed by direct numerical solution of* (3.9)*. Stationary stripes are stable between the tongue boundary $\gamma = \nu$ and γ_S but nucleate from unlocked oscillations only between $\gamma = \nu$ and γ_N. When $\gamma < \gamma_S$, the stripes are unstable to uniform oscillations. The parameter values are the same as in Figure* 5*.*

$\gamma_S < \gamma < \nu$ show the existence of a zero front-velocity line. We identify this line with the boundary $\gamma = \gamma_N$. For $\gamma > \gamma_N$, the stationary stripe state, $(0, B_k)$, invades the uniform oscillation state, $(B_0, 0)$, and in this sense is dominant. In the context of (3.2), this invasion takes the form of stripe nucleation, as Figure 6(b) shows. The stripes nucleate stripe by stripe at the boundary of the growing stationary pattern domain. The stripes are also unstable to transverse perturbations (zigzag instability). Note the similarity to the experimental labyrinth formation shown in Figure 3(b).

The analysis of the amplitude equations (3.9) also provides the amplitude of the stationary stripes. The amplitude of the stripes is given by $|B_k| = \sqrt{\eta/c_1}$ (see (3.11)). Far from the Hopf bifurcation where $\mu \sim O(1)$, $|B_k|$ can be of order unity even at $\gamma = \gamma_N$. This explains the large amplitude values of the stationary stripe patterns in the range $\gamma_N < \gamma < \nu$.

4. Conclusions. Both the experiments and the complex Ginzburg–Landau equation produce nonequilibrium labyrinthine patterns through two different mechanisms: a transverse instability of fronts between locked states and a nucleation of stripes from an unlocked oscillating state. Unlike previous studies of phase-locked labyrinthine patterns, the labyrinths are not small amplitude patterns modulating one of the phase-locked states [1]. Resonant labyrinths persist even outside the 2:1 tongue of uniform phase-locked states in the complex Ginzburg–Landau model, indicating that the boundary for resonant patterns in extended oscillating systems need not coincide with that of a single forced oscillator. The large amplitude of the labyrinths both inside and outside the tongue boundary is a consequence of the large distance of the system from the Hopf bifurcation.

The labyrinthine patterns are found on only one side of the 2:1 resonance tongue both in the experiments and in the forced complex Ginzburg–Landau equation. In the experiments, they are found on the side of the tongue closest to the 3:1 resonance tongue. In (3.2), the side of the tongue is determined by the sign of the parameter α.

Similar labyrinthine patterns have been observed in numerical solutions of the forced Brusselator reaction-diffusion system [18] but without a description of the underlying mechanism. Those results show similar features, such as the labyrinths forming in a region on only one side of the resonance tongue. Since the complex Ginzburg–Landau equation we study is a generic model, we expect to find the same mechanisms for labyrinthine pattern formation in other 2:1 resonant oscillatory systems.

Quantitative comparison between the experiment and the model is difficult because the chemical kinetics and diffusion coefficients are not well known. The parameter values in a complex Ginzburg–Landau model depend on these quantities. The two mechanisms of labyrinthine pattern formation are also expected to be found in liquid crystal systems with dynamics described by (3.2) [10, 9].

Acknowledgments. We thank Hezi Yizhaq, Erez Gilad, and Valery Petrov for helpful discussions and comments.

REFERENCES

[1] P. Coullet and K. Emilsson, *Strong resonances of spatially distributed oscillators: A laboratory to study patterns and defects*, Phys. D, 61 (1992), pp. 119–131.

[2] P. Coullet, J. Lega, B. Houchmanzadeh, and J. Lajzerowicz, *Breaking chirality in nonequilibrium systems*, Phys. Rev. Lett., 65 (1990), pp. 1352–1355.

[3] A. De Wit, *Spatial patterns and spatiotemporal dynamics in chemical systems*, Advances in Chemical Physics, 109 (1999), pp. 435–513.

[4] A. De Wit, D. Lima, G. Dewel, and P. Borckmans, *Spatiotemporal dyanmics near a codimension-two point*, Phys. Rev. E (3), 54 (1996), pp. 261–271.

[5] G. Dewel, S. Métens, M. Hilali, P. Borckmans, and C. B. Price, *Resonant patterns through coupling with a zero mode*, Phys. Rev. Lett., 74 (1995), pp. 4647–4650.

[6] C. Elphick, A. Hagberg, and E. Meron, *A phase front instability in periodically forced oscillatory systems*, Phys. Rev. Lett., 80 (1998), pp. 5007–5010.

[7] C. Elphick, A. Hagberg, and E. Meron, *Multiphase patterns in periodically forced oscillatory systems*, Phys. Rev. E (3), 59 (1999), pp. 5285–5291.

[8] C. Elphick, A. Hagberg, E. Meron, and B. Malomed, *On the origin of traveling pulses in bistable systems*, Phys. Lett. A, 230 (1997), pp. 33–37.

[9] T. Frisch and J. M. Gilli, *Excitability and defect-mediated turbulence in nematic liquid crystal*, J. Phys. II France, 5 (1995), pp. 561–572.

[10] T. Frisch, S. Rica, P. Coullet, and J. M. Gilli, *Spiral waves in liquid crystal*, Phys. Rev. Lett., 72 (1994), pp. 1471–1474.

[11] J. M. Gambaudo, *Perturbation of a Hopf-bifurcation by an external time-periodic forcing*, J. Differential Equations, 57 (1985), pp. 172–199.

[12] J. A. Glazier and A. Libchaber, *Quasi-periodicity and dynamical systems: An experimentalist's view*, IEEE Trans. Circuits Systems, 35 (1988), pp. 790–809.

[13] F. Haenssler and L. Rinderer, *Statique et dynamique de l'état intermédiaire des supraconducteurs du type* I, Helv. Phys. Acta, 40 (1967), pp. 659–687.

[14] A. Hagberg and E. Meron, *From labyrinthine patterns to spiral turbulence*, Phys. Rev. Lett., 72 (1994), pp. 2494–2497.

[15] T. Kawagishi, T. Mizuguchi, and M. Sano, *Points, walls, and loops in resonant oscillatory media*, Phys. Rev. Lett., 75 (1995), pp. 3768–3771.

[16] H. Kidachi, *On mode interactions in reaction diffusion equation with nearly degenerate bifurcations*, Progr. Theoret. Phys., 63 (1980), pp. 1152–1169.

[17] K. J. Lee and H. L. Swinney, *Lammelar structures and self-replicating spots in a reaction-diffusion system*, Phys. Rev. E (3), 51 (1995), pp. 1899–1915.

[18] A. L. Lin, M. Bertram, K. Martinez, H. L. Swinney, A. Ardelea, and G. F. Carey, *Resonant phase patterns in a reaction-diffusion system*, Phys. Rev. Lett., 84 (2000), pp. 4240–4243.

[19] K. Martinez, A. L. Lin, R. Kharrazian, X. Sailer, and H. L. Swinney, *Resonance in periodically inhibited reaction-diffusion systems*, Phys. D, 2 (2002), pp. 168–169.

[20] S. Métens, G. Dewel, P. Borckmans, and R. Engelhardt, *Pattern selection in bistable systems*, Europhys. Lett., 37 (1997), pp. 109–114.

[21] G. E. Molau, ed., *Colloidal and Morphological Behavior of Block and Graft Copolymers*, Plenum Press, New York, 1971.

[22] Q. Ouyang and H. L. Swinney, *Transition from a uniform state to hexagonal and striped Turing patterns*, Nature (London), 352 (1991), pp. 610–612.

[23] H.-K. Park, *Frequency locking in spatially extended systems*, Phys. Rev. Lett., 86 (2001), pp. 1130–1133.

[24] V. Petrov, Q. Ouyang, and H. L. Swinney, *Resonant pattern formation in a chemical system*, Nature, 388 (1997), pp. 655–657.

[25] M. Seul and D. Andelman, *Domain shapes and patterns: The phenomenology of modulated phases*, Science, 267 (1995), pp. 476–483.

[26] D. Walgraef, *Spatio-Temporal Pattern Formation*, Springer-Verlag, New York, 1997.

[27] Additional movie formats available at http://math.lanl.gov/Labyrinth/.

SIAM J. Applied Dynamical Systems
Vol. 1, No. 2, pp. 248–270

Numerical Bifurcation Analysis for Multisection Semiconductor Lasers[*]

Jan Sieber[†]

Abstract. We investigate the dynamics of a multisection laser implementing a delayed optical feedback experiment where the length of the cavity is comparable to the length of the laser. First, we reduce the traveling-wave model with gain dispersion (a hyperbolic system of PDEs) to a system of ODEs describing the semiflow on a local center manifold. Then we analyze the dynamics of the system of ODEs using numerical continuation methods (AUTO). We explore the plane of the two parameters—feedback phase and feedback strength—to obtain a bifurcation diagram for small and moderate feedback strength. This diagram permits us to understand the roots of a variety of nonlinear phenomena observed numerically and experimentally such as, e.g., self-pulsations, excitability, hysteresis, or chaos, and to locate them in the parameter plane.

Key words. semiconductor lasers, delayed optical feedback, numerical bifurcation analysis

AMS subject classifications. 34C60, 78A60

PII. S1111111102401746

1. Introduction.

Semiconductor lasers subject to delayed optical feedback show a variety of nonlinear effects. Self-pulsations, excitability, coexistence of several stable regimes, and chaotic behavior have been observed in both experiments and numerical simulations [5], [13], [19], [24], [28]. Multisection lasers allow us to design and control these feedback effects and permit their application, e.g., in optical data transmission, processing, and recovery [24], [34].

If mathematical modeling is to be helpful in guiding this difficult and expensive design process, it has to meet two criteria that contradict each other. On one hand, the model should be accurate, and its parameters should be directly accessible for experimenters. Typically, only very complex models allow for that, e.g., systems of PDEs.

On the other hand, the modeling should permit insight into the nature of the nonlinear effects. This is often impossible using only simulations, i.e., performing the experiment at the computer. Only a detailed bifurcation analysis allows us to find coexisting stable regimes, unstable objects which are boundaries of coexisting attracting regions, and bifurcations of higher codimension which are a common source for various nonlinear phenomena. Unfortunately, there exist numerical bifurcation tools only for low-dimensional ODEs [11], [16] and very restricted classes of PDEs, e.g., delay-differential equations [12].

In this paper, we start from the traveling-wave model [2], [18], [29], which resolves the light amplitude E within the laser spatially in the longitudinal direction but treats the carrier density n as a spatially sectionwise averaged quantity. For illustrative purposes, we restrict

[*] Received by the editors February 1, 2002; accepted for publication by B. Krauskopf June 18, 2002; published electronically November 6, 2002. This work was supported by the Sonderforschungsbereich 555 "Komplexe Nichtlineare Prozesse" of the Deutsche Forschungsgemeinschaft.

http://www.siam.org/journals/siads/1-2/40174.html

[†] Humboldt University, Berlin, Germany, and Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D-10117 Berlin, Germany (sieber@wias-berlin.de).

our analysis to a particular multisection laser configuration implementing a classical delayed optical feedback experiment (see Figure 1). This way, the model has the structure

$$\begin{aligned} \dot{E} &= H(n)E, \\ \dot{n} &= \varepsilon\left(f(n) - g(n)[E,E]\right), \end{aligned} \tag{1.1}$$

where the first equation is a hyperbolic linear system of PDEs for E which is nonlinearly coupled with one scalar ODE for n. This model is particularly well adapted to multisection lasers: It is sufficiently accurate to keep track of the effects caused by the longitudinal resonator structure within the laser. On the other hand, the parameter ε is small. This slow-fast structure permits us to derive analytically low-dimensional systems of ODEs (*mode approximations* [2], [5], [34]) which approximate the semiflow on a local center manifold [26], [32]. These ODEs are accessible for classical numerical bifurcation analysis tools like AUTO. This way we meet both requirements: we present reasonably complete bifurcation diagrams in the physically relevant parameters for a theoretically, and numerically [23], justified approximation of the full system (1.1) of PDEs .

One further remark about the choice of the model: another very popular model for the investigation of delayed optical feedback effects is the Lang–Kobayashi system [13], [17], [19], [30], [28]. It also has the structure of (1.1). Hence it can be treated by the model reduction methods presented in this paper, too. However, we use the traveling-wave model as it fits better to the setting of a multisection laser.

The outline of the paper is as follows: In section 2, we will describe the traveling-wave model in more detail and reiterate the theorem of [26] concerning the model reduction to mode approximations. In section 3, we perform a numerical bifurcation analysis of the *single-mode approximation* (a two-dimensional system of ODEs), which is a valid approximation if the feedback strength is sufficiently small. The primary bifurcation parameters are the feedback strength η and the feedback phase φ. In section 4, we derive an appropriately posed two-mode approximation (a four-dimensional system of ODEs) in the vicinity of a point where the critical eigenvalue of $H(n)$ has algebraic multiplicity two. Furthermore, we perform a numerical bifurcation analysis of this system, thus completing the bifurcation diagram up to higher levels of feedback. However, this system shows complicated dynamics such that the bifurcation analysis remains incomplete. In section 5, we give a brief summary and an outlook to further investigations.

2. The mathematical model.

2.1. The traveling-wave model with gain dispersion.

We consider the geometric configuration presented in Figure 1. Let $\psi(t) \in \mathbb{L}^2([0,L];\mathbb{C}^2)$ describe the spatially resolved complex amplitude of the optical field, which is split into a forward and backward traveling wave. Let $p(t) \in \mathbb{L}^2([0,L];\mathbb{C}^2)$ be the corresponding nonlinear polarization [4], [25]. Denote the one-dimensional spatial variable by $z \in [0,L]$ (the longitudinal direction in the laser). The scalar $n(t) \in \mathbb{R}$ describes the spatially averaged carrier density in the first section S_1. Then the traveling-wave model with gain dispersion [4], [25], [26] poses an initial-boundary-value problem for ψ, p, and n which reads as follows (see the appendix for the physical interpretation

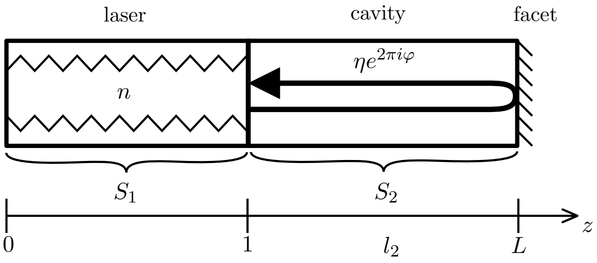


Figure 1. *Geometric configuration for the case of a two-section laser. The DFB section S_1 (DFB=distributed feedback, i.e., $\kappa(z) \neq 0$ in system (2.1)) acts as a laser. Its spatially averaged carrier density n is a scalar dependent variable. The other section acts as a cavity and provides delayed optical feedback of strength η and phase φ.*

and further references):

$$\begin{aligned}\frac{d}{dt}\psi(t,z) &= \sigma\partial_z\psi(t,z) + \beta(n(t),z)\psi(t,z) - i\kappa(z)\sigma_c\psi(t,z) + \rho(z)p(t,z),\\ \frac{d}{dt}p(t,z) &= (i\Omega_r(z) - \Gamma(z))\cdot p(t,z) + \Gamma(z)\psi(t,z),\\ \frac{d}{dt}n(t) &= I - \frac{n(t)}{\tau} - P[(G(n(t)) - \rho_1)(\psi(t),\psi(t))_1 + \rho_1 \operatorname{Re}(\psi(t),p(t))_1],\end{aligned} \tag{2.1}$$

where $\sigma = \left(\begin{smallmatrix}-1 & 0\\ 0 & 1\end{smallmatrix}\right)$, $\sigma_c = \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$, and ψ satisfies the reflection boundary conditions

$$\psi_1(t,0) = r_0\psi_2(t,0), \quad \psi_2(t,L) = r_L\psi_1(t,L), \tag{2.2}$$

where $|r_0|, |r_L| < 1$, and $r_0 r_L \neq 0$. For brevity, we introduced the notation

$$(\psi,\varphi)_1 = \int_0^1 \psi(z)^*\varphi(z)\,dz$$

for ψ, $\varphi \in \mathbb{L}^2([0,L];\mathbb{C}^2)$ in (2.1).

The coefficients $\beta \in \mathbb{C}$, $\kappa \in \mathbb{R}$, $\rho \in \mathbb{R}$, $\Omega_r \in \mathbb{R}$, and $\Gamma \in \mathbb{R}$ ($\rho \geq 0$, $\Gamma > 0$) are sectionwise spatially constant functions. We refer to their value in section S_k by appending the according index, e.g., ρ_1. Moreover, β_1 depends on n in the following way:

$$\beta_1(n) = \beta_1^0 + (1 + i\alpha_1)G(n) - \rho_1, \tag{2.3}$$

where $\beta_1^0 \in \mathbb{C}$, $\operatorname{Re}\beta_1^0 < 0$, and $\alpha_1 > 0$. We assume that $G(n)$ is affine:

$$G(n) = g_1 \cdot (n-1), \qquad \text{where } g_1 > 0. \tag{2.4}$$

It is obvious that system (2.1) has the form (1.1), where $E = (\psi, p)$. The linear differential operator $H(n)$ is defined by

$$H(n)\begin{pmatrix}\psi\\ p\end{pmatrix} = \begin{pmatrix}\sigma\partial_z + \beta(n) - i\kappa\sigma_c & \rho\\ \Gamma & i\Omega_r - \Gamma\end{pmatrix}\begin{pmatrix}\psi\\ p\end{pmatrix} \tag{2.5}$$

and acts from

$$Y := \{(\psi, p) \in \mathbb{H}^1([0,L];\mathbb{C}^2) \times \mathbb{L}^2([0,L];\mathbb{C}^2) : \psi \text{ satisfying (2.2)}\} \tag{2.6}$$

into $X = \mathbb{L}^2([0,L];\mathbb{C}^4)$. The coefficients κ, Γ, $\beta(n)$, Ω_r, and ρ are bounded linear operators in $\mathbb{L}^2([0,L];\mathbb{C}^2)$ defined by the corresponding coefficients in (2.1). The hermitian form $g(n)[E,E]$ is defined by

$$g(n)\left[\begin{pmatrix}\psi_1\\p_1\end{pmatrix}, \begin{pmatrix}\psi_2\\p_2\end{pmatrix}\right] = \int_0^1 (\psi_1^*(z), p_1^*(z)) \begin{pmatrix} G(n)-\rho_1 & \frac{1}{2}\rho_1 \\ \frac{1}{2}\rho_1 & 0 \end{pmatrix} \begin{pmatrix}\psi_2(z)\\p_2(z)\end{pmatrix} dz.$$

Finally, we define the small parameter ε and the function $f(n)$ in (1.1) by $\varepsilon f(n) = I - n/\tau$, observing that I and τ^{-1} are of order $O(10^{-2})$ (see Table 1 and the appendix).

Time t and space z are scaled in system (1.1) such that the speed of light within the device is 1 and the length of section S_1 is 1. Moreover, n is measured in multiples of the *transparency density* (the zero of $G(n)$), and E is scaled such that the factor P in (2.1) is actually ε. In this scaling, we typically have $\Gamma_1 \gg 1$, whereas the real parts of β are of order $O(1)$.

Furthermore, we have $\kappa_2 = \rho_2 = 0$ in the particular situation considered in this paper: the feedback is nondispersive (see Figure 1 and the introduction). Hence the amount η and the phase φ of the feedback from section S_2 can be varied by changing the modulus and the phase of r_L, and we can set $\beta_2 = \Omega_{r,2} = \Gamma_2 = 0$ without loss of generality.

Remark. Several of the assumptions are made only to simplify the presentation. The computations presented in this paper can also be done in a more general framework, e.g., more sections, other coefficients depending on n, or nonaffine but monotone $G(n)$.

2.2. Model reduction. Under the assumptions of section 2.1, the following statements hold for the evolution system (1.1) (see [26] for proofs in a more general context).

Theorem 2.1 (existence of semiflow). *Let $(E^0, n^0) \in V := \mathbb{L}^2([0,L];\mathbb{C}^4) \times \mathbb{R}$. Then system (1.1) generates a strongly continuous semiflow $S(t, (E^0, n^0))$ in V which depends C^∞ smoothly on its initial values and on all parameters for all $t \geq 0$. If $E^0 \in Y$, then $S(t, (E^0, n^0))$ is a classical solution of (1.1); i.e., $(E(t), n(t)) = S(t, (E^0, n^0))$ is continuously differentiable with respect to t, and $E(t) \in Y$ for all $t \geq 0$.*

This theorem is a direct consequence of the theory of strongly continuous semigroups [22] and an a priori estimate exploiting the dissipativity of system (2.1). It guarantees that (1.1) is indeed an infinite-dimensional dynamical system; i.e., its solutions exist for all positive times. The next statement investigates the spectrum of $H(n)$ for fixed n and the strongly continuous group $T(t)$ in X generated by H.

Theorem 2.2 (spectral properties of $H(n)$). *Let*

$$\xi > \xi_- := \max\left\{-\Gamma_1, \frac{1}{L}\operatorname{Re}\left(\beta_1 + \frac{1}{2}\log(r_0 r_L)\right)\right\}.$$

Then X can be decomposed into two $T(t)$-invariant closed subspaces $X = X_+ \oplus X_-$, where X_+ is at most finite-dimensional and spanned by the eigenvectors and generalized eigenvectors associated to the eigenvalues of H in the right half-plane $\{\lambda : \operatorname{Re}\lambda \geq \xi\}$. The restriction of $T(t)$ to X_- is bounded according to

$$\|T(t)|_{X_-}\| \leq Me^{\xi t} \qquad \textit{for } t \geq 0$$

in any norm which is equivalent to the X-norm where the constant M depends on the particular choice of the norm. If $\kappa_1 \neq 0$ or $\rho_1 > 0$, the subspace X_+ is nontrivial for sufficiently small $|r_0 r_L|$.

Moreover, we know that the eigenvalues of the operator $H(n)$ can be computed as roots of its characteristic function $h(\cdot, n)$.

Lemma 2.3 (computation of eigenvalues of $H(n)$). *A complex number $\lambda > \xi_-$ is an eigenvalue of $H(n)$ if and only if*

$$0 = h(\lambda, n) = (\eta e^{2\pi i\varphi - 2\lambda l_2}, -1) T_1(1; \lambda, n) \begin{pmatrix} r_0 \\ 1 \end{pmatrix}.$$

Here, we denoted

$$T_1(z; \lambda, n) = \frac{e^{-\gamma z}}{2\gamma} \begin{pmatrix} \gamma + \mu + e^{2\gamma z}(\gamma - \mu) & i\kappa_1 \left(1 - e^{2\gamma z}\right) \\ -i\kappa_1 \left(1 - e^{2\gamma z}\right) & \gamma - \mu + e^{2\gamma z}(\gamma + \mu) \end{pmatrix}$$

for $z \in [0, 1]$, where $\mu = \lambda - \rho_1 \Gamma_1 (\lambda - i\Omega_{r,1} + \Gamma_1)^{-1} - \beta_1(n)$ and $\gamma = \sqrt{\mu^2 + \kappa_1^2}$.

If $n < 1$, all eigenvalues λ of $H(n)$ are in the left half-plane $\{\operatorname{Re} \lambda < 0\}$. For increasing n, finitely many of them will cross the imaginary axis if $|r_0 r_l|$ is small, and $\kappa_1 \neq 0$ according to Theorem 2.2. Denote the smallest n where $q \geq 1$ eigenvalues λ of $H(n)$ are on the imaginary axis by n_0. Typically, the value n_0 is referred to as *threshold* carrier density. Choose $\xi < 0$ such that all other non purely-imaginary eigenvalues of $H(n_0)$ lie to the left of the line $\{\operatorname{Re} \lambda = \xi\}$. We denote the space X_+ of complex dimension q according to Theorem 2.2 by $X_c(n)$, and we define the spectral projection $P_c(n)$ for $H(n)$ onto $X_c(n)$. $P_c(n)$ depends smoothly on n in a neighborhood of n_0. Let $B(n)$ be a smooth basis of $X_c(n)$.

According to [26], the following local center manifold theorem holds in the vicinity of n_0.

Theorem 2.4 (model reduction). *Let $k > 2$ be an integer number, and let $C > 0$. Let $\varepsilon_0 > 0$ be sufficiently small and U be a sufficiently small neighborhood of n_0 (depending on C and k). Define the balls*

$$\begin{aligned} \mathcal{B} = &\ \{(E_c, n) \in \mathbb{C}^q \times \mathbb{R} : \|E_c\| < C, n \in U\} \subset \mathbb{C}^q \times \mathbb{R} \text{ and} \\ \mathcal{N} = &\ \{(E, n) \in X \times \mathbb{R} : \|E\| < C, n \in U\} \subset X \times \mathbb{R}. \end{aligned}$$

Then there exists a manifold $\mathcal{C}$ with the following properties:

(i) Representation. *$\mathcal{C}$ can be represented as the graph of a map from $\mathcal{B}$ into $\mathcal{N}$ which maps $(E_c, n) \in \mathcal{B}$ to $(B(n)E_c + \varepsilon\nu(E_c, n, \varepsilon), n)$, where $\nu : \mathcal{B} \times (0, \varepsilon_0) \to X$ is C^k with respect to all arguments. Denote the E-component of $\mathcal{C}$ by $E_X(E_c, n, \varepsilon) = B(n)E_c + \varepsilon\nu(E_c, n, \varepsilon) \in X$.*

(ii) Invariance. *$\mathcal{C}$ is $S(t, \cdot)$-invariant relative to $\mathcal{N}$ if $\varepsilon < \varepsilon_0$.*

(iii) Exponential attraction. *Let (E, n) be such that $S(t, (E, n)) \in \mathcal{N}$ for all $t \geq 0$. Then there exists an $(E_c, n_c) \in \mathcal{B}$ such that, for some $M > 0$,*

$$\|S(t, (E, n)) - S(t, (E_X(E_c, n_c, \varepsilon), n_c))\| \leq M e^{\xi t} \qquad \text{for all } t \geq 0, \tag{2.7}$$

and $\xi < 0$ is as defined above.

(iv) Flow on the manifold. *The values $\nu(E_c, n, \varepsilon)$ are in Y, and $P_c \nu = 0$ for all $(E_c, n, \varepsilon) \in \mathcal{B} \times (0, \varepsilon_0)$. The flow on $\mathcal{C}$ is differentiable with respect to t and governed by the system of ODEs*

$$\begin{aligned} \frac{d}{dt} E_c =&\ H_c(n)E_c + \varepsilon a_1(E_c, n, \varepsilon)E_c + \varepsilon^2 a_2(E_c, n, \varepsilon)\nu, \\ \frac{d}{dt} n \ =&\ \varepsilon F(E_c, n, \varepsilon), \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} H_c(n) =&\ B(n)^{-1} H(n) P_c(n) B(n), \\ a_1(E_c, n, \varepsilon) =&\ -B(n)^{-1} P_c(n) \partial_n B(n) F(E_c, n, \varepsilon), \\ a_2(E_c, n, \varepsilon) =&\ B^{-1}(n) P_c(n) \partial_n P_c(n) F(E_c, n, \varepsilon), \\ F(E_c, n, \varepsilon) =&\ f(n) - g(n) \left[E_X(E_c, n, \varepsilon), E_X(E_c, n, \varepsilon) \right]. \end{aligned}$$

System (2.8) is symmetric with respect to rotation $E_c \to E_c e^{i\varphi}$, and ν satisfies the relation $\nu(e^{i\varphi} E_c, n, \varepsilon) = e^{i\varphi} \nu(E_c, n, \varepsilon)$ for all $\varphi \in [0, 2\pi)$.

This theorem is based on the general results in [6], [7], [8], [33]. We observe that the term $\nu(E_c, n, \varepsilon)$ enters E_X with a factor ε in front. Hence, ν enters system (2.8) with a factor of order $O(\varepsilon^2)$. Consequently, the replacement of ν by 0 is a regular perturbation of (2.8) preserving the rotational symmetry of (2.8). The approximate system is called *mode approximation* and reads

$$\begin{aligned} \frac{d}{dt} E_c =&\ H_c(n)E_c + \varepsilon a(E_c, n)E_c, \\ \frac{d}{dt} n \ =&\ \varepsilon F_0(E_c, n), \end{aligned} \tag{2.9}$$

where

$$\begin{aligned} H_c(n) =&\ B(n)^{-1} H(n) P_c(n) B(n), \\ a(E_c, n) =&\ -B(n)^{-1} P_c(n) \partial_n B(n) F_0(E_c, n), \\ F_0(E_c, n,) =&\ f(n) - g(n) \left[B(n)E_c, B(n)E_c \right]. \end{aligned}$$

The matrix $H_c(n)$ is a representation of $H(n)$ restricted to its critical subspace $X_c(n)$ in some basis $B(n)$. The matrix H_c depends on the particular choice of $B(n)$, but its spectrum coincides with the critical spectrum of $H(n)$. The term $\varepsilon a E_c$ appears since the space X_c depends on time t. Any normally hyperbolic invariant manifold (e.g., fixed point, periodic orbit, invariant torus) that is present in the dynamics of (2.9) persists under the perturbation $O(\varepsilon^2)\nu$. Hence it is also present in system (2.8), describing the flow on the invariant manifold $\mathcal{C}$, and in the semiflow of the complete system (1.1). Furthermore, its hyperbolicity and the exponential attraction toward $\mathcal{C}$ ensure its continuous dependence on small parameter perturbations.

Remark. System (2.9) still depends on ε. Hence we should investigate how the transversal Lyapunov exponents depend on ε for any normally hyperbolic invariant manifold found in the subsequent analysis of (2.9). It was pointed out in [32] that system (2.9) is conservative for

Table 1

Choice of parameters for the bifurcation diagrams presented in sections 3 and 4.

$l_1 = 1$	$l_2 = 1.136$	$r_0 = 10^{-5}$	$r_L = \eta e^{2\pi i\varphi}$
$\beta_1^0 = -0.275$	$\beta_2^0 = 0$	$\kappa_1 = 3.96$	$\kappa_2 = 0$
$g_1 = 2.145$	$g_2 = \rho_2 = 0$	$\alpha_1 = 5$	$\rho_1 = 0.44$
$\Gamma_1 = 90$	$\Omega_{r,1} = -20$	$I = 6.757 \cdot 10^{-3}$	$\tau = 3.59 \cdot 10^2$

$q = 1$ at $\varepsilon = 0$. Thus we must expect that the normal hyperbolicity of most objects is of order $o(1)$ for $\varepsilon \to 0$.

However, we will perform the bifurcation analysis for (2.9) only for one fixed "physically realistic" ε, checking the eigenvalues of relative equilibria and the Floquet multipliers of modulated waves numerically.

2.3. Particular choice of parameters.

The mode approximations (2.9) derived in section 2.2 permit detailed studies of their long-time behavior since they are low-dimensional ODEs. The particular form of system (2.9) depends on the number q of critical eigenvalues on the imaginary axis at the threshold n_0. We restrict our interest to cases where the number q of critical eigenvalues of H is at most 2.

Furthermore, we adjust the relative resonance frequency of the material, $\Omega_{r,1}$, in the following manner: The solitary section S_1 with zero facet reflectivities, gain dispersion, and feedback, i.e., $\rho_1 = r_0 = r_L = 0$, is symmetric with respect to reflection. Thus, if $H(n)$ has the eigenvalue $\lambda + i\,\mathrm{Im}\,\beta_1(n)$, it also has the eigenvalue $\bar{\lambda} + i\,\mathrm{Im}\,\beta_1(n)$. Typically, a pair of eigenvalues becomes critical having the frequencies $\mathrm{Im}\,\lambda_{1,2} \approx \mathrm{Im}\,\beta_1(n_0) \pm \kappa_1$. The frequency region $(\mathrm{Im}\,\beta_1 - \kappa_1, \mathrm{Im}\,\beta_1 + \kappa_1)$ is usually referred to as the *stopband* of the active section. Hence the solitary section S_1 can have *on-states* (i.e., rotating-wave solutions, relative equilibria of (1.1)) at both ends of the stopband. We break the reflection symmetry by choosing $\rho_1 > 0$ and $\Omega_{r,1}$ outside of the stopband frequency region. Then the slope of the gain curve (see the appendix) favors frequencies closer to $\Omega_{r,1}$ such that the solitary active section S_1 has a distinct stable on-state at the threshold n_0.

We choose the feedback phase and strength as our primary bifurcation parameters and use numerical continuation methods [11] to explore the bifurcation diagram in the two-parameter plane keeping all other parameters fixed according to Table 1. We can do so by varying the absolute value η and the phase φ of $r_L := \eta e^{2\pi i\varphi}$, setting $\beta_2 = 0$ without loss of generality. This is in contrast to the experiments, where β_2 is varied by changing the current in the feedback section, but r_L is kept fixed. Hence the experimenters cannot vary the parameters φ and η independently in a continuous manner.

In order to obtain the coefficients of (2.9), we have to compute the critical eigenvalues, their corresponding eigenvectors, and the adjoint eigenvectors.

The critical eigenvalues are roots of the characteristic function h of $H(n)$ defined in Lemma 2.3. We include n, η, and φ as parameters to emphasize that the eigenvalue λ_j depends on them (see section 2.2):

$$0 = h(\lambda_j, n, \eta, \varphi) = (\eta e^{2\pi i\varphi - 2\lambda_j l_2}, -1) T_1(1; \lambda_j, n) \begin{pmatrix} r_0 \\ 1 \end{pmatrix}.$$

The eigenvector $v_j = (\psi_j, p_j)$ corresponding to λ_j and its adjoint $v_j^\dagger = (\psi_j^\dagger, p_j^\dagger)$ are defined up to a scaling by (see [3], [34] for the adjoint)

$$\begin{pmatrix}\psi_j\\ p_j\end{pmatrix} = \begin{pmatrix} T(z,0;\lambda_j,n)\left(\begin{smallmatrix} r_0\\ 1\end{smallmatrix}\right), \\ \frac{\Gamma}{\lambda_j - i\Omega_r + \Gamma} T(z,0;\lambda_j,n)\left(\begin{smallmatrix} r_0\\ 1\end{smallmatrix}\right)\end{pmatrix}, \qquad \begin{pmatrix}\psi_j^\dagger\\ p_j^\dagger\end{pmatrix} = \begin{pmatrix} \begin{pmatrix}\bar\psi_{j,2}\\ \bar\psi_{j,1}\end{pmatrix} \\ \frac{\rho}{\Gamma}\begin{pmatrix}\bar p_{j,2}\\ \bar p_{j,1}\end{pmatrix}\end{pmatrix}. \tag{2.10}$$

In (2.10), $T(z,0;\lambda_j,n) = T_1(z;\lambda_j,n)$ if $z \le 1$, and

$$T(z,0;\lambda_j,n) = \begin{pmatrix} e^{-\lambda_j z} & 0\\ 0 & e^{\lambda_j z}\end{pmatrix} T_1(1;\lambda_j,n) \quad \text{if } z \ge 1.$$

3. The single-mode case.

3.1. Definition of the system.

First, we consider the generic case, where a single eigenvalue λ of $H(n)$ is on the imaginary axis ($q = 1$) at $n = n_0$. Since we have chosen the configuration of the active section S_1 such that it has only one stable on-state at $\eta = 0$, this model is certainly valid for sufficiently small η in the vicinity of n_0. Since λ is uniformly isolated, it depends smoothly on n and all parameters. The basis $B(n)$ is the eigenvector $v = (\psi, p)$ associated to λ, and the projection $B(n)^{-1}P_c(n)$ is the corresponding adjoint eigenvector $v^\dagger$. The term a in (2.9) vanishes if we scale v and $v^\dagger$ such that $\left(v, v^\dagger\right) = 1$ for all n under consideration. Moreover, we can decouple the phase of the complex quantity E_c in (2.9) due to the rotational symmetry of the system. Hence we have to analyze a system for $S = |E_c|^2$, n, and λ which reads as follows:

$$\dot S = 2\operatorname{Re}(\lambda) S, \tag{3.1}$$

$$\dot n = \varepsilon\left(I_1 - n - R(\lambda,n,\varphi,\eta)S\right), \tag{3.2}$$

$$0 = h(\lambda,n,\varphi,\eta), \tag{3.3}$$

where $\varepsilon = \tau^{-1}$, $I_1 = I\tau$, and the coefficient R is defined as follows:

$$R(\lambda,n,\varphi,\eta) = [G(n) - \rho_1 + \operatorname{Re}\chi_1(\lambda)](\psi,\psi)_1^2$$

(see the appendix for the definition of χ).

On-states are equilibria of (3.1)–(3.3), whereas periodic solutions of (3.1)–(3.3) represent quasi-periodic solutions of (2.9). This type of modulated rotating-wave solution is typically referred to as *self-pulsation*. Several analytic and computational results have been obtained previously about the existence regions of self-pulsations and their synchronization properties using the single-mode approximation [2], [3], [5], [31], [34].

System (3.1)–(3.3) is a differential-algebraic equation (DAE). Its inherent dynamical system is two-dimensional. Standard numerical continuation software such as, e.g., AUTO [11] is not able to treat DAEs directly. However, we can easily convert (3.1)–(3.3) to an equivalent explicit system of ODEs by changing (3.3) to

$$\dot\lambda = -\frac{\partial_n h(\lambda,n,\varphi,\eta)\dot n + c\cdot h(\lambda,n,\varphi,\eta)}{\partial_\lambda h(\lambda,n,\varphi,\eta)} \tag{3.4}$$

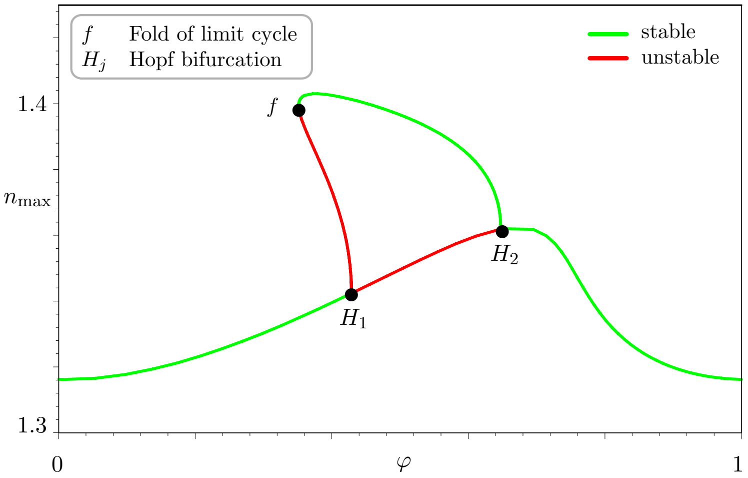


Figure 2. *Bifurcation diagram for $\eta = 0.1$. We report the n-component for the on-state and the maximum of the n-component for the periodic solutions.*

if λ is an isolated simple root of h. For sufficiently large $c > 0$, (3.1), (3.2), and (3.4) is a four-dimensional ODE which has a stable invariant manifold defined by $h = 0$. On this invariant manifold, the flow is identical to the flow of (3.1)–(3.3). The transformation of (3.3) to (3.4) is sometimes referred to as Baumgarte regularization [10].

Remark. There should exist a curve md in the parameter plane (φ, η) which bounds the range of validity of the single-mode model in the following sense: the function $\lambda(n)$ defined implicitly by (3.3) has a discontinuous derivative $\partial_n \lambda(n)$ along some line $n = n_{md}$ in the phase plane for $(\varphi, \eta) \in md$, i.e., $\partial_\lambda h(\lambda(n), n, \varphi, \eta) = 0$ (λ is a double root of h). Hence we expect the continuation of families of periodic orbits or equilibria to fail if it crosses the curve md.

3.2. The bifurcation diagram.

We explore the dynamics numerically in the two-parameter plane (φ, η), choosing the other parameters according to the example presented in [35] (see Table 1). The procedure is as follows: First, we choose a very small feedback level $\eta = 0.1$ and report the smallest n_0 such that $H(n_0)$ possesses an eigenvalue on the imaginary axis for $\varphi = 0$, i.e., $h(\cdot, n_0, 0, \eta)$ has a purely imaginary root. This is $n_0 = 1.316194$ for the parameter situation outlined in Table 1. The corresponding root of h is $\lambda = -1.632607i$. We consider only this eigenvalue λ of H and its eigenvector ψ in (3.3) in this section. There exists an equilibrium with $S = (I_1 - n_0)/R(\lambda, n_0, 0, \eta)$ and $n = n_0$. In the next step, we report on how this equilibrium changes its location in phase space and its stability under variation of φ (see Figure 2). Note that the location of the equilibrium coincides exactly with the location of the corresponding rotating-wave solution of the complete system (1.1).

Adding η as a free parameter, we compute the curve of Hopf points and the curve of saddle-nodes (folds) of limit cycles in the two-parameter plane (φ, η) (see the full bifurcation diagram Figure 4). The saddle-node curve of limit cycles and the Hopf curve meet at a stable generalized Hopf point GH_1 [14]. Both ends of the Hopf curve meet at the point $MD = (\varphi_0, \eta_0)$. The point MD is situated on the curve md. Moreover, the equilibrium of system (3.1)–(3.3) is situated on the line $n = n_{md}$ for the parameters (φ_0, η_0). We refer to it

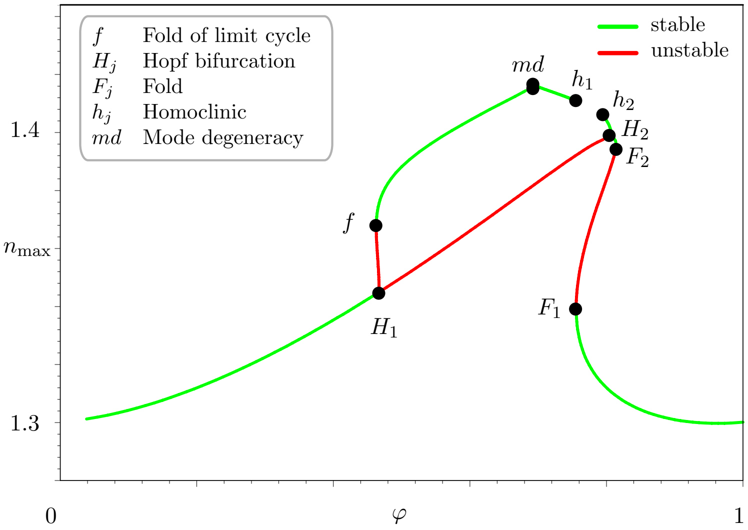


Figure 3. *Bifurcation diagram at $\eta = 0.2$. We report the n-component for the on-state and the maximum of the n-component for the periodic solutions. The family of periodic orbits is discontinuous at md.*

as an equilibrium with *mode degeneracy* as λ is an eigenvalue of $H(n)$ of algebraic multiplicity 2. We explore the vicinity of this equilibrium using the two-mode approximation in section 4 since the number q of critical eigenvalues of $H(n)$ is 2 near MD.

Starting from one of the Hopf points at $\eta = 0.2$, we draw another one-dimensional bifurcation diagram, keeping $\eta = 0.2$ fixed but varying φ. The numerical result is shown in Figure 3.

The line of equilibria has folded twice at F_1 and F_2 in saddle-node bifurcations. The family of periodic orbits starting from H_2 is stable and ends in a homoclinic bifurcation at h_2. The family of periodic orbits starting from H_1 is unstable and collides with a stable branch in the saddle-node bifurcation of limit cycles f. Continuing the stable branch of periodic solutions starting from f, we approach a discontinuity of system (3.1)–(3.3) at md. Hence there is no *continuous* family of periodic orbits between h_1 and f.

In the last step, we continue the saddle-node points F_1 and F_3 and the homoclinic orbit[1] h_2, varying both parameters, η and φ. Figure 4 shows the complete two-parameter bifurcation diagram, and Figure 5 shows the corresponding symbolic phase portrait sketches. We note that the two saddle-node bifurcations observed in Figure 3 emanate from a cusp bifurcation at CU in Figure 4. Moreover, the curve of homoclinic bifurcations in the (φ, η)-plane starting from h_2 at $\eta = 0.2$ turns back to $\eta = 0.2$ at h_1 (see Figure 3). Along this curve, the approach of the homoclinic orbit toward the saddle changes (see Figure 5): There is a parameter path of central saddle-nodes on closed orbits between two noncentral saddle-nodes on closed orbits (NC_1–NC_2). The region 6 in the vicinity of the central saddle-node on a closed orbit is a classical excitability scenario [15].

Remark. The approximation of homoclinic orbits by periodic orbits of large periods is

[1] Actually, we continued periodic orbits of a fixed large period.

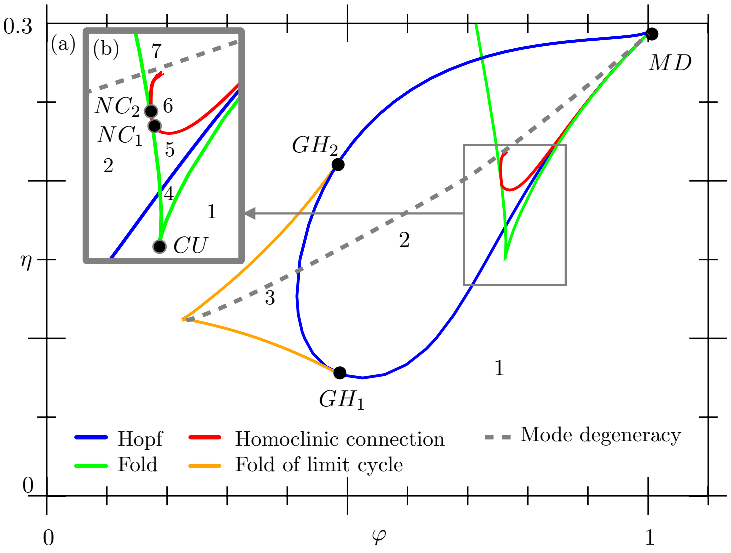


Figure 4. *Bifurcation diagram in the two-parameter plane* (φ, η)*. Codimension 2 bifurcations are marked by black points and labels. GH refers to a generalized Hopf bifurcation, CU to a cusp bifurcation, and NC to a noncentral saddle-node on a closed orbit. MD is an equilibrium where* $\partial_\lambda h = 0$ *in* (3.4)*. The family of periodic orbits is discontinuous along the dashed line because* $\partial_\lambda h(\lambda, n)$ *becomes zero on the periodic orbits.*

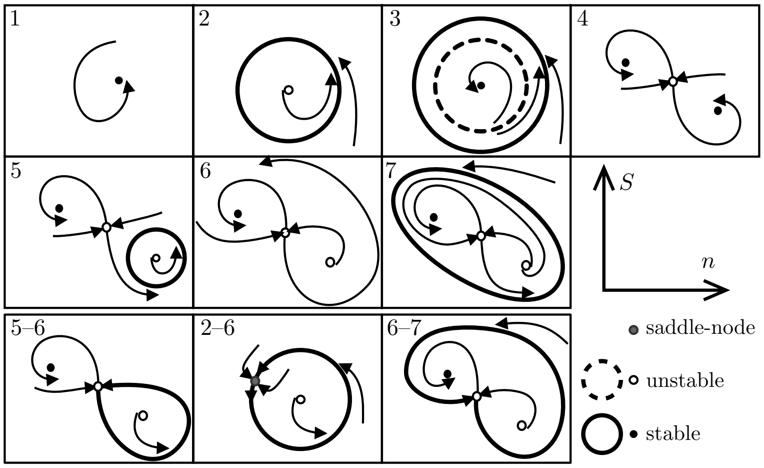


Figure 5. *Symbolic phase-portrait sketches for the regions* 1 *to* 7 *of Figure* 4*. In addition, we display the global bifurcations between the regions* 2, 5, 6, *and* 7*. Black points symbolize stable equilibria, and white points are unstable equilibria. If an unstable equilibrium is a saddle, we sketch the stable and unstable manifolds. The additional arrows show the behavior of the vector field. The* $S - n$ *coordinate cross gives a rough orientation.*

very accurate if the equilibrium is a saddle with eigenvalues distant from the imaginary axis. However, the order of approximation is worse in the vicinity of homoclinics to saddle-nodes. For this reason, we have used the HOMCONT part of AUTO to compute the points NC_1 and NC_2. HOMCONT utilizes projection boundary conditions to approximate the homoclinic orbit [9], [11].

The continuation of the curve of homoclinic bifurcations in the plane (φ, η) ends at some distance after (above) the point NC_2 in Figure 4, where the homoclinic orbit approaches a point in the phase plane where λ is degenerate.

The curve of homoclinic bifurcations, the Hopf curve, and the saddle-node curve in Figure 4 approach each other and become tangent at MD. However, this does not imply the presence of a Takens–Bogdanov bifurcation since the model is discontinuous in the vicinity of MD. Indeed, the Hopf frequency along the Hopf curve increases toward MD.

Remark. Figure 5 assumes completeness of the bifurcation diagram Figure 4. The problem of completeness of Figure 4 is not investigated here. There may be additional nested pairs of stable and unstable limit cycles. See, e.g., [21], [32] for a treatment of the single-mode laser as a perturbation of a conservative oscillator of order $O(\sqrt{\varepsilon})$. In [25], a result about the uniqueness of the stable limit cycle is obtained by imposing conditions on the shape of the coefficients λ and R in system (3.1), (3.2).

4. The two-mode case—unfolding of the mode degeneracy in the parameters φ and η.

4.1. Definition of $B(n)$ and $P_c(n)$—elimination of the absolute phase.

In this section, we explore in detail the neighborhood of MD from Figure 4, completing the bifurcation diagram for larger feedback levels η. Let n_0 be a threshold carrier density, where $H(n_0)$ has a dominating eigenvalue λ of algebraic multiplicity 2 on the imaginary axis. An n_0 of this type exists, e.g, in the point $MD = (\varphi_0, \eta_0)$ of the parameter plane (φ, η) (see Figure 4). Hence the complex dimension q of the critical subspace X_c is 2, and the real dimension of the invariant manifold $\mathcal{C}$ is 5 (see section 2.2). In order to obtain the coefficients of system (2.9), we have to construct a basis $B(n)$ and a projector $P_c(n)$ which depend smoothly on n, and the bifurcation parameters η and φ in the vicinity of n_0, φ_0, and η_0.

Let λ_1 and λ_2 be the two roots of $h(\cdot, n, \varphi, \eta)$ which coincide and are situated on the imaginary axis if $n = n_0$, $\varphi = \varphi_0$, and $\eta = \eta_0$. We denote the corresponding eigenvectors (ψ_j, p_j) by v_j and the adjoint eigenvectors by $v_j^\dagger$ $(j = 1, 2)$. The vectors v_j and $v_j^\dagger$ are defined in (2.10). Hence v_j and $v_j^\dagger$ depend analytically on λ_j.

We introduce the quantities

$$\theta := \frac{1}{2}(\lambda_1 + \lambda_2), \qquad \mu := \frac{1}{4}(\lambda_1 - \lambda_2)^2$$

and define the basis $B = [u_1, u_2]$ by

$$u_1 = \frac{v_1 - v_2}{\lambda_1 - \lambda_2}, \qquad u_2 = \frac{1}{2}(v_1 + v_2). \tag{4.1}$$

θ and μ depend smoothly on n, φ, and η in the vicinity of the degeneracy MD. The vectors u_1 and u_2 do not change if we permute the eigenvalues λ_1 and λ_2. Moreover, they are smooth with respect to n, φ, and η and uniformly linearly independent around the degeneracy point. We denote the ψ-component of u_j by ξ_j $(j = 1, 2)$. The representation of H in the basis B reads

$$Hu_1 = \theta u_1 + u_2, \qquad Hu_2 = \mu u_1 + \theta u_2. \tag{4.2}$$

Hence the representation of H with respect to the basis B is smooth in the degeneracy point.

Furthermore, we define the following functionals for $x \in \mathbb{L}^2([0,L];\mathbb{C}^4)$ and $y \in \mathbb{L}^2([0,L];\mathbb{C}^2)$:

$$\begin{aligned}
P_1 x &= \frac{1}{2}\left[\frac{\lambda_1-\lambda_2}{(v_1^\dagger, v_1)}(v_1^\dagger, x) + \frac{\lambda_2-\lambda_1}{(v_2^\dagger, v_2)}(v_2^\dagger, x)\right], & P_2 x &= \frac{(v_1^\dagger, x)}{(v_1^\dagger, v_1)} + \frac{(v_2^\dagger, x)}{(v_2^\dagger, v_2)},\\
\Theta_1 y &= (G(n) - \rho_1 + \chi_1(\lambda_1))(\psi_1, y)_1, & \Theta_2 y &= (G(n) - \rho_1 + \chi_1(\lambda_2))(\psi_2, y)_1,\\
Q_1 y &= \frac{\Theta_1 y - \Theta_2 y}{\bar\lambda_1 - \bar\lambda_2}, & Q_2 y &= \frac{1}{2}\left(\Theta_1 y + \Theta_2 y\right).
\end{aligned}$$

(See the appendix for the definition of χ_1.) P_1 and P_2, as well as Q_1 and Q_2, are not affected by a permutation of λ_1 and λ_2. Since $P_j u_k = \delta_{jk}$, the functionals P_j depend smoothly on n and are uniformly linearly independent around the degeneracy. We define

$$P_c x = [P_1 x]u_1 + [P_2 x]u_2.$$

Using these definitions of B and P_c, system (2.9) reads as follows ($E_c = (x_1, x_2) \in \mathbb{C}^2$):

$$\begin{aligned}
\dot x_1 &= \theta(n)x_1 + \mu(n)x_2 - \dot n(P_1(n)\partial_n u_1(n)x_1 + P_1(n)\partial_n u_2(n)x_2),\\
\dot x_2 &= x_1 + \theta(n)x_2 - \dot n(P_2(n)\partial_n u_1(n)x_1 + P_2(n)\partial_n u_2(n)x_2),\\
\dot n &= \varepsilon\left[I_1 - n - \operatorname{Re}\left(|x_1|^2 Q_1\xi_1 + |x_2|^2 Q_2\xi_2 + \bar x_1 x_2 Q_1 \xi_2 + \bar x_2 x_1 Q_2 \xi_1\right)\right].
\end{aligned}\tag{4.3}$$

Finally, we observe that we can eliminate the absolute phase of the vector (x_1, x_2) due to the rotational symmetry of system (4.3). We introduce the quantities

$$r = |x_1|^2 - |x_2|^2 \qquad (r \in \mathbb{R}), \qquad\qquad \zeta = \bar x_1 x_2 \qquad (\zeta \in \mathbb{C}).$$

Using r and ζ, we can recover the quantities

$$|x_1|^2 = \frac{1}{2}(\sqrt{r^2 + 4|\zeta|^2} + r), \qquad\qquad |x_2|^2 = \frac{1}{2}(\sqrt{r^2 + 4|\zeta|^2} - r).$$

Hence system (4.3) has a four-dimensional subsystem for r, ζ, and n which reads as follows:

$$\begin{aligned}
\dot r &= 2\operatorname{Re}\left(b_{11}|x_1|^2 - b_{22}|x_2|^2 + (b_{12} - \bar b_{21})\zeta\right),\\
\dot \zeta &= b_{21}|x_1|^2 + \bar b_{12}|x_2|^2 + (\bar b_{11} + b_{22})\zeta,\\
\dot n &= \varepsilon\left[I_1 - n - \operatorname{Re}\left(|x_1|^2 Q_1\xi_1 + |x_2|^2 Q_2\xi_2 + \zeta Q_1\xi_2 + \bar\zeta Q_2 \xi_1\right)\right],
\end{aligned}\tag{4.4}$$

where the functions $b_{ij}(r,\zeta,n)$ are defined as

$$\begin{aligned}
b_{11}(r,\zeta,n) &= \theta(n) - \dot n P_1(n)\partial_n u_1(n), & b_{12}(r,\zeta,n) &= \mu(n) - \dot n P_1(n)\partial_n u_2(n),\\
b_{21}(r,\zeta,n) &= 1 - \dot n P_2(n)\partial_n u_1(n), & b_{22}(r,\zeta,n) &= \theta(n) - \dot n P_2(n)\partial_n u_2(n).
\end{aligned}$$

The quantities $P_i\partial_n u_j$ depend on n, μ, and θ and can be computed by the chain rule:

$$P_i \partial_n u_j = P_i(\partial_n\theta\partial_\theta + \partial_n\mu\partial_\mu + \partial_n)u_j.$$

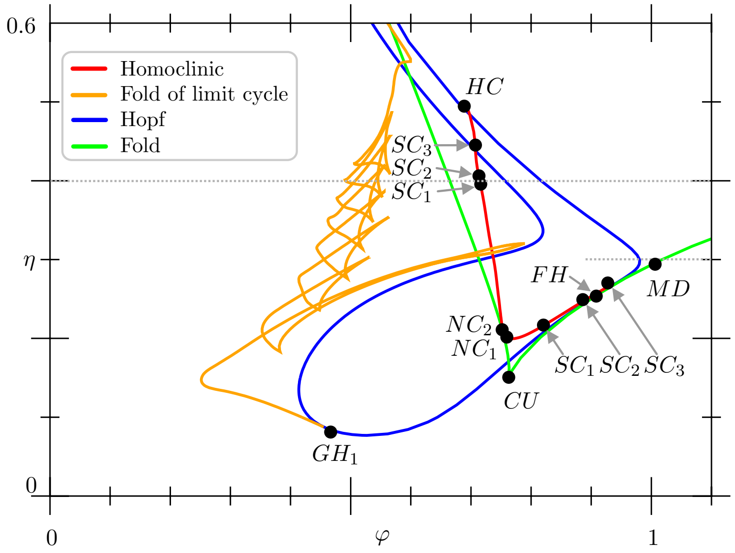


Figure 6. *Bifurcation curves of Figure 4 recomputed in the (φ, η)-plane using model (4.4), (4.5). The one-dimensional bifurcation diagrams of Figures 8 and 9 have been obtained along the horizontal dotted lines. This diagram should be compared to Figure 4. Along the curve of homoclinic bifurcations, the points SC_j mark changes of saddle quantities. FH marks a fold-Hopf interaction (see text and Figure 7).*

The functions $\theta(n)$ and $\mu(n)$ are given only implicitly by the root curves of $h(\cdot, n)$. We introduce θ and μ as new variables satisfying the differential equations

$$(4.5) \qquad \dot{\theta} = \frac{1}{2}(\dot{\lambda}_1 + \dot{\lambda}_2), \qquad \dot{\mu} = \frac{1}{2}(\dot{\lambda}_1 - \dot{\lambda}_2)(\lambda_1 - \lambda_2)$$

in order to put the system in a form that is recognized by AUTO. In (4.5), the equations for $\dot{\lambda}_j$ are constructed in the same manner as in (3.4). Assembling (4.4) and (4.5), we obtain a system of dimension 8, where all coefficients depend smoothly on n, φ, and η. This system has a four-dimensional uniformly attracting invariant manifold, where $\theta = \theta(n)$ and $\mu = \mu(n)$. We restrict our bifurcation analysis to this invariant manifold.

4.2. The bifurcation diagram 2. We explore the (φ, η)-plane in the same manner as in section 3. First, we redraw the curves of the single-mode bifurcation diagram Figure 4 in the plane (φ, η) in Figure 6. We observe that the curves are in good quantitative agreement for smaller η. The curve of the folds even coincides exactly. Notable differences are as follows.

1. The curve of the folds of limit cycles does not connect to the Hopf curve in a generalized Hopf point GH_2 in Figure 6, but it turns toward larger η undergoing a sequence of cusps.

2. In Figure 6, the Hopf curve and the fold curve meet each other in FH in a fold-Hopf interaction of type "$s = 1, \theta < 0$" according to [16]. A small neighborhood of FH in parameter space is depicted in Figure 7. The time must be reversed compared to [16] since the torus bifurcation in Figure 7 is supercritical. We observe that the equilibrium with mode degeneracy is no longer a special point except that it is situated on the fold curve.

The curve of homoclinic bifurcations depicted in the single-mode bifurcation diagram Figure 4 can be continued in both directions for increasing η. However, the characteristic

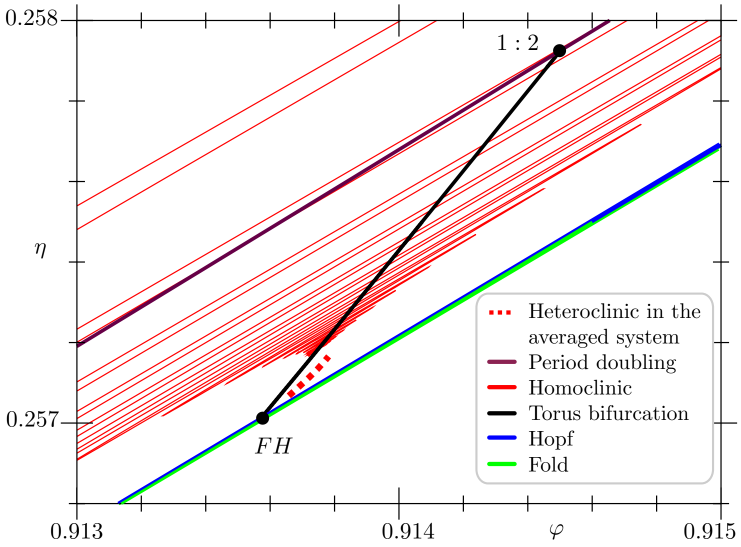


Figure 7. *Local (incomplete) bifurcation diagram in the vicinity of the fold-Hopf interaction FH in the (φ, η)-plane. There is a $1:2$ resonance at the point $1:2$. The dotted line has not been actually computed. It is inserted in the picture to simplify the identification of the appropriate fold-Hopf interaction type in, e.g.,* [16].

quantities of the linearized system at the saddle change along the line (see Figure 6): At SC_1, we have three leading stable eigenvalues. For higher η, a complex pair of eigenvalues becomes dominant in the negative half-plane, turning the homoclinic orbit into a saddle-focus homoclinic connection. Denote the real part of this complex pair of eigenvalues by σ_s and the real part of the (real) unstable eigenvalue by σ_u. Between SC_1 and SC_2, we have $\nu := -\sigma_u/\sigma_s < 1$.

At SC_2, the saddle becomes neutral, i.e., $\nu = 1$, and, at SC_3, we have $\nu = 2$. Shilnikov's theorems [16] imply that there is one stable limit cycle in the vicinity of the homoclinic bifurcation before SC_2, i.e., along NC_1–SC_2 and NC_2–SC_2, but there are infinitely many limit cycles in the vicinity of the homoclinic bifurcation after SC_2. Furthermore, infinitely many of these limit cycles are stable between SC_2 and SC_3. We observe that one end of the curve of homoclinic bifurcations approaches the fold-Hopf interaction FH in an oscillatory manner (see the curve of homoclinic bifurcations in Figure 7).

We must reach the upper end HC of the curve of homoclinic bifurcations (see Figure 6) before the saddle having the homoclinic connection undergoes a subcritical Hopf bifurcation. At the point HC, there is a heteroclinic connection between the saddle and the small Hopf cycle which is also of saddle type (two unstable and one stable Floquet multipliers).

Figure 6 gives a complete overview concerning the equilibria of system (4.4), (4.5), i.e., their number, the number of their stable and unstable directions, and their local bifurcations. In order to get more information about the periodic orbits and global bifurcations, we draw one-dimensional bifurcation diagrams varying φ at $\eta = 0.3$ (see Figure 8) and $\eta = 0.4$ (see Figure 9 (a)) by continuing the periodic orbits from the Hopf bifurcations.

In Figure 8, we observe that the family of periodic orbits emerging at the saddle Hopf point undergoes a sequence of folds and period doublings approaching a homoclinic connection to a

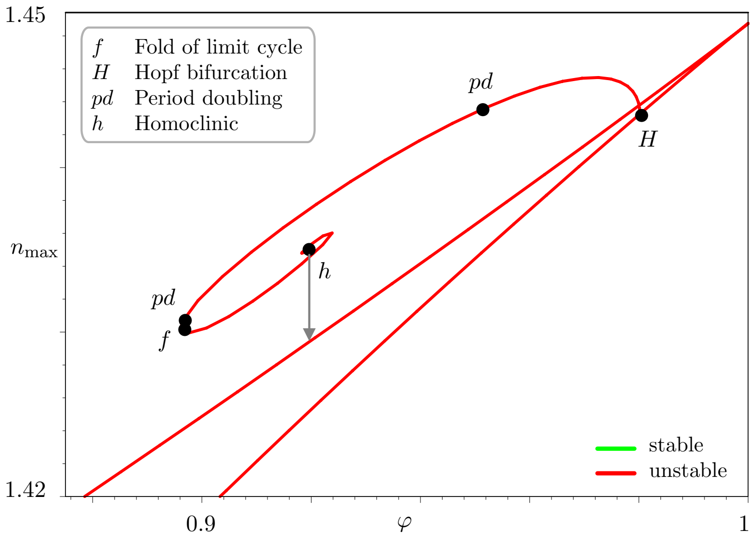


Figure 8. *Bifurcation diagram for $\eta = 0.3$ for a part of the full phase period for φ. We report the n-component for the on-state and the maximum of the n-component for the periodic solutions. The arrow points to the saddle approached by the homoclinic h. This saddle is of focus-focus type. Only the first period doublings and folds are reported.*

focus-focus. Parts of the family are stable according to Shilnikov's theorems [16] because the linearization at the focus-focus has a negative saddle value; i.e., the real part of the unstable eigenvalues is smaller than the modulus of the real part of the stable eigenvalues.

In Figure 9 (a), the two Hopf points are connected by a "small bridge" of periodic solutions undergoing a torus bifurcation at t. We observe that both Hopf points are close to the point X, where the saddle equilibrium and the former node equilibrium have equal n-component. At X, there are two eigenvalues of $H(n)$ with different imaginary parts on the imaginary axis in the equilibria. The periodic orbits between H_1 and H_2 have an angular velocity corresponding approximately to the difference between these two eigenvalues. Typically, this type of self-pulsation is called mode beating or mixed-mode self-pulsation; see [13] and references therein. These oscillations are of great interest because their frequency can be of order $O(1)$ in our scaling. In our case, the angular velocity of the oscillations increases from 0.2 for $\eta = 0.2$ to 0.4 for $\eta = 0.6$ along the channel between the two Hopf curves in Figure 6 which corresponds to frequencies between 11 and 22 GHz.

The family of periodic orbits emerging from the homoclinic bifurcation (the red curve of Figure 6) is now disconnected from the family of equilibria. It folds several times and approaches a transversal homoclinic intersection of a periodic orbit. We have depicted this end of the family enlarged in Figure 9 (b) and (c). Figure 9 (d) displays how a typical periodic orbit close to the homoclinic intersection looks.

Adding η as a free parameter, we continue the homoclinic bifurcations, the first period doublings and folds of limit cycles of Figure 8, and the torus bifurcation of Figure 9 in two parameters. The resulting bifurcation curves are depicted in Figure 10. We observe the following.

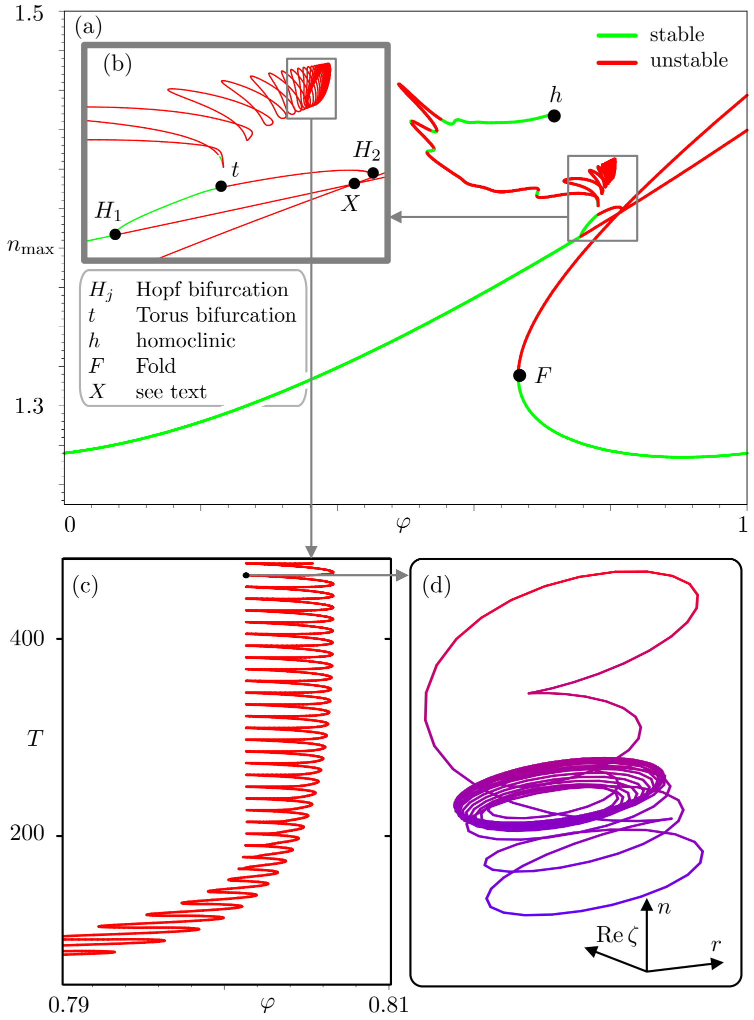


Figure 9. *Bifurcation diagram for* $\eta = 0.4$. *Diagram* (b) *shows the framed region enlarged. In* (a) *and* (b), *we report the* n*-component for the on-state and the maximum of the* n*-component for the periodic solutions. In* (c), *we report the period for the family approaching a transversal homoclinic intersection.* (d) *shows a projection of the phase portrait of one of the periodic orbits.*

1. The curve of homoclinic bifurcations forms a closed loop. The stable eigenvalues become real for smaller η: There are focus-focus to saddle-focus transitions at SC_1 and SC_2 in Figure 10. However, the saddle value is negative along the whole loop. Hence there are infinitely many stable periodic orbits in the vicinity of the loop of homoclinics.

2. The curve of period doublings forms a closed loop. It meets the torus bifurcation curve starting from the fold-Hopf interaction FH and the torus bifurcation curve passing t in Figure 9 in 1 : 2 resonances (see Figures 7 and 10). There are infinitely many other period doubling curves in the vicinity of the loop of homoclinics.

3. The curve of folds of limit cycles ends in an unstable generalized Hopf bifurcation [14] at GH_2 in one direction. There are infinitely many curves of folds of limit cycles in the

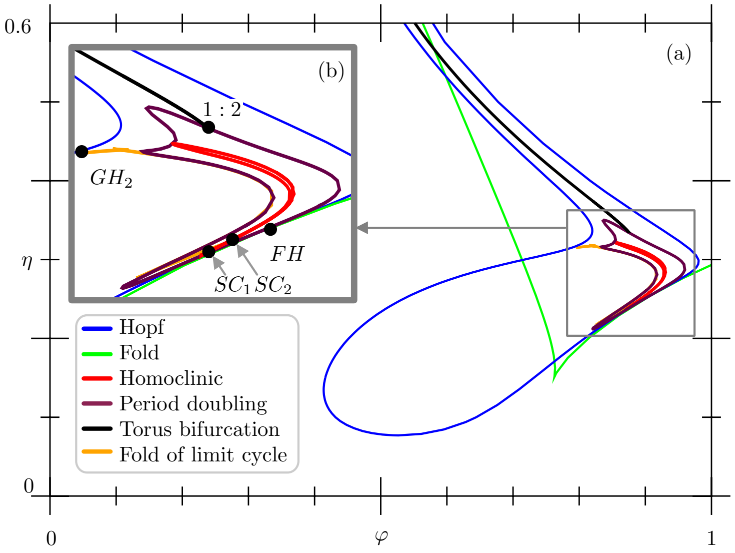


Figure 10. *Location of the homoclinic, period doubling, and torus bifurcation. The Hopf curve and the saddle-node curve are also drawn for orientation.* (b) *shows the framed region enlarged. The vicinity of the fold-Hopf interaction FH is shown in Figure 7. The curve of folds of limit cycles is partially obscured by the period doubling curve due to its proximity (see, e.g., Figure 8).*

vicinity of the loop of homoclinic bifurcations.

5. Conclusions—outlook. In this paper, we performed a numerical bifurcation analysis for a system of ODEs describing delayed optical feedback phenomena in a semiconductor laser with short cavity. We constructed the model analytically in advance by reducing the traveling-wave model with gain dispersion [4], [25], a singularly perturbed system of PDEs, to a local center manifold [26], [32]. The bifurcation parameters were the phase φ and the strength η of the delayed optical feedback.

In the first step, we analyzed the single-mode dynamics for small η, observing Hopf instabilities and two fold curves of the equilibria. The single-mode self-pulsations emerging at the Hopf bifurcation typically have an angular velocity of order $O(\sqrt{\varepsilon})$ or less as they approach a homoclinic bifurcation (see the region bordered by the dashed gray line in Figure 11). A certain part of the homoclinic is actually a closed orbit to a saddle-node, implying excitability [15] in the vicinity of the homoclinic bifurcation (see the yellow region in Figure 11).

In the second step, we extended our analysis to the neighborhood of the mode degeneracy point MD detected in the single-mode analysis using an appropriately posed two-mode system (a system of ODEs of dimension 4). We observed that there is a fold-Hopf interaction close to the point MD in the (φ, η)-plane and that the homoclinic orbits change into saddle-focus connections for larger η, implying complicated dynamics according to Shilnikov's theorems. Furthermore, we found another type of self-pulsation often referred to as mode beating or mixed-mode self-pulsation [13] (see light gray region in Figure 11). The angular velocity of the mixed-mode oscillations is related to the difference of the imaginary parts of two critical eigenvalues of H. Hence it can be of order $O(1)$, which makes this type of oscillation

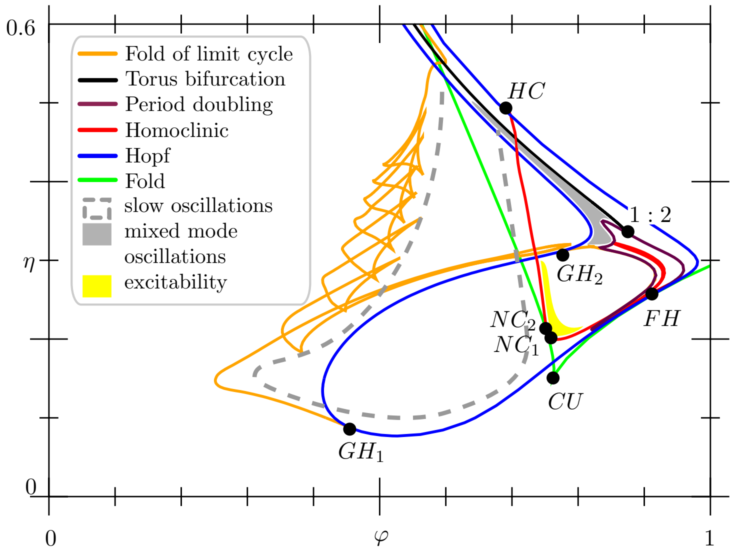


Figure 11. *Bifurcation diagram in the* (phi, η)*-plane—overview. The bifurcations of codimension* 2*, marked by black points and labels, are explained in detail in sections* 3.2 *and* 4.2*. Some regions of interest for practical applications are pointed out in particular.*

particularly interesting for applications.

In addition, we detected a torus bifurcation of the mixed-mode self-pulsations that ends in a strong 1 : 2 resonance. The period doubling bifurcation crossing this strong resonance is only the first one in an infinite sequence accumulating to a curve of homoclinic bifurcations of focus-focus type. Again, Shilnikov's theorems imply the existence of stable complicated dynamics in the vicinity of these homoclinic bifurcations. Figure 11 assembles all pictures concerning the (φ, η)-plane.

In summary, the map in the (φ, η)-plane (see Figure 11) shows the roots and borders of many delayed optical feedback phenomena observed experimentally and numerically [24], [35]. The diagram contains points (e.g., the fold-Hopf interaction and the strong resonance) and curves (e.g., homoclinic connections to saddle-foci or foci-foci) which imply the presence of complicated dynamics. Hence the diagram remains incomplete. However, several phenomena of particular practical interest have been detected and precisely located, e.g., single-mode and mixed-mode self-pulsations and excitability.

In the future, we will investigate other interesting experimental configurations (e.g., dispersive feedback, active feedback). Moreover, we will study some of the bifurcations of codimension 2 of Figure 11 in more detail and compare the bifurcation diagrams to simulation results for the complete PDE system (1.1). Moreover, it is interesting to note that first examinations of the Lang–Kobayashi system using the methods outlined in this paper lead to bifurcation diagrams of the same structure. This points to the mode degeneracy of $H(n)$, which is common to both models as the organizing center. Hence it is worth studying the normal form for laser equations close to a mode degeneracy proposed by [32] in detail.

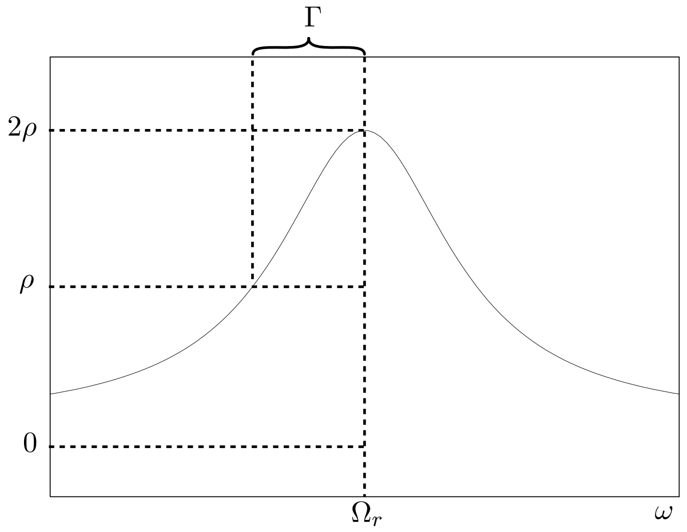


Figure 12. *Shape of the Lorentzian* $2\,\mathrm{Re}\,\chi(i\omega)$ *for* $\omega \in \mathbb{R}$ *and visualization of its parameters (see Table 2).*

6. Appendix. Physical interpretation of the traveling-wave equations—discussion of typical parameter ranges.

System (2.1) is well known as the traveling-wave model describing longitudinal dynamical effects in semiconductor lasers (see [4], [18], [29] for further references). Results of numerical simulations have been presented in [2], [3], [4], [5], [24].

The quantities ψ and p describe the complex optical field E in a spatially modulated waveguide:

$$E(\vec{r},t) = E(x,y) \cdot (\psi_1(t,z)e^{i\omega_0 t - i\frac{\pi}{\Lambda}z} + \psi_2(t,z)e^{i\omega_0 t + i\frac{\pi}{\Lambda}z}).$$

The complex amplitudes $\psi_{1,2}(t,z)$ are the longitudinally slowly varying envelopes of E. The transversal space directions are x and y, the longitudinal direction is z, and $\vec{r} = (x,y,z)$. For periodically modulated waveguides, Λ is the longitudinal modulation wavelength. The central frequency is $\omega_0/(2\pi)$, and $E(x,y)$ is the dominant transversal mode of the waveguide.

The equation $\dot{E} = H(n)E$ for an uncoupled waveguide ($\kappa = 0$), a monochromatic lightwave in the forward direction $e^{i\omega t}\psi_1(z)$, and a constant carrier density n imply a spatial shape of the power $|\psi_1|^2$ according to

$$\partial_z|\psi_1(z)|^2 = (2\,\mathrm{Re}\,\beta(z) + 2\,\mathrm{Re}\,\chi(i\omega,z))|\psi_1(z)|^2, \tag{6.1}$$

where

$$\chi(i\omega,z) = \frac{\rho(z)\Gamma(z)}{i\omega - i\Omega_r(z) + \Gamma(z)}. \tag{6.2}$$

$2\,\mathrm{Re}\,\chi(i\omega,z))$ is a Lorentzian intended to fit the gain curve of the waveguide material (see Figure 12). Hence $\dot{E} = HE$ produces gain dispersion; i.e., the spatial growth rate of the wave $e^{i\omega t}\psi(z)$ depends on its frequency ω. The variable $p(t,z)$ reports the internal state of the

Table 2

Ranges and explanations of the variables and coefficients appearing in (2.1)–(2.4)*. See also* [4], [27] *to inspect their relations to the originally used physical quantities and scales.*

	Typical range	Explanation
$\psi(t,z)$	$\mathbb{C}^2$	optical field, forward and backward traveling wave
$i \cdot p(t,z)$	$\mathbb{C}^2$	nonlinear polarization for the forward and backward traveling wave
$n(t)$	$(\underline{n}, \infty)$	spatially averaged carrier density in section S_1
$\operatorname{Im} \beta_k^0$	$\mathbb{R}$	frequency detuning
$\operatorname{Re} \beta_k^0$	< 0, $(-10, 0)$	decay rate due to internal losses
α_H	$(0, 10)$	negative of line-width enhancement factor
g_1	≈ 1	differential gain in S_1
κ_k	$(-10, 10)$	real coupling coefficients for the optical field ψ
ρ_k	$[0, 1)$	ρ_k is maximum of the gain curve
Γ_k	$O(10^2)$	half width of half maximum of the gain curve
$\Omega_{r,k}$	$O(10)$	resonance frequency
I	$O(10^{-2})$	current injection in section S_1
τ	$O(10^2)$	spontaneous lifetime for the carriers
P	$(0, \infty)$	scale of (ψ, p) (can be chosen arbitrarily)
r_0, r_L	$\mathbb{C}$, $\lvert r_0 \rvert, \lvert r_L \rvert < 1$	facet reflectivities

gain filter. See [4], [27] for more details. The Lorentzian gain filter is also used by [1], [18], and [20]. Since the coefficients ρ, Γ, and Ω are supposed to be spatially sectionwise constant, $\chi(\lambda, z) = \chi_1(\lambda)$ for z in section S_1.

The equation for $\dot{n}$ in (2.1) is a simple rate equation for the spatially averaged carrier density. It accounts for the current I, the spontaneous recombination $-n/\tau$, and the stimulated recombination. See Table 2 for typical ranges of the quantities.

REFERENCES

[1] E. A. Avrutin, J. H. Marsh, and J. M. Arnold, *Modelling of semiconductor laser structures for passive harmonic mode locking at terahertz frequencies*, Int. J. Optoelectronics, 10 (1995), pp. 427–432.

[2] U. Bandelow, *Theorie longitudinaler Effekte in* 1.55 *μm Mehrsektions DFB-Laserdioden*, Ph.D. thesis, Humboldt-Universität Berlin, Berlin, Germany, 1994.

[3] U. Bandelow, L. Recke, and B. Sandstede, *Frequency regions for forced locking of self-pulsating multi-section DFB lasers*, Opt. Comm., 147 (1998), pp. 212–218.

[4] U. Bandelow, M. Wolfrum, M. Radziunas, and J. Sieber, *Impact of gain dispersion on the spatio-temporal dynamics of multisection lasers*, IEEE J. Quantum Electronics, 37 (2001), pp. 183–189.

[5] U. Bandelow, H. J. Wünsche, B. Sartorius, and M. Mhrle, *Dispersive self Q-switching in DFB-lasers: Theory versus experiment*, IEEE J. Selected Topics in Quantum Electronics, 3 (1997), pp. 270–278.

[6] P. W. Bates, K. Lu, and C. Zeng, *Existence and persistence of invariant manifolds for semiflows in Banach space*, Mem. Amer. Math. Soc., 135 (1998).

[7] P. W. Bates, K. Lu, and C. Zeng, *Persistence of overflowing manifolds for semiflow*, Comm. Pure Appl. Math., 52 (1999), pp. 983–1046.

[8] P. W. Bates, K. Lu, and C. Zeng, *Invariant foliations near normally hyperbolic invariant manifolds for semiflows*, Trans. Amer. Math. Soc., 352 (2000), pp. 4641–4676.

[9] W. Beyn, *The numerical computation of connecting orbits in dynamical systems*, IMA J. Numer. Anal., 10 (1990), pp. 379–405.

[10] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, North–Holland, New York, 1989.

[11] E. J. Doedel, A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. Sandstede, and X. Wang, AUTO97: *Continuation and Bifurcation Software for Ordinary Differential Equations*, Tech. Rep., Department of Computer Science, Concordia University, Montreal, Canada; available via anonymous ftp from http://ftp.cs.concordia.ca from the directory pub/doedel/auto, 1998.

[12] K. Engelborghs, *DDE-BIFTOOL: A Matlab Package for Bifurcation Analysis of Delay Differential Equations*, Report TW 305, Katholieke Universiteit Leuven, Leuven, The Netherlands, 2000.

[13] T. Erneux, F. Rogister, A. Gavrielides, and V. Kovanis, *Bifurcation to mixed external cavity mode solutions for semiconductor lasers subject to external feedback*, Opt. Comm., 183 (2000), pp. 467–477.

[14] W. Govaerts, Y. A. Kuznetsov, and B. Sijnave, *Numerical methods for the generalized Hopf bifurcation*, SIAM J. Numer. Anal., 38 (2000), pp. 329–346.

[15] E. M. Izhikevich, *Neural excitability, spiking and bursting*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 10 (2000), pp. 1171–1266.

[16] Y. Kuznetsov, *Elements of Applied Bifurcation Theory*, Springer-Verlag, New York, 1995.

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

[18] D. Marcenac, *Fundamentals of Laser Modelling*, Ph.D. thesis, University of Cambridge, Cambridge, UK, 1993.

[19] J. Mork, B. Tromborg, and J. Mark, *Chaos in semiconductor lasers with optical feedback: Theory and experiment*, IEEE J. Quantum Electronics, 28 (1992), pp. 93–108.

[20] C. Z. Ning, R. A. Indik, and J. V. Moloney, *Effective Bloch equations for semiconductor lasers and amplifiers*, IEEE J. Quantum Electronics, 33 (1997), pp. 1543–1550.

[21] G. L. Oppo and A. Politi, *Toda potentials in laser equations*, Z. Phys., 59 (1985), pp. 111–150.

[22] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Appl. Math. Sci. 44, Springer-Verlag, New York, 1983.

[23] M. Radziunas and H.-J. Wünsche, *Dynamics of Multi-Section DFB Semiconductor Laser: Traveling Wave and Mode Approximation Models*, Preprint 713, WIAS, Berlin, Germany, 2002; submitted to SPIE.

[24] M. Radziunas, H.-J. Wünsche, B. Sartorius, O. Brox, D. Hoffmann, K. Schneider, and D. Marcenac, *Modeling self-pulsating DFB lasers with integrated phase tuning section*, IEEE J. Quantum Electronics, 36 (2000), pp. 1026–1034.

[25] J. Sieber, *Longitudinal Dynamics of Semiconductor Lasers*, Report 20, WIAS, Berlin, Germany, 2001.

[26] J. Sieber, *Longtime Behavior of the Traveling-Wave Model for Semiconductor Lasers*, Preprint 743, WIAS, Berlin, Germany, 2002; SIAM J. Appl. Dynamical Systems, submitted.

[27] J. Sieber, U. Bandelow, H. Wenzel, M. Wolfrum, and H.-J. Wünsche, *Travelling Wave Equations for Semiconductor Lasers with Gain Dispersion*, Preprint 459, WIAS, Berlin, Germany, 1998.

[28] A. A. Tager and K. Petermann, *High-frequency oscillations and self-mode locking in short external-cavity laser diodes*, IEEE J. Quantum Electronics, 30 (1994), pp. 1553–1561.

[29] B. Tromborg, H. E. Lassen, and H. Olesen, *Travelling wave analysis of semiconductor lasers*, IEEE J. Quantum Electronics, 30 (1994), pp. 939–956.

[30] B. Tromborg, J. H. Osmundsen, and H. Olesen, *Stability analysis for a semiconductor laser in an external cavity*, IEEE J. Quantum Electronics, 20 (1984), pp. 1023–1032.

[31] V. Tronciu, H.-J. Wünsche, J. Sieber, K. Schneider, and F. Henneberger, *Dynamics of single mode semiconductor lasers with passive dispersive reflectors*, Opt. Comm., 182 (2000), pp. 221–228.

[32] D. Turaev, *Fundamental obstacles to self-pulsations in low-intensity lasers*, Preprint 629, WIAS, Berlin, Germany, 2001; SIAM J. Appl. Math., submitted.

[33] A. Vanderbauwhede and G. Iooss, *Center manifold theory in infinite dimensions*, in Dynamics Reported, Vol. 1, Springer-Verlag, New York, 1992, pp. 125–163.

[34] H. Wenzel, U. Bandelow, H.-J. Wünsche, and J. Rehberg, *Mechanisms of fast self pulsations in two-section DFB lasers*, IEEE J. Quantum Electronics, 32 (1996), pp. 69–79.

[35] H. J. Wünsche, O. Brox, M. Radziunas, and F. Henneberger, *Excitability of a semiconductor laser by a two-mode homoclinic bifurcation*, Phys. Rev. Lett., 88 (2002).

SIAM J. Applied Dynamical Systems
Vol. 1, No. 2, pp. 271–302

Higher Order Modulation Equations for a Boussinesq Equation[*]

C. Eugene Wayne[†] and J. Douglas Wright[†]

Abstract. In order to investigate corrections to the common KdV approximation to long waves, we derive modulation equations for the evolution of long wavelength initial data for a Boussinesq equation. The equations governing the corrections to the KdV approximation are explicitly solvable, and we prove estimates showing that they do indeed give a significantly better approximation than the KdV equation alone. We also present the results of numerical experiments which show that the error estimates we derive are essentially optimal.

Key words. Boussinesq equation, KdV approximation, modulation equations, water wave problem

AMS subject classifications. 76B15, 35Q51, 35Q53

PII. S1111111102411298

1. Introduction.

Modulation, or amplitude, equations are approximate, often explicitly solvable, model equations derived—usually through asymptotic analysis and the method of multiple time scales—to model more complicated physical situations. Although these equations have been used for over a century, only lately has there been an attempt to rigorously relate solutions of the modulation equations to the original physical problem. In particular, through the work of Craig [8], Kano and Nishida [13], Kalyakin [12], Schneider [21], Ben Youssef and Colin [1], and Schneider and Wayne [22], [23], the validity of Korteweg-de Vries (KdV) equations as a leading order approximation to the evolution of long wavelength water waves and to a number of other dispersive partial differential equations has been established.

While the KdV approximation is extremely useful due to its simplicity and the fact that the KdV equation can be explicitly solved by the inverse scattering transform, both experimentally and numerically one observes departures from the predictions of the KdV equation. Our goal in this paper is to derive modulation equations which govern corrections to the KdV model. In the present paper, we will not work with the full water wave problem but rather will study modulation equations for long wavelength solutions of the Boussinesq equation:

$$\begin{gathered}\theta_{tt} - \theta_{xx} = (\theta^2)_{xx} + \theta_{ttxx}, \\ x \in \mathbb{R},\, t \geq 0,\, \theta(x,t) \in \mathbb{R}.\end{gathered} \tag{1.1}$$

Our motivation for studying this equation is twofold. First, the Boussinesq equation was originally derived as a model equation for water waves, and, as our ultimate goal is to derive corrections to the KdV approximation to water waves, we regard the study of (1.1) as a useful first step in understanding the much more complicated water wave situation. We note that

[*] Received by the editors July 17, 2002; accepted for publication by G. Iooss September 6, 2002; published electronically December 31, 2002. This work was supported in part by NSF grant DMS-0103915.
http://www.siam.org/journals/siads/1-2/41129.html

[†] Department of Mathematics and Statistics and Center for BioDynamics, Boston University, Boston, MA 02215 (cew@math.bu.edu, jdoug@math.bu.edu).

Schneider's analysis of the KdV approximation for (1.1) in [21] served as a template for the analysis of the water wave problem in [22].

Our second justification for deriving second order modulation equations for (1.1) is that these modulation equations serve as a sort of normal form for more complicated PDEs, and as such we expect that the modulation equations which describe corrections to the KdV approximation to (1.1) will govern corrections to the KdV approximation in more complicated situations as well. Thus the results on existence, uniqueness, and other properties of the modulation equations we derive in this paper should also be of use in more complicated situations that we plan to treat in the future.

We now describe in more detail our results. It is convenient to rewrite (1.1) as a system of two first order equations. As in [21], we introduce new variables

$$\begin{aligned} u(x,t) &= \frac{1}{2}(\theta(x,t) - \lambda^{-1}\theta_t(x,t)), \\ v(x,t) &= \frac{1}{2}(\theta(x,t) + \lambda^{-1}\theta_t(x,t)), \end{aligned} \tag{1.2}$$

where λ is a skew-symmetric multiplication operator in Fourier space defined by $\widehat{\lambda u} = (ik/\sqrt{1+k^2})\hat{u}$. Note that, for $\lambda^{-1}\theta_t$ to be well defined, we must have $\hat{\theta}_t(0,t) = 0$. That is, the average value of θ_t should be zero. We note that, if $\theta(x,t)$ is a solution of (1.1), we have that the average value of $\theta_t(x,t)$ is a constant of the motion. Thus, if the initial condition for θ_t has zero average, $\hat{\theta}_t(0,t) = 0$ will remain zero for all time. Furthermore, as discussed in [21], assuming that the initial condition $\theta_t(x,0)$ has zero average is not unnatural considering the origin of (1.1). Thus we will make that assumption so that the change of variables (1.2) is well defined.

Taking time derivatives of u and v, we find

$$\partial_t \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -\lambda(u+v)^2 \\ \lambda(u+v)^2 \end{pmatrix}. \tag{1.3}$$

Not only is (1.3) convenient from a mathematical point of view, but, as we shall see, u and v have the physical interpretation of being the left and right moving parts of the solution.

We turn now to the assumptions on the initial conditions of (1.3). The KdV equation is an approximation of small amplitude and long wavelength motions, and thus we will assume that the initial conditions of (1.1) are of this form. More precisely, fix a constant $C_I > 0$ and assume the following.

Hypothesis 1. *There exist U_0, V_0 with*

$$\max\left\{\|U_0\|_{H^\sigma(4)\cap H^{\sigma+9}}, \|V_0\|_{H^\sigma(4)\cap H^{\sigma+9}}\right\} < C_I$$

such that the initial conditions of (1.3) are of the form

$$u(x,0) = \epsilon^2 U_0(\epsilon x), \quad v(x,0) = \epsilon^2 V_0(\epsilon x),$$

where ϵ is small, and σ will be fixed in the statement of the main theorems.

Here, $H^\sigma(m) = \{f | (1+x^2)^{m/2} f \in H^\sigma\}$. The norm on this weighted Sobolev space is given by $\|f\|_{H^\sigma(m)} = \|(1+x^2)^{m/2} f\|_{H^\sigma}$. We use these spaces because we are interested in solutions which are in some (weak) sense "localized." In particular, any small perturbation of the known soliton solutions to the KdV equation will satisfy this localization property.

Remark 1. We can, of course, recover the initial conditions for (1.1) from Hypothesis 1 via (1.2), and we see that the initial conditions expressed in the θ variables are also of small amplitude long wavelength form.

According to the KdV approximation results of [21], long wavelength solutions of (1.1) split into two pieces—one a right moving wave train and one a left moving wave train. Each of these wave trains evolves according to a KdV equation, and there is no interaction between the left and right moving pieces. One might expect two types of corrections to such an approximation:

- corrections due the fact that the left and right moving wave trains will interact at a higher order,
- corrections due to the fact that, even in the case of a purely right (or left) moving wave train, solutions to the Boussinesq equation are not exactly described by solutions to the KdV equation.

Both of these types of corrections are apparent in our results, and, in fact, the corrections to the KdV approximation are a sum of solutions of two types of modulation equations—an inhomogeneous transport equation and a linearized KdV equation which can be seen (roughly speaking) as reflecting these two sources of corrections.

To incorporate these two types of corrections, we add to the KdV wave trains, which we denote by U and V (since they represent the leading terms in u and v, respectively), additional functions A and B and F and G. These functions then satisfy the modulation equations

$$\begin{aligned} \partial_T U &= -\frac{1}{2}\partial_{X_-}^3 U - \frac{1}{2}\partial_{X_-} U^2, \\ \partial_T V &= \ \frac{1}{2}\partial_{X_+}^3 V + \frac{1}{2}\partial_{X_+} V^2, \end{aligned} \tag{1.4}$$

$$\begin{aligned} \partial_\tau A + \partial_X A &= -\frac{1}{2}\partial_{X_+} V^2(X+\tau,\epsilon^2\tau) - \partial_X\left(U(X-\tau,\epsilon^2\tau)V(X+\tau,\epsilon^2\tau)\right), \\ \partial_\tau B - \partial_X B &= \ \frac{1}{2}\partial_{X_-} U^2(X-\tau,\epsilon^2\tau) + \partial_X\left(U(X-\tau,\epsilon^2\tau)V(X+\tau,\epsilon^2\tau)\right), \end{aligned} \tag{1.5}$$

and

$$\begin{aligned} \partial_T F &= -\partial_{X_-}(UF) - \frac{1}{2}\partial_{X_-}^3 F + J^1, \\ \partial_T G &= \ \partial_{X_+}(VG) + \frac{1}{2}\partial_{X_+}^3 G + J^2, \end{aligned} \tag{1.6}$$

where $T = \epsilon^3 t$, $\tau = \epsilon t$, $X = \epsilon x$, and $X_\pm = X \pm \tau$.

The first of these pairs of equations is simply the KdV approximation. The second and third pairs give rise to the corrections to the KdV approximation. We note that the terms J^1 and J^2 which appear in (1.6) are inhomogeneous terms which are made up of a combination of sums and products of the solutions to (1.4), (1.5), and their derivatives (see (2.8)–(2.9)).

There is some freedom in how we choose the initial data for the modulation equations. For simplicity, we assume that $U(X,0)=U_0(X)$ and $V(X,0)=V_0(X)$ and choose zero initial data for (1.5) and (1.6), i.e., $A(X,0)=B(X,0)=F(X,0)=G(X,0)=0$.

That the KdV equation has solutions for all times with initial data of the type described is well known. In particular, one has the following theorem (see [22]).

Theorem 1.1. *Let $\sigma \geq 4$. Then, for all $C_0, T_0 > 0$, there exists $C_1 > 0$ such that, if U, V satisfy* (1.4) *with initial conditions U_0, V_0, and*

$$\max\{\|U_0\|_{H^\sigma(4)\cap H^{\sigma+9}}, \|V_0\|_{H^\sigma(4)\cap H^{\sigma+9}}\} < C_0, \tag{1.7}$$

then

$$\sup_{T\in[0,T_0]} \left\{\|U(\cdot,T)\|_{H^\sigma(4)\cap H^{\sigma+8}}, \|V(\cdot,T)\|_{H^\sigma(4)\cap H^{\sigma+8}}\right\} < C_1. \tag{1.8}$$

On the other hand, it is less clear that solutions of (1.5) and (1.6) will remain bounded over the very long time scales necessary for the KdV approximation. Thus the first significant technical result of this paper is the following proposition.

Proposition 1.2. *Fix $T_0 > 0$ and $\sigma > 21/2$. Suppose that U_0, V_0 satisfy* (1.7) *and U, V, A, B, F, and G satisfy* (1.4)–(1.6)*; then there exists a constant C_2 such that the solutions of* (1.5) *and* (1.6) *satisfy the estimates*

$$\sup_{\tau\in[0,T_0\epsilon^{-2}]} \{\|A(\cdot,\tau)\|_{H^{\sigma-3}}, \|B(\cdot,\tau)\|_{H^{\sigma-3}}\} \leq C_2,$$
$$\sup_{T\in[0,T_0]} \left\{\|F(\cdot,T)\|_{\tilde{H}}, \|G(\cdot,T)\|_{\tilde{H}}\right\} \leq C_2,$$

where $\tilde{H} = H^{\sigma-5}\cap H^{\sigma-9}(2)$.

With this preliminary result in hand, we can now state our principal result.

Theorem 1.3. *Fix T_0, $C_I > 0$, and $\sigma \geq 13$. Suppose U, V, A, B, F, and G satisfy* (1.4)–(1.6)*. Then there exist $\epsilon_0 > 0$ and $C_F > 0$ such that, if the initial conditions for* (1.3) *satisfy Hypothesis* 1*, then, for $\epsilon \in (0,\epsilon_0)$, we have that the unique solution $\bar{u}$ to* (1.3) *satisfies*

$$\|\bar{u}(\cdot,t) - \bar{w}(\cdot,t)\|_{H^{\sigma-13}\times H^{\sigma-13}} \leq C_F\epsilon^{11/2}$$

for $t\in[0,T_0\epsilon^{-3}]$, where

$$\bar{w}(x,t) = \epsilon^2 \begin{pmatrix} U(X_-,T) \\ V(X_+,T) \end{pmatrix} + \epsilon^4 \begin{pmatrix} A(X,\tau)+F(X_-,T) \\ B(X,\tau)+G(X_+,T) \end{pmatrix}.$$

Given this result and the change of variables (1.2), we can immediately rewrite this approximation theorem in terms of the original variables. Define

$$\begin{aligned}\theta_{app}(x,t) = \ & \epsilon^2\left(U(\epsilon(x-t),\epsilon^3 t) + V(\epsilon(x+t),\epsilon^3 t)\right) \\ & + \epsilon^4\left(A(\epsilon x,\epsilon t) + B(\epsilon x,\epsilon t)\right) \\ & + \epsilon^4\left(F(\epsilon(x-t),\epsilon^3 t) + G(\epsilon(x+t),\epsilon^3 t)\right).\end{aligned}$$

Corollary 1.4. *Fix T_0, $C_I > 0$, $\sigma \geq 13$. Suppose U, V, A, B, F, and G satisfy* (1.4)–(1.6). *Then there exist $\epsilon_0 > 0$ and $C_F > 0$ such that, if the initial conditions for* (1.3) *satisfy Hypothesis* 1*, then, for $\epsilon \in (0, \epsilon_0)$, the unique solution $\theta(x,t)$ to* (1.1) *satisfies*

$$\|\theta(\cdot,t) - \theta_{app}(\cdot,t)\|_{H^{\sigma-13}} \leq C_F \epsilon^{11/2}$$

for $t \in [0, T_0\epsilon^{-3}]$.

Remark 2. The initial conditions for (1.1) are obtained from those of (1.3) simply by inverting the transformation (1.2).

Remark 3. In Schneider's paper [21], he proves a similar theorem which shows that the H^s error made in approximating (1.3) by the KdV equations alone is of $O(\epsilon^3)$.

The remainder of the paper is devoted to the proof of Theorem 1.3 and Proposition 1.2. In the next section, we give a formal derivation of (1.4)–(1.6). In section 3, we study the existence of solutions to (1.5) and (1.6) and prove Proposition 1.2. Section 4 is the technical heart of the paper and contains the proof of Theorem 1.3. The proof follows the general approach for justifying modulation equations laid out in [14], but controlling the higher order approximation requires fairly extensive technical modifications. In section 5, we present the results of a variety of numerical computations related to Corollary 1.4. These computations give insight into several aspects of the second order approximation. First, it allows us to estimate how large the values of ϵ_0 and C_F in Corollary 1.4 are. They also show that the order of ϵ in the error estimates (i.e., 11/2) is apparently optimal. Finally, in the concluding section, we discuss other work on second order corrections to the KdV approximation, both rigorous and nonrigorous, and how it relates to our own results.

2. Formal derivation of the modulation equations.

One can derive from (1.3) a system of KdV equations via the method of multiple time scales—this was done in [21], for example. We extend that calculation in this section to include the approximating equations for the next order correction.

To derive the modulation equations, we first make the ansatz

$$\begin{pmatrix} u(x,t) \\ v(x,t) \end{pmatrix} = \epsilon^2 \begin{pmatrix} U(X_-,T) \\ V(X_+,T) \end{pmatrix} + \epsilon^4 \begin{pmatrix} A(X,\tau) + F(X_-,T) \\ B(X,\tau) + G(X_+,T) \end{pmatrix} + O(\epsilon^6), \tag{2.1}$$

where $\tau = \epsilon t$, $T = \epsilon^3 t$, $X = \epsilon x$, $X_- = X - \tau$, and $X_+ = X + \tau$. The two new time variables are the "multiple time scales" spoken of earlier. For convenience, we will also denote $\bar{u} = (u,v)^t$, $\bar{U} = (U,V)^t$, $\bar{A} = (A,B)^t$, and $\bar{F} = (F,G)^t$.

It may seem somewhat odd that the $O(\epsilon^4)$ correction consists of a sum of functions as opposed to a single function. The reason for this is that, for our first order approximation terms, U and V, we are assuming that u and v exhibit only unidirectional motion (right and left, respectively). A and B, loosely, correct for the effect of the interaction of right and left moving waves, and they evolve on the fast time scale, τ. There are also unidirectional second order effects, which we represent with F and G. Their functional form is the same as that of the first order terms.

In a moment, we will insert (2.1) into (1.3), but first we compute the effect of the operator λ on long wavelength data. Define a function W by $w(x) = W(X)$. We wish to compute $\lambda w(x)$

and see how that relates to $W(X)$. The long wavelength Fourier transform variable is denoted by $K = k/\epsilon$. Since ϵ is small, we will (formally) approximate the effect of $\hat{\lambda}(k) = \hat{\lambda}(\epsilon K)$ by the first few terms of its Maclaurin series.

$$\begin{aligned}
\lambda w(x) &= \mathfrak{F}^{-1}\left\{\hat{\lambda}(k)\hat{w}(k)\right\}(x) \\
&= \int e^{ikx}\hat{\lambda}(k)\epsilon^{-1}\hat{W}(k/\epsilon)dk \\
&= \int e^{iKX}\hat{\lambda}(\epsilon K)\hat{W}(K)dK \\
&= \mathfrak{F}^{-1}\left\{\hat{\lambda}(\epsilon K)\hat{W}(K)\right\}(X) \\
&= \mathfrak{F}^{-1}\left\{\left(\epsilon(iK) + \frac{1}{2}\epsilon^3(iK)^3 + \frac{3}{8}\epsilon^5(iK)^5 + O(\epsilon^7)\right)\hat{W}(K)\right\}(X) \\
&= \left(\epsilon\partial_X + \frac{1}{2}\epsilon^3\partial_X^3 + \frac{3}{8}\epsilon^5\partial_X^5 + O(\epsilon^7)\right)W(X).
\end{aligned}$$

It is important to note that this approximation is only formally good to $O(\epsilon^7)$.

Now we insert this approximation for λ and the ansatz into (1.3). This is a necessarily messy procedure. To reduce the notation, anything formally $O(\epsilon^9)$ or higher is (more or less) disregarded. Also, an additional term is added to the ansatz of the form

$$\epsilon^6\bar{S} = \epsilon^6\begin{pmatrix} S^1(X,\tau) \\ S^2(X,\tau) \end{pmatrix}. \tag{2.2}$$

While this term will be treated in much the same way as the other terms in the ansatz, it should be noted that this is not truly part of the next order correction. It will, however, be quite useful when we prove the approximation is a good one.

We must re-express the partial derivatives in (1.3) in terms of the new coordinates. By the chain rule, we have

$$\partial_t = -\epsilon\partial_{X_-} + \epsilon\partial_{X_+} + \epsilon\partial_\tau + \epsilon^3\partial_T.$$

Spatial derivatives of terms of the form $f(X_-)g(X_+)$ or $f(X)g(X_\pm)$ are denoted by ∂_X, though all other spatial derivatives are denoted with respect to the appropriate coordinate.

So we get on the left-hand side of (1.3)

$$\begin{aligned}
\partial_t\begin{pmatrix} u(x,t) \\ v(x,t) \end{pmatrix} &= \epsilon^3\begin{pmatrix} -\partial_{X_-}U(X_-,T) \\ \partial_{X_+}V(X_+,T) \end{pmatrix} \\
&+ \epsilon^5\begin{pmatrix} \partial_T U(X_-,T) + \partial_\tau A(X,\tau) - \partial_{X_-}F(X_-,T) \\ \partial_T V(X_+,T) + \partial_\tau B(X,\tau) + \partial_{X_+}G(X_+,T) \end{pmatrix} \\
&+ \epsilon^7\begin{pmatrix} \partial_T F(X_-,T) + \partial_\tau S^1(X,\tau) \\ \partial_T G(X_+,T) + \partial_\tau S^2(X,\tau) \end{pmatrix}.
\end{aligned} \tag{2.3}$$

Now we must compute the right-hand side of (1.3). A routine calculation yields

$$\begin{aligned}
RHS = \epsilon^3 & \begin{pmatrix} -\partial_{X_-} U \\ \partial_{X_+} V \end{pmatrix} \\
& + \epsilon^5 \begin{pmatrix} -\frac{1}{2}\partial_{X_-}^3 U - \partial_X A - \partial_{X_-} F \\ \frac{1}{2}\partial_{X_+}^3 V + \partial_X B + \partial_{X_+} G \end{pmatrix} \\
& + \epsilon^5 \begin{pmatrix} -\frac{1}{2}\partial_{X_-} U^2 - \frac{1}{2}\partial_{X_+} V^2 - \partial_X (UV) \\ \frac{1}{2}\partial_{X_-} U^2 + \frac{1}{2}\partial_{X_+} V^2 + \partial_X (UV) \end{pmatrix} \\
& + \epsilon^7 \begin{pmatrix} -\frac{1}{2}\partial_X^3 A - \frac{1}{2}\partial_{X_-}^3 F - \frac{3}{8}\partial_{X_-}^5 U \\ \frac{1}{2}\partial_X^3 B + \frac{1}{2}\partial_{X_+}^3 G + \frac{3}{8}\partial_{X_+}^5 V \end{pmatrix} \\
& + \epsilon^7 \begin{pmatrix} -\partial_X(UA) - \partial_{X_-}(UF) - \partial_X(UB) - \partial_X(UG) \\ +\partial_X(UA) + \partial_{X_-}(UF) + \partial_X(UB) + \partial_X(UG) \end{pmatrix} \\
& + \epsilon^7 \begin{pmatrix} -\partial_X(VA) - \partial_X(VF) - \partial_X(VB) - \partial_{X_+}(VG) \\ +\partial_X(VA) + \partial_X(VF) + \partial_X(VB) + \partial_{X_+}(VG) \end{pmatrix} \\
& + \epsilon^7 \begin{pmatrix} -\frac{1}{4}\partial_{X_-}^3 U^2 - \frac{1}{4}\partial_{X_+}^3 V^2 - \frac{1}{2}\partial_X^3 (UV) - \partial_X S^1 \\ +\frac{1}{4}\partial_{X_-}^3 U^2 + \frac{1}{4}\partial_{X_+}^3 V^2 + \frac{1}{2}\partial_X^3 (UV) + \partial_X S^2 \end{pmatrix} + O(\epsilon^9).
\end{aligned} \tag{2.4}$$

So we see that we can satisfy (1.3) formally to $O(\epsilon^5)$ by taking

$$\begin{aligned}
\partial_T U &= -\frac{1}{2}\partial_{X_-}^3 U - \frac{1}{2}\partial_{X_-} U^2, \\
\partial_T V &= \ \ \frac{1}{2}\partial_{X_+}^3 V + \frac{1}{2}\partial_{X_+} V^2,
\end{aligned} \tag{1.4}$$

and

$$\begin{aligned}
\partial_\tau A + \partial_X A &= -\frac{1}{2}\partial_{X_+} V^2(X+\tau, \epsilon^2\tau) - \partial_X U(X-\tau, \epsilon^2\tau) V(X+\tau, \epsilon^2\tau), \\
\partial_\tau B - \partial_X B &= \ \ \frac{1}{2}\partial_{X_-} U^2(X-\tau, \epsilon^2\tau) + \partial_X U(X-\tau, \epsilon^2\tau) V(X+\tau, \epsilon^2\tau).
\end{aligned} \tag{1.5}$$

Equations (1.4) are a pair of uncoupled KdV equations. That their solutions provide the first order approximation to long wavelength solutions of (1.1) was proven in [21]. Solutions to the KdV equations are known to exist and to be bounded over a long time scale (see Theorem 1.1).

System (1.5) is a set of inhomogeneous transport equations driven by the solutions to the KdV equations. We can write an explicit formula for the solutions.

Corollary 2.1. *The solutions to* (1.5) *are given by*

$$\begin{aligned}
A(X,\tau) &= \frac{1}{4}V^2(X-\tau, 0) - \frac{1}{4}V^2(X+\tau, \epsilon^2\tau) \\
&\quad + \alpha(X-\tau, \epsilon^2\tau) + A_1(X,\tau) \\
B(X,\tau) &= \frac{1}{4}U^2(X+\tau, 0) - \frac{1}{4}U^2(X-\tau, \epsilon^2\tau) \\
&\quad + \beta(X+\tau, \epsilon^2\tau) + B_1(X,\tau),
\end{aligned} \tag{2.5}$$

where

$$\begin{aligned} A_1(X,\tau) &= \frac{\epsilon^2}{4}\int_0^\tau \partial_T V^2(X-\tau+2s,\epsilon^2 s)ds, \\ B_1(X,\tau) &= \frac{\epsilon^2}{4}\int_0^\tau \partial_T U^2(X+\tau-2s,\epsilon^2 s)ds, \end{aligned} \tag{2.6}$$

$$\begin{aligned} \alpha(X_-,T) &= -\epsilon^{-2}\int_0^T \partial_{X_-}\left(U(X_-,s)V(X_-+2\epsilon^{-2}s,s)\right)ds, \\ \beta(X_+,T) &= \epsilon^{-2}\int_0^T \partial_{X_+}\left(U(X_+-2\epsilon^{-2}s,s)V(X_+,s)\right)ds. \end{aligned} \tag{2.7}$$

Proof. The proof follows directly from Lemmas 3.2 and 3.3, which appear in the next section. They can also be verified by inserting the expressions for A and B back into (1.5). Furthermore, as we prove below, in spite of the prefactor of ϵ^{-2}, α and β remain $O(1)$ for all $0 \le T \le T_0$, for any T_0. ■

The terms of $O(\epsilon^7)$ in (2.4) give rise to two sets of linear evolution equations, one for F and G and one for S^1 and S^2. Both are inhomogeneous systems of equations due to the presence of terms involving U, V, A, B, and their derivatives. We have some freedom in the way we split up the inhomogeneous terms between these equations, and we attempt to group them in such a way that it is easy to estimate the resulting solutions over the long time scales relevant for the approximation problem. In particular, we will break A and B up as in the explicit solutions above. We have

$$\begin{aligned} \partial_T F &= -\,\partial_{X_-}(UF) - \frac{1}{2}\partial^3_{X_-}F + J^1, \\ \partial_T G &= \ \ \partial_{X_+}(VG) + \frac{1}{2}\partial^3_{X_+}G + J^2, \end{aligned} \tag{1.6}$$

where the inhomogeneous terms J^1 and J^2 are given by

$$\begin{aligned} J^1(X_-,T) = &-\frac{3}{8}\partial^5_{X_-}U(X_-,T) - \frac{1}{4}\partial^3_{X_-}U^2(X_-,T) \\ &+\frac{1}{4}\partial_{X_-}U^3(X_-,T) - \frac{1}{8}\partial^3_{X_-}V^2(X_-,0) \\ &-\partial_{X_-}\left(U(X_-,T)\left(\frac{1}{4}V^2(X_-,0)+\alpha(X_-,T)\right)\right) \\ &-\frac{1}{2}\partial^3_{X_-}\alpha(X_-,T), \end{aligned} \tag{2.8}$$

$$\begin{aligned} J^2(X_+,T) = &\ \ \frac{3}{8}\partial^5_{X_+}V(X_+,T) + \frac{1}{4}\partial^3_{X_+}V^2(X_+,T) \\ &-\frac{1}{4}\partial_{X_+}V^3(X_+,T) + \frac{1}{8}\partial^3_{X_+}U^2(X_+,0) \\ &+\partial_{X_+}\left(V(X_+,T)\left(\frac{1}{4}U^2(X_+,0)+\beta(X_+,T)\right)\right) \\ &+\frac{1}{2}\partial^3_{X_+}\beta(X_+,T). \end{aligned} \tag{2.9}$$

The additional terms S^1 and S^2 should satisfy

$$\begin{aligned}\partial_\tau S^1 + \partial_X S^1 =& J^1_{ct} + J^1_d + J^1_{sp},\\ \partial_\tau S^2 - \partial_X S^2 =& J^2_{ct} + J^2_d + J^2_{sp},\end{aligned} \tag{2.10}$$

where

$$\begin{aligned}J^1_{ct} = & -\partial_X U(X_-,\epsilon^2\tau)\left(G(X_+,\epsilon^2\tau) + \beta(X_+,\epsilon^2\tau)\right)\\ & - \partial_X U(X_-,\epsilon^2\tau)\left(\frac{1}{4}U^2(X_+,0) - \frac{1}{4}V^2(X_+,\epsilon^2\tau)\right)\\ & - \partial_X\left(V(X_+,\epsilon^2\tau)F(X_-,\epsilon^2\tau)\right) - \frac{1}{2}\partial_X^3\left(U(X_-,\epsilon^2\tau)V(X_+,\epsilon^2\tau)\right)\\ & - \partial_X\left(V(X_+,\epsilon^2\tau)\left(\frac{1}{4}V^2(X_-,0) + \alpha(X_-,\epsilon^2\tau) - \frac{1}{4}U^2(X_-,\epsilon^2\tau)\right)\right),\end{aligned}$$

$$J^1_d = -\partial_X\left(V(X_+,\epsilon^2\tau)G(X_+,\epsilon^2\tau)\right) - \frac{1}{8}\partial_X^3 V^2(X_+,\epsilon^2\tau) + \frac{1}{4}\partial_X V^3(X_+,\epsilon^2\tau),$$

$$\begin{aligned}J^1_{sp} = & -\frac{1}{2}\partial_X^3 A_1(X,\tau) - \partial_X\left((U(X_-,\epsilon^2\tau) + V(X_+,\epsilon^2\tau))(A_1(X,\tau) + B_1(X,\tau))\right)\\ & - \partial_X\left(V(X_+,\epsilon^2\tau)\left(\frac{1}{4}U^2(X_+,0) + \beta(X_+,\epsilon^2\tau)\right)\right),\end{aligned}$$

$$\begin{aligned}J^2_{ct} = & \partial_X V(X_+,\epsilon^2\tau)\left(F(X_-,\epsilon^2\tau) + \alpha(X_-,\epsilon^2\tau)\right)\\ & \partial_X(X_+,\epsilon^2\tau)\left(\frac{1}{4}V^2(X_-,0) - \frac{1}{4}U^2(X_-,\epsilon^2\tau)\right)\\ & + \partial_X\left(U(X_-,\epsilon^2\tau)G(X_+,\epsilon^2\tau)\right) + \frac{1}{2}\partial_X^3\left(U(X_-,\epsilon^2\tau)V(X_+,\epsilon^2\tau)\right)\\ & + \partial_X\left(U(X_-,\epsilon^2\tau)\left(\frac{1}{4}U^2(X_+,0) + \beta(X_+,\epsilon^2\tau) - \frac{1}{4}V^2(X_+,\epsilon^2\tau)\right)\right),\end{aligned}$$

$$J^2_d = \partial_X\left(U(X_-,\epsilon^2\tau)F(X_-,\epsilon^2\tau)\right) + \frac{1}{8}\partial_X^3 U^2(X_-,\epsilon^2\tau) - \frac{1}{4}\partial_X U^3(X_-,\epsilon^2\tau),$$

$$\begin{aligned}J^2_{sp} = & \frac{1}{2}\partial_X^3 B_1(X,\tau) + \partial_X\left(U(X_-,\epsilon^2\tau) + V(X_+,\epsilon^2\tau)(A_1(X,\tau) + B_1(X,\tau))\right)\\ & + \partial_X\left(U(X_-,\epsilon^2\tau)\left(\frac{1}{4}V^2(X_-,0) + \alpha(X_-,\epsilon^2\tau)\right)\right).\end{aligned}$$

The inhomogeneous terms are grouped according to the means by which we will control their contribution to the growth of S^1 and S^2 in the following section. The subscripts "ct," "d," and "sp" refer, respectively, to inhomogeneities which are cross-terms, perfect derivatives, or terms which require special consideration.

Equations (1.6) are our second set of modulation equations for the terms of $O(\epsilon^4)$ in our long wavelength approximation. Since they are linearized inhomogeneous KdV equations, linearized about a KdV solution, they are in principle explicitly solvable [18]. However, the form of the solution that results is quite complicated (see [19] and [11]), and thus it requires some effort to show that these solutions remain uniformly bounded in the norms which we use to bound the errors. As we noted above, the functions S^1 and S^2 do not actually form a part of the approximation at $O(\epsilon^4)$; however, we will show that they remain bounded over the time scales of interest as a part of controlling the error in our approximation.

3. Estimates on the solutions to the modulation equations. Before showing that the approximation is a good one, we must first show that the solutions to the modulation equations are tractable in their own right. Keeping in mind that our goal is to show that the approximation to (1.3) is good for a long time, we need to show that solutions to the modulation equations are bounded on the appropriate time scale, that is, for $t \sim O(\epsilon^{-3})$. First, we remark on Theorem 1.1.

Notice that, since $T = \epsilon^3 t$, this theorem states that we have bounded solutions of (1.4) for $t \in [0, T_0/\epsilon^3]$, as we had hoped. Moreover, since the solutions to (1.4) appear in the other modulation equations (often as inhomogeneities), that they are reasonably smooth and of rapid decay is crucial to showing that the other modulation equations are solvable over a long time and are of appropriate size. In particular, we will henceforth take U_0, $V_0 \in H^\sigma(4)$, where σ will be suitably large. We now state and prove a number of lemmas.

The first set of lemmas concerns the solutions to inhomogeneous transport equations with zero initial conditions. From the method of characteristics, we have explicit formulas for solutions.

Lemma 3.1. *Suppose*

$$\partial_\tau u \pm \partial_X u = f(X,\tau), \quad u(X,0) = 0,$$

with $\|f(X,\tau)\|_{H^s} \leq C$ *for* $\tau \in [0, T_0\epsilon^{-2}]$. *Then* $\|u(\cdot,\tau)\|_{H^s} \leq C\epsilon^{-2}$ *for* $\tau \in [0, T_0\epsilon^{-2}]$.

Proof. We have

$$u(X,\tau) = \int_0^\tau f(X \mp \tau \pm s, s)\, ds.$$

A naive estimate on the integral proves the result. ■

Lemma 3.2. *Suppose*

$$\partial_\tau u \pm \partial_X u = \partial_X f(X \pm \tau, \epsilon^2\tau), \quad u(X,0) = 0.$$

Then

$$u(X,\tau) = \pm\frac{1}{2}\left(f(X \pm \tau, \epsilon^2\tau) - f(X \mp \tau, 0)\right) \mp \frac{\epsilon^2}{2}\int_0^\tau \partial_T f(X \mp \tau \pm 2s, \epsilon^2 s)\, ds. \tag{3.1}$$

Also, if $\|f(\cdot,T)\|_{H^s} \leq C$ *and* $\|\partial_T f(\cdot,T)\|_{H^s} \leq C$ *for* $T \in [0,T_0]$, *then* $\|u(\cdot,\tau)\|_{H^s} \leq C$ *for* $\tau \in [0, T_0\epsilon^{-2}]$.

Proof. One can check this result explicitly. The estimate on the norm follows as in Lemma 3.1. ■

Lemma 3.3. *Suppose*

$$\partial_\tau u \pm \partial_X u = l(X+\tau, \epsilon^2 t) r(X-\tau, \epsilon^2\tau), \quad u(X,0)=0,$$

with $\|l(\cdot,T)\|_{H^s(4)} \leq C$ *and* $\|r(\cdot,T)\|_{H^s(4)} \leq C$ *for* $T \in [0,T_0]$*; then*

$$u(X,\tau) = \upsilon(X \mp \tau, \epsilon^2\tau)$$

with $\|\upsilon(\cdot,T)\|_{H^s(2)} \leq C$ *for* $T \in [0,T_0]$ *(that is, for* $\tau \in [0, T_0\epsilon^{-2}]$*). The constant* C *is uniform in* ϵ*.*

Proof. See the appendix. ■

Remark 4. If l and r are taken to be in $H^s(2)$, a similar proof shows that υ is in H^s over the long time scale.

Remark 5. Since the proof of the lemma does not make explicit use of the slow time scale dependence of the inhomogeneous factors l and r, the proof is still valid if the right-hand side is of the form $l(X+\tau)r(X-\tau,\epsilon^2\tau)$, $l(X+\tau,\epsilon^2\tau)r(X-\tau)$ or $l(X+\tau)r(X-\tau)$.

Remark 6. A general study of the growth of solutions of the transport equation and related linear equations that arise in the justification of modulation equations was recently completed by Lannes [15].

With these results, we can now prove the estimate for A and B in Proposition 1.2. That is, we have the following corollary.

Corollary 3.4. *If* U_0, V_0 *satisfy* (1.7) *with* $\sigma > 4$ *and* U*,* V*,* A*, and* B *satisfy* (1.4)–(1.5)*, then*

$$\sup_{\tau\in[0,T_0\epsilon^{-2}]} \{\|A(\cdot,\tau)\|_{H^{\sigma-3}}, \|B(\cdot,\tau)\|_{H^{\sigma-3}}\} \leq C.$$

Proof. From Corollary 2.1, we know the form of A and B. By Lemma 3.3, we have α and β uniformly bounded in $H^{\sigma-1}(2)$ over the long time scale. Also, by Lemma 3.2, we see that A_1 and B_1 are in the same space as $\partial_T U^2$ and $\partial_T V^2$. U and V satisfy the KdV equations (1.4), so we lose three space derivatives for the one time derivative here. That is, A_1 and B_1 are uniformly bounded in $H^{\sigma-3}$ for the long time scale. ■

We will occasionally be using an alternate, but equivalent, norm on $H^s(2)$. It is

$$|f|_{H^s(2)} = \sum_{j=0}^{s} \|(1+x^2)\partial_x^j f(x)\|_{L^2}.$$

The associated inner product is denoted by $\langle\cdot,\cdot\rangle_{H^s(2)}$.

Lemma 3.5. *For* $s > 3/2$*, if* $u \in H^s$*, then*

$$\langle u\partial_x f, f\rangle_{H^s} \leq C|u|_{H^s}|f|^2_{H^s}.$$

Proof. The proof is similar to and simpler than that of Lemma 3.6, which follows. ■

Lemma 3.6. *For* $s > 3/2$*, if* $u \in H^s(2)$*, then*

$$\langle u\partial_x f, f\rangle_{H^s(2)} \leq C|u|_{H^s(2)}|f|^2_{H^s(2)}.$$

Proof. See the appendix. ■

Lemma 3.7. *For $f \in H^s(2) \cap H^{s+4}$,*

$$(f, \partial_x^3 f)_{H^s(2)} \leq C(\|f\|^2_{H^s(2)} + \|f\|^2_{H^{s+4}}).$$

Proof. See the appendix. ■

We can now prove the estimates on F and G in Proposition 1.2. That is, we have the following lemma.

Lemma 3.8. *If U_0, V_0 satisfy (1.7) with $\sigma > 21/2$ and U, V, A, B, F, and G satisfy (1.4)–(1.6), then F and G satisfy the estimates*

$$\sup_{T \in [0,T_0]} \left\{ \|F(\cdot,T)\|_{\tilde{H}}, \|G(\cdot,T)\|_{\tilde{H}} \right\} \leq C,$$

where $\tilde{H} = H^{\sigma-5} \cap H^{\sigma-9}(2)$.

Proof. The proof follows from Lemmas 3.5–3.7 and Gronwall's inequality. We show the details for F. The case for G is entirely analogous. We take the definition of the inner product on $\tilde{H}$ to be $(\cdot,\cdot)_{\tilde{H}} = (\cdot,\cdot)_{H^{\sigma-9}(2)} + (\cdot,\cdot)_{H^{\sigma-5}}$.

The inhomogeneity J^1 is in $H^{\sigma-5}(2)$. (The term $\partial^5_{X\_} U$ causes the loss of derivatives.) So we take the inner product of (1.6) with F, apply Lemmas 3.5–3.7, and arrive at

$$\partial_T \|F\|^2_{\tilde{H}} \leq C(\|F\|_{\tilde{H}} + \|F\|^2_{\tilde{H}}) \leq C(1 + \|F\|^2_{\tilde{H}}).$$

An application of Gronwall's inequality yields, for $T \in [0, T_0]$,

$$\|F\|^2_{\tilde{H}}(T) \leq CTe^{CT},$$

which concludes the proof of the lemma and also of Proposition 1.2. ■

We now turn our eyes to the set of equations (2.10). As there are many terms driving these equations, many different techniques are used to show that the equations do not blow up over the long time scale. We are aided in this task by the above lemmas, though certain terms will need special consideration.

Lemma 3.9. *Suppose U, V, A, B, F, G, S^1, and S^2 satisfy (1.4)–(1.6) and (2.10); then S^1 and S^1 satisfy the estimates*

$$\sup_{\tau \in [0, T_0\epsilon^{-2}]} \left\{ \|S^1(\cdot,\tau)\|_{H^{\sigma-10}}, \|S^2(\cdot,\tau)\|_{H^{\sigma-10}} \right\} \leq C.$$

Proof. We shall treat the equation for S^1 here. The situation for S^2 is completely analogous. Since equations (2.10) are linear, we can consider the inhomogeneity term by term. First, we notice that we can apply Lemma 3.3 to bound the growth coming from all terms in J^1_{ct}, while Lemma 3.2 suffices to control all terms coming from J^1_d. Thus these terms cause no growth over the long time scale. We now take a moment to discuss the smoothness of these terms. The least smooth term in J^1_d is $\partial_X(V(X+\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$. When we apply Lemma 3.2, we need to examine the smoothness of $\partial_T(V(X+\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$. Now G is uniformly bounded in $H^{\sigma-5}$ (from Lemma 3.8), and, since G satisfies a linearized KdV equation, we have $\partial_T(V(X+\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$ uniformly bounded in $H^{\sigma-8}$. However, the least smooth term in J^1_{ct} is the term $\partial_X(U(X-\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$, which is uniformly bounded in $H^{\sigma-10}(2)$. Thus at best S^1 is uniformly bounded in $H^{\sigma-10}$.

Each term in J^1_{sp} will require some special consideration. These terms are

$$-\partial_X(V(X+\tau,\epsilon^2\tau)\beta(X+\tau,\epsilon^2\tau)), \tag{3.2}$$

$$-\partial_X(V(X+\tau,\epsilon^2\tau)U^2(X+\tau,0)), \tag{3.3}$$

$$C\partial_X^3 A_1(X,\tau), \tag{3.4}$$

$$C\partial_X\left((U(X-\tau,\epsilon^2\tau)+V(X+\tau,\epsilon^2\tau))(A_1(X,\tau)+B_1(X,\tau))\right). \tag{3.5}$$

Terms (3.2) and (3.3) are treated with slight variations on Lemmas 3.2 and 3.3. The technique by which (3.4) and (3.5) are dealt with relies primarily on the prefactor of ϵ^2, which appears in the definition of the functions A_1 and B_1. Unfortunately, each computation is rather messy.

In the case of the first of these, we apply Lemma 3.2 and get

$$\begin{aligned} S(X,\tau) = &-\frac{1}{2}(V(X+\tau,\epsilon^2\tau)\beta(X+\tau,\epsilon^2\tau)-V(X-\tau,0)\beta(X-\tau,0)) \\ &+\frac{\epsilon^2}{2}\int_0^\tau \partial_T V(X-\tau+2s,\epsilon^2 s)\beta(X-\tau+2s,\epsilon^2 s)ds \\ &+\frac{\epsilon^2}{2}\int_0^\tau V(X-\tau+2s,\epsilon^2 s)\partial_T\beta(X-\tau+2s,\epsilon^2 s)ds. \end{aligned}$$

The first three terms are easily bounded by the techniques discussed previously. (Namely, we replace $\partial_T V$ with the right-hand side of the KdV equation and use naive bounds.) However, when we replace $\partial_T\beta$, we lose the prefactor of ϵ^2. That is, from (7.1) in the proof of Lemma 3.3, we have

$$\partial_T\beta(X_+,T)=\epsilon^{-2}\partial_X(U(X_+-2T\epsilon^{-2},T)V(X_+,T)).$$

We make this substitution into the last term of (3) to get

$$\frac{1}{2}\int_0^\tau V(X-\tau+2s,\epsilon^2 s)\partial_X(U(X-\tau,\epsilon^2 s)V(X-\tau+2s,\epsilon^2 s))ds.$$

Notice that in this integral we have only terms that lie in the weighted Sobolev spaces, and we can use the same techniques used in the proof of Lemma 3.3 to control this term.

The term (3.3) is very nearly of the form needed to apply Lemma 3.2. The only difference is that there is no dependence on $\epsilon^2\tau$ in one of the terms. The ideas are essentially the same here as in the proof of Lemma 3.2. Consider

$$\partial_\tau S+\partial_X S=-\partial_X(V(X+\tau,\epsilon^2\tau)U^2(X+\tau,0)).$$

The solution to this equation is given by

$$\begin{aligned} S(X,\tau) = &-\frac{1}{2}\left\{U^2(X+\tau,0)V(X+\tau,\epsilon^2\tau)-U^2(X-\tau,0)V(X-\tau,\epsilon^2\tau)\right\} \\ &+\frac{\epsilon^2}{2}\int_0^\tau U^2(X-\tau+2s,0)\partial_T V(X-\tau+2s,\epsilon^2 s)ds. \end{aligned}$$

If one replaces $\partial_T V(X-\tau+2s,\epsilon^2\tau)$ in the integral by the right-hand side of the KdV equation and then takes naive norms, we find that this term is also controllable.

We now turn our attention to the final two terms which involve the functions A_1 and B_1. The calculations here are quite messy, though the ideas are straightforward. We replace $\partial_T V$ with the right-hand side of the KdV equation and then apply a number of the same techniques used in proving Lemmas 3.2 and 3.3. The factor of ϵ^2 present in the definitions of A_1 and B_1 is crucial. Consider

$$\begin{aligned}
&\partial_\tau S+\partial_X S\\
&\quad = C\partial_X^3 A_1(X,\tau)\\
&\quad = C\epsilon^2\partial_X^3\int_0^\tau \partial_T V^2(X-\tau+2s,\epsilon^2 s)ds\\
&\quad = C\epsilon^2\partial_X^2\int_0^\tau \left(\partial_s(\partial_T V^2(X-\tau+2s,\epsilon^2 s))-\epsilon^2\partial_T^2 V(X-\tau+2s,\epsilon^2 s)\right)ds\\
&\quad = C\epsilon^2\partial_X^2\left\{\partial_T V^2(X+\tau,\epsilon^2\tau)-\partial_T V^2(X-\tau,0)\right\}\\
&\qquad +\ C\epsilon^4\partial_X^2\int_0^\tau \partial_T^2 V^2(X-\tau+2s,\epsilon^2 s)ds.
\end{aligned}$$

For ease of notation, we will let $P(X,\tau)=C\epsilon^2\partial_X^2\int_0^\tau\partial_T^2 V^2(X-\tau+2s,\epsilon^2 s)ds$. Notice that, by taking naive estimates on this function, we have that $\|P\|_{H^s}\le C$ for $\tau\in[0,T_0\epsilon^{-2}]$. Thus we apply Lemma 3.1 to this equation to find that S is bounded on the long time interval.

In order to deal with (3.5), we will rewrite $\partial_T V^2$ and $\partial_T U^2$. That is,

$$\begin{aligned}
\partial_T V^2 &= 2V\partial_T V\\
&= V(\partial_X^3 V+\partial_X V^2)\\
&= \partial_X\left(V\partial_X^2 V-\frac{1}{2}(\partial_X V)^2+\frac{2}{3}V^3\right)\\
&= \partial_X\tilde{V},
\end{aligned}$$

where $\tilde{V}=V\partial_X^2 V-1/2(\partial_X V)^2+2/3V^3$. A similar calculation yields $\partial_T U^2=\partial_X\tilde{U}$, where $\tilde{U}=-U\partial_X^2 U+1/2(\partial_X U)^2-2/3U^3$. Notice that $\tilde{U},\tilde{V}\in H^{\sigma-2}$ for $T\in[0,T_0]$, since they lose at most two derivatives in comparison with U and V. Similarly, we have $\partial_T\tilde{U}\in H^{\sigma-5}$. So consider the equation

$$\begin{aligned}
&\partial_\tau S + \partial_X S \\
&\quad = C\partial_X \left(U(X-\tau,\epsilon^2\tau) + V(X+\tau,\epsilon^2\tau)\right)(A_1(X,\tau)+B_1(X,\tau)) \\
&\quad = C\epsilon^2\partial_X\Bigg[\left(U(X-\tau,\epsilon^2\tau) + V(X+\tau,\epsilon^2\tau)\right) \\
&\qquad\qquad \times \int_0^\tau \partial_T\left(V^2(X-\tau+2s,\epsilon^2 s) + U^2(X+\tau-2s,\epsilon^2 s)\right) ds\Bigg] \\
&\quad = C\epsilon^2\Bigg[\partial_X\left(U(X-\tau,\epsilon^2\tau) + V(X+\tau,\epsilon^2\tau)\right) \\
&\qquad\qquad \times \int_0^\tau \partial_X\big(\tilde{V}(X-\tau+2s,\epsilon^2 s) + \tilde{U}(X+\tau-2s,\epsilon^2 s)\big)ds\Bigg] \\
&\quad = C\epsilon^2\partial_X\Bigg[\left(U(X-\tau,\epsilon^2\tau) + V(X+\tau,\epsilon^2\tau)\right) \\
&\qquad\qquad \times \Bigg(\tilde{V}(X+\tau,\epsilon^2\tau) - \tilde{V}(X-\tau,0) - \epsilon^2\int_0^\tau \partial_T\tilde{V}(X-\tau+2s,\epsilon^2 s)ds \\
&\qquad\qquad + \tilde{U}(X-\tau,\epsilon^2\tau) - \tilde{U}(X+\tau,0) - \epsilon^2\int_0^\tau \partial_T\tilde{U}(X+\tau-2s,\epsilon^2 s)ds\Bigg)\Bigg]
\end{aligned}$$

Notice that, by taking naive estimates, the terms $Q^1(X,\tau) = \epsilon^2\int_0^\tau \partial_T\tilde{V}(X-\tau+2s,\epsilon^2 s)\ ds$ and $Q^2(X,\tau) = \epsilon^2\int_0^\tau \partial_T\tilde{U}(X+\tau-2s,\epsilon^2 s)\ ds$ are uniformly bounded in $H^{\sigma-5}$ over the long time scale. Thus we apply Lemma 3.1 to the above equation and find that this term is well behaved over the long time scale. ■

4. The validity of the approximation.

In this section, we prove that the approximation to a true solution of (1.3) made by the ansatz is in fact a good one by completing the proof of Theorem 1.3.

Proof of Theorem 1.3. To prove this theorem, we shall need a number of lemmas.

Lemma 4.1. *If $\Phi \in H^{s+1}$, then, for $\epsilon < 1$,*

$$\|\lambda\Phi(\epsilon\cdot)\|_{H^s} \le C\epsilon^{1/2}\|\Phi\|_{H^{s+1}}.$$

Proof. The proof here is analogous to the proof of the following lemma. ■

Lemma 4.2. *Let $T_1(y) = y$, $T_3(y) = y + 1/2y^3$, and $T_5(y) = y + 1/2y^3 + 3/8y^5$. Then, for $j = 1, 3, 5$, if $\Phi(X) \in H^{s+j+2}$, we have, for $\epsilon < 1$,*

$$\|\lambda\Phi(\epsilon\cdot) - T_j(\partial_x)\Phi(\epsilon\cdot)\|_{H^s} \le C\epsilon^{j+3/2}\|\Phi(\cdot)\|_{H^{s+j+2}}.$$

Proof. See the appendix. ■

Now suppose that there is a solution to (1.3) of the form

$$\bar{u}(x,t) = \epsilon^2\bar{\Psi}(x,t) + \epsilon^{11/2}\bar{R}(x,t), \tag{4.1}$$

where

$$\epsilon^2\bar{\Psi}(x,t) = \epsilon^2\bar{U} + \epsilon^4(\bar{A}+\bar{F}) + \epsilon^6\bar{S} \tag{4.2}$$

and $\bar{R} = (R^1(x,t), R^2(x,t))^t$. We consider the term $\bar{R}$ to be the error in our approximation. Substituting (4.1) into (1.3), we find that $\bar{R}$ must satisfy the equation

$$\begin{aligned}\partial_t \begin{pmatrix} R^1 \\ R^2 \end{pmatrix} = \begin{pmatrix} -\lambda & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} R^1 \\ R^2 \end{pmatrix} + \epsilon^2 \begin{pmatrix} -\lambda(\Psi^1+\Psi^2)(R^1+R^2) \\ \lambda(\Psi^1+\Psi^2)(R^1+R^2) \end{pmatrix} \\ + \frac{\epsilon^{11/2}}{2}\begin{pmatrix} -\lambda(R^1+R^2)^2 \\ \lambda(R^1+R^2)^2 \end{pmatrix} + \epsilon^{-11/2}\mathrm{Res}[\epsilon^2\bar{\Psi}],\end{aligned} \tag{4.3}$$

where

$$\mathrm{Res}[\epsilon^2\bar{\Psi}] = -\partial_t \begin{pmatrix} \epsilon^2\Psi^1 \\ \epsilon^2\Psi^2 \end{pmatrix} + \begin{pmatrix} -\lambda & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} \epsilon^2\Psi^1 \\ \epsilon^2\Psi^2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -\lambda(\epsilon^2\Psi^1+\epsilon^2\Psi^2)^2 \\ \lambda(\epsilon^2\Psi^1+\epsilon^2\Psi^2)^2 \end{pmatrix}. \tag{4.4}$$

We have selected our modulation equations precisely so that this term is small. By taking the time derivative of $\bar{\Psi}$ and then making substitutions from the modulation equations, we find that

$$\begin{aligned}&\mathrm{Res}[\epsilon^2\bar{\Psi}] \\ &\quad = \epsilon^2\begin{pmatrix} (T_5(\partial_x)-\lambda)U \\ -(T_5(\partial_x)-\lambda)V \end{pmatrix} \\ &\quad + \epsilon^4\begin{pmatrix} (T_3(\partial_x)-\lambda)\left(A+F+\frac{1}{2}(U+V)^2\right) \\ -(T_3(\partial_x)-\lambda)\left(B+G+\frac{1}{2}(U+V)^2\right) \end{pmatrix} \\ &\quad + \epsilon^6\begin{pmatrix} (T_1(\partial_x)-\lambda)\left((U+V)(A+F+B+G)+S^1\right) \\ -(T_1(\partial_x)-\lambda)\left((U+V)(A+F+B+G)+S^1\right) \end{pmatrix} \\ &\quad + \epsilon^8\begin{pmatrix} -\lambda\left(2(U+V)(S^1+S^2)+(A+F+B+G)^2\right) \\ \lambda\left(2(U+V)(S^1+S^2)+(A+F+B+G)^2\right) \end{pmatrix} \\ &\quad + 2\epsilon^{10}\begin{pmatrix} -\lambda\left((A+F+B+G)(S^1+S^2)\right) \\ \lambda\left((A+F+B+G)(S^1+S^2)\right) \end{pmatrix} \\ &\quad + \epsilon^{12}\begin{pmatrix} -\lambda\left((S^1+S^2)^2\right) \\ \lambda\left((S^1+S^2)^2\right) \end{pmatrix}.\end{aligned} \tag{4.5}$$

Note that we have suppressed the variables on which the functions depend for brevity. While the algebra that goes into showing this is lengthy, it should be noted that this step is accomplished by undoing the algebra that goes into deriving the modulation formally (see (2.3) and (2.4)).

Notice that, in the above expression, all functions are of long wavelength form. Thus we can apply Lemmas 4.1 and 4.2 to prove the following result.

Lemma 4.3. *Under the hypotheses of Theorem 1.3, the residual satisfies the estimate*

$$\sup_{t\in[0,T_0\epsilon^{-3}]} \|\mathrm{Res}[\epsilon^2\bar{\Psi}]\|_{H^{\sigma-13}\times H^{\sigma-13}} \leq C\epsilon^{17/2}.$$

Notice that the loss of three more derivatives is caused by the application of Lemma 4.2 to the term in the fourth line of (4.5) since $S^j,\ j = 1, 2$, are uniformly bounded in $H^{\sigma-10}$.

We also need the following fact.

Lemma 4.4.

$$\left(R^2 - R^1, \lambda[(\Psi^1 + \Psi^2)(R^1 + R^2)]\right)_{H^s} \leq -\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^s} + C\epsilon^3\|\bar{R}\|_{H^s \times H^s}.$$

Proof. See the appendix. ■

We wish to keep the norm of $\bar{R}$ from growing too much over the long time scale. That is, if we can show that $\|\bar{R}\|$ is $O(1)$ for $t \in [0, T_0\epsilon^{-3}]$, we will have shown that our approximation is good.

The first term on the right-hand side of (4.3) will not cause any growth in the norm since $(f, \lambda f)_{H^s} = 0$. The third term has the prefactor of $\epsilon^{11/2}$, which will assist in controlling it, and we know from Lemma 4.3 that the residual is small.

If we tried to control solutions of (4.3) by applying a Gronwall-type estimate to the time derivative of $(f, f)_{H^s}$, the second term would result in growth of the norm, which would destroy our estimate over the time scale of interest. To avoid this problem, we introduce a new energy functional which yields a norm equivalent to the $H^s \times H^s$ norm but which does not suffer from this sort of uncontrolled growth.

Thus we define

$$E_s^2(\bar{R}) = \frac{1}{2}\left(\|\bar{R}\|^2_{H^s \times H^s} + \epsilon^2(R^1 + R^2, (\Psi^1 + \Psi^2)(R^1 + R^2))_{H^s}\right). \tag{4.6}$$

That this norm is equivalent to the standard norm on $H^s \times H^s$ can be seen by applying the Cauchy–Schwarz inequality to the inner product, provided that we have $\epsilon^2\|\Psi^1 + \Psi^2\|_{H^s} < 1$. Thus we use without further comment

$$\frac{1}{C}\|\bar{R}\|_{H^s \times H^s} \leq E_s(\bar{R}) \leq C\|\bar{R}\|_{H^s \times H^s}.$$

We now state and prove a useful lemma.

Lemma 4.5. *Set $s > 0$. Suppose $f(x), g(x) \in H^s$ and $\gamma(X) \in H^{s+1}$, where $X = \epsilon x$. Then*

$$|(f(\cdot), \gamma(\epsilon\cdot)g(\cdot))_{H^s} - (g(\cdot), \gamma(\epsilon\cdot)f(\cdot))_{H^s}| \leq C\epsilon\|f\|_{H^s}\|g\|_{H^s}\|\gamma\|_{W^{s,\infty}}.$$

Proof. See the appendix. ■

We now have all of the tools needed to finish the proof of the theorem.

$$\begin{aligned}
&\partial_t E^2_{\sigma-13}(\bar{R}) \\
&\quad= \frac{1}{2}\partial_t \|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}} + \frac{\epsilon^2}{2}\partial_t \left(R^1+R^2, (\Psi^1+\Psi^2)(R^1+R^2)\right)_{H^{\sigma-13}} \\
&\quad= (R^1, \partial_t R^1)_{H^{\sigma-13}} + (R^2, \partial_t R^2)_{H^{\sigma-13}} \\
&\qquad+ \frac{\epsilon^2}{2}\left(\partial_t(R^1+R^2), (\Psi^1+\Psi^2)(R^1+R^2)\right)_{H^{\sigma-13}} \\
&\qquad+ \frac{\epsilon^2}{2}\left((R^1+R^2), (\Psi^1+\Psi^2)\partial_t(R^1+R^2)\right)_{H^{\sigma-13}} \\
&\qquad+ \frac{\epsilon^2}{2}\left((R^1+R^2), \partial_t(\Psi^1+\Psi^2)(R^1+R^2)\right)_{H^{\sigma-13}} \\
&\quad\le (R^1, \partial_t R^1)_{H^{\sigma-13}} + (R^2, \partial_t R^2)_{H^{\sigma-13}} \\
&\qquad+ \epsilon^2\left(\partial_t(R^1+R^2), (\Psi^1+\Psi^2)(R^1+R^2)\right)_{H^{\sigma-13}} \\
&\qquad+ C\epsilon^3\|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\qquad+ C\epsilon^3\|\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}}\|\partial_t\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\quad= \epsilon^2\left(R^2-R^1, \lambda[(\Psi^1+\Psi^2)(R^1+R^2)]\right)_{H^{\sigma-13}} \\
&\qquad+ \epsilon^2\left(\partial_t(R^1+R^2), (\Psi^1+\Psi^2)(R^1+R^2)\right)_{H^{\sigma-13}} \\
&\qquad+ \epsilon^{11/2}\left(R^2-R^1, \lambda[(R^1+R^2)^2]\right)_{H^{\sigma-13}} \\
&\qquad+ \epsilon^{-11/2}(\bar{R}, \mathrm{Res}[\bar{\Psi}])_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\qquad+ C\epsilon^3\|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\qquad+ C\epsilon^3\|\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}}\|\partial_t\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\quad\le + C\epsilon^{11/2}\|\bar{R}\|^3_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\qquad+ C\epsilon^3\|\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\qquad+ C\epsilon^3\|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}} \\
&\quad\le C\epsilon^3 E^2_{\sigma-13}(\bar{R}) + C\epsilon^3 E_{\sigma-13}(\bar{R}) + C\epsilon^{11/2}E^3_{\sigma-13}(\bar{R}).
\end{aligned}$$

We now state another lemma, which proves that the approximation is good over the long time interval.

Lemma 4.6. *Given $C > 0$, $T_0 > 0$, there exists $\epsilon_0 > 0$ such that, if $\epsilon \in (0, \epsilon_0)$ and*

$$\dot{\eta}(T) \le C(1 + \eta(T) + \epsilon^{5/2}\eta^{3/2}(T)), \quad \eta(0) = 0, \tag{4.7}$$

for $T \in [0, T_0]$, then $\eta(T) \le 2CT_0 e^{2CT_0}$ for $T \in [0, T_0]$.

Proof. See the appendix. ■

We apply Lemma 4.6 to the equation for the energy of the remainder and find that

$$E_{\sigma-13}(\bar{R}(\cdot, t)) \le C \tag{4.8}$$

for $t \in [0, T_0\epsilon^{-3}]$. Note that, given this a priori estimate, proving the existence and uniqueness of solutions of (4.3) is a standard exercise.

We now use the equivalence of $E_{\sigma-13}$ to the typical norm on $H^{\sigma-13}$, (4.8), and Lemma 4.1 to find, for $t \in [0, T_0\epsilon^{-3}]$,

$$\begin{aligned} \|\bar{u}(\cdot,t) - \bar{w}(\cdot,t)\|_{H^{\sigma-13}\times H^{\sigma-13}} &= \|\bar{u}(\cdot,t) - \epsilon^2\bar{\Psi}(\cdot,t) + \epsilon^6\bar{S}(\epsilon\cdot,\epsilon t)\|_{H^{\sigma-13}\times H^{\sigma-13}} \\ &= \|\epsilon^{11/2}\bar{R}(\cdot,t) + \epsilon^6\bar{S}(\epsilon\cdot,\epsilon t)\|_{H^{\sigma-13}\times H^{\sigma-13}} \\ &\leq C\epsilon^{11/2} \end{aligned} \tag{4.9}$$

This completes the proof. ■

5. Some numerics.

In this section, we show the results of some numerical simulations. We performed these numerics to gain insight into the qualitative nature of the higher order approximation, to estimate the values of the constants C_F and ϵ_0 that appear in Theorem 1.3, and to validate the results of said theorem.

We will choose the initial conditions of the system so that we can use the known solitary wave solutions to the KdV equation. We shall solve the Boussinesq equation (1.3) numerically. Though techniques are known for finding explicit solutions to the linearized KdV equation (see [11] and [18]), the resulting expressions are quite complicated, and so we also solve (1.6) numerically.

One may wonder why we should even bother computing higher order modulation equations if we have to solve them numerically. In our situation, numerically computing solutions to the Boussinesq equation is not particularly more complicated or time intensive than finding solutions to the linearized KdV equations. However, our goal is to apply these same ideas to derive corrections to the KdV approximation for the water wave problem, whose numerical solution is a much more difficult task. We expect the same modulation equations to hold in these more general and complicated systems. Thus, for the water wave problem, numerically solving the modulation equations should result in a great reduction in the complexity of the numerics.

The solutions of (1.1) and (1.6) are numerically computed using methods which are largely based around the pseudospectral techniques for Matlab used in [20]. Since our equations are relatively simple, Matlab, though slower than other languages (C or Fortran, for example), performs adequately rapidly. The techniques used are largely built around the use of the fast Fourier transform (FFT) to compute the various operators and derivatives and the use of an iterative technique to compute the nonlinear terms in (1.1) and the term $\partial_X(UF)$ in (1.6). It is implicit in the time step.

As noted previously, we use the known explicit solutions to (1.4). Where possible, we find explicit solutions for the various terms of A and B. The notable exception to this is in the computation of α and β, which we compute via routine trapezoidal rule techniques.

We first consider the head-on collision of two solitary waves. Note that the head-on collision will take place in the initial variable θ and not in either the u or v variables. This is because we have (formally) decomposed the system into left and right moving waves when we rewrite the system as (1.3). We therefore take initial conditions such that U and V will evolve as the well-known sech-squared solitary wave solutions to (1.4). That is, we take

$$\begin{aligned} u(x,0) &= 6\epsilon^2\text{sech}^2(\epsilon x - 10), \\ v(x,0) &= 6\epsilon^2\text{sech}^2(\epsilon x + 10) \end{aligned}$$

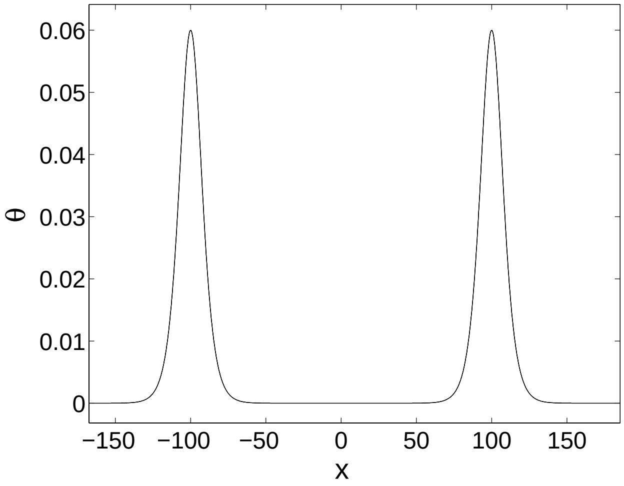


Figure 1. *Initial profile for head-on collision.* $\epsilon = 0.1$.

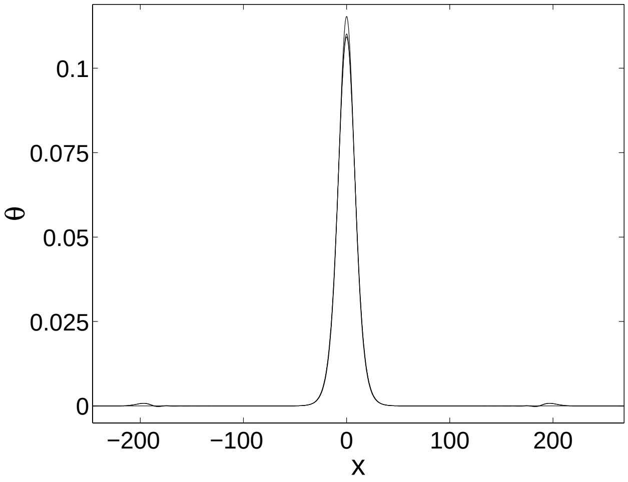


Figure 2. *Head-on collision.* $\epsilon = 0.1$.

as initial conditions.

Figures 1 and 2 show the solution to the Boussinesq equation, as well as the KdV approximation and the second order correction on the same plot, at the start and at the collision. Here $\epsilon = 0.1$. We remark on several features of the Boussinesq equation that are not reflected in the KdV approximation but are present in the second order correction.

First, in the KdV approximation, during the collision, the two waves add in linear superposition. (This can be seen as the KdV equations evolve independently.) However, the solutions to the Boussinesq equation do not display this simple linear property during the collision; the total height of the wave is slightly less than the sum of the two heights of the two waves independently. The second order correction does a notably better job at displaying this feature (see Figure 3). The second feature we notice is the presence of "shadow waves"

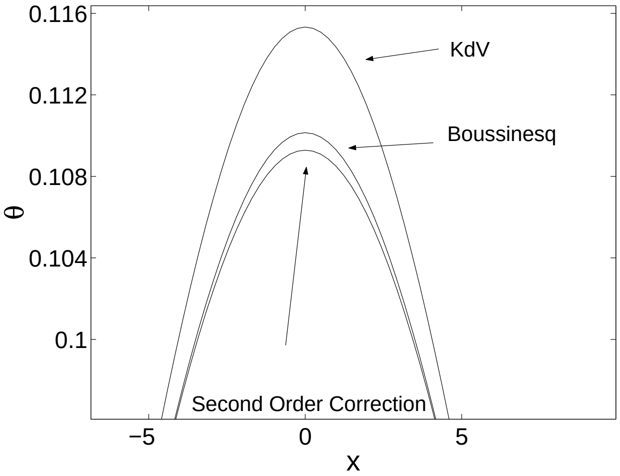


Figure 3. *Close-up of peak of waves during the head-on collision.* $\epsilon = 0.1$.

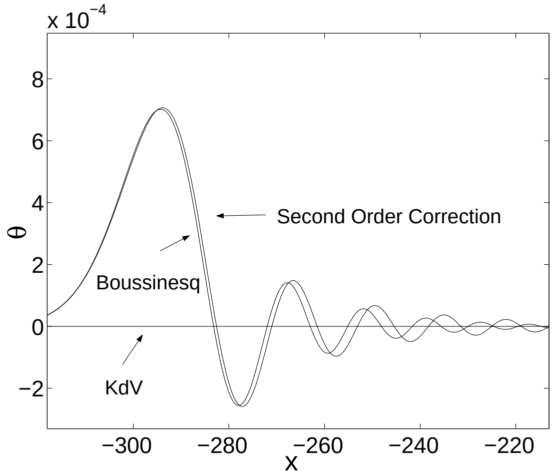


Figure 4. *A "shadow wave" and dispersive wave train in the head-on collision.* $\epsilon = 0.1$.

with dispersive wave trains (see Figure 4) in the solution to (1.1). These are not present in the KdV approximation but are seen in the second order correction.

From these pictures, we see that the second order correction is in fact doing a better job than simply the KdV approximation alone. In order to quantify this, we computed the solution for a variety of values of ϵ and computed the value of the L^2 and L^∞ error of the KdV and second order approximations. The time to collision is of $O(\epsilon^{-1})$, and on this time scale and slightly beyond, the maximum error occurs during the collision.

Figures 5 and 6 display log-log plots of the L^2 and L^∞ error versus ϵ, respectively. The slopes of these lines are the order of the correction. We expect the order of the correction in the L^∞ norm to be a half power greater than that in the L^2 norm due to the scaling of the spatial variable. We note that we have used only those values of $\epsilon \leq 0.1$ in computing these

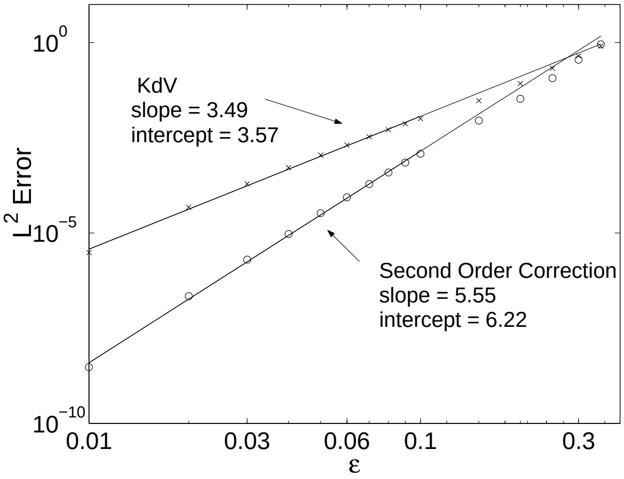


Figure 5. $\sup \|u - w\|_{L^2}$ *versus* ϵ *for head-on collision.*

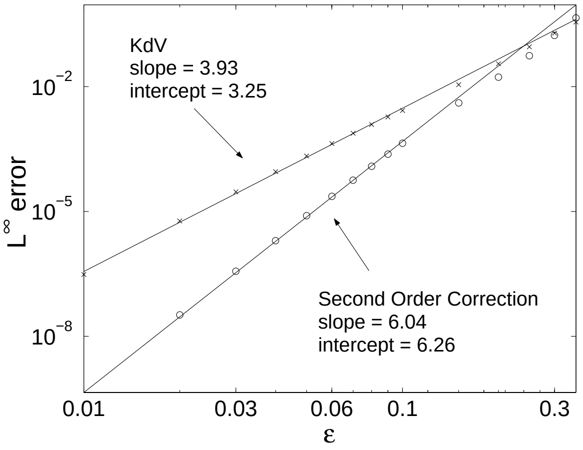


Figure 6. $\sup \|u - w\|_{L^\infty}$ *versus* ϵ *for head-on collision.*

slopes, as we expect the error estimates to hold if ϵ is sufficiently small. Table 1 summarizes the results.

From this we see that our estimate of the error made in approximating the true solution by the second order approximation is optimal, in terms of powers of ϵ. By taking note of the y-intercept of these lines, we can get an estimate of the value of the constant C_F in each case; see Table 2. Unfortunately, these values of the constant are quite large for the second order correction. We also note that it is not so much the actual value of the leading coefficient that matters, as it is the location (in ϵ) at which the second order correction and the KdV correction return the same error, that is, graphically, where the lines in Figures 5 and 6 cross.

The next simulation was that of right moving overtaking waves. We take initial data such that U will evolve as the famous two-soliton solution to (1.4). Since we are not interested in

Table 1

Order of the approximation, numerically computed, for the head-on collision.

	L^2	L^∞
KdV	3.49	3.93
KdV + second order correction	5.55	6.04

Table 2

Value of C_F, numerically computed, for the head-on collision.

	L^2	L^∞
KdV	35.5	25.8
KdV + second order correction	503	523

left moving waves, we take initial data for v to be zero. Note that v does not remain zero, however, due to the coupling.

Unlike the previous situation, the time scale of the overtaking wave collision is $O(\epsilon^{-3})$. To observe the entirety of the collision, we take $T_0 = 8$. We also observe that the error in the approximation is largest at the end of the interval $[0, T_0\epsilon^{-3}]$. From the proof of Lemma 4.6, one can see that, as T_0 increases, ϵ_0 decreases. This requires smaller values of ϵ, which in turn necessitates running the simulation for a longer period of time.

Figures 7, 8, and 9 display the values of u and the approximations at various times during the collision. As in the case of the head-on collision, the second order correction picks up the presence of a dispersive wave, which is not seen in the KdV approximation (see Figure 10).

It is well known that, in the two-soliton interaction, the waves are phase-shifted after the collision. (That is, the faster wave is further ahead after the collision than it would have been had no interaction taken place, and the slower wave falls behind in a similar fashion.) Overtaking waves in the Boussinesq equation share this feature, though with a different phase shift. This can be seen in Figure 9, where the KdV approximation is leading the Boussinesq solution. The second order correction noticeably "fixes" this problem. In Figure 11, we plot the locations of the peaks. Note that this figure reflects the fact that the numerics are computed in a moving reference frame (moving to the right with unit velocity).

In Figure 12, we plot the error in the phase shifts for the two approximations versus ϵ. Notice that the slope for the second order correction is steeper than that of the KdV approximation.

In Figures 13 and 14, we plot the maximum of the L^2 and L^∞ error for the two approximations versus ϵ on a log-log plot (as we did for the head-on interaction earlier). We summarize the results in Tables 3 and 4.

6. Conclusions.

We conclude by briefly surveying other work on the derivation of higher order modulation equations for water waves and related systems.

For the actual water wave equations, there have been a number of studies of corrections to the KdV approximation to water waves spanning the spectrum from nonrigorous asymptotic expansions [4], [5], [24], [2] to numerical solutions of the equations of motion and comparison with the KdV predictions [3],[25], [9] to experimental investigations [16], [7]. We concentrate here on the theoretical studies since they have the closest connection to our work. In the

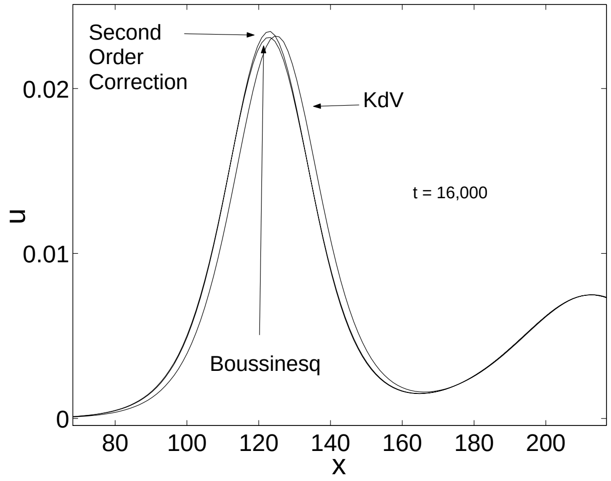


Figure 7. *Overtaking wave before the collision.* $\epsilon = 0.05$.

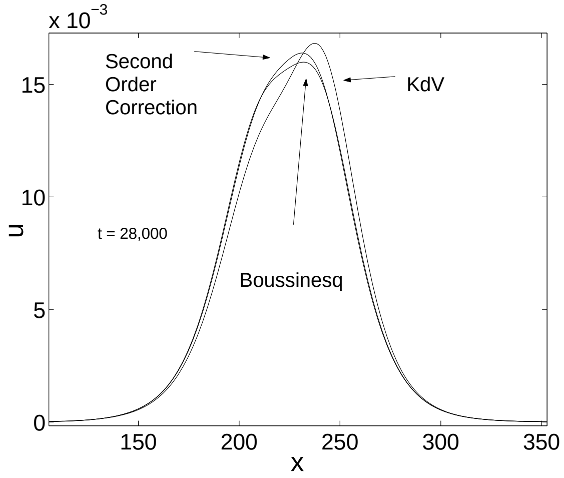


Figure 8. *Overtaking wave at the collision.* $\epsilon = 0.05$.

investigations of Byatt-Smith [4], [5] and Su and Mirie [24] the focus is on the head-on collision of solitary waves. This has several consequences. First, the authors assume that the initial conditions are of a special form, namely, a pair of counterpropagating solitary waves. The higher order corrections to the solution then exploit this special form by including not only a correction to the amplitude of the solution but a phase shift for each wave as it undergoes the collision. This is a very reasonable hypothesis in these physical circumstances but one which cannot easily be adapted to the more general type of initial conditions considered in our work. Furthermore, since these papers consider specifically the head-on collision of solitary waves, they are concerned with events which occur on relatively short time scales (i.e., time scales of $O(\frac{1}{\epsilon})$ in our scaling). As noted in [5, p. 503], these expansions are not uniformly valid in time, and it is not clear whether or not their solutions could be controlled over time scales of $O(\frac{1}{\epsilon^3})$.

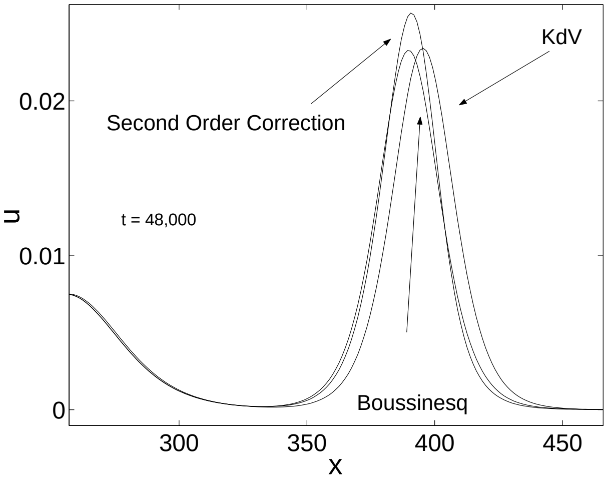


Figure 9. *Overtaking wave after the collision.* $\epsilon = 0.05$.

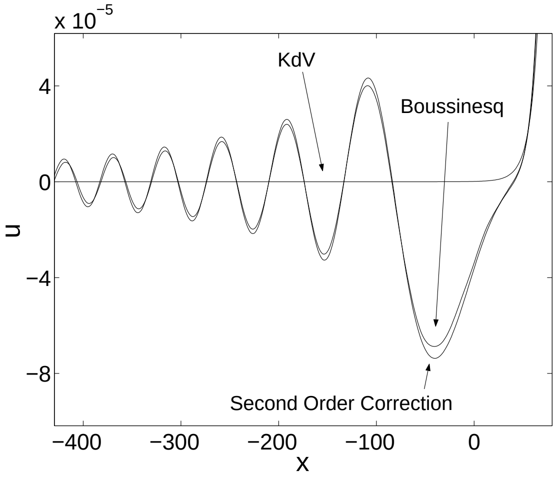


Figure 10. *Dispersive wave for the overtaking wave.*

It is worth noting that, in spite of the differences between our approach and those discussed here, Byatt-Smith [5] also finds that corrections to the amplitude of the solitary wave evolve according to the linearized KdV equation.

An alternative approach to improve the KdV approximation to water waves is to work directly with a Boussinesq approximation to the water wave problem, as done by Bona and Chen in [2]. Note that, in such an approach, the modulation equations depend on the small parameter ϵ, whereas our goal was derive a hierarchy of modulation equations which are independent of ϵ. Thus the work in the present paper is quite different than that in [2].

Another set of papers by Sachs [19], Zho and Su [26], and Hărăguş-Courcelle, Nicholls, and Sattinger [10] considers corrections to the KdV approximation for unidirectional motion. The first two of these papers study this question in the context of water waves, while [10]

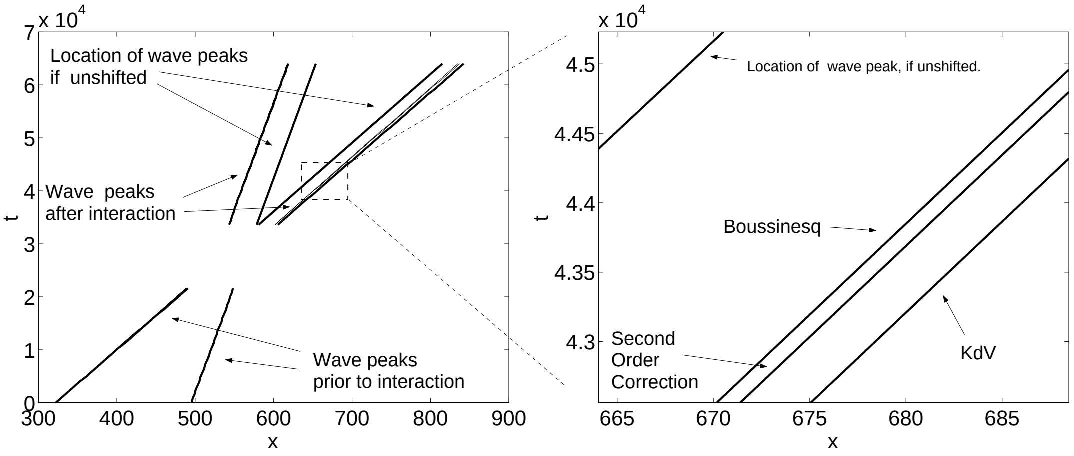


Figure 11. *Wave peak locations.* $\epsilon = 0.05$.

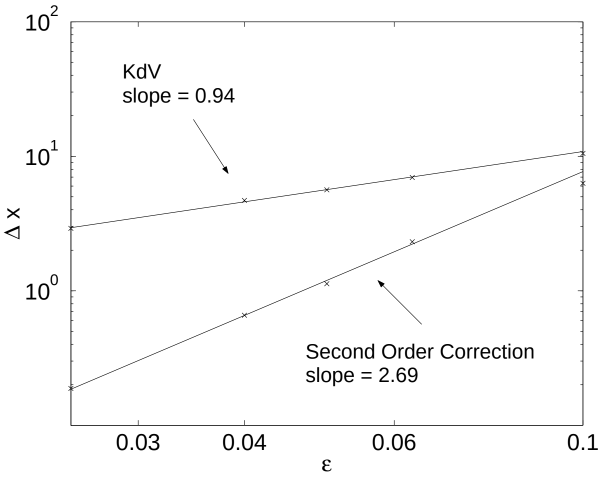


Figure 12. *Error in phase shift versus* ϵ.

studies the KdV approximation to solutions of the Euler–Poisson equations. The focus of these papers (particularly [19] and [10]) is rather different than ours, however. Both derive an inhomogeneous linearized KdV equation for the correction to the KdV approximation. However, rather than deriving rigorous estimates of the difference between the approximate solutions provided by the model equations and the true solutions, they focus on the nature of the solutions of the linearized inhomogeneous KdV equation. In particular, Sachs [19] shows that, if one linearizes about the N-soliton solution of the KdV equation, the resulting inhomogeneous equation has solutions which have no secular growth. In [10], the authors obtain explicit solutions of the linearized KdV equation, particularly for the case in which one linearizes about the two-soliton solution of the KdV equation. Hărăguş-Courcelle, Nicholls, and Sattinger then compare the approximation they obtain to numerically computed solutions

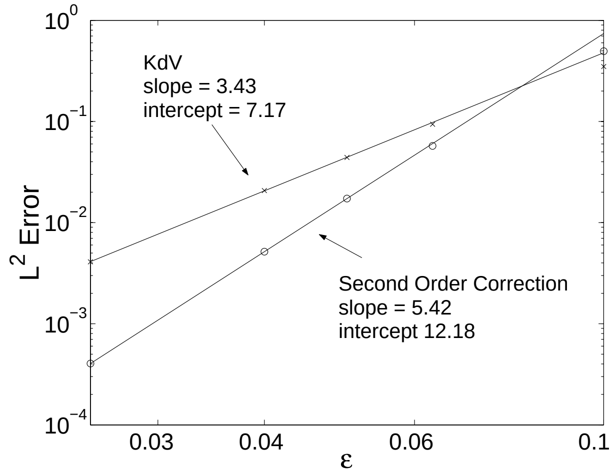


Figure 13. $\sup \|u - w\|_{L^2}$ *versus* ϵ *for overtaking wave collision.*

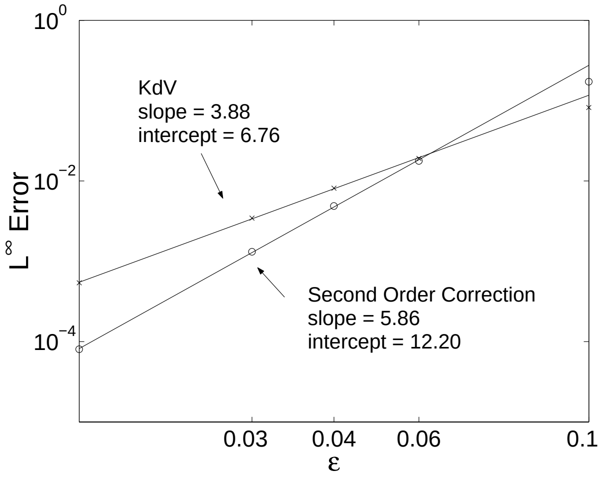


Figure 14. $\max \|u - w\|_{L^\infty}$ *versus* ϵ *for overtaking wave collision.*

of the Euler–Poisson equation, and they find that the addition of the solution of the linearized KdV equation to the approximation given by the two-soliton solution of the KdV equation does a significantly better job of approximating the solution of the Euler–Poisson equation. In particular, they note that the prediction of the phase shift that occurs when a “fast” traveling wave overtakes a slower one is significantly better when the second order correction is included. This effect is also present in our approximation—see Figure 12.

7. Appendix.

Proof of Lemma 3.3. We consider the case with the “minus” sign on the left-hand side for simplicity. The other case is analogous. First, we change variables to $X_+ = X + \tau$, $T = \epsilon^2\tau$,

Table 3

Order of the approximation, numerically computed, for the overtaking collision.

	L^2	L^∞
KdV	3.43	3.88
KdV + second order correction	5.42	5.86

Table 4

Value of C_F, numerically computed, for the overtaking collision.

	L^2	L^∞
KdV	1300	860
KdV + second order correction	19,500	19,900

and $\upsilon(X_+,T)=u(X,\tau)$. Under this change, we get the equation

$$\partial_T \upsilon(X_+,T)=\epsilon^{-2} l(X_+,T) r(X_+-2T\epsilon^{-2},T). \tag{7.1}$$

This can be solved by integrating with respect to the variable T. We get

$$\upsilon(X_+,T)=\epsilon^{-2}\int_0^T l(X_+,s) r(X_+-2s\epsilon^{-2}2,s)ds.$$

Now we multiply by the appropriate weight and take norms

$$\begin{aligned}
&(1+X_+^2)|\upsilon(X_+,T)| \\
&\quad\leq \epsilon^{-2}\int_0^T (1+X_+^2)|l(X_+,s)||r(X_+-2s\epsilon^{-2},s)|\ ds \\
&\quad\leq \epsilon^{-2}\int_0^T (1+X_+^2)|l(X_+,s)|(1+(X_+-2s\epsilon^{-2})^2)|r(X_+-2s\epsilon^{-2},s)|\ ds \\
&\quad= \epsilon^{-2}\int_0^T \frac{(1+X_+^2)^2|l(X_+,s)|(1+(X_+-2s\epsilon^{-2})^2)^2|r(X_+-2s\epsilon^{-2},s)|}{(1+X_+^2)(1+(X_+-2s\epsilon^{-2})^2)}\ ds \\
&\quad\leq \epsilon^{-2}\int_0^T \frac{(1+X_+^2)^2|l(X_+,s)|(1+(X_+-2s\epsilon^{-2})^2)|r(X_+-2s\epsilon^{-2},s)|}{(1+(2s\epsilon^{-2})^2)}\ ds.
\end{aligned}$$

Now take the H^s norm of each side of this equation, and find that

$$\begin{aligned}\|v(\cdot,T)\|_{H^s(2)} &\leq \|l\|_{H^s(4)}\|r\|_{H^s(4)}\epsilon^{-2}\int_0^T \frac{1}{(1+(2s\epsilon^{-2})^2)}ds\\ &\leq C\|l\|_{H^s(4)}\|r\|_{H^s(4)}\arctan(2T\epsilon^{-2})\\ &\leq C\|l\|_{H^s(4)}\|r\|_{H^s(4)}. \qquad \blacksquare\end{aligned}$$

Proof of Lemma 3.6. In this proof, we use the alternate inner product of $H^s(2)$.

$$\begin{aligned}\langle u\partial_x f, f\rangle_{H^s(2)} &= \langle u\partial_x f, f\rangle_{H^{s-1}(2)} + ((1+x^2)\partial_x^s(u\partial_x f),(1+x^2)\partial_x^s f)_{L^2}\\ &\leq |u\partial_x f|_{H^{s-1}(2)}|f|_{H^{s-1}(2)} + \sum_{j=0}^{s-1} c_{sj}((1+x^2)^2\partial_x^{s-j}u\partial_x^{j+1}f,\partial_x^s f)_{L^2}\\ &\qquad + \int (1+x^2)^2 u\partial_x^{s+1}f\partial_x^s f dx\\ &\leq C|u|_{H^s(2)}|f|^2_{H^s(2)} + \frac{1}{2}\int (1+x^2)^2 u\partial_x(\partial_x^s f)^2 dx\\ &\leq C|u|_{H^s(2)}|f|^2_{H^s(2)} - \frac{1}{2}\int \partial_x((1+x^2)^2 u)(\partial_x^s f)^2 dx\\ &\leq C|u|_{H^s(2)}|f|^2_{H^s(2)} - \frac{1}{2}\int \partial_x u((1+x^2)\partial_x^s f)^2 dx\\ &\qquad - 2\int xu(1+x^2)(\partial_x^s f)^2 dx\\ &\leq C|u|_{H^s(2)}|f|^2_{H^s(2)}. \qquad \blacksquare\end{aligned}$$

Proof of Lemma 3.7. In this proof, we use the standard norm on $H^s(2)$.

$$\begin{aligned}(f,\partial_x^3 f)_{H^s(2)} &= -6(x^2 f,\partial_x f)_{H^s} - 6((1+x^2)f, x\partial_x^2 f)_{H^s}\\ &\leq C(\|f\|_{H^s(2)}\|f\|_{H^{s+1}}) + C(\|f\|_{H^s(2)}\|x\partial_x^2 f\|_{H^s}).\end{aligned}$$

So now consider

$$\|x\partial_x^2 f\|^2_{H^s} = \|x\partial_x^2 f\|^2_{H^{s-2}} + \|\partial_x^{s-1}(x\partial_x^2 f)\|^2_{L^2} + \|\partial_x^s(x\partial_x^2 f)\|^2_{L^2}.$$

We now treat the last term in the above, as the middle term can be handled in a similar fashion and the first is easily dealt with.

$$\begin{aligned}\|\partial_x^s(x\partial_x^2 f)\|^2_{L^2} &\leq C\|\partial_x^{s+1}f\|^2_{L^2} + \|x\partial_x^{s+2}f\|_{L^2}\\ &\leq C\|f\|^2_{H^{s+4}} + \int x^2\partial_x^{s+2}f\partial_x^{s+2}f dx\\ &\leq C\|f\|^2_{H^{s+4}} + \int x^2\partial_x^{s+4}f\partial_x^s f dx + 4\int x\partial_x^{s+3}f\partial_x^s f dx\\ &\qquad + 2\int \partial_x^{s+2}f\partial_x^s f dx\\ &\leq C(\|f\|^2_{H^{s+4}} + \|f\|_{H^{s+4}}\|f\|_{H^s(2)}).\end{aligned}$$

This estimate completes the proof. ■

Proof of Lemma 4.2. Notice that the polynomials T_j are the first, third, and fifth order polynomial expansions of $y/\sqrt{1-y^2}$ about $y=0$. Moreover, note that only odd powers appear in the expansion. So, by Taylor's theorem, there is a constant C such that $|iy/\sqrt{1+y^2}-T_j(iy)| \leq C|y|^{j+2}$.

We shall now use the Fourier transform version of the Sobolev norms in the following computation, which concludes the proof. Consider

$$\begin{aligned}
&\|\lambda\Phi(\epsilon\cdot) - T_j(\partial_x)\Phi(\epsilon\cdot)\|^2_{H^s}\\
&\quad= \int (1+k^2)^s|\widehat{\lambda\Phi(\epsilon x)} - T_j\widehat{(\partial_x)\Phi(\epsilon x)}|^2 dk\\
&\quad= \epsilon^{-2}\int (1+k^2)^s\left|\left(\frac{ik}{\sqrt{1+k^2}} - T_j(ik)\right)\hat{\Phi}(k/\epsilon)\right|^2 dk\\
&\quad\leq C\epsilon^{-2}\int (1+k^2)^s|k^{j+2}\hat{\Phi}(k/\epsilon)|^2 dk, \quad K=k/\epsilon,\\
&\quad= C\epsilon^{2j+3}\int (1+(\epsilon K)^2)^s|K^{j+2}\hat{\Phi}(K)|^2 dK\\
&\quad\leq C\epsilon^{2j+3}\int (1+K^2)^s|K^{j+2}\hat{\Phi}(K)|^2 dK\\
&\quad\leq C\epsilon^{2j+3}\|\partial_X^{j+2}\Phi\|^2_{H^s}\\
&\quad\leq C\epsilon^{2j+3}\|\Phi\|^2_{H^{s+j+2}}. \qquad \blacksquare
\end{aligned}$$

Proof of Lemma 4.4.

$$\begin{aligned}
&\big(R^2-R^1, \lambda[(\Psi^1+\Psi^2)(R^1+R^2)]\big)_{H^s}\\
&\quad= \big(-\lambda(R^2-R^1), (\Psi^1+\Psi^2)(R^1+R^2)\big)_{H^s}\\
&\quad\leq -\big(\partial_t(R^1+R^2), (\Psi^1+\Psi^2)(R^1+R^2)\big)_{H^s}\\
&\quad+ \epsilon^{-11/2}\big(\mathrm{Res}[\bar{\Psi}]^1 + \mathrm{Res}[\bar{\Psi}]^2, (\Psi^1+\Psi^2)(R^1+R^2)\big)_{H^s}\\
&\quad\leq -\big(\partial_t(R^1+R^2), (\Psi^1+\Psi^2)(R^1+R^2)\big)_{H^s} + C\epsilon^3\|\bar{R}\|_{H^s\times H^s}. \qquad \blacksquare
\end{aligned}$$

Proof for Lemma 4.5. We shall be using the fact that, by the Sobolev embedding theorem, we have γ and its first s derivatives in L^∞.

$$\begin{aligned}
&|(f(x), \gamma(\epsilon x)g(x))_{H^s} - (g(x), \gamma(\epsilon x)f(x))_{H^s}|\\
&\quad= \left|\sum_{j=0}^{s}\big(\partial_x^j f(x), \partial_x^j(\gamma(\epsilon x)g(x))\big)_{L^2} - \big(\partial_x^j g(x), \partial_x^j(\gamma(\epsilon x)f(x))\big)_{L^2}\right|\\
&\quad= \left|\sum_{j=0}^{s}\sum_{l=0}^{j} c_{jl}\big\{\big(\partial_x^j f(x), \partial_x^l(\gamma(\epsilon x))\partial_x^{j-l}g(x)\big)_{L^2} - \big(\partial_x^j g(x), \partial_x^l(\gamma(\epsilon x))\partial_x^{j-l}f(x)\big)_{L^2}\big\}\right|
\end{aligned}$$

$$\begin{aligned}
&= \left| \sum_{j=0}^{s} \sum_{l=0}^{j} c_{jl} \big(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x), \partial_x^l (\gamma(\epsilon x)) \big)_{L^2} \right| \\
&= \epsilon \left| \sum_{j=0}^{s} \sum_{l=1}^{j} c_{jl} \big(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x), \epsilon^{l-1} \partial_X^l \gamma(\epsilon x) \big)_{L^2} \right| \\
&= \epsilon \left| \sum_{j=0}^{s} \sum_{l=1}^{j} c_{jl} \int \big(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x) \big) \big(\epsilon^{l-1} \partial_X^l \gamma(\epsilon x) \big) dx \right| \\
&\le \epsilon \|\gamma\|_{W^{s,\infty}} \sum_{j=0}^{s} \sum_{l=1}^{j} c_{jl} \int |\big(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x) \big)| dx \\
&\le C\epsilon \|\gamma\|_{W^{s,\infty}} \|f\|_{H^s} \|g\|_{H^s}. \qquad \blacksquare
\end{aligned}$$

Proof of Lemma 4.6. Functions which obey the inequality are bounded above by solutions to the family of ODEs

$$\dot{\eta}(T;\epsilon) = C(1 + \eta(T;\epsilon) + \epsilon^{5/2} \eta^{3/2}(T;\epsilon)), \quad \eta(0;\epsilon) = 0,$$

and so we prove the result for these equations.

By separation of variables, we have that $\eta(T;0) = e^{CT} - 1$. We notice that, for fixed T, $\eta(T;\epsilon)$ is a continuous and increasing function of ϵ. This follows since solutions of ODEs depend smoothly on their parameters and the right-hand side of the differential equations is increasing in ϵ.

Thus, by the intermediate value theorem, there exists ϵ_0 such that $\eta(T_0;\epsilon_0) = \epsilon_0^{-5}$. Moreover, since ϵ^{-5} is a decreasing function for $\epsilon > 0$, we have $\eta(T_0;\epsilon) \le \epsilon^{-5}$ for $\epsilon \in (0, \epsilon_0)$. We further note that, for fixed ϵ, η is continuous and increasing in T. So we have $\eta(T;\epsilon) \le \epsilon^{-5}$ for $T \in [0, T_0]$ and $\epsilon \in (0, \epsilon_0)$.

Thus we have

$$\dot{\eta}(T;\epsilon) \le C(1 + 2\eta(T;\epsilon)), \quad \eta(0;\epsilon) = 0,$$

for $T \in [0, T_0]$ and $\epsilon \in (0, \epsilon_0)$. We apply Gronwall's inequality to this to prove the result. ■

Acknowledgments. The authors thank D. Nicholls and D. Sattinger for several discussions of their work [10] prior to publication. The second author also thanks G. Schneider for numerous discussions of the derivation and justification of modulation equations and H. Segur for discussions of higher order approximations to water waves. They would also like to thank the referees for their helpful comments.

REFERENCES

[1] W. Ben Youssef and T. Colin, *Rigorous derivation of Korteweg-de Vries-type systems from a general class of nonlinear hyperbolic systems*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 873–911.

[2] J. L. Bona and M. Chen, *A Boussinesq system for two-way propagation of nonlinear dispersive waves*, Phys. D, 116 (1998), pp. 191–224.

[3] J. L. Bona, W. G. Pritchard, and L. R. Scott, *An evaluation of a model equation for water waves*, Philos. Trans. Roy. Soc. London Ser. A, 302 (1981), pp. 457–510.

[4] J. G. B. Byatt-Smith, *An integral equation for unsteady surface waves and a comment on the Boussinesq equation*, J. Fluid Mech., 49 (1971), pp. 625–633.

[5] J. G. B. Byatt-Smith, *The reflection of a solitary wave by a vertical wall*, J. Fluid Mech., 197 (1988), pp. 503–521.

[6] J. G. B. Byatt-Smith, *The head-on interaction of two solitary waves of unequal amplitude*, J. Fluid Mech., 205 (1989), pp. 573–579.

[7] M. J. Cooker, P. D. Weidman, and D. S. Bale, *Reflection of a high-amplitude solitary wave at a vertical wall*, J. Fluid Mech., 342 (1997), pp. 141–158.

[8] W. Craig, *An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits*, Comm. Partial Differential Equations, 10 (1985), pp. 787–1003.

[9] J. D. Fenton and M. M. Rienecker, *A Fourier method for solving nonlinear water-wave problems: Application to solitary-wave interactions*, J. Fluid Mech., 118 (1982), pp. 411–443.

[10] M. Hărăguş-Courcelle, D. P. Nicholls, and D. H. Sattinger, *Solitary wave interactions of the Euler–Poisson equations*, SIAM J. Appl. Math., submitted; also available online from http://www.math.usu.edu/~dhs (2002).

[11] M. Hărăguş-Courcelle and D. H. Sattinger, *Inversion of the linearized Korteweg-de Vries equation at the multi-soliton solutions*, Z. Angew. Math. Phys., 49 (1998), pp. 436–469.

[12] L. A. Kalyakin, *Long-wave asymptotics. Integrable equations as the asymptotic limit of nonlinear systems*, Uspekhi Mat. Nauk, 44 (1989), pp. 5–34, 247.

[13] T. Kano and T. Nishida, *A mathematical justification for Korteweg-de Vries equation and Boussinesq equation of water surface waves*, Osaka J. Math., 23 (1986), pp. 389–413.

[14] P. Kirrmann, G. Schneider, and A. Mielke, *The validity of modulation equations for extended systems with cubic nonlinearities*, Proc. Roy. Soc. Edinburgh Sect. A, 122 (1992), pp. 85–91.

[15] D. Lannes, *Secular Growth Estimates for Hyperbolic Systems*, preprint, Université de Bordeaux I, Talanec, France, 2002.

[16] T. Maxworthy, *Experiments on collisions between solitary waves*, J. Fluid Mech., 76 (1976), pp. 177–185.

[17] J. W. Miles, *Obliquely interacting solitary waves*, J. Fluid Mech., 79 (1977), pp. 157–169.

[18] R. L. Sachs, *Completeness of derivatives of squared Schrödinger eigenfunctions and explicit solutions of the linearized KdV equation*, SIAM J. Math. Anal., 14 (1983), pp. 674–683.

[19] R. L. Sachs, *A justification of the KdV approximation to first order in the case of N-soliton water waves in a canal*, SIAM J. Math. Anal., 15 (1984), pp. 468–489.

[20] D. H. Sattinger and Y. Li, *Matlab Codes for Nonlinear Dispersive Wave Equations*, http://www.math.usu.edu/~dhs (1998).

[21] G. Schneider, *The long wave limit for a Boussinesq equation*, SIAM J. Appl. Math., 58 (1998), pp. 1237–1245.

[22] G. Schneider and C. E. Wayne, *The long-wave limit for the water wave problem.* I. *The case of zero surface tension*, Comm. Pure Appl. Math., 53 (2000), pp. 1475–1535.

[23] G. Schneider and C. E. Wayne, *The rigorous approximation of long-wavelength capillary-gravity waves*, Arch. Ration. Mech. Anal., 162 (2002), pp. 247–285.

[24] C. H. Su and R. M. Mirie, *On head-on collisions between solitary waves*, J. Fluid Mech., 98 (1980), pp. 509–525.

[25] C. H. Su and R. M. Mirie, *Collisions between two solitary waves.* II. *A numerical study*, J. Fluid Mech., 115 (1982), pp. 475–492.

[26] Q. Zho and C.-H. Su, *Overtaking collision between two solitary waves*, Phys. Fluids, 29 (1986), pp. 2113–2123.