J. Fluid Mech. (2010), *vol.* 662, *pp.* 36–65.
doi:10.1017/S0022112010002879

Empirical pseudo-balanced model reduction and feedback control of weakly nonlinear convection patterns

A. C. OR† AND J. L. SPEYER

Department of Aerospace and Mechanical Engineering, University of California, Los Angeles,
CA 90095, USA

(Received 13 April 2009; revised 16 May 2010; accepted 17 May 2010;

An empirical model reduction method is performed on nonlinear transient convection patterns near the threshold, using a pseudo-inverse-based projective method called the pseudo-balanced proper orthogonal decomposition (PBPOD). These transient patterns are large-scale amplitude/phase modulations in convection rolls, obtained by prescribing selected spatial input-shape functions. For the nonlinear convection patterns modelled, PBPOD appears to be very effective. Using the nonlinear front example, PBPOD is compared with other existing methods, such as POD and linearized BPOD. The limitations of the methods are discussed. Using a complex prescribed input, complex disturbances are generated and the outputs of the open-loop responses are compared between the original and the reduced-order models. The agreement is good. A feedback-control study is performed using the nonlinear front example. The controller is built by the pseudo-balanced reduced-order model, with a low-order nonlinear estimator. Closed-loop simulations show that the nonlinear travelling fronts can be effectively damped out by the feedback-control actions.

Key words: Bénard convection, low-dimensional models, pattern formation

1. Introduction

There are a number of studies in using feedback control to suppress Rayleigh–Bénard convection in an unstable, static layer of fluid (Howle 1997; Tang & Bau 1998; Or, Cortelezzi & Speyer 2001; Or & Speyer 2003, 2005). However, there exists a significant gap in the degree of stabilization between theoretical predictions and experimental results. Theoretical models assume a spatial resolution refined to the cellular level. Convection disturbances are assumed to be homogeneous in the layer. In experiments, however, disturbances can be associated with large-scale phase and amplitude modulation patterns (see Cross & Hohenberg 1993).

In practice, controller arrays targeting the cellular level of resolution are too large for experimental implementation. Model order reduction of convection patterns in the spanwise direction is of interest because it allows a faster and more compact controller array to be built. In previous studies, model reduction was performed only in the cross-plane direction. Therefore, an effective model reduction method in the spanwise direction is desirable to have.

---

† Email address for correspondence: or@ucla.edu

Near the stability threshold, the large-scale modulation patterns are governed by a nonlinear envelope equation (Cross & Hohenberg 1993). In the model, the solutions do not depend on the cross-plane coordinate. An input/output version of this equation is derived in the Appendix. The goal here is twofold. First, it is important to demonstrate that, for the nonlinear large-scale dynamics, a low-order model can be constructed to approximate the original full-order model for the given input and output effectively. The low-order model is obtained from an empirical, nonlinear-based pseudo-inverse projective model reduction technique called pseudo-balanced proper orthogonal decomposition (PBPOD), in contrast to the linear-based balanced proper orthogonal decomposition (BPOD). For the development of BPOD, we refer to Lall, Marsden & Glavaski (2002) and Rowley (2005). In this study, the effectiveness of PBPOD is demonstrated. Several representative patterns are generated from both the original and reduced-order models and compared. In particular, the nonlinear front pattern is used for a comparison between the PBPOD and other methods, namely, proper orthogonal decomposition (POD) and BPOD with balanced modes extracted from a linearized model, or the linearized BPOD. The limitations of the methods involved are discussed. The open-loop performance is first demonstrated. Using a complex spatial pattern and an arbitrarily prescribed input signal, the output generated by the complex open-loop responses between the original model and the pseudo-balanced reduced-order model (PBROM) are compared. The agreement is excellent.

Second, after demonstrating that we can effectively build a reduced-order model, we perform a closed-loop study of the nonlinear feedback control of the large-scale convection patterns. It is understood that building a controller design can be a very elaborate effort. Here, for illustration, a direct intuitive approach is used. We build a reduced-order, nonlinear estimator using a steady-state Kalman gain based on a linear-quadratic-Gaussian-estimator model. The nonlinear estimator is used in conjunction with a simple proportional feedback-control law. The nonlinear travelling front is again employed as an example. The closed-loop simulations show that the nonlinear front disturbances can be damped out effectively.

This paper is organized as follows. The numerical discretization of the continuous system is highlighted in § 2. In § 3, the empirical pseudo-inverse-based projective model reduction technique, PBPOD, is described in detail. In § 4, all the results in comparing PBPOD to the original models are presented. In § 5, the feedback-control study is given. Section 6 provides a conclusion. The input/output version of the amplitude equation is derived in the Appendix.

2. The nonlinear input/output amplitude equation

According to (A 25), the amplitude equation with a prescribed input term is given by

$$\tau_0 \partial_t A = A + \xi_0^2 A_{XX} - \chi_0 |A|^2 A + \gamma b(X,t), \tag{2.1}$$

where amplitude $A(X,t)$ depends on a stretched coordinate, X $(0 \leqslant X \leqslant L)$, and a slow time, t $(0 \leqslant t < \infty)$. The Appendix provides a full derivation of the equation. (In the Appendix, T is the slow time and t is the physical time.)

The Appendix provides expressions for coefficients τ_0, ξ_0, χ_0 and γ. For the purpose of analysis, a rescaling of (2.1) is desirable to reduce the number of parameters. In order to avoid using too many symbols, we retain the same variable notations in the following canonical equations:

$$\partial_t A = A + A_{XX} - |A|^2 A + b(X,t). \tag{2.2}$$

The normalized variables of (2.2) are $(t/\tau_0) \rightarrow t$, $(X/\xi_0) \rightarrow X$, $(A/\chi_0^{1/2}) \rightarrow A$, $(\chi_0^{1/2}\gamma b) \rightarrow b$ and $(\chi_0^{1/2} y_n) \rightarrow y_n$, where the rescaled groups are indicated on the left-hand side of the arrows.

The partial differential system is discretized and the fast Fourier transform (FFT) is employed, with $A(X_n, t) \rightarrow A(k_m, t)$; and its inverse discrete Fourier transform (IFFT), $A(k_m, t) \rightarrow A(X_n, t)$. The fundamental wavenumber is $\varDelta_k = 2\pi/L$. Also, for a single-input–single-output configuration (SISO), we separate the spatial and time components in the input, that is, $b(X, t) \rightarrow b(X)u(t)$. Upon discretization, we obtain a set of N-coupled, nonlinear ordinary differential equations,

$$\dot{\boldsymbol{x}} = \boldsymbol{L}\boldsymbol{x} - \boldsymbol{N}(\boldsymbol{x}) + \boldsymbol{B}u. \tag{2.3}$$

The state vector, $\boldsymbol{x}$, consists of the Fourier coefficients as functions of time. In the treatment of the cubic nonlinear term, $\boldsymbol{N}(\boldsymbol{x})$, we employ a well-known pseudospectral technique, which calculates nonlinear terms by multiplications of the discrete dependent variables in physical space at collocation points (Gottlieb & Orszag 1981). We refer to § 3.3 for the equations describing the cubic nonlinear term.

The output corresponds to a single spatial-point sensor measurement at the mid-plane of the layer, at the nth collocation point (see (A 26) of the Appendix)

$$y_n(t) = g_0\left(\tfrac{1}{2}\right) A(X_n, t). \tag{2.4}$$

We denote the FFT version of the equation by

$$y = \boldsymbol{C}\boldsymbol{x}, \tag{2.5}$$

where $\boldsymbol{C}$ is determined by the explicit transformation of FFT and the value of g_0 at the mid-plane of the layer, $z = 1/2$.

We note that for nonlinear systems, the input-shape functions used to stimulate the low-order dynamics of interest are not necessarily the input-shape functions to be used for feedback control of the dynamics. For example, a function $b(X, t)$ consisting of a temporal impulse and a spatial pulse generates the nonlinear travelling fronts. However, the shape function to be used for the feedback control of the fronts is a different $b(X, t)$ that depends on the PBPOD modes of the dynamics (see § 5 for more detail).

3. Empirical model reduction for input/output systems

The proper orthogonal decomposition (Holmes, Lumley & Berkooz 1998), commonly referred to as POD, is a snapshot-based technique to extract the low-order modal structure to be used for a low-order projection of a large dynamical system. For input/output systems, POD has been extended to include balancing between the inputs and outputs (Lall *et al.* 2002; Rowley 2005). Rowley referred to the method as BPOD. While the method has many variants, a glimpse of examples can be found in Wilcox & Peraire (2002), Or & Speyer (2008) and Barbagallo, Sipp & Schmid (2009).

3.1. *Empirical balanced model reduction for linear time-invariant state-space systems*

For theoretical discussion, we refer to large linear time-invariant (LTI) dynamical systems as the following state-space systems (SS):

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}u, \quad y = \boldsymbol{C}\boldsymbol{x}. \tag{3.1}$$

Balanced realization of (3.1) seeks a similarity transformation for the balancing between controllability and observability (see Chen 1970). Finding a full-rank transformation is laborious for a large dynamical system.

The snapshot-based empirical method seeks approximations of the controllability and the observability Gramians. Analytically, these quantities are defined, respectively, as $\boldsymbol{G}_c = \int_0^\infty \exp(\boldsymbol{A}t)\boldsymbol{B}\boldsymbol{B}^+ \exp(\boldsymbol{A}^+t)\,\mathrm{d}t$ and $\boldsymbol{G}_o = \int_0^\infty \exp(\boldsymbol{A}^+t)\boldsymbol{C}^+\boldsymbol{C} \exp(\boldsymbol{A}t)\,\mathrm{d}t$. A balanced realization is achieved using the full-rank transformation $\boldsymbol{T}$, with

$$\boldsymbol{S}\boldsymbol{G}_c\boldsymbol{S}^+ = \boldsymbol{\Sigma}, \quad \boldsymbol{T}^+\boldsymbol{G}_o\boldsymbol{T} = \boldsymbol{\Sigma}, \quad \boldsymbol{S} \equiv \boldsymbol{T}^{-1}. \tag{3.2}$$

The diagonal matrix $\boldsymbol{\Sigma}$ has its entries ordered in descending magnitude. The entries provide a measure of the energy content of the modes and is referred to as the Hankel singular values (HSVs).

Many large dynamical systems display only a few significant HSVs. These systems are good candidates for the empirical methods. Rowley (2005) demonstrates that controllability and observability data blocks, $\boldsymbol{X}$ and $\boldsymbol{Y}$, can be constructed using a smaller number of snapshots. The empirical Gramians are defined by the data blocks as

$$\boldsymbol{W}_c = \boldsymbol{X}\boldsymbol{X}^+, \quad \boldsymbol{W}_o = \boldsymbol{Y}\boldsymbol{Y}^+, \tag{3.3}$$

where the empirical Gramians are good approximations of the exact ones, that is, $\boldsymbol{W}_c \approx \boldsymbol{G}_c$ and $\boldsymbol{W}_o \approx \boldsymbol{G}_o$. Scherpen (1993) provides a good description on the subject of Gramians, controllability and observability.

Rowley (2005) constructed the product $\boldsymbol{Y}^+\boldsymbol{X}$ for balancing. This quantity is identified as the generalized Hankel matrix in (3.12) (see Or & Speyer 2008). For the linear time-invariant state-space system (LTISS), $\boldsymbol{H} = \boldsymbol{Y}^+\boldsymbol{X}$ can be expressed in closed form with elements $H_{ij} = y_j(t_i) = \boldsymbol{C}\exp\{\boldsymbol{A}(t_i + t_j)\}\boldsymbol{B}$. This relationship is the basis of balancing and can be used to derive other variant algorithms for BPOD (see Ma, Ahuja & Rowley 2009). For the LTISS, BPOD can be understood as the algorithm of finding two sets of vectors, or modes (far less than those necessary to span the entire full space) that simultaneously diagonalizes the pair of Gramians. The two sets are typically distinct for a non-normal system. Assuming both $N \times N_c$ (with $N_c \ll N$) blocks $\boldsymbol{X}$, $\boldsymbol{Y}$ are available, then the SVD of $\boldsymbol{H}$ yields $\boldsymbol{H} = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^+$, with $\boldsymbol{Y}^+\boldsymbol{X} = \boldsymbol{H}$, the SVD can be expressed as

$$\left(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{U}_1^+\boldsymbol{Y}^+\right)\left(\boldsymbol{X}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2}\right) = \boldsymbol{I}. \tag{3.4}$$

The singular values in the SVD of $\boldsymbol{H}$ are referred to as the HSVs. The above expression contains two transformations which have an inverse relationship, $\boldsymbol{S}_1\boldsymbol{T}_1 = \boldsymbol{I}$, with

$$\boldsymbol{S}_1 = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{U}_1^+\boldsymbol{Y}^+, \quad \boldsymbol{T}_1 = \boldsymbol{X}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2}. \tag{3.5}$$

Transforming the Gramians in (3.3) by direct substitutions, it can be verified that

$$\boldsymbol{S}_1\boldsymbol{W}_c\boldsymbol{S}_1^+ = \boldsymbol{\Sigma}_1, \quad \boldsymbol{T}_1^+\boldsymbol{W}_o\boldsymbol{T}_1 = \boldsymbol{\Sigma}_1. \tag{3.6}$$

We called the above results upper-block balancing. In BPOD, balancing refers to the even stronger conditions. It can be verified by direct substitution that

$$\boldsymbol{W}_c\boldsymbol{S}_1^+ = \boldsymbol{T}_1\boldsymbol{\Sigma}_1, \quad \boldsymbol{W}_o\boldsymbol{T}_1 = \boldsymbol{S}_1^+\boldsymbol{\Sigma}_1. \tag{3.7}$$

Equation (3.7) yields the following eigenvalue relationships by direct substitutions,

$$(\boldsymbol{W}_c\boldsymbol{W}_o)\boldsymbol{T}_1 = \boldsymbol{T}_1\boldsymbol{\Sigma}_1^2, \quad (\boldsymbol{W}_o\boldsymbol{W}_c)\boldsymbol{S}_1^+ = \boldsymbol{S}_1^+\boldsymbol{\Sigma}_1^2. \tag{3.8}$$

The above relationship indicates that $\boldsymbol{S}_1$ and $\boldsymbol{T}_1$ are, respectively, the left and right eigenvectors of the product of the Gramians, $\boldsymbol{W}_c\boldsymbol{W}_o$, since $\boldsymbol{W}_o\boldsymbol{W}_c$ is the transpose of $\boldsymbol{W}_c\boldsymbol{W}_o$. If both empirical Gramians are available, alternatively, $\boldsymbol{S}_1$ and $\boldsymbol{T}_1$ can be obtained from the normalized eigenvectors by determining a weighting coefficient for each eigenvector.

Depending on the sizes of the HSVs, BPOD allows further truncation. One can extract the first N_r (where $N_r < N_c$) row vectors from $\boldsymbol{S}_1$ and the first N_r column vectors from $\boldsymbol{T}_1$ to form $(\boldsymbol{S}_r, \boldsymbol{T}_r)$. The full-order LTISS, $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ can be balanced and reduced to $(\boldsymbol{S}_r\boldsymbol{A}\boldsymbol{T}_r, \boldsymbol{S}_r\boldsymbol{B}, \boldsymbol{C}\boldsymbol{T}_r)$. BPOD preserves this symmetric property because the LTISS utilizes both controllability and observability data, whereas the latter requires adjoint state information. Based on BPOD, Rowley (2005) demonstrated that another pair of basis vectors $(\boldsymbol{S}_2, \boldsymbol{T}_2)$ can be constructed based on $(\boldsymbol{S}_1, \boldsymbol{T}_1)$ to complete the full set. Let

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1 \\ \boldsymbol{S}_2 \end{bmatrix}, \quad \boldsymbol{T} = [\boldsymbol{T}_1 \quad \boldsymbol{T}_2], \tag{3.9}$$

and impose the constraint $\boldsymbol{S}\boldsymbol{T} = \boldsymbol{T}\boldsymbol{S} = \boldsymbol{I}$. One can either find the null-space of $\boldsymbol{S}_1^+$ or $\boldsymbol{T}_1$. If we perform SVD on $\boldsymbol{S}_1^+$ and obtain the null-space as $\boldsymbol{T}_2$, then $\boldsymbol{T} = [\boldsymbol{T}_1\,\boldsymbol{T}_2]$ and $\boldsymbol{S} = \boldsymbol{T}^{-1}$. BPOD produces block diagonalization of the Gramians:

$$\boldsymbol{S}\boldsymbol{W}_c\boldsymbol{S}^+ = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_c \end{bmatrix}, \quad \boldsymbol{T}^+\boldsymbol{W}_o\boldsymbol{T} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_o \end{bmatrix}. \tag{3.10}$$

3.2. *Empirical pseudo-balanced model reduction applicable to nonlinear systems*

For flow-control applications, what makes the empirical-based model reduction technique attractive is that the number of dynamics states are typically much greater than the number of snapshots. In this paper, we consider a pseudo-inverse projective method for model reduction for input/output nonlinear systems of the following form:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{B}u, \quad y(t) = \boldsymbol{C}\boldsymbol{x}. \tag{3.11}$$

Consider integrating the input dynamics equation above using the Dirac-impulse function as the initial condition for $u(t)$. The input vector $\boldsymbol{B}$ is generated by a certain input-shape function $b(X)$. A number of snapshots is collected to form the empirical controllability Gramian (for its form, see (3.3) above). After performing a singular-value decomposition (SVD) on this Gramian, it is determined that N_c (where $N_c \ll N$) eigenvalues are energetic enough to be retained. Then, using each of the N_c eigenvectors, $\boldsymbol{X}_j$, also known as the input-stimulated POD functions, as the initial condition to integrate (3.11) in time, with $u(t) \equiv 0$, we obtain N_c output values, $y_j(t_i)$. In general, POD functions can be stimulated by any initial conditions. The set of snapshot vectors and the corresponding pseudo-inverse form the POD-based model reduction transformations. We construct the $N_o \times N_c$ generalized Hankel matrix, $\boldsymbol{H}$, expressed as:

$$\boldsymbol{H} = \begin{bmatrix} y_1(t_1) & y_2(t_1) & \dots & y_{N_c}(t_1) \\ y_1(t_2) & y_2(t_2) & \dots & y_{N_c}(t_2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ y_1(t_{N_o}) & y_2(t_{N_o}) & \dots & y_{N_c}(t_{N_o}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_1^+ \\ \boldsymbol{Y}_2^+ \\ \cdot \\ \cdot \\ \boldsymbol{Y}_{N_o}^+ \end{bmatrix} [\boldsymbol{X}_1\boldsymbol{X}_2\dots\boldsymbol{X}_{N_c}]. \tag{3.12}$$

The above equation can be written as $\boldsymbol{H}=\boldsymbol{Y}^{+}\boldsymbol{X}$. For nonlinear systems (3.11), $\boldsymbol{X}$ and $\boldsymbol{H}$ are empirical data and are known. However, $\boldsymbol{Y}^{+}$ is not to be known explicitly. We retain the form (3.12) in the understanding that it is a natural extension of the LTISS. In LTISS, $\boldsymbol{Y}$ is the observability data generated via the adjoint equation (see in the following section), which has no parallel for nonlinear systems e.g. (3.11).

3.2.1. *PBPOD for LTISS and the limitation of left upper-block balancing*

For nonlinear applications, with (3.11), $\boldsymbol{Y}$, and therefore the empirical observability Gramian $\boldsymbol{W}_o$, is not available from the empirical data. We will perform a projective-type balancing, using BPOD as a guide, by utilizing only information available from $\boldsymbol{X}$ and $\boldsymbol{H}$. In this context, it is assumed that we have the controllability Gramian but not the observability Gramian. Some authors linearize the system to obtain the adjoint equations at selected points along the nonlinear trajectory. Then, the adjoint equations are time-varying. This procedure can be elaborate. There are promising concepts and algorithms developed in the applied mathematics arena, such as by Krener & Hunt (2009) and by Scherpen (1993). These ideas exploit the path of extending balancing results based on LTISS to nonlinear structures. However, the complexity of the concepts and algorithms are significantly higher than PBPOD. These complex schemes are of interest, but are too demanding for flow control applications.

In this paper, we attempt to construct $\boldsymbol{Y}$ based solely on the knowledge of $\boldsymbol{H}$ and $\boldsymbol{X}$ alone. The method is aided by pseudo-inversion and projection. To differentiate our projective pseudo-inverse-based method from BPOD, we called this method PBPOD, or projective BPOD. Since no adjoint states are propagated, an *a priori* assumption in PBPOD is that the 'balanced' modes are both controllable and observable. All the important observable modes are within the subspace of the eigenvectors of the controllability Gramian. Omitted modes that are supposed to be highly observable, but not very controllable, will not be retained in PBPOD.

When using LTISS for the evaluation of PBPOD, it is assumed that both $\boldsymbol{X}$ and $\boldsymbol{Y}$, and therefore the empirical Gramians $\boldsymbol{W}_c$ and $\boldsymbol{W}_o$, are known *a priori*. Then, PBPOD seeks to find a pair of transformations, $(\boldsymbol{S}_{p1}, \boldsymbol{T}_1)$. While $\boldsymbol{T}_1$ is picked to be the same as that of BPOD, $\boldsymbol{S}_{p1}$ has to be constructed based only on the knowledge of $\boldsymbol{X}$ and $\boldsymbol{H}$, with no knowledge of $\boldsymbol{Y}$ assumed. The imperfections of balancing will be assessed. There are two methods to construct $\boldsymbol{S}_{p1}$: (i) use $\boldsymbol{Y}_p^{+}=\boldsymbol{H}\boldsymbol{X}^{\#}$ to replace $\boldsymbol{Y}^{+}$ and (ii) use simply $\boldsymbol{S}_{p1}=\boldsymbol{T}_1^{\#}$ where the superscript # denotes the Penrose pseudo-inverse. For instance, for the $N \times N_c$ matrix $\boldsymbol{X}$, its Penrose pseudo-inverse is defined by the $N_c \times N$ matrix, $\boldsymbol{X}^{\#}=(\boldsymbol{X}^{+}\boldsymbol{X})^{-1}\boldsymbol{X}^{+}$, so that $\boldsymbol{X}^{\#}\boldsymbol{X}=\boldsymbol{I}$. In either case, it can be shown that the left-upper blocks of the Gramians are simultaneously diagonalized, with the same HSVs as in BPOD,

$$\boldsymbol{S}_{p1}\boldsymbol{W}_c\boldsymbol{S}_{p1}^{+} = \boldsymbol{\Sigma}_1, \quad \boldsymbol{T}_1^{+}\boldsymbol{W}_o\boldsymbol{T}_1 = \boldsymbol{\Sigma}_1. \tag{3.13}$$

This result is equivalent to BPODs given in (3.6). However, one of the two stronger conditions in (3.7) does not hold, which has an extra term. In the case of PBPOD, (3.8) has to be modified by the extra term.

For LTISS, PBPOD achieves left upper-block balancing but not block-diagonal balancing because, unlike BPOD, (3.13) holds, but the second of the stronger conditions in (3.7) does not hold.

From the last expression of $\boldsymbol{S}_{p1}$ on the right-hand side of (3.22), or from (3.23), we have

$$\boldsymbol{W}_c\boldsymbol{S}_{p1}^{+} = \boldsymbol{X}\boldsymbol{X}^{+}(\boldsymbol{X}^{\#})^{+}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{1/2} = \boldsymbol{X}(\boldsymbol{X}^{\#}\boldsymbol{X})^{+}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{1/2} = \boldsymbol{X}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_1 = \boldsymbol{T}_1\boldsymbol{\Sigma}_1, \tag{3.14}$$

noting that $\boldsymbol{X}^{\#}\boldsymbol{X} = \boldsymbol{I}$. By the same token, we have

$$\boldsymbol{W}_o\boldsymbol{T}_1 = \boldsymbol{Y}\boldsymbol{Y}^{+}\boldsymbol{X}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2} = \boldsymbol{Y}\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^{+}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2} = \boldsymbol{S}_1^{+}\boldsymbol{\Sigma}_1 \neq \boldsymbol{S}_{p1}^{+}\boldsymbol{\Sigma}_1, \tag{3.15}$$

though, in both situations $\boldsymbol{T}_1^{+}\boldsymbol{S}_1^{+} = \boldsymbol{T}_1^{+}\boldsymbol{S}_{p1}^{+} = \boldsymbol{I}$. By virtue of the constraint of (3.13), one expects

$$\boldsymbol{W}_o\boldsymbol{T}_1 = \boldsymbol{S}_{p1}^{+}\boldsymbol{\Sigma}_1 + \boldsymbol{E}. \tag{3.16}$$

The error term $\boldsymbol{E}$ has to be in the null-space of $\boldsymbol{T}_1^{+}$, that is, $\boldsymbol{T}_1^{+}\boldsymbol{E} = \boldsymbol{0}$. The limitation in PBPOD is manifested in the lack of block diagonalization of the second equation of (3.10).

For PBPOD, we construct the full set of transformations similar to BPOD, $(\boldsymbol{S}_p, \boldsymbol{T})$, using (3.9). Replacing $\boldsymbol{S}_1$ by $\boldsymbol{S}_{p1}$ and $\boldsymbol{S}$ by $\boldsymbol{S}_p$, we observe that $\boldsymbol{S}_p\boldsymbol{W}_c\boldsymbol{S}_p^{+}$ is block diagonal, but $\boldsymbol{T}^{+}\boldsymbol{W}_o\boldsymbol{T}$ is not. Based on (3.5), we observe $\boldsymbol{T}_1^{+}\boldsymbol{S}_{p2}^{+} = \boldsymbol{0}$. Therefore, $\boldsymbol{E}$ lies in the subspace $\boldsymbol{S}_{p2}^{+}$. Write $(\boldsymbol{W}_o\boldsymbol{T}_1 - \boldsymbol{S}_{p1}^{+}\boldsymbol{\Sigma}_1) = \boldsymbol{E} = \boldsymbol{S}_{p2}^{+}\boldsymbol{q}$, where $\boldsymbol{q}$ forms the coefficients of the column vectors of $\boldsymbol{S}_{p2}^{+}$. The non-vanished cross-diagonal blocks correspond to the term $\boldsymbol{T}_2^{+}\boldsymbol{W}_o\boldsymbol{T}_1$ and its transpose. We have

$$\boldsymbol{T}_2^{+}\boldsymbol{W}_o\boldsymbol{T}_1 = \boldsymbol{T}_2^{+}\boldsymbol{E} = \boldsymbol{T}_2^{+}\boldsymbol{S}_{p2}^{+}\boldsymbol{q} = \boldsymbol{q}. \tag{3.17}$$

One can proceed to derive analogous expressions of (3.8), which will yield extra terms, indicating that the pair $(\boldsymbol{S}_{p1}, \boldsymbol{T}_1)$ are not exact eigenvectors of the product of the Gramians. It is true that the rows and columns of $(\boldsymbol{S}_{p1}, \boldsymbol{T}_1)$ are not eigenvectors, but they can be considered as eigenvectors 'contaminated' with components from the modes that have low energy content, and therefore, are being truncated.

3.2.2. *Proof of the left upper-block balancing*

In this section, we provide proof of the first term of (3.13) with methods (i) and (ii). In the end, we will show that the two methods are basically equivalent. The second term of (3.13) remains valid, if we assume $\boldsymbol{Y}$ and in turn $\boldsymbol{W}_o$ is available, as it will be available for the case of LTISS.

To prove method (i), we note that $\boldsymbol{H} = \boldsymbol{Y}^{+}\boldsymbol{X} = \boldsymbol{Y}_p^{+}\boldsymbol{X}$. In method (i), $\boldsymbol{S}_{p1}$ is constructed by using the same expression as in BPOD,

$$\boldsymbol{S}_{p1} = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{U}_1^{+}\boldsymbol{Y}_p^{+}. \tag{3.18}$$

Using (3.18) and noting $\boldsymbol{W}_c = \boldsymbol{X}\boldsymbol{X}^{+}$, we have

$$\left.\begin{aligned} \boldsymbol{S}_1\boldsymbol{W}_c\boldsymbol{S}_1^{+} &= \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{U}_1^{+}\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^{+}(\boldsymbol{X}^{+}\boldsymbol{X})^{-1}\boldsymbol{X}^{+}(\boldsymbol{X}\boldsymbol{X}^{+}) \\ \boldsymbol{X}(\boldsymbol{X}^{+}\boldsymbol{X})^{-1}\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{U}_1^{+}\boldsymbol{U}_1\boldsymbol{\Sigma}_1^{-1/2} &= \boldsymbol{\Sigma}_1. \end{aligned}\right\} \tag{3.19}$$

This completes the proof of method (i).

To prove method (ii), we start with the matrix identity $\boldsymbol{T}_1^{\#}\boldsymbol{T}_1 = \boldsymbol{I}$, after substituting $\boldsymbol{T}_1 = \boldsymbol{X}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2}$, we invert $\boldsymbol{V}_1$ and $\boldsymbol{\Sigma}_1^{-1/2}$ so that they appear on the right-hand side of the equation. Note that $\boldsymbol{V}_1\boldsymbol{V}_1^{+} = \boldsymbol{I}$ is permitted if $\boldsymbol{V}_1$ is a square Hermitian matrix. The matrix identity can be rewritten as

$$\boldsymbol{T}_1^{\#}\boldsymbol{X} = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{V}_1^{+}. \tag{3.20}$$

By definition, $\boldsymbol{W}_c = \boldsymbol{X}\boldsymbol{X}^{+}$. Multiplying both sides by $\boldsymbol{T}_1^{\#}$ and substituting (3.20), we obtain

$$\boldsymbol{T}_1^{\#}\boldsymbol{W}_c(\boldsymbol{T}_1^{\#})^{+} = (\boldsymbol{T}_1^{\#}\boldsymbol{X})(\boldsymbol{T}_1^{\#}\boldsymbol{X})^{+} = \boldsymbol{\Sigma}_1. \tag{3.21}$$

This completes the proof of method (ii) by using $\boldsymbol{S}_{p1} = \boldsymbol{T}_1^{\#}$.

The proof of method (ii) requires $\boldsymbol{U}_1$ and $\boldsymbol{V}_1$ to be square Hermitian matrices, but the proof of method (i), which resembles BPOD, does not. In BPOD and in method (i) of PBPOD, one can pick the N_r-dominant HSV from the N_c-by-N_c $\boldsymbol{H}$ ($N_r < N_c$) for block-balancing. In other words, one can perform truncation and then balancing. In this case, $\boldsymbol{U}_1$ and $\boldsymbol{V}_1$ are rectangular, N_c-by-N_r matrices. In method (ii), however, one has to perform balancing on the order of N_c, before performing truncation. With a clear understanding of the differences, there is no need to introduce additional notations associated with $\boldsymbol{S}_{p1}$ and $\boldsymbol{T}_1$ to distinguish the two methods of PBPOD. In each case, the Gramian transformations yield the left upper-block balancing relationship (3.6).

Methods (i) and (ii) are in fact equivalent. For balancing at the order of N_c, it can be shown that both methods (i) and (ii) are equivalent. Starting with method (i), we have

$$\boldsymbol{S}_{p1} = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{U}_1^{+}\boldsymbol{H}\boldsymbol{X}^{\#} = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{U}_1^{+}\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^{+}\boldsymbol{X}^{\#} = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{V}_1^{+}\boldsymbol{X}^{\#}. \tag{3.22}$$

Now, starting with method (ii), provided that $\boldsymbol{V}_1$ is a square matrix and is unitary, we have

$$\boldsymbol{S}_{p1} = \boldsymbol{T}_1^{\#} = \left(\boldsymbol{X}\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2}\right)^{\#} = \left(\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1/2}\right)^{-1}\boldsymbol{X}^{\#} = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{V}_1^{+}\boldsymbol{X}^{\#}. \tag{3.23}$$

Both (3.22) and (3.23) produce the same balanced transformation for order N_c.

3.3. *Nonlinear reduced-order model by pseudo-balanced proper orthogonal decomposition*

We take the original model, described by (2.3) and (2.5), to construct the PBROM. For convenience, we denote the FFT and IFFT transformations as $\boldsymbol{F}$ and $\boldsymbol{F}_i$, respectively. PBPOD gives the pair of transformations $\boldsymbol{S}_{p1}$ and $\boldsymbol{T}_1$. The state vector $\boldsymbol{x}$ of the original model is then approximated by its projective version, $\boldsymbol{x}_1$, where

$$\boldsymbol{x}_1 = \boldsymbol{T}_1\boldsymbol{x}_r, \tag{3.24}$$

where $\boldsymbol{x}_r$ is the PBROM state vector.

The cubic nonlinear term is treated by the pseudospectral technique as mentioned in §2 (Gottlieb & Orszag 1981). Multiplications of the nonlinear term are performed in the physical space at the collocation points. In the physical space, if $\boldsymbol{a}_1$ approximates $\boldsymbol{a}$, where $\boldsymbol{a} = \boldsymbol{F}_i\boldsymbol{x}$, we have

$$\boldsymbol{a}_1 = \boldsymbol{F}_i\boldsymbol{T}_1\boldsymbol{x}_r. \tag{3.25}$$

The approximated version of the nonlinear term in Fourier space is

$$\mathcal{N}_1 = \boldsymbol{F}\boldsymbol{\Lambda}_{a_1}^{*}\boldsymbol{\Lambda}_{a_1}\boldsymbol{a}_1, \tag{3.26}$$

where the asterisk denotes the complex conjugate. The matrix $\boldsymbol{\Lambda}_{a_1}$ denotes a diagonal matrix whose diagonal entries are given by the vector $\boldsymbol{a}_1$. The nonlinear PBROM corresponding to the N_r-dimensional subspace with state vector $\boldsymbol{x}_r(t)$, is given by

$$\dot{\boldsymbol{x}}_r = (\boldsymbol{S}_{p1}\boldsymbol{L}\boldsymbol{T}_1)\boldsymbol{x}_r - \boldsymbol{S}_{p1}\mathcal{N}_1 + \boldsymbol{S}_{p1}\boldsymbol{B}u, \quad y(t) = \boldsymbol{C}\boldsymbol{T}_1\boldsymbol{x}_r. \tag{3.27}$$

4. Results on pseudo-balanced proper orthogonal decomposition

The pseudo-balanced reduced-order model is based on PBPOD. We choose three cases corresponding to different pattern shapes for a SISO model to demonstrate the performance of the PBROM by comparing the envelope snapshots between the original and reduced-order models. Case (i) consists of patterns of nonlinear front propagation, case (ii) with real envelopes generated by stable and unstable wavenumber shapes and case (iii) with complex envelopes generated by stable and unstable wavenumber shapes. Next, we perform an open-loop simulation using the same arbitrary input signal to compare the outputs between the original and the PBROM. Lastly, before completing this section, we compare the performance of PBPOD to two other existing methods, POD and BPOD, applied to a linearized model about a certain equilibrium of the nonlinear system. We refer to this as the linearized BPOD.

We start by evaluating the PBPOD method using the cases (i)–(iii). A comparison of the snapshots between the original full-order model and the PBROM is performed, by separately integrating (2.3) and (3.27) in time. The performance of the PBROM is determined by how well the reconstructed snapshots from the PBROM compared with the original snapshots. Four cases are considered. The first three cases provide comparisons of snapshots to demonstrate PBROM performance. A different input shape, $b(X)$, is used for each PBROM. The fourth case consists of two open-loop simulations to determine how well the PBROM's reconstructed transient responses compared with the original model's transient responses for a given input signal in time.

The PBPOD uses the following parameters. The total number of integration steps is N_t; the number of integration steps between successive snapshots is set fixed at N_i; and the number of snapshots, $N_c = N_t/N_i$, is set to be equal to 25 for all runs. Since the singular values of the diagonalized Gramian entries are modified by the nonlinear algorithm, we refer to these as the modified HSVs (MHSVs), as opposed to the HSVs in the LTISS. Thus, 25 MHSVs are present in the upper-left blocks of the two Gramians. After the upper-left blocks are balanced, the number of states of the PBROM is further truncated from N_c to N_r. Here, N_r is the lowest number of PBPOD modes retained in the PBROM for it to remain a good approximation of the original model.

To avoid over-crowdedness of lines in the figures, all except for case (i), only the odd numbers of snapshots are plotted. All the snapshots are sampled at same time intervals of both original and reduced-order models. The implementation of PBPOD is automatic and systematic. No heuristic assumptions are involved in the order-reduction process. Using FFT, the nonlinear partial differential system is discretized to a set of $N = 512$ nonlinear-coupled ordinary differential equations. This set is referred to as the original model. The 512 states are the Fourier coefficients. The resolution used here is significantly larger than what is required. Except for case (iii), second run, a moderate original model size of $N = 64$ Fourier modes is actually sufficient for the approximation of the envelope curves. A large N is used for demonstrating a dramatic reduction in order and to ensure that all details are captured. The spatial increment of collocation points is given by $\Delta = 0.05$. For $N = 512$, it provides a layer length of $L = 25.6$. A semi-implicit Euler integration scheme is used, with an integration step size of $\Delta t = 0.005$.

4.1. *Nonlinear front propagation*

Equations (2.3) and (2.5) of §2 provide the simulation system. In the system, input vector $\boldsymbol{B}$ corresponds to a rectangular pulse-shaped function, with the pulse centred at the origin ($X_0 = 0$). The output vector $\boldsymbol{C}$ corresponds to the mid-layer temperature at the collocation point with $X = 0$. A rectangular pulse shape provides a very rich harmonic content. When used as an initial condition, the convection pattern evolves into a front propagation seen in many experiments.

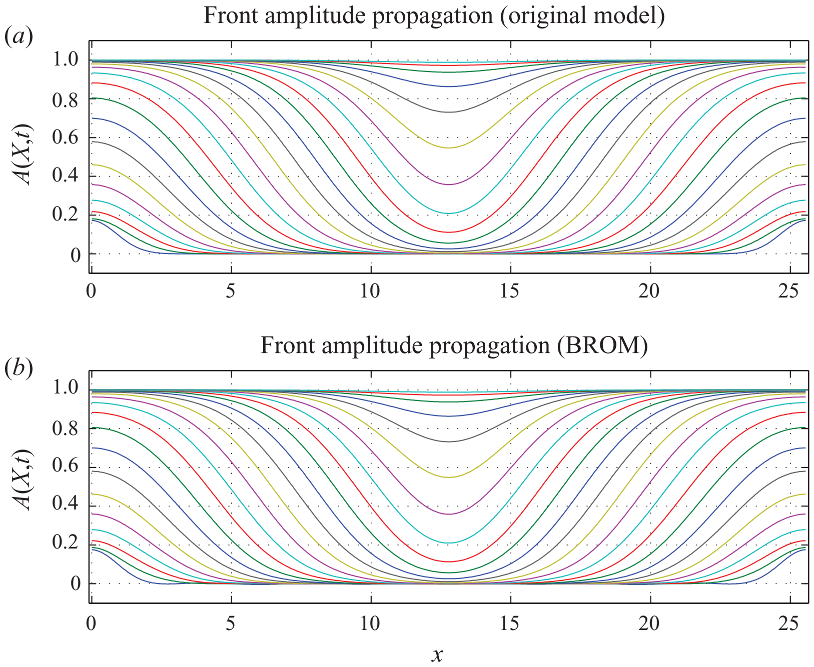


FIGURE 1. (Colour online) Comparison of front propagation in original and reduced-order models.

When linearized about the unstable equilibrium, $A(X,t)=0$, the normal modes consist of the sinusoidal functions. In the nonlinear solutions, the Fourier modes are coupled via the cubic nonlinear term $\boldsymbol{N}(\boldsymbol{x})$. For the rectangular-pulse shape, all the coefficients in the input vector $\boldsymbol{B}$ have comparable magnitudes. The output vector $\boldsymbol{C}$ also has non-sparse coefficients. Both $\boldsymbol{B}$ and $\boldsymbol{C}$ are complex. Due to the periodicity and the wrap-around property, the rectangular pulse can be wrap-around using two rectangular half-pulses positioned initially at the two ends, $X=0$ and $X=L$, respectively. Each half-pulse has width $2\varDelta$ and height $0.1/3\varDelta=0.67$.

In time, the two half-pulses move towards each other and merge at the mid-point of the layer. In case (i), we use $N_t=2000$, $N_i=80$ for both the original model and the PBROM. The total number of snapshots is $N_c=25$. The original model has $N=512$ states and the PBROM has $N_r=8$ states.

In figure 1, the family of curves represent a total of 25 snapshots of a developing front envelope, captured at sampled times with intervals of $80\Delta t$. Given a long enough time, the front curves tend to the unstable equilibrium, $A(X_n,\infty)=1$. In figure 1, the agreement between the propagating front envelopes from the PBROM (*b*) and those from the original model (*a*) is striking, apart from some small distortions at the base and two ends of the PBROM plot. It suggests that the PBROM is a very good approximation of the original model for the given input. The performance of the PBROM is not sensitive to the sensor location as long as it stays away from the end points.

As the first case presented, certain PBPOD diagnostics appear to be useful. For case (i), the discrete-time square Hankel matrix $\boldsymbol{H}$ has dimension $N_c=25$. Given the PBPOD balancing of the upper block of the Gramians, both the $N_c\times N$ transformation matrix $\boldsymbol{S}_{p1}$ and the $N\times N_c$ matrix $\boldsymbol{T}_1$ are constructed (where $N=512$, $N_c=25$) based on the pseudo-balancing theory of § 3. Then, we perform further truncation to keep

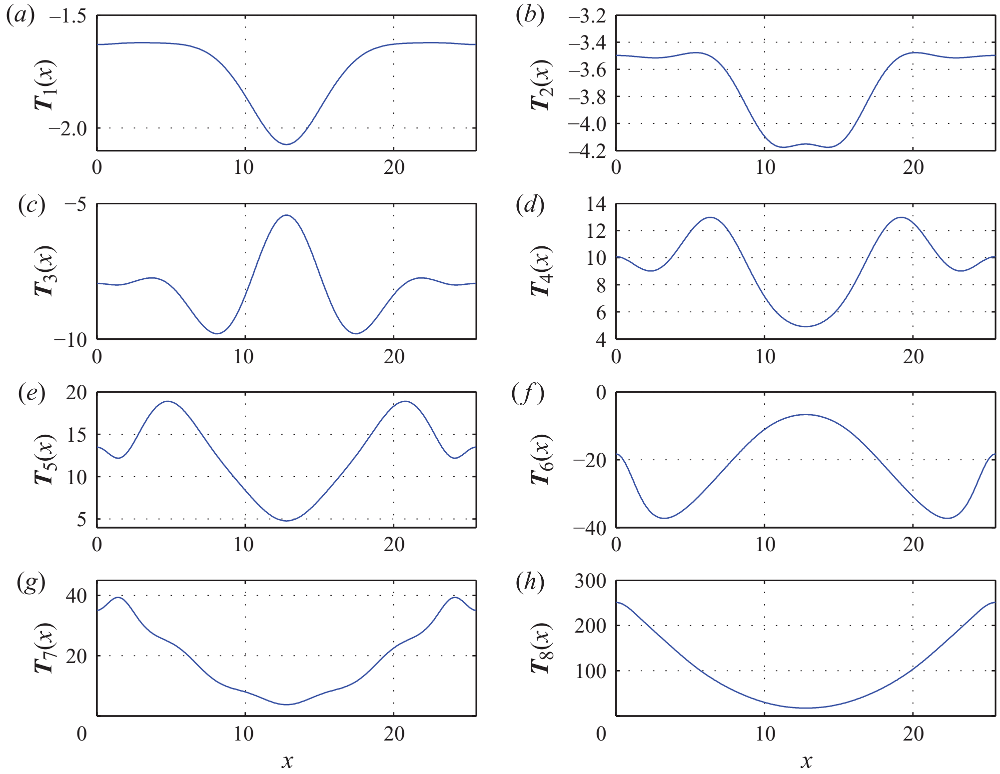


FIGURE 2. (Colour online) The first eight PBPOD basis functions of the evolving fronts (order: (*a*)–(*h*)).

N_r MHSVs. This means that $\boldsymbol{S}_{p1}$ is truncated to its first N_r rows and $\boldsymbol{T}_1$ to its first N_r columns ($N_r = 8$). As stated in § 3, we have used the same notations, $\boldsymbol{S}_{p1}$ and $\boldsymbol{T}_1$, before and after the truncation for simplicity. The IFFT of $\boldsymbol{T}_1$ gives eight basis functions to span the reduced-order space. The shapes of these basis functions are shown in figure 2, labelled T. All turn out to have an even parity. These functions are in different orders of amplitudes and are not mutually orthogonal to one another. Their shapes are significantly different from those of sinusoids. At least 64 or more sinusoidal modes are required to expand the evolved fronts, but these 8 basis functions are sufficient to approximate the fronts very well.

The pseudo-inverse, $\boldsymbol{S}_{p1}$, provides the other set of projections needed. These functions are shown in figure 3, labelled S. These basis functions are highly oscillatory in a smaller scale. For the LTISS example (Or & Speyer 2008), the pseudo-inverse functions can be directly compared with the corresponding set obtained using BPOD (by integrating the adjoint equation). They look very different from one another. The difference is consistent with the property that $\boldsymbol{S}_{p1}$ can only diagonalize the upper-block of the Gramians, but $\boldsymbol{S}_1$ can block-diagonalize the Gramians.

It is determined that $N_r = 8$ is the minimum for PBROM without causing the fronts to develop significant distortions.

4.2. *Envelope solutions stimulated by real input sinusoids*

In the next two cases, (ii) and (iii), we consider two simulations, one at a small wavenumber, K_s (stable), and the other at a larger wavenumber, K_u (unstable).

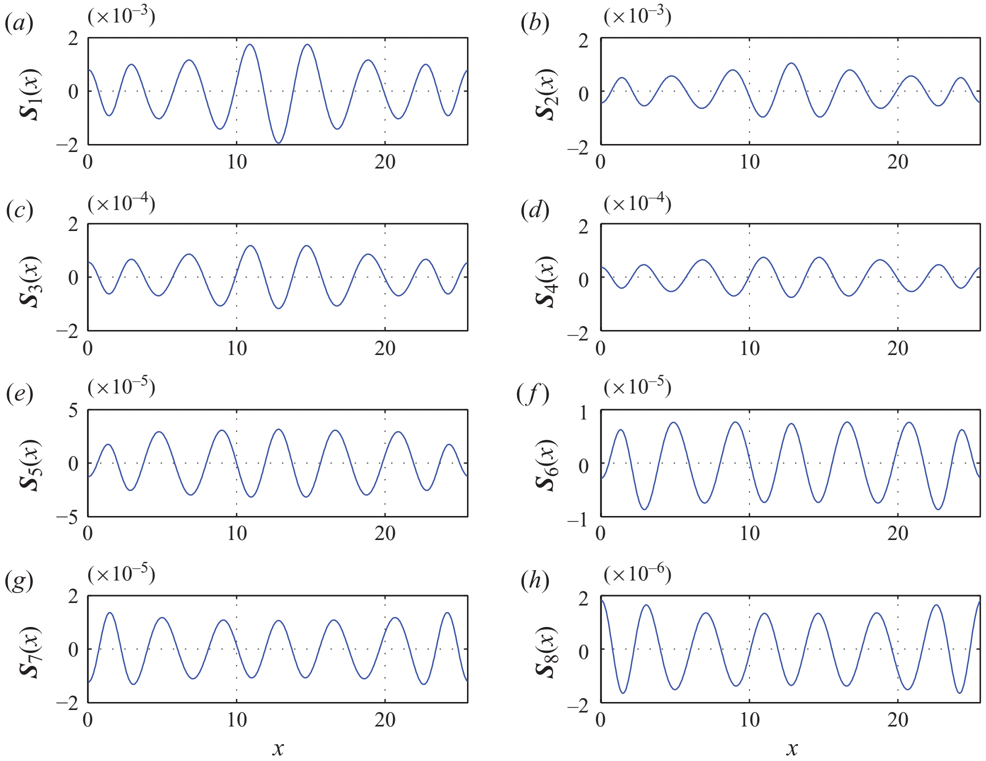


FIGURE 3. (Colour online) The corresponding pseudo-inverse basis functions of the evolving fronts (order: (*a*)–(*h*)).

For both wavenumbers, linear disturbances grow. However, the stable (unstable) wavenumber refers to the stability of the finite-amplitude solutions admissible as steady states. A stable (unstable) wavenumber is within (outside) the Eckhaus stability boundary. For the discrete system, the wavenumber is not a continuous parameter, but assumes discrete values. For $N=512$, we pick $K_s=\pm\Delta_k$ and $K_u=\pm 4\Delta_k$, where Δ_k is defined in §2. We have $1-K_s^2=0.939$ and $1-K_u^2=0.032$. For case (ii), a real spatial cosine shape is an even function $(1-K^2)^{1/2}/2N\exp(\mathrm{i}KX)+\text{c.c.}$, where $K=K_s$ for the first run and K_u for the second run. FFT is performed to generate the input vector $\boldsymbol{B}$. The output vector $\boldsymbol{C}$ remains the same as in the case (i). For a single wavenumber, $\boldsymbol{B}$ is highly sparse. We simulate the background noises by padding zero entries of $\boldsymbol{B}$ with real, small zero-mean random Gaussian numbers, with standard deviation of 10^{-4}, but with a known seed in the simulations.

Note that a real sinusoid envelope is not a steady-state solution of the amplitude equation. A real envelope solution implies that $A(X,t)$ has a zero phase. If we express the complex amplitude with a real magnitude and a real phase, i.e. $A(X,t)=|A|\exp \mathrm{i}\Theta$, then no spatial change in phase can occur in the evolving solution. For zero-phase solutions, at least three equilibria are identified, $A=0$ (static no-motion state) and $A=\pm 1$ (homogeneous convection state). There exist other equilibria as well (see figure 4).

We use $N_t=3750$, $N_i=150$ and 25 snapshots are obtained for the PBPOD. Figure 4 shows how the initial shape evolves from an unstable static equilibrium to a stable

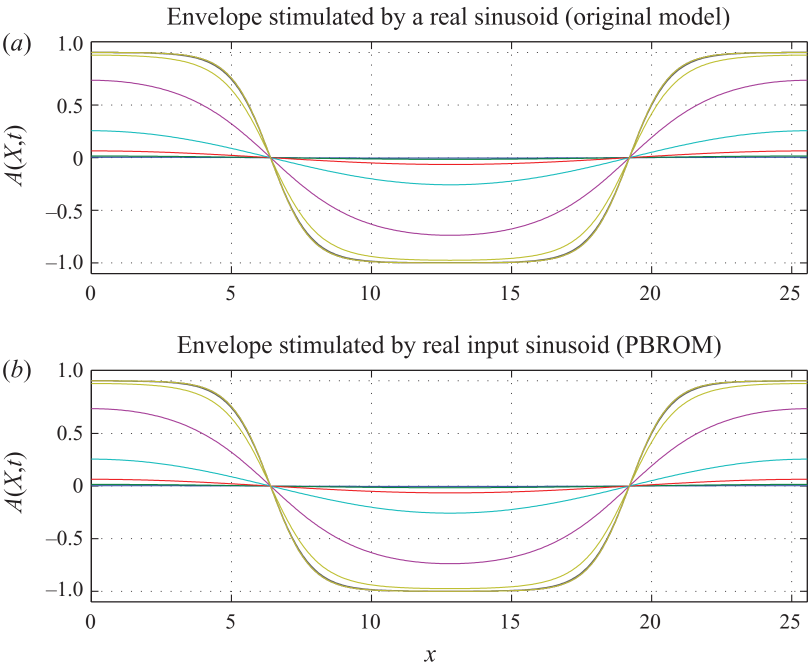


FIGURE 4. (Colour online) Comparison of the original envelope solution with the PBROM solution (stable wavenumber).

equilibrium with spatial variation with the shape of an inverted hat-shaped envelope. Figure 4 shows that the snapshots of the PBROM (*b*) agree very well with those of the original model (*a*). The PBROM retains only four states ($N_r = 4$). All four PBPOD basis functions resemble the hat shape, despite the difference in the signs and amplitudes (not shown here). The corresponding pseudo-inverse basis functions are very oscillatory, as in the case (i). If we do not use these four PBPOD basis functions, but use four sinusoids instead, the evolved envelopes shown in figure 4(*b*) do not agree with those depicted in figure 4(*a*).

In the second run, with an unstable wavenumber, the transition from the unstable to the stable equilibrium is captured over the same period of time. Again, we use $N_t = 3750$, $N_i = 150$ and 25 snapshots are generated. With background noises, the sinusoid is unstable and will evolve into a shape which is affected strongly by the Eckhaus instability. In figure 5, the envelope solution evolves from $A(X, 0) = 0$ to $A(X, \infty) = -1$. Unlike the stable-wavenumber case, this solution requires more number of reduced states for resolution. For $N_r = 15$, the reduced-order solution agrees with that of the original model solution very well. For fewer than 15 states, noticeable differences occur in the two plots.

4.3. *Envelope solutions stimulated by complex input sinusoids*

As in the case (ii), the input stimulation corresponds to a sinusoid; however, a complex sinusoid is used this time. Wavenumbers K_s and K_u are used for the first and second run, respectively. Unlike case (ii), the complex sinusoid means that the Fourier coefficients do not occur in conjugate pairs. The only non-zero complex coefficient is given by $(1 - K^2)^{1/2}/N \exp(\mathrm{i}KX)$. If the factor $1/N$ is removed, then this is actually an

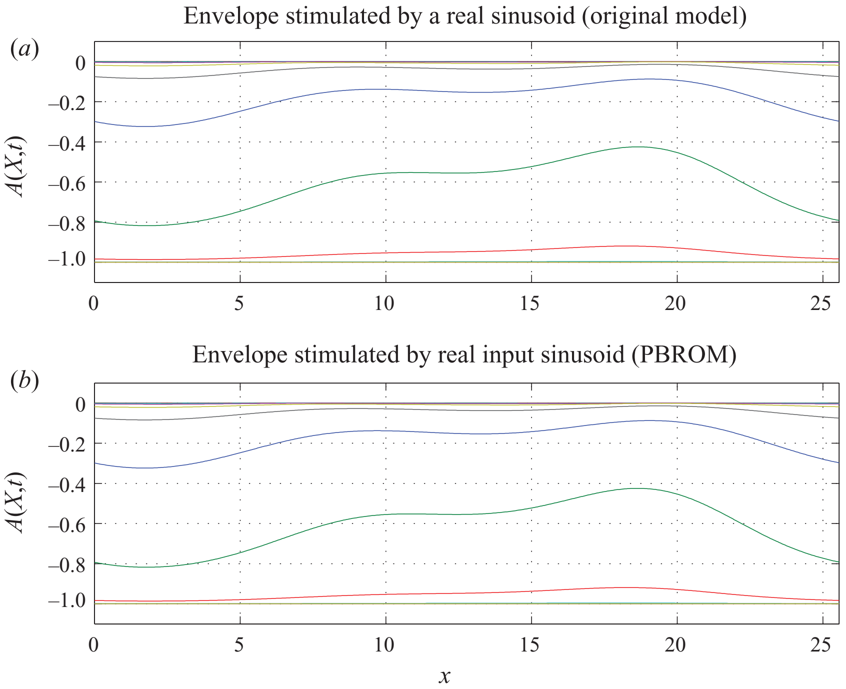


FIGURE 5. (Colour online) Comparison of the original envelope solution with the PBROM solution (unstable wavenumber).

exact solution of the amplitude equation. The complex envelope shape modulation, with the basic rolls $\exp(\mathrm{i}k_c x)$, corresponds to the modulated rolls $\exp(\mathrm{i}(k_c + \epsilon K)x)$ (see Appendix). The modulation is a shift in the wavenumber of the rolls from its critical value by ϵK.

Two runs are performed for the complex case, one at the stable wavenumber $K = K_s$ and the other at the unstable wavenumber $K = K_u$. The input vector $\boldsymbol{B}$ and output vector $\boldsymbol{C}$ are generated in the same manner as the real cases. The zeros in $\boldsymbol{B}$ are padded with small complex random numbers as in the case (ii), to mimic background noise. Unlike the real cases, the complex envelope solutions evolve more slowly in time. Therefore, a longer integration time and a larger snapshot sampling interval are used. We use $N_t = 10\,000$ and $N_i = 400$ for the two cases. A total of 25 snapshots are obtained for each case. In case (iii), we use $N_r = 4$ for the stable-wavenumber case, but $N_r = 20$ for the unstable-wavenumber case.

For the stable wavenumber, figure 6 shows that the envelopes approach the equilibrium in a large-scale complex sinusoid. This run shows that a PBROM with four complex states is remarkably accurate in capturing the original solution. As in case (ii), the four PBPOD basis functions suggest that these functions are highly and mutually non-orthogonal. For the unstable wavenumber, figure 7 shows envelopes that are very irregular and clustered. Convergence is very slow. The trend suggests that the solution will eventually reach an equilibrium with the real part of $A(X, \infty)$ settled at -1, but with the imaginary part of $A(X, \infty)$ settled at an inverted hat-shaped envelope. For this run, the PBROM requires 15 states. The agreement between the original model and the PBROM is remarkable. Unlike the cases of real envelopes,

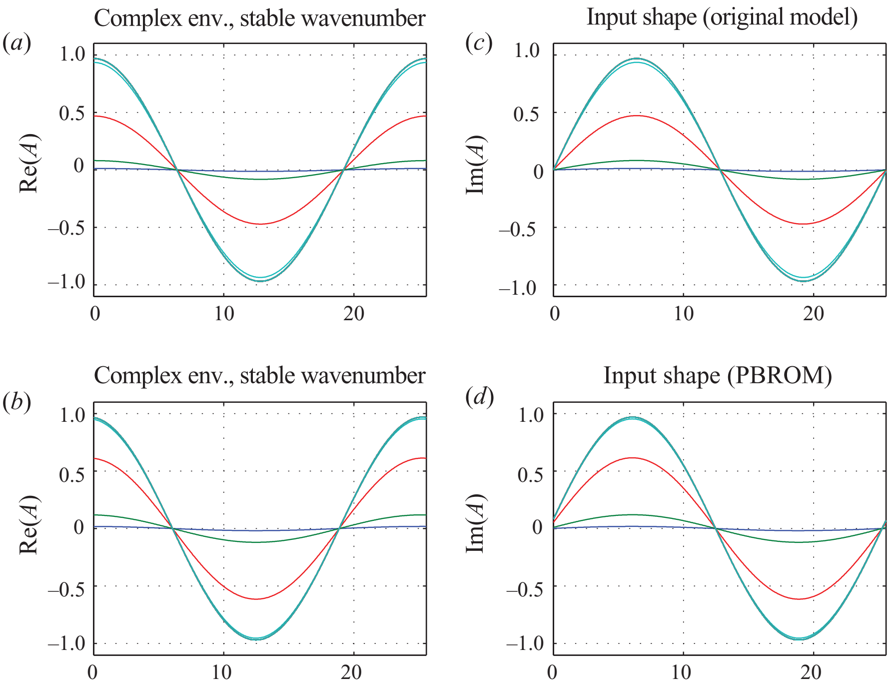


FIGURE 6. (Colour online) Comparison of the envelope solution between the original model and the PBROM (stable wavenumber).

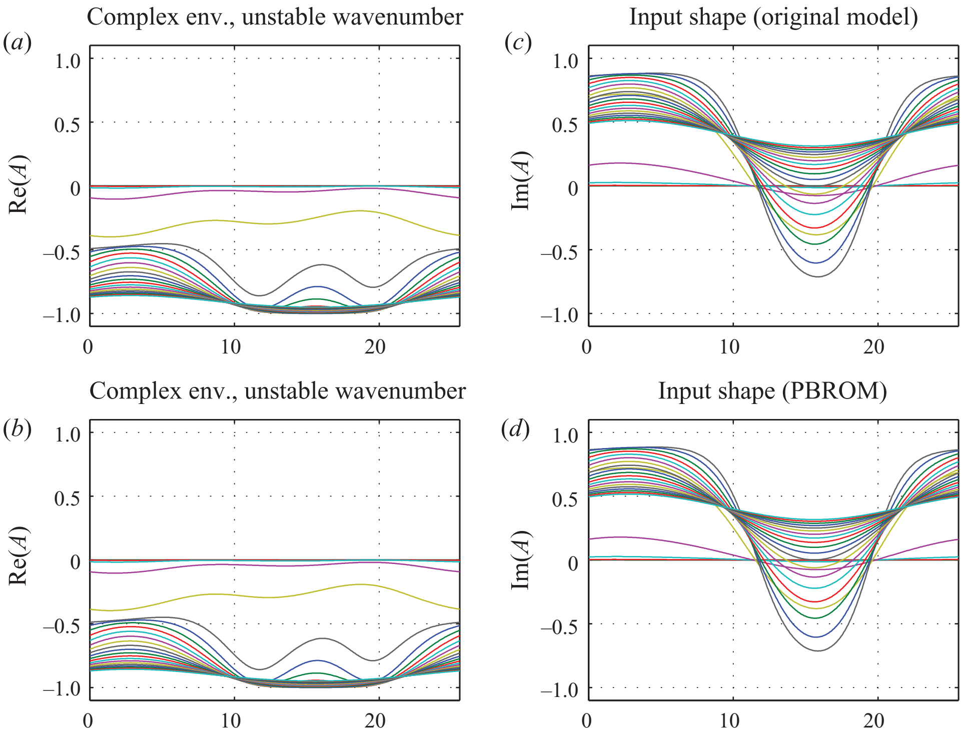


FIGURE 7. (Colour online) Comparison of envelope solution between the original model and the PBROM (unstable wavenumber).

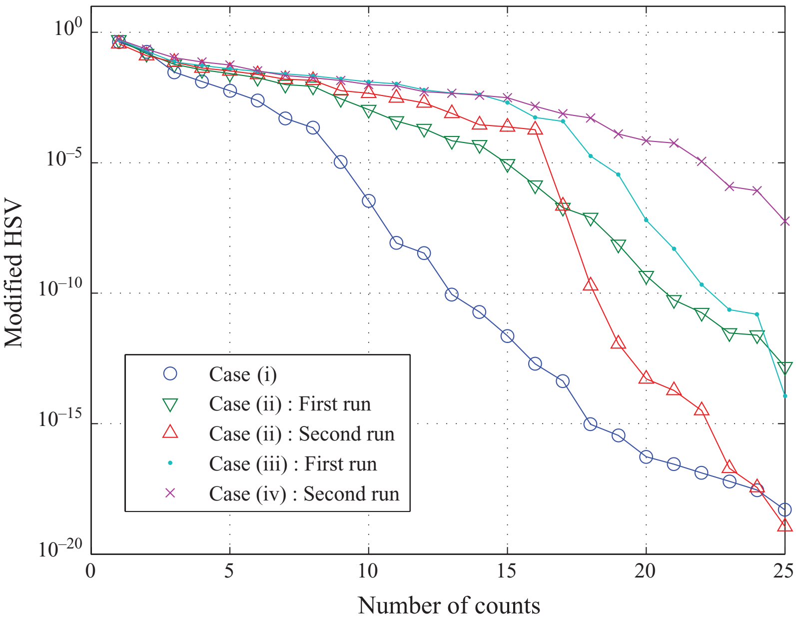


FIGURE 8. (Colour online) Modified Hankel singular values captured in the three cases (total of five runs).

the complex envelopes can shift the wavenumber of the cellular pattern away from its critical value, k_c.

After completing five simulation runs for cases (i)–(iii), it is of interest to summarize the modified Hankel singular values associated with the five PBPODs. Unlike the eigenvalues, MHSVs provide a more global measure that incorporates the effects of input stimulation and output measurement, in addition to the dynamics. Figure 8 shows five curves, each containing the magnitude distribution of the 25 MHSVs. Note that the larger the magnitude of the MHSV in each curve, the more controllable and observable is the basis function that is associated with the MHSV.

The MHSV curve of the fronts has a steeper slope than the rest, indicating a faster roll-off of the energy content of individual modes. For the fronts, the first few modes carry most of the energy. For sinusoids in the second run in case (iii), the energy content is more evenly distributed among all the modes. The second run in case (ii) shows a sharp change of characteristics of the slope of the MHSVs after the sixteenth MHSV. The choice of having to keep 15 states for the PBROM appears to be consistent with the behaviour of the MHSV curve. For case (iii), the second run shows a relatively flat MHSV slope compared with the rest of the runs. In this run, the PBROM needs 20 states for a good approximation.

Nonlinear systems in general do not obey the principle of superposition. As a result, the norm of the input vector can have an impact on the output errors. In case (i), we have already compared the envelope snapshots generated by a rectangular pulse. We consider this case to have a scale factor (SF) equal to 1. Now, we consider multiplying the input vector by an SF of 0.1 to perform the same run, and then by an SF of 10. We compare the snapshots for the same integration period. The maximal error between the original model envelopes and the corresponding envelopes in PBROM occurs at $X=0$. We plot the error at $X=0$. In figure 9, at $t=0$ and for $\mathrm{SF}=10$,

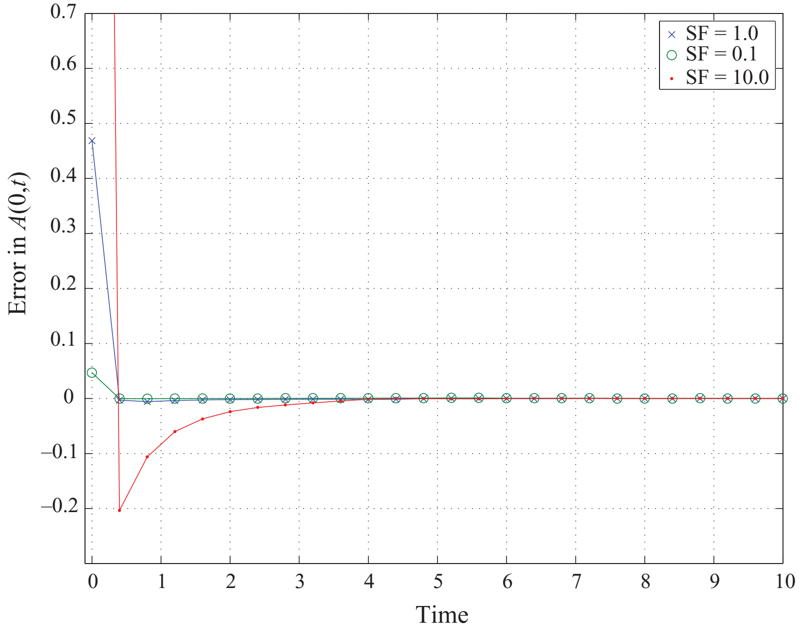


FIGURE 9. (Colour online) Errors of $A(X,t)$ in the PBROM at $X=0$ as a function of time.

the error is 4.68, well outside the box of the figure. The high value of this error is largely due to the non-smoothness of the initial rectangular pulse. While a larger SF initial condition causes significantly larger transient response in the output error, all rapidly tend to zero, as $A(0,t)$ in each case approaches the stable equilibrium value of 1. We conclude that an SF uncertainty in the initial condition does not affect the performance of the PBROM.

4.4. *Comparison of open-loop simulations*

We make a comparison of the open-loop responses based on case (iii), the complex envelopes. Using the same $\boldsymbol{B}$ and $\boldsymbol{C}$ as before and for the same prescribed input signal $u(t)$, we compare the outputs between the original model and the PBROM. For demonstration, we arbitrarily pick the following $u(t)$ with two harmonic frequencies,

$$u(t) = 2\cos(\pi t) - 0.5s\sin(\pi t) + 0.5\cos(2\pi t) - 1.2\sin(2\pi t). \tag{4.1}$$

At $t=0$, we set the initial conditions of the original model and PBROM to zero. We integrate both models forward in time and plot the evolved envelope amplitude as functions of time at three chosen locations. Figure 10 shows the real (*a*–*c*) and imaginary parts (*d*–*f*). The original-model output response is shown in solid line and the PBROM output response in dashed line. The responses are shown at $X=0$ (*a* and *d*), $X=L/4$ (*b* and *e*) and $X=L/2$ (*c* and *f*). In all three locations, the output responses between the original model (solid lines) and the PBROM (dashed lines) agree very well in that the lines in the plots are indistinguishable. Due to the fact that the richness in frequency content of input is limited, the effectiveness of PBROM demonstrated by open-loop simulations may not guarantee its effectiveness in the feedback control simulations.

4.5. *Comparison of PBPOD with other methods*

The standard POD-based reduced-order model (PROM) appears to be the simplest and easiest to construct, as long as the initial conditions can be chosen to extract

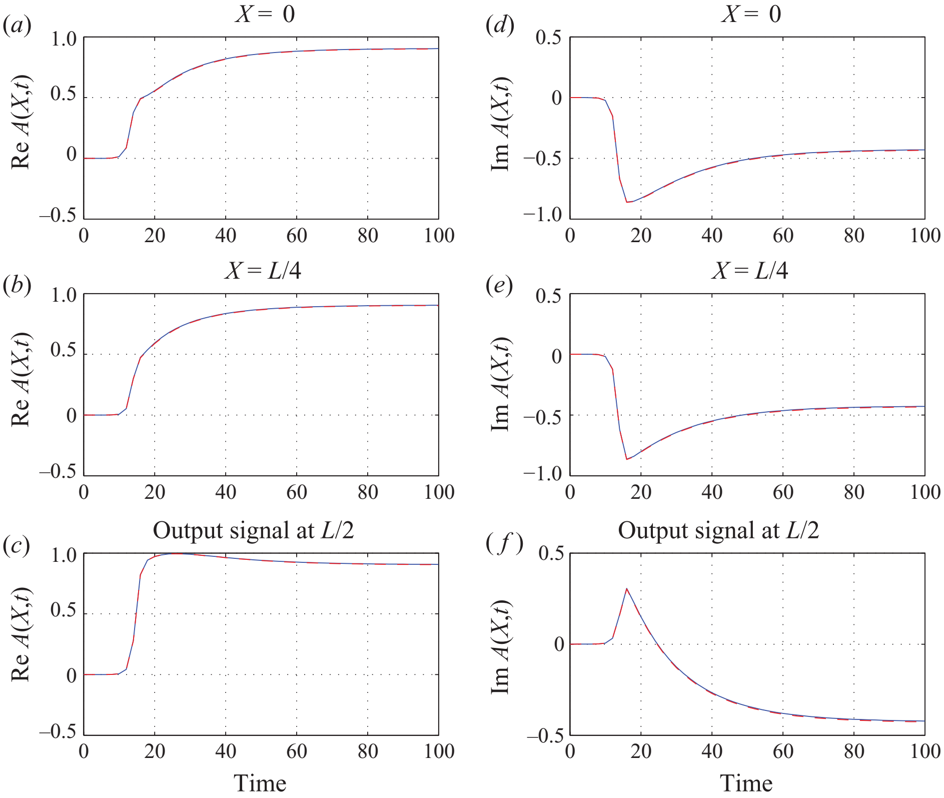


FIGURE 10. (Colour online) Comparison of amplitude response between the original model and PBROM for a complex sinusoid shape input and harmonic signal $u(t)$.

the right set of POD modes. To illustrate the input-based POD, we use case (i), the nonlinear front propagation, for a comparison study. Using the method described in § 3.2, the empirical controllability Gramian yields a set of N_c eigenvectors, $\boldsymbol{X}_j$ ($j = 1, 2, \ldots, N_c$), which forms the input-based POD modes. This set of modes and its transposes are used for projecting the original nonlinear system to form the PROM in the reduced-order space. Three cases are studied. Using the original input vector $\boldsymbol{B}$, we show in figure 11 the PROM envelopes. Figure 11(*a*) shows the front envelopes, which match very well with those shown in figure 1. This run used six POD modes for projection. The sensitivity with respect to an input SF is investigated next, since for nonlinear systems, such a sensitivity can exist. We consider a large input magnified by SF = 10, again using six POD modes. Apart from some visible wavy distortions at the bottom wall, the front envelopes basically preserve their shape (see figure 11*b*). Keeping SF = 10 but increasing to 25 POD modes, we notice that the bottom wavy distortions disappear as the resolution is improved. The three plots of figure 11 highlight the effectiveness of POD for the model reduction of the dynamics of the nonlinear fronts. Although POD does not involve balancing between controllability and observability, POD is likely to be adequate for the convection problem because its linearized dynamics is self-adjoint.

Next, we compare PBPOD with the linearized BPOD. The BPOD modes for the projection are obtained by Rowley's method (2005). Without extending the BPOD method to a linear time-varying system, we consider only the BPOD modes obtained from a linearized system about the unstable equilibrium, namely, $A_s = A(X, t) = 0$.

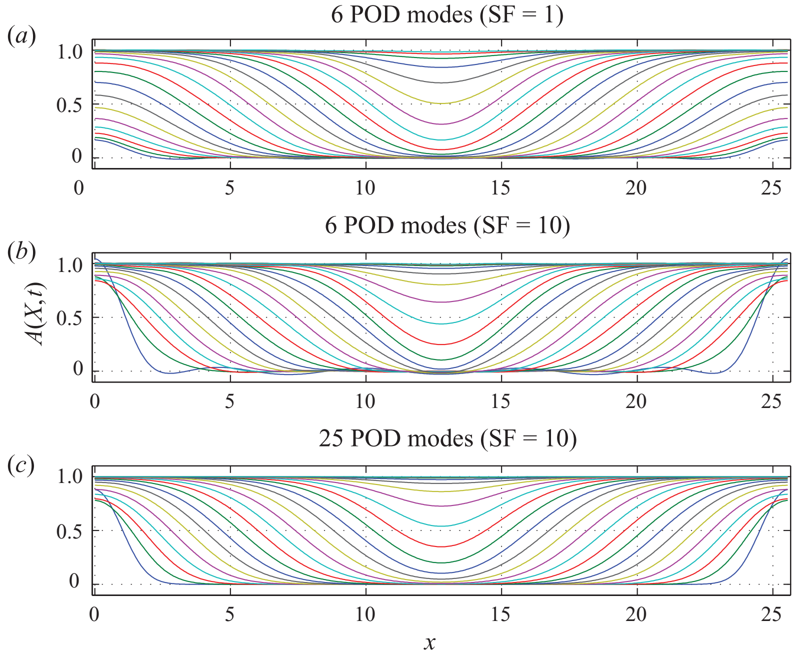


FIGURE 11. (Colour online) Reduced-order front simulations using POD with initial conditions $\boldsymbol{B}$ (*a*) and 10 times $\boldsymbol{B}$ (*b*).

This approach is suggested by Rowley (2005). It is noted that for small-order systems (see Lall *et al.* 2002), one can effectively obtain the empirical data for the construction of the full-order quantity analogous to the controllability Gramian and form the full-order transformation $\boldsymbol{T}$. In the full-order case, the transformation $\boldsymbol{S}=\boldsymbol{T}^{-1}$. Therefore, no adjoint or observability data need to be computed independently. The approximate approach of using the balanced modes from a linearized state about an equilibrium of the nonlinear systems was suggested by Rowley (2005). However, it should be noted that this procedure is not exact in the BPOD sense for the LTISS. Therefore, we refer to the resulting reduced-order model as a linearized balanced nonlinear reduced-order model (LBNROM), as opposed to BROM (see table 1).

For the amplitude equation, the linear state matrix is diagonal. The balanced modes are actually the Fourier modes. We use the same integration and sampling parameters as in the nonlinear front propagation. Figure 12 shows three runs in which each run uses a different number of BPOD modes. In case (*a*), we use only six BPOD modes. The envelopes become wavy when approaching the stable equilibrium, that is, $A(X,t)=1$. In case (*b*), the number of BPOD modes is increased to 10. The wavy behaviour subsides considerably. Lastly, in case (*c*), we use 25 BPOD modes. The wavy behaviour near the edges simply disappears with the improved resolution.

5. Feedback control of the nonlinear front propagation

The closed-loop study can require an elaborate effort. Instead of building a complete linear-quadratic-Gaussian-type controller, we take a straightforward approach here for demonstration purposes only. In the controller, instead of building a linear-quadratic regulator, a control law with a scalar proportional gain is used (see Howle 1997; Tang & Bau 1998). The control law is used in cascade with a nonlinear estimator

	All Applicable to the LTISS	Applicability to nonlinear systems
POD	Galerkin-projection-based. No balancing between input and output. May not work well for non-normal systems.	For nonlinear systems the controllable data block can be easily constructed similarly to those of linear systems. The POD modes are readily extracted.
BPOD	Block-diagonal balancing between controllability and observability. Good for non-normal systems. Provides a good approximation for balanced realization if few energetic modes dominate the dynamics.	BPOD is extended to time-periodic linear systems: see Ma, Rowley & Tadmor (2010). For nonlinear systems, one has to compute the full-order transformation and its inversion (see Lall *et al.* 2002), or use linearized BPOD or pseudo-balancing.
Linearized BPOD	Linearized BPOD is the same as BPOD when applied to LTISS	Perform BPOD on the linearized model about a certain equilibrium state to obtain BPOD modes for projecting the nonlinear system, as suggested by Rowley (2005). Its effectiveness depends on BPOD mode shapes.
PBPOD	Left upper-block balancing between controllability and observability. Hybrid between POD and BPOD. Performance for non-normal systems not demonstrated in this paper.	Expected to give a better performance than POD since it balances the left upper block of the Gramians. Observability block data or observability Gramian is unavailable.

TABLE 1. Comparing the main features of POD, BPOD and PBPOD.

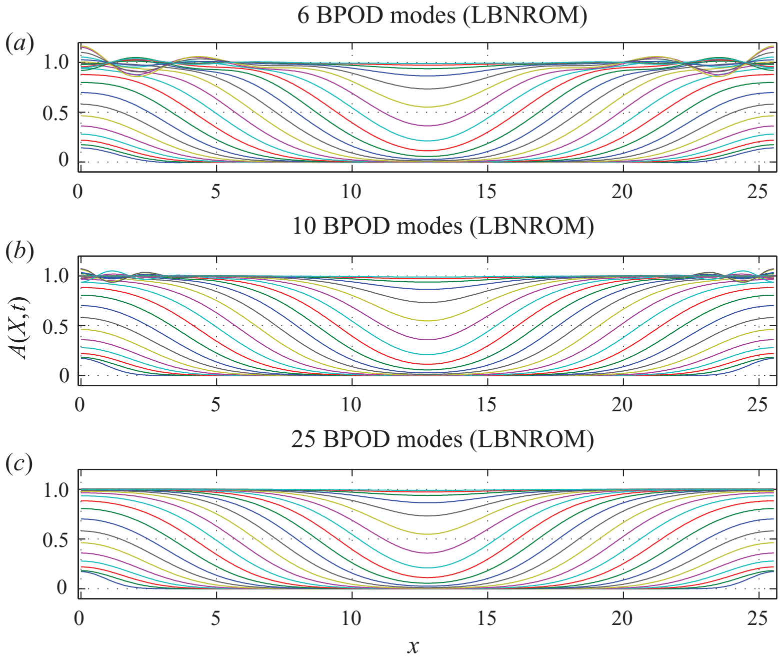


FIGURE 12. (Colour online) Comparison of three LBNROMs obtained using BPOD modes generated from linearized models at various stages of the nonlinear front propagation.

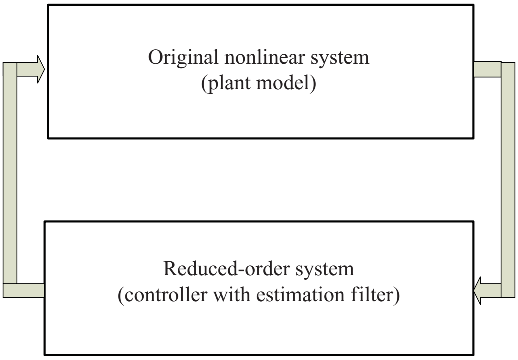


FIGURE 13. (Colour online) Feedback-control loop block diagram.

based on the reduced-order model (see our block diagram of figure 13). In this section the original model to be controlled is referred to as the plant model (upper block). The same pulse shape imposed at $t = 0$ as in § 4.1 is used as the initial condition here. However, for the current control design, we found that it is necessary to extend the measurements from one collocation point to multiple collocation points. In the development of the PBPOD, we assume that the envelope is measured only at one point. Here, we assume that measurements are available at N_r points (assumed to be evenly positioned within the end points of the layer). This number of output channels is picked because N_r represents the order of the PBROM. The output information is fed into the nonlinear, reduced-order controller (lower block). For a multiple-input–multiple-output (MIMO) configuration with N_r channels, we replace $\boldsymbol{B}u(t)$ in the SISO case of the previous sections by $\boldsymbol{B}\boldsymbol{u}(t)$, where $\boldsymbol{B}$ is now an $N \times N_r$ input matrix and $\boldsymbol{u}(t)$ is an N_r-channel control signal vector. The pseudo-balanced reduced order model developed from the previous section is now employed to build the controller, which consists of a nonlinear Kalman filter using a steady-state linear quadratic estimator (LQE) gain. The Kalman gain is computed based on a linearized model about $X = 0$.

The plant model is described by

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{B}\boldsymbol{u}, \tag{5.1}$$

where $\boldsymbol{x}$ is the true plant state vector, and according to (2.3), we have

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{L}\boldsymbol{x} - \boldsymbol{N}(\boldsymbol{x}). \tag{5.2}$$

In the plant model, we assume that the input matrix is given by

$$\boldsymbol{B} = \boldsymbol{T}_1, \tag{5.3}$$

where $\boldsymbol{T}_1$ represents the PBPOD modes (see (3.13) in § 3). The control law represents a feedback of all the reduced-order states with equal weightings,

$$\boldsymbol{u}(t) = -K_p \hat{\boldsymbol{x}}_r, \tag{5.4}$$

where K_p is a scalar constant and $\hat{\boldsymbol{x}}_r$ is the reduced-order estimate state vector of order N_r. We denote $\boldsymbol{B}\hat{\boldsymbol{x}}_r$ by $\hat{\boldsymbol{x}}_h$, which represents a reconstruction of the truth state vector based on the BPOD modes obtained from the BPOD algorithm. According to the control law, it is important to note that the control term relies solely on the reduced-order information fed to the plant. The quantity $\hat{\boldsymbol{x}}_h$ in Fourier space is inverse-transformed to $A_h(X,t)$ in physical space, which can then be converted into the actuation to be applied at the lower wall.

The nonlinear, reduced-order estimator is described by

$$\dot{\hat{\boldsymbol{x}}}_r = \boldsymbol{f}_r(\hat{\boldsymbol{x}}_r) - K_p\hat{\boldsymbol{x}}_r + \boldsymbol{K}_f(\boldsymbol{x}_m - \hat{\boldsymbol{x}}_h). \tag{5.5}$$

According to (3.27), we have

$$\boldsymbol{f}_r(\hat{\boldsymbol{x}}_r) = (\boldsymbol{S}_{p1}\boldsymbol{L}\boldsymbol{T}_1)\hat{\boldsymbol{x}}_r - \boldsymbol{S}_{p1}\boldsymbol{N}_1. \tag{5.6}$$

The second term on the right-hand side of (5.5) is $-K_p\hat{\boldsymbol{x}}_r$ because $\boldsymbol{S}_{p1}\boldsymbol{B}=\boldsymbol{I}$. The last term on the right-hand side involves $\boldsymbol{K}_f$, a steady-state $N_r \times N$ Kalman gain computed from an LQE model, in which the state matrix used in the filter is obtained by linearization of $\boldsymbol{f}_r(\hat{\boldsymbol{x}}_r)$ of (5.5) about the unstable equilibrium $A(X,t)=0$. The $N\times 1$ measurement vector $\boldsymbol{x}_m$ is generated as follows. Consider the output of the envelope values at N_r evenly spaced collocation points within the layer, that is, $A(X_j,t)$, where $j=1,\ldots,N_r$; N_r is the order of the reduced-order model. Referring to (3.25), the explicit Fourier relationship between the envelope values and the reduced-order states is $\boldsymbol{a}_1=\boldsymbol{F}_i\boldsymbol{T}_1\boldsymbol{x}_r$. Here, we consider $\boldsymbol{x}_r$ as the truth version of the estimated state vector $\hat{\boldsymbol{x}}_r$. We denote the N_r entries of $\boldsymbol{a}_1$ at which the envelope is measured by a vector $\boldsymbol{a}_r$, and the explicit form of the $N_r \times N$ inverse Fourier transformation matrix by $\boldsymbol{F}_{ir}$. Upon inverting the matrix, we obtain $\boldsymbol{x}_r=(\boldsymbol{F}_{ir}\boldsymbol{T}_1)^{-1}\boldsymbol{a}_r$. Since $\boldsymbol{x}\approx\boldsymbol{T}_1\boldsymbol{x}_r$, we have derived the measured counterpart of $\hat{\boldsymbol{x}}_h$ as $\boldsymbol{x}_m$, given by

$$\boldsymbol{x}_m = \boldsymbol{T}_1(\boldsymbol{F}_{ir}\boldsymbol{T}_1)^{-1}\boldsymbol{a}_r. \tag{5.7}$$

The alternative form of the nonlinear, reduced-order estimator of (5.5) is

$$\dot{\hat{\boldsymbol{x}}}_r = \boldsymbol{f}_r(\hat{\boldsymbol{x}}_r) - K_p\hat{\boldsymbol{x}}_r + \boldsymbol{K}_f\boldsymbol{T}_1((\boldsymbol{F}_{ir}\boldsymbol{T}_1)^{-1}\boldsymbol{a}_r - \hat{\boldsymbol{x}}_r), \tag{5.8}$$

and $\boldsymbol{a}_r$ contains the actual measured envelope values.

In the controller model, PBROM has $N_r=8$ pseudo-balanced modes, which are constructed from the original model of order $N=512$ based on the PBPOD method. The pair of transformation matrices ($\boldsymbol{S}_{p1}$, $\boldsymbol{T}_1$) obtained using PBPOD are precomputed and stored to be used for building the estimator model. The $N_r \times N$ steady-state Kalman gain matrix is determined using the MATLAB linear quadratic estimator (LQE) function. Both noise covariance matrices are assumed to be diagonal with entries 10^{-2}. The process and measurement noises are assumed to be uncorrelated. In the plots, the envelope amplitude $A(X,t)$ of the original model corresponds to the true state vector, $\boldsymbol{x}$, whereas the PBROM-reconstructed amplitude based on estimation, $A_h(X,t)$, corresponds to $\hat{\boldsymbol{x}}_h$, which depends on the reduced-order estimated state vector, $\boldsymbol{x}_r$.

We first show that $A_h(X,t)$ tracks $A(X,t)$ by letting $A_h(X,0)=0$ at $t=0$, corresponding to the estimator having no *a priori* knowledge of the plant initial condition. In this case we let $\hat{\boldsymbol{x}}_r(0)=\boldsymbol{0}$. We zero out the proportional gain ($K_p=0$) as well. The length of the simulation consists of 10 time units, with a time step of 0.0005. The simulation has 2×10^4 steps. The snapshots are sampled at an interval of 800 steps. A total of 25 snapshots are generated and shown in the plots. Figure 14 (*a,b*) corresponds to the evolved envelopes of the plant model and estimator model, respectively. Figure 14(*c*) shows the estimator error. In figure 14(*a–c*), the last of the 25 envelopes is shown in a dotted line to highlight the final envelope shape in the simulation. The estimator error converges to zero.

Next, we study two cases in the form of closed-loop responses, in both cases we use $K_p=2$. In the first case, no estimator state knowledge is assumed so that the initial

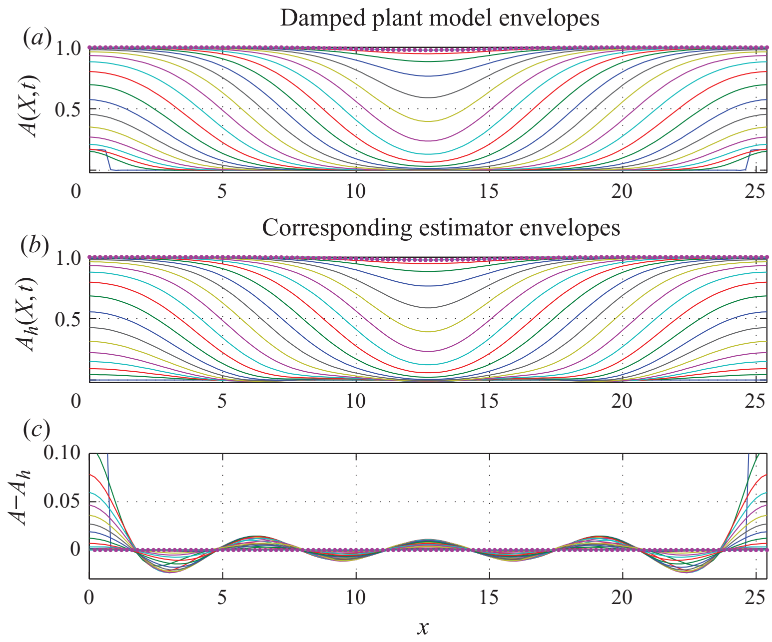


FIGURE 14. (Colour online) The performance of the Kalman filter.

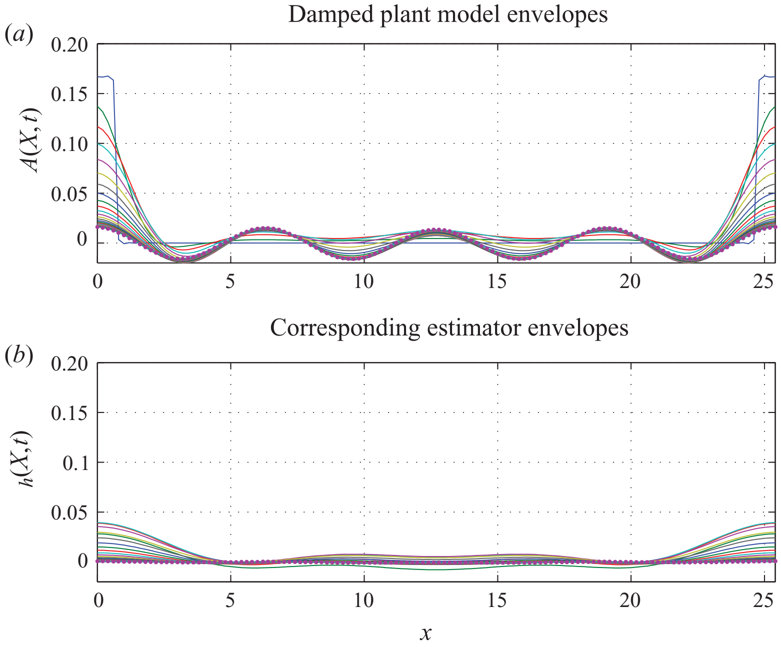
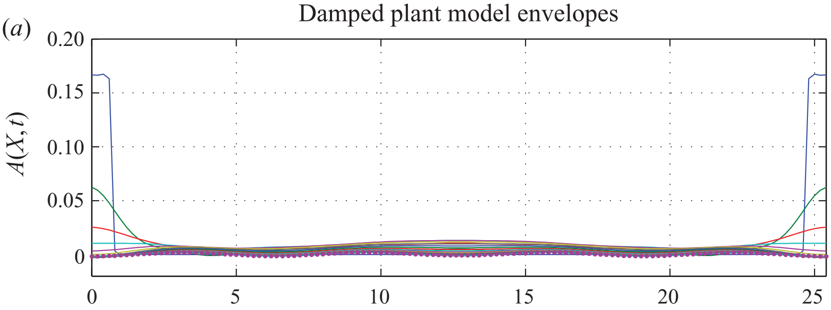


FIGURE 15. (Colour online) Feedback-control responses with zero initial estimator states.

estimate states are set to zero at $t = 0$. Figure 15 shows that the envelope in the plant model starts to damp, but then settles to a residual behaviour that shows a small spatial oscillation (a). The reconstructed envelope of the estimator, $A_h(X, t)$, initially

Damped plant model envelopes



Corresponding estimator envelopes

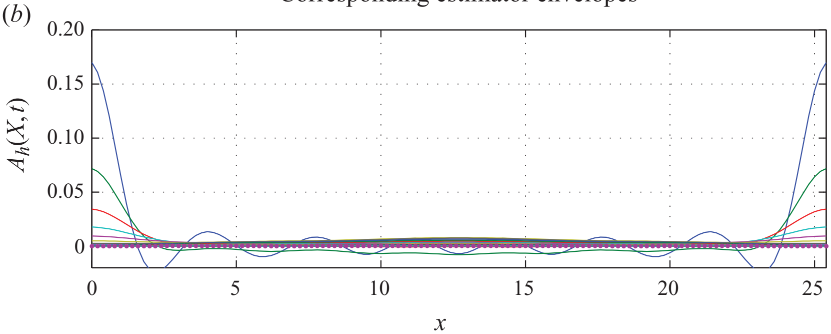


FIGURE 16. (Colour online) Feedback-control responses with initial impulsive estimator states.

starts to grow, but then appears to damp to zero (see *b*). The residual oscillatory behaviour persists even by increasing K_p to 10 times.

In the second case, we assume a certain knowledge of the initial condition in the plant model. The rectangular pulse used initially to generate the fronts is used to obtain an initial condition for the estimator, with a scalar factor to increase the pulse peak by 30 % to allow for a certain margin of uncertainty. With some knowledge of the plant initial condition, figure 16 shows a substantial improvement in the residual estimator behaviour. The oscillatory residual in the plant envelopes disappears. The envelopes converge to zero.

6. Conclusion

We have investigated the order-reduction of the dynamical model governing the spanwise amplitude modulation of convection patterns by the PBPOD. For performance, different representative amplitude modulation patterns have been considered. A few PBPOD modes generated by the method are capable of capturing the envelope shapes very well by utilizing a small number of snapshots in the sampled data.

Using the front propagation as an example, we have compared the performance of PBPOD to POD, as well as PBPOD to the linearized BPOD. For an initial condition magnified by a scale factor, PBPOD performs better than POD when using the same small number of projective modes. For the nominal initial condition, or when using more POD modes, both PBPOD and POD perform very well. When using the same small number of projective modes, PBPOD and POD outperform the linearized BPOD, in which the balanced modes correspond to the Fourier modes. However, as the number of linearized BPOD modes is increased, the linearized BPOD performance is greatly improved.

The open-loop simulations are used to compare the outputs between the original model and the PBROM, using the same prescribed control signal input. The only drawback of the open-loop simulations is that the input signal cannot be rich enough in frequency content as to guarantee the desired closed-loop responses if a reduced-order controller is used.

Finally, we provide a feedback-control study. The knowledge of the envelope shape is assumed at several collocation points at the plant output. A nonlinear low-order estimator is used to estimate the envelope shape based on the small number of PBPOD modes obtained. For demonstration, instead of building a sophisticated control law, we only use a simple proportional feedback relationship. Using the nonlinear front example, we show that the controller is effective in damping out the nonlinear disturbance envelopes by the closed-loop actions. The performance of feedback control is further enhanced if the estimator assumes a certain knowledge about the initial disturbance shapes of the plant model. If the nonlinear filter starts with a zero estimate, then a small residual oscillation will remain in the plant under the closed-loop actions.

We thank Professor Clancy Rowley for his in-depth review of the submitted paper. The earlier version contains a few key presentation errors of our method. With Professor Rowley's help, § 3 of this paper provides an enhanced description of the model reduction methods. This work was supported by the Air Force Office of Scientific Research under contract FA9550-05-C-0031.

Appendix. The input/output amplitude equation

A.1. *The small-amplitude expansions*

The derivation provides boundary-value solutions at each order, which are valid for the free–free (FF) as well as rigid–rigid (RR) boundary conditions. Well-known derivations in the literature, for example, by Newell & Whitehead 1969 (NW) and by Cross & Hohenberg 1993 (CH), only treat the FF case. The present derivation also incorporates the input and output elements.

The Boussinesq equations for two-dimensional convection rolls at infinite Prandtl number (Pr) is considered,

$$\left.\begin{aligned} &\partial_t \boldsymbol{v} + \boldsymbol{L}\boldsymbol{v} = \boldsymbol{N}(\boldsymbol{v}), \ \text{where } \boldsymbol{L} = \begin{bmatrix} \nabla^4 & R\nabla_1^2 \\ -1 & -\nabla^2 \end{bmatrix}, \\ &\boldsymbol{N}(\boldsymbol{v}) = \begin{bmatrix} 0 \\ -u\partial_x\theta - w\partial_z\theta \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} w \\ \theta \end{bmatrix}. \end{aligned}\right\} \tag{A 1}$$

The horizontal velocity field, $u(x, z, t)$, is obtained if the vertical velocity field $w(x, z, t)$ is known, through the continuity equation

$$\nabla \cdot \boldsymbol{u} = \partial_x u + \partial_z w = 0. \tag{A 2}$$

The boundary conditions for the velocity at both walls are either of RR or FF type, respectively,

$$\left.\begin{aligned} w(x, z, t) = \partial_z w(x, z, t) = 0 \quad &\text{at} \quad z = 0, 1, \\ w(x, z, t) = \partial_{zz} w(x, z, t) = 0 \quad &\text{at} \quad z = 0, 1. \end{aligned}\right\} \tag{A 3}$$

The temperature conditions are isothermal at the upper wall ($z=1$) and control prescribed at the lower wall ($z=0$),

$$\theta(x,z,t)=0 \quad \text{at } z=1. \tag{A 4}$$

The lower-wall temperature condition incorporate at a control $\theta_c(x,y,t)$ ($z=0$),

$$\theta(x,z,t)=\theta_c(x,t) \quad \text{at } z=0. \tag{A 5}$$

Linear stability analysis determines a parabolic-shaped neutral curve, $R=R_0(k)$, with a minimum at the critical wavenumber and Rayleigh number, (k_c, R_c). For $R>R_c$, $R=R_c+\epsilon^2R_c$ defines a small parameter ϵ, which is of the order of the symmetrical half-bandwidth of the unstable wavenumber, k, by allowing a scaling $k=k_c+\epsilon K$. To describe waveforms associated with the unstable band, we introduce a stretched coordinate $X=\epsilon x$ and a slow-time $T=\epsilon^2 t$. Therefore, we express $\partial_x \rightarrow \partial_x+\epsilon\partial_X$, $\partial_t \rightarrow \epsilon^2\partial_T$, ∇^2 by $\nabla^2+\epsilon 2\partial_x\partial_X+\epsilon^2\partial_{XX}$, and ∇^4 by $\nabla^4+\epsilon 4\partial_x\partial_X\nabla^2+\epsilon^2(4\partial_{xx}\partial_{XX}+2\partial_{XX}\nabla^2)+\cdots$ and so on.

CH expands upon (A 1). Equation (A 1) is expanded in the form $\boldsymbol{L}=\boldsymbol{L}_0+\epsilon\boldsymbol{L}_1+\boldsymbol{L}_2+\cdots$ and $\boldsymbol{v}=\epsilon\boldsymbol{v}_0+\epsilon^2\boldsymbol{v}_1+\epsilon^3\boldsymbol{v}_2+\cdots$, respectively. The difficulty with this approach occurs at the $O(\epsilon^2)$ balance, with $\boldsymbol{L}_0\boldsymbol{v}_1=-\boldsymbol{L}_1\boldsymbol{v}_0+\boldsymbol{N}_0$. We substitute a horizontal dependence of $\exp \mathrm{i}k_cx$ and the vertical dependence gives $\boldsymbol{v}_0=[f_0(z), g_0(z)]$. The adjoint solution is $\tilde{\boldsymbol{v}}=[f_0(z), k_c^2R_cg_0(z)]$. At the critical point, we obtain $\langle\tilde{\boldsymbol{v}}'\boldsymbol{L}_1\boldsymbol{v}_0\rangle=0$, but noting that $\boldsymbol{L}_1\boldsymbol{v}_0 \neq \boldsymbol{0}$ constitutes a resonant term. To determine the $O(\epsilon^2)$ solution, we go back to the linear balance, $\boldsymbol{L}_0\boldsymbol{v}_0=\boldsymbol{0}$, and differentiate this equation with respect to k; we obtain $\boldsymbol{L}_1\boldsymbol{v}_0=-\boldsymbol{L}_0\partial_k\boldsymbol{v}_0$, where $\boldsymbol{L}_1=\partial_k\boldsymbol{L}_0$ at $k=k_c$. Therefore, at the critical point, the $O(\epsilon^2)$ equation is $\boldsymbol{L}_0(\boldsymbol{v}_1+\partial_k\boldsymbol{v}_0)=\boldsymbol{N}_0$. Thus, $\boldsymbol{v}_1$ will consist of two terms. The first term is generated by the nonlinear effect. The second term is contributed by a term $\partial_k\boldsymbol{v}_0$. When carried on to $O(\epsilon^3)$ balance, with the equation $\boldsymbol{L}_0\boldsymbol{v}_2=-\boldsymbol{L}_2\boldsymbol{v}_0-\boldsymbol{L}_1\boldsymbol{v}_1+\boldsymbol{N}_1$, the second term will contribute to $-\boldsymbol{L}_1\boldsymbol{v}_1$. Significantly, more algebraic steps are involved to carry this second term, which turns out to affect only the diffusive-term coefficient of the amplitude equation.

In NW, (A 1) is converted to a sixth-order scalar PDE in a single variable, in w (or equivalently, in θ),

$$\mathscr{L}w = R\nabla_1^2(\partial_t\theta+u\partial_x\theta+w\partial_z\theta), \tag{A 6}$$

where $\mathscr{L}=(R\nabla_1^2-\nabla^6)$ with appropriate boundary conditions.

We expand $\mathscr{L}$, R and θ in ϵ about the critical point,

$$\left.\begin{aligned} \mathscr{L} &= \mathscr{L}_0+\epsilon\mathscr{L}_1+\epsilon^2\mathscr{L}_2+\cdots, \\ R &= R_c+\epsilon^2R_c, \\ w = \epsilon w_0+\epsilon^2w_1+\epsilon^3w_2+\cdots, &\quad \theta=\epsilon\theta_0+\epsilon^2\theta_1+\epsilon^3\theta_2+\cdots. \end{aligned}\right\} \tag{A 7}$$

Note that if R is expanded about an arbitrary point, then its expansion will contain odd-parity terms as well. Note that $R=R_c+\epsilon^2R_c$ serves as a definition for ϵ here. The boundary conditions for inverting the sixth-order operator $\mathscr{L}_0$ are

$$w_0=\Delta^2w_0=0, \quad \text{at } z=0,1, \tag{A 8}$$

and

$$\partial_{zz}w_0=0 \quad \text{at } z=0,1 \quad \text{for FF} \quad \text{or} \quad \partial_zw_0=0 \quad \text{at } z=0,1 \quad \text{for RR}. \tag{A 9}$$

The linear balance is determined from $\mathscr{L}_0 w_0 = 0$, which gives a solution

$$\left.\begin{aligned} u_0 &= -\frac{1}{\mathrm{i}k_c} A(X,t) f_0'(z) \exp \mathrm{i}k_c x + \text{c.c.}, \\ w_0 &= A(X,t) f_0(z) \exp \mathrm{i}k_c x + \text{c.c.}, \\ \theta_0 &= A(X,t) g_0(z) \exp \mathrm{i}k_c x + \text{c.c.}, \end{aligned}\right\} \tag{A 10}$$

where $f'(z)$ denotes the z-derivative of $f(z)$ and c.c. stands for complex conjugate, u_0 is determined from w_0 via continuity equation. With the FF boundary condition and with the maximum of $f_0(z)$ (assumed positive) normalized to one, we have $f_0(z) = \sin(\pi z)$, for $0 \leqslant z \leqslant 1$ and $g_0(z) = 2/(3\pi^2)\sin(\pi z)$. It is well known that we have $R_c = 27\pi^4/4$ and $k_c = \pi/\sqrt{2}$. With RR boundary condition, $f_0(z)$ and $g_0(z)$ are real functions to be solved numerically, with $R_c = 1707.762$ and $k_c = 3.117$, approximately.

The second-order balance in ϵ gives

$$\mathscr{L}_0 w_1 = -\mathscr{L}_1 w_0 - k_c^2 R_c (u_0 \partial_x \theta_0 + w_0 \partial_z \theta_0). \tag{A 11}$$

We note that the solution is different for the same-order expansion given by (A 1). We expect $\mathscr{L}_1 w_0 = 0$ is satisfied at (k_c, R_c), because by differentiating the characteristic equation we have $\mathscr{L}_0(k) = 0$ (that defines the neutral curve) with respect to k, we have

$$\mathscr{L}_0'(k) = -2kR_0(k) - k^2 R_0'(k) - 3\Delta^2(k)\Delta'(k) = \mathscr{L}_1(k) - k^2 R_0'(k) \equiv 0, \tag{A 12}$$

where we have $\Delta(k) = (\mathrm{d}^2/\mathrm{d}z^2 - k^2)$. At $k = k_c$, we have $R_0'(k_c) = 0$ because it is the minimum, thus we have $\mathscr{L}_1(k_c) = 0$.

The $O(\epsilon^2)$ solution is given by

$$\left.\begin{aligned} &u_1 = -\frac{1}{\mathrm{i}2k_c} A^2 f_1'(z) \exp \mathrm{i}2k_c x + \text{c.c.}, \quad w_1 = A^2 f_1(z) \exp \mathrm{i}2k_c x + \text{c.c.}, \\ &\theta_1 = |A|^2 g_{1m}(z) + A^2 g_1(z) \exp \mathrm{i}2k_c x + \text{c.c.}, \end{aligned}\right\} \tag{A 13}$$

where g_{1m} represents the x-mean temperature solution. For FF, $f_1 \equiv g_1 \equiv 0$ is satisfied exactly, and $g_{1m} = 1/(3\pi^3)\sin 2\pi z$. For RR boundary conditions, f_1 and g_1 do not vanish, but their contribution to the $O(\epsilon^3)$ solvability condition is about two orders of magnitude smaller than the contribution by the x-mean term.

For the $O(\epsilon^3)$ balance, we have

$$\mathscr{L}_0 w_2 = -\mathscr{L}_2 w_0 - k_c^2 R_c \partial_T \theta_0 - k_c^2 R_c (u_1 \partial_x \theta_0 + w_1 \partial_z \theta_0 + u_0 \partial_x \theta_1 + w_0 \partial_z \theta_1)_1 + \cdots, \tag{A 14}$$

where $(.)_1$ denotes resonant terms only (proportional to $\exp \mathrm{i}k_c x$) generated in the expression (.). Note that the term $\mathscr{L}_1 w_1$ does not contain any resonant terms and can be discarded. From the expansion, $\mathscr{L}_2(k_c)$ has the following form:

$$\mathscr{L}_2(k_c) = -k_c^2 R_c - 0.5\mathscr{R}_2, \quad \mathscr{R}_2(k_c) = \left(-R_c - 12k_c^2\Delta + 3\Delta^2\right)\partial_{XX}. \tag{A 15}$$

Note that $\mathscr{R}_2 f_0$ is equal to $R_c'' f_0$, where we have $R_c'' = (\mathrm{d}^2 R_0(k)/\mathrm{d}k^2)_{k=k_c}$. For FF, we have $R_c'' = 36\pi^2$.

A.2. *Derivation of the coefficients of the amplitude equation*

Without the control input, the $O(\epsilon^3)$ balance yields the following amplitude equation as the solvability condition:

$$k_c^2 R_c \langle f_0 g_0 \rangle \partial_T A = k_c^2 R_c \langle f_0^2 \rangle A + \tfrac{1}{2} k_c^2 R_c'' \langle f_0^2 \rangle A_{XX} - k_c^2 R_c \langle (\beta_1 + \beta_2) f_0 \rangle |A|^2 A, \tag{A 16}$$

where $\beta_1(z)$ is associated with the mean temperature profile, and $\beta_2(z)$ is associated with the resonant term generated by the second harmonics, i.e. proportional to

Coefficient/Reference	BC	NW	CH	SLB	Current
$g'_{1m}(0)$	FF	$2/(3\pi^2)$	$2/(3\pi^2)$	–	$2/(3\pi^2)$
	RR	–	–	–	0.040906
τ_0	FF	$2/(3\pi^2)$	$2/(3\pi^2)$	–	$2/(3\pi^2)$
	RR	–	–	–	0.05014
ξ_0^2	FF	$8/(3\pi^2)$	$8/(3\pi^2)$	–	$8/(3\pi^2)$
	RR	–	–	–	0.19973
χ_0	FF	$1/(3\pi^2)$	$8/(3\pi^2)$	–	$1/(3\pi^2)$
	RR	–	–	–	0.03009
$(Nu-1)R_c/$ $(R-R_c)$	FF	–	1/0.5000	–	1/0.5000
	RR	–	–	1/0.6994*	1/0.74522
γ	FF	–	–	–	$4/(3\pi)$
	RR	–	–	–	0.31364

TABLE 2. Comparison of the coefficients of the amplitude equation.

$\exp \mathrm{i}2kx$. Their expressions are, respectively,

$$\beta_1(z) = f_0 g'_{1m}, \;\; \beta_2(z) = 0.5 f'_1 g_0 + 2 f'_0 g_1 + f_1 g'_0 + f_0 g'_1. \tag{A 17}$$

The amplitude equation is typically expressed in the standard form,

$$\tau_0 \partial_T A = A + \xi_0^2 A_{XX} - \chi_0 |A|^2 A, \tag{A 18}$$

with the coefficients given by

$$\tau_0 = \langle f_0 g_0 \rangle / \langle f_0^2 \rangle, \; \xi_0^2 = 0.5 R''_c / k_c^2 R_c, \; \chi_0 = -\langle f_0 (\beta_1 + \beta_2) \rangle / \langle f_0^2 \rangle. \tag{A 19}$$

Note that for FF, $\beta_2(z) \equiv 0$, but for RR, $\beta_2(z)$ contributes. The coefficients from NW, CH and the current results are tabulated in table 2.

The numerical scheme developed for computing the various orders of solution is identical for FF and RR except for the boundary conditions that distinguish between the two. The scheme used 75 Chebyshev modes for the z-dependence functions. The factor of 8 difference in χ_0 between CH and ours is due to the different scaling factor for normalizing f_0. We have $f_0 = 1$ at maximum, whereas CH has $2\sqrt{2}$ instead.

Convection heat transport is defined by the Nusselt number, Nu,

$$Nu - 1 = \epsilon^2 |A|^2 |g'_{1m}(0)|. \tag{A 20}$$

For steady rolls modulation, i.e. $A = (1 - K^2)^{1/2} \exp(\mathrm{i}KX)$, (A 18) gives

$$|A|^2 = \frac{1 - K^2}{\chi_0}. \tag{A 21}$$

By definition $\epsilon^2 = (R - R_c)/R_c$. Eliminating ϵ and $|A|^2$,

$$\frac{(Nu - 1)R_c}{R - R_c} = \frac{g'_{1m}(0)}{\chi_0}(1 - K^2). \tag{A 22}$$

We obtain $(Nu-1)R_c/(R-R_c) = 1/0.74522$ for RR. If we neglect β_2 the value becomes 1/0.74636. Schlüter, Lortz & Busse (1967) give a slightly different $(Nu-1)R_c/(R-R_c) = 1/0.6994$, which appeared in many references, for example, Drazin & Reid (1982). This value for FF can be derived analytically, equal to 2 exactly, which checks our numerical scheme.

A.3. *Derivation of the input/output elements of the amplitude equation*

The amplitude equation governs envelope modulation of the convection cells. Therefore, the control should affect the wave band about the critical point rather than shift the critical point. Inspecting the solvability conditions indicates that the control affects the dynamics at $O(\epsilon^3)$ balance. In contrast, the stability threshold shifts according to the $O(\epsilon^0)$ balance (see Or *et al.* 2001).

To incorporate the input and output, the results from the previous section have to be modified to incorporate the following $O(\epsilon^3)$ control temperature boundary condition, (A 5), at $z=0$,

$$\theta_2(x,0,t) = \theta_c(x,t) = \epsilon^3 b(X,T)\exp(\mathrm{i}k_c x) + \text{c.c.}, \tag{A 23}$$

where $b(X,T)$ is the input-shape function prescribed at the lower wall. The lower order functions remain unchanged, the high-order function, $g_2(z)$, satisfies $g_2(0)=1$ and $g_2(1)=0$. We consider the $O(\epsilon^3)$ balance of (A 1), noting that the control appears only as a non-homogeneous term associated with the linear term $\boldsymbol{L}_0\boldsymbol{v}_2$. Multiplying both sides of the $O(\epsilon^3)$ equation by the adjoint solution, $\tilde{\boldsymbol{v}}$ and integrating over volume, we then evaluate $\langle \tilde{\boldsymbol{v}}^T \boldsymbol{L}_0 \boldsymbol{v}_2 \rangle$, by integration by parts. The explicit form of this term is

$$\left\langle f_0\left(f_2^{iv} - 2k_c^2 f_2'' + k_c^4 f_2 - k_c^2 R_c g_2\right)\right\rangle + \left\langle k_c^2 R_c g_0\left(-f_2 - \left(g_2'' - k_c^2 g_2\right)\right)\right\rangle. \tag{A 24}$$

Integration by parts of the above integrals and application of the boundary conditions result in the only non-vanishing contribution from the control boundary condition, (A 23), $-k_c^2 R_c g_0(0)' g_2(0)$. We define $\gamma = k_c^2 R_c g_0(0)' g_2(0)$; then for FF we obtain $\gamma = 4/3\pi$, and for RR we obtain $\gamma = 0.31364$. With control input, the amplitude equation (A 16) becomes

$$\tau_0 \partial_T A = A + \xi_0^2 A_{XX} - \chi_0 |A|^2 A + \gamma b. \tag{A 25}$$

The term $b(X,T)$ provides the prescribed temporal-spatial input. Equivalently, using the sixth-order PDE (A 14), we evaluate the integral $< w_0 \mathscr{L}_0 w_2 >$, where in (A 23), the expression $\theta_2 = \theta_c$ becomes

$$\theta_2(x,0,t) = \Delta^2(k) w_2(x,0,t) = \theta_c(x,t) = \epsilon^3 b(X,T)\exp(\mathrm{i}k_c x) + \text{c.c.} \tag{A 26}$$

We obtain the same control term, $-k_c^2 R_c g_0'(0) g_2(0)$.

The output elements correspond to the mid-layer temperature at an array of evenly spaced collocation points,

$$y_j(T) = g_0\left(\tfrac{1}{2}\right) A(X_j, T) + \text{c.c.}, \quad j = 1, 2, \ldots, N. \tag{A 27}$$

The set of collocation points is defined using the discrete Fourier transform (see § 2). A single output uses only a single collocation point.

REFERENCES

Barbagallo, A., Sipp, D. & Schmid, P. 2009 Closed-loop control of an open cavity flow. Part 2. Reduced-order models based on balanced truncation. *J. Fluid Mech.* **641**, 1–50.

Chen, C. T. 1970 *Introduction to Linear System Theory*. Rinehart and Winston.

Cross, M. C. & Hohenberg, P. C. 1993 Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851–1112.

Drazin, P. G. & Reid, W. H. 1982 *Hydrodynamic Stability*. Cambridge Monographs on Mechanics and Applied Mechanics, Cambridge University Press.

Gottlieb, D. & Orszag, S. A. 1981 *Numerical Analysis of Spectral Methods: Theory and Applications*. SIAM.

Holmes, P., Lumley, J. L. & Berkooz, G. 1998 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press.

Howle, L. E. 1997 Linear stability analysis of controlled Rayleigh–Bénard convection using shadowgraphic measurement. *Phys. Fluids* **9**, 3111–3113.

Lall, S., Marsden, J. E. & Glavaski, S. 2002 A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Intl J. Robust Nonlinear Control* **12**, 519–535.

Ma, Z., Ahuja, S. & Rowley, C. W. 2009 Reduced order models for control of fluids using the eigensystem realization algorithm. *Theor. Comput. Fluid Dyn.* (in press).

Ma, Z., Rowley, C. W. & Tadmor, G. 2010 Snapshot-based balanced truncation for linear time-periodic systems. *IEEE Trans. Autom. Control* **55** (2), 469–473.

Newell, A. C. & Whitehead, J. A. 1969 Finite bandwidth, finite amplitude convection. *J. Fluid Mech.* **38**, 279–303.

Or, A. C., Cortelezzi, L. & Speyer, J. L. 2001 Robust feedback control of Rayleigh–Bénard convection. *J. Fluid Mech.* **437**, 175–202.

Or, A. C. & Speyer, J. L. 2003 Active suppression of finite-amplitude Rayleigh–Bénard convection. *J. Fluid Mech.* **483**, 111–128.

Or, A. C. & Speyer, J. L. 2005 Gain-scheduled controller for the suppression of convection at high Rayleigh number. *Phys. Rev.* E **71**, 046302-1.

Or, A. C. & Speyer, J. L. 2008 Model reduction of input-output dynamical systems by proper orthogonal decomposition. *J. Guid. Control Dyn.* **31**, 322–328.

Rowley, C. W. 2005 Model reduction for fluids, using balanced proper orthogonal decomposition. *Intl J. Bifurcation Chaos* **15**, 997–1013.

Scherpen, J. M. A. 1993 Balancing for nonlinear systems. *Syst. Control Lett.* **21**, 143–153.

Schlüter, A., Lortz, D. & Busse, F. H. 1965 On the stability of steady finite amplitude convection. *J. Fluid Mech.* **23**, 129–144.

Tang, J. & Bau, H. H. 1998 Experiments on the stabilization of the no-motion state of a fluid layer heated from below and cooled above. *J. Fluid Mech.* **363**, 153–171.

Willcox, K. & Peraire, J. 2002 Balanced model reduction via the proper orthogonal decomposition. *AIAA J.* **40**, 2323–2330.