Optimal Galerkin approximations of partial differential equations using principal interaction patterns

F. Kwasniok*
Max-Planck-Institut für Meteorologie, Bundesstrasse 55, D-20146 Hamburg, Germany
(Received 2 December 1996)

A method of constructing minimal systems of ordinary differential equations modeling the dynamics of nonlinear partial differential equations is presented. Characteristic spatial structures called principal interaction patterns are extracted from the system according to a nonlinear variational principle based on a dynamical optimality criterion and used as basis functions in a Galerkin approximation. The potential of the method is illustrated using the Kuramoto-Sivashinsky equation as an example. As to the number of modes required to capture the dynamics of the complete system a reduced model based on principal interaction patterns yields a considerable improvement on more conventional approaches using Sobolev eigenfunctions or Karhunen-Loève modes as basis functions and is far more efficient than a dynamical description based on Fourier modes. [S1063-651X(97)14105-8]

PACS number(s): 05.45.+b, 02.60.−x, 02.70.−c

I. INTRODUCTION

The standard approach to the numerical study of the dynamical behavior of partial differential equations (PDEs) *a priori* vested with an infinite number of degrees of freedom consists in their approximation by finite-dimensional dynamical systems. Given the well known fact that the dynamics of PDEs are often confined to attractor sets of relatively low dimension, the construction of minimal dynamical models capturing the principal properties of the complete system is an interesting task. A class of frequently used approximation schemes is formed by the Galerkin methods. The dynamical field is expanded into a finite set of time-independent global basis functions; projection of the PDE onto these basis functions yields a finite-dimensional system of ordinary differential equations (ODEs) for the time evolution of the expansion coefficients. The efficiency of such a dynamical description, i.e., the number of degrees of freedom required to capture the dynamics of the PDE, depends crucially on a proper choice of the basis functions. The most traditional approach lies in using eigenfunctions of a suitably chosen linear differential operator, commonly Fourier modes, as basis functions. Despite the compact and elegant mathematical framework of these Fourier-Galerkin methods, they often provide a system of equations that is much larger than the true dimensionality of the system since Fourier modes are completely general and not adapted to the particular system under consideration. Hence information on the dynamics of the system has to be incorporated into the choice of the basis functions in order to arrive at a dynamical description reflecting more closely the intrinsic dimensionality of the system.

Up to now modes obtained from a Karhunen-Loève (KL) decomposition (also referred to as principal component analysis, empirical orthogonal function analysis, or proper orthogonal decomposition) are widely used as basis functions in a Galerkin approximation of a PDE in order to arrive at a low-dimensional model [1–6]. The KL eigenfunctions allow for an optimal spatial representation of an attractor in high- or even infinite-dimensional phase space by a linear subspace of given dimension in a mean least-squares sense. The KL approach has been extended to the Sobolev eigenfunctions by Kirby [7]. However, the optimality criterion defining the KL modes and the Sobolev eigenfunctions, respectively, does not refer to the time evolution of the truncated system obtained when projecting the PDE onto these modes. Thus, as has been first pointed out by Hasselmann [8], a methodology referring simultaneously to spatial as well as temporal features of the system by taking into account the dynamics of the reduced model in order to define the basic spatial patterns may be the even more efficient (although the considerably more cumbersome) approach when searching for a minimal model. Following the proposition of Hasselmann, a general algorithm for reducing a high-dimensional Fourier-Galerkin approximation to a low-dimensional dynamical system has been derived and illustrated in the context of a geophysical fluid system by Kwasniok [9]. A set of spatial structures is obtained by minimizing the error in the time derivative between the complete system and the reduced system in a mean least-squares sense and used as a basis in a dynamical description. The same methodology has been used by Achatz *et al.* [10] and Achatz and Schmitz [11] in order to arrive at a low-dimensional model of baroclinic wave life cycles, a nonlinear oscillatory phenomenon in the field of dynamic meteorology. A similar spatiotemporal strategy for constructing reduced dynamical models of complex systems in the vicinity of critical points based on the theory of synergetics has been derived by Uhl *et al.* [12,13] and used for the analysis of experimental data by Jirsa *et al.* [14] and Friedrich and Uhl [15].

In the present study the methodology outlined by Kwasniok [9] is developed further by deriving an improved optimality criterion for defining the basis functions taking into account the nonlinear time evolution of the reduced system over a finite time interval rather than only the local time

*FAX: +49-40-41173298. Electronic address: kwasniok@dkrz.de

derivative. Reduced models of the Kuramoto-Sivashinsky equation are constructed using this method. The question is addressed to what extent these low-dimensional systems are capable of capturing the principal properties of the complete system.

The paper is organized as follows: in Sec. II the methodology for deriving a reduced model from a PDE is outlined in general. In Sec. III the model system used as an illustrative example is introduced and the methodology is applied to this particular system. Then the results are given and discussed. The paper is concluded in Sec. V. In the two Appendices the numerical details of the algorithm are given.

II. METHODOLOGY

We start out with a partial differential equation

$$\frac{\partial u}{\partial t} = \mathcal{D}(u), \tag{1}$$

where u is a function of n-dimensional space and time:

$$u = u(x,t), \quad x = (x_1,\ldots,x_n) \tag{2}$$

and $\mathcal{D}$ is a nonlinear differential operator, which is assumed to be polynomial in u and its spatial derivatives. Equation (1) is considered on a bounded spatial domain $\Omega \subset \mathcal{R}^n$ subject to appropriate boundary conditions. Nearly all nonlinear PDEs treated in theoretical physics (the Navier-Stokes equations and various derivates, the Kuramoto-Sivashinsky equation, the Ginzburg-Landau equation, the Korteweg–de Vries equation, numerous reaction-diffusion equations, etc.) fall into this class. The phase space of Eq. (1) is some infinite-dimensional Hilbert space $\mathcal{H}(\Omega)$ of sufficiently smooth functions from Ω into the real numbers. A scalar product is introduced in $\mathcal{H}$, which usually is of the form

$$[g,h] = \int_\Omega \mathcal{L}(g)\mathcal{L}(h)dx \tag{3}$$

for two arbitrary functions $g,h \in \mathcal{H}$ with a suitably chosen linear operator $\mathcal{L}$.

The system reduction is now achieved in two steps: first, the PDE is cast into a finite-dimensional dynamical description using a standard Fourier-Galerkin procedure. Second, a variational principle is applied in this finite- but high-dimensional phase space in order to identify a low-dimensional subspace optimally suited for the construction of a reduced model.

A. Spectral basis

Usually a basis of $\mathcal{H}$ is given by an infinite, denumerable, complete set $\{f_\mu ; \mu \in \mathcal{N}\}$ of eigenfunctions of some appropriately chosen self-adjoint linear differential operator depending on the spatial domain Ω and the boundary conditions. Commonly, the functions f_μ are Fourier modes. Considering the finite-dimensional subspace $\mathcal{F}$ spanned by the first N functions

$$\mathcal{F} = \mathrm{Span}\{f_1,\ldots,f_N\}, \tag{4}$$

u is expanded into a truncated series

$$u_F = \sum_{\mu=1}^{N} u_\mu f_\mu . \tag{5}$$

Here and in the following the subscript F signifies quantities associated with the Fourier-Galerkin approximation of Eq. (1). Insertion of the expansion of Eq. (5) into Eq. (1) and projection onto the adjoint functions yields an autonomous system of N first-order ordinary differential equations for the time evolution of the expansion coefficients:

$$\dot{u}_\mu = \left[f_\mu^*, \mathcal{D}\left(\sum_{\nu=1}^{N} u_\nu f_\nu \right) \right], \quad \mu = 1,\ldots,N \tag{6}$$

The adjoint functions are defined as the functions $f_\mu^* \in \mathcal{F}$ satisfying

$$[f_\mu^*, f_\nu] = \delta_{\mu\nu}, \quad \mu,\nu = 1,\ldots,N. \tag{7}$$

Often the functions f_μ form an orthonormal set (having $f_\mu^* = f_\mu$ then). But sometimes it may be desirable to consider several scalar products involving different spatial derivatives that correspond to particular physical quantities, e.g., kinetic energy in the case of a fluid system (cf. [9] and Sec. III of this paper). Therefore the more general notation involving the adjoint modes is used here. In the following we assume the truncation limit N to be large enough that the dynamical system of Eq. (6) captures the long-term behavior of the PDE of Eq. (1) monitored by statistical quantities (e.g., moments or Fourier spectra) and dynamical characteristics (e.g., Lyapunov exponents or attractor dimensions) within sufficient accuracy; all properties of Eq. (1) will be identified with the corresponding properties of Eq. (6).

It is convenient to separate the dynamical field u_F into the mean state and the deviation from it:

$$u_F = \langle u_F \rangle + \hat{u}_F = \sum_{\mu=1}^{N} (\langle u_\mu \rangle + \hat{u}_\mu) f_\mu \tag{8}$$

$\langle\ \rangle$ denotes the ensemble average over the attractor on which the asymptotic motion of the dynamical system of Eq. (6) resides. Equation (6) then becomes

$$\dot{\hat{u}}_\mu = \left[f_\mu^*, \mathcal{D}\left(\langle u_F \rangle + \sum_{\nu=1}^{N} \hat{u}_\nu f_\nu \right) \right], \quad \mu = 1,\ldots,N. \tag{9}$$

B. Principal interaction patterns

We now consider an L-dimensional subspace $\mathcal{P} \subset \mathcal{F}$ spanned by only a limited number of linearly independent spatial modes p_i, which will be called principal interaction patterns (PIPs):

$$\mathcal{P} = \mathrm{Span}\{p_1,\ldots,p_L\}, \quad L \ll N. \tag{10}$$

In order to simplify the notation from now on greek indices always run from 1 to N, latin ones from 1 to L if not explicitly indicated otherwise. Each PIP p_i is a linear combination of the standard basis functions:

$$p_i = \sum_\mu P_{\mu i} f_\mu , \tag{11}$$

P being the $(N\times L)$ matrix with the vectors of Fourier coefficients of the PIPs as its columns. The anomaly field $\hat{u}_F$ is expanded into a series of PIPs:

$$u_{\mathrm{PIP}}=\langle u_F\rangle+\hat{u}_{\mathrm{PIP}}=\langle u_F\rangle+\sum_i z_i p_i . \tag{12}$$

The subscript PIP indicates quantities associated with the PIP approximation of Eq. (1). A linear projection operator $\mathcal{Z}^{\mathcal{P}}$ is defined that operates from $\mathcal{H}$ into $\mathcal{P}$ and maps a function $g\in\mathcal{H}$ onto its projection onto PIP space

$$g\rightarrow\mathcal{Z}^{\mathcal{P}}(g)=\sum_i [p_i^*,g]p_i \tag{13}$$

with the adjoint patterns $p_i^*\in\mathcal{P}$ defined by

$$[p_i^*,p_j]=\delta_{ij} . \tag{14}$$

A reduced model is given by the projection of Eq. (1) onto PIP space

$$\frac{\partial\hat{u}_{\mathrm{PIP}}}{\partial t}=\mathcal{Z}^{\mathcal{P}}(\mathcal{D}(u_{\mathrm{PIP}})), \tag{15}$$

which is equivalent to the system of modal equations for the amplitudes of the PIPs:

$$\dot{z}_i=\left[p_i^*,\mathcal{D}\left(\langle u_F\rangle+\sum_j z_j p_j\right)\right]. \tag{16}$$

Note that Eqs. (10)–(16) are completely general. A truncated model as given in Eq. (16) can be derived for any L-dimensional subspace $\mathcal{P}\subseteq\mathcal{F}$ spanned by L arbitrary linearly independent spatial modes p_i. Especially for $p_i=f_i$ and $L=N$ one returns to the system of Eq. (9).

C. Uniqueness of the patterns

If $\mathcal{D}(u)$ is polynomial in u and its spatial derivatives of maximum degree l the system of modal equations for the PIP amplitudes [Eq. (16)] takes the form

$$\dot{z}_i=K_i^{(0)}+\sum_{j_1}K_{ij_1}^{(1)}z_{j_1}+\sum_{j_1,j_2}K_{ij_1j_2}^{(2)}z_{j_1}z_{j_2}+\cdots+\sum_{j_1,\ldots,j_l}K_{ij_1\ldots j_l}^{(l)}z_{j_1}\cdots z_{j_l}, \tag{17}$$

where $K^{(i)}$ are the tensors of coupling coefficients determined by the spatial structure of the patterns and the operator $\mathcal{D}$. It is now obvious that the PIPs are only determined to within a linear transformation. Consider a matrix P representing a set of patterns $\{p_1,\ldots,p_L\}$ and an arbitrary regular linear transformation T in L-dimensional space. The transformed matrix is then $P'=PT$ representing a set of patterns $\{p_1',\ldots,p_L'\}$ with $p_i'=\Sigma_j T_{ji}p_j$; the transformed expansion coefficients are $z_i'=\Sigma_j T_{ij}^{-1}z_j$. One then has $\hat{u}_{\mathrm{PIP}}=\Sigma_i z_i p_i=\hat{u}'_{\mathrm{PIP}}=\Sigma_i z_i' p_i'$. Transforming the coupling coefficients $K^{(i)}$ according to the rules for tensors of $(i+1)$th order

$$K'^{(i)}_{j_1\cdots j_{i+1}}=\sum_{k_1,\ldots,k_{i+1}}T^{-1}_{j_1k_1}T_{k_2j_2}\cdots T_{k_{i+1},j_{i+1}}K^{(i)}_{k_1\ldots k_{i+1}} \tag{18}$$

one obtains a system of differential equations of the same form as Eq. (17) for the amplitudes of the transformed patterns:

$$\dot{z}_i'=K_i'^{(0)}+\sum_{j_1}K_{ij_1}'^{(1)}z_{j_1}'+\sum_{j_1,j_2}K_{ij_1j_2}'^{(2)}z_{j_1}'z_{j_2}'+\cdots+\sum_{j_1,\ldots,j_l}K_{ij_1\ldots j_l}'^{(l)}z_{j_1}'\cdots z_{j_l}'. \tag{19}$$

Now $\partial\hat{u}_{\mathrm{PIP}}/\partial t=\Sigma_i\dot{z}_i p_i=\partial\hat{u}'_{\mathrm{PIP}}/\partial t=\Sigma_i\dot{z}_i' p_i'$ holds. Hence the dynamical systems of Eqs. (17) and (19) are equivalent in the respect that $\hat{u}_{\mathrm{PIP}}$ and $\hat{u}'_{\mathrm{PIP}}$ remain identical if they are identical at some initial time and also in the long-term behavior all statistical and dynamical properties of $\hat{u}_{\mathrm{PIP}}$ and $\hat{u}'_{\mathrm{PIP}}$ are identical. Especially the error function χ, which will be introduced in Sec. II D is invariant under the transformation $\chi(P';\tau_{\max})=\chi(P;\tau_{\max})$. Put differently, P and P' are associated with the same projection operator $\mathcal{Z}^{\mathcal{P}}$ and thus lead to equivalent reduced systems.

One may eliminate this gauge freedom by referring to some normal form for the matrix of patterns. One possible way to do so is (in analogy to the KL modes) to impose the constraints so that the patterns form an orthonormal set and that their amplitudes are pairwise uncorrelated with the patterns ordered according to descending mean squared amplitude:

$$[p_i,p_j]=\sum_{\mu,\nu}M_{\mu\nu}P_{\mu i}P_{\nu j}=\delta_{ij}, \tag{20}$$

$$\langle z_i z_j\rangle=\sum_{\mu,\nu}(M\Gamma M)_{\mu\nu}P_{\mu i}P_{\nu j}=\delta_{ij}\lambda_i^{\mathrm{PIP}},\quad \lambda_i^{\mathrm{PIP}}>\lambda_{i+1}^{\mathrm{PIP}} . \tag{21}$$

The symmetric, positive definite metric M is the representation of the scalar product in the space $\mathcal{F}$:

$$M_{\mu\nu}=[f_\mu,f_\nu] \tag{22}$$

and Γ is the covariance matrix of the Fourier components of $\hat{u}_F$:

$$\Gamma_{\mu\nu}=\langle\hat{u}_\mu\hat{u}_\nu\rangle. \tag{23}$$

Often the metric M is diagonal; in particular, for all scalar products considered in Sec. III in the context of the Kuramoto-Sivashinsky equation this is the case. Moreover, some convention concerning the sign of the patterns has to be adopted. From now on the PIPs are always assumed to be orthonormal in space ($p_i^*=p_i$) and uncorrelated in time. The orthonormality in space facilitates the formulation of the variational principle in Sec. II D and greatly reduces the computational effort involved in the corresponding numerical minimization procedure (cf. Sec. II D and Appendix A).

D. The variational principle

Consider some initial condition $\hat{u}_F^0=\Sigma_\mu \hat{u}_\mu^0 f_\mu$ characterized by the Fourier coefficients $\hat{u}_\mu^0$ at initial time $t=0$. Let $\hat{u}_F^\tau=\Sigma_\mu \hat{u}_\mu^\tau f_\mu$ be the state obtained when integrating the high-dimensional Fourier-Galerkin approximation forward in time from $t=0$ to $t=\tau$ with initial condition $\hat{u}_F^0$. Consider now the projection of the initial condition onto PIP space $\hat{u}_{\mathrm{PIP}}^0=\Sigma_i z_i^0 p_i$ with $z_i^0=[p_i,\hat{u}_F^0]$. Let $\hat{u}_{\mathrm{PIP}}^\tau=\Sigma_i z_i^\tau p_i$ be the state obtained when integrating the reduced system forward in time from $t=0$ to $t=\tau$ with initial condition $\hat{u}_{\mathrm{PIP}}^0$. $\hat{u}_{\mathrm{PIP}}^\tau$ is different from the projection of the complete system onto the PIP space $\mathcal{Z}^{\mathcal{P}}(\hat{u}_F^\tau)=\Sigma_i \tilde{z}_i^\tau p_i$ with $\tilde{z}_i^\tau=[p_i,\hat{u}_F^\tau]$. An error function is introduced that measures the spatially and temporally integrated squared error between the state given by the full model and that given by the reduced model normalized by the spatially integrated variance of the system:

$$Q(\hat{u}_F^0,P;w)=\frac{1}{\mathrm{Var}}\int_0^\infty [\hat{u}_{\mathrm{PIP}}^\tau-\hat{u}_F^\tau,\hat{u}_{\mathrm{PIP}}^\tau-\hat{u}_F^\tau]w(\tau)d\tau, \tag{24}$$

$$\mathrm{Var}=\langle[\hat{u}_F,\hat{u}_F]\rangle=\sum_\mu \langle\hat{u}_\mu^2\rangle. \tag{25}$$

$w(\tau)$ is an arbitrary non-negative weight function satisfying $\int_0^\infty w(\tau)d\tau=1$. In the present study $w(\tau)$ is taken to be

$$w(\tau)=\begin{cases}\dfrac{1}{\tau_{\max}} & \text{for} \quad 0\leqslant\tau\leqslant\tau_{\max}\\ 0 & \text{for} \quad \tau>\tau_{\max}\end{cases} \tag{26}$$

with the free parameter $\tau_{\max}>0$. For N large enough Q is independent of N if $\tau_{\max}$ is not too large. Taking the ensemble average over all initial conditions $\hat{u}_F^0$ on the attractor yields an error function that depends only on the pattern set and the parameter $\tau_{\max}$:

$$\chi(P;\tau_{\max})=\langle Q(\hat{u}_F^0,P;\tau_{\max})\rangle. \tag{27}$$

The optimal set of patterns is determined by minimizing the error function χ. In practice the ensemble average is replaced by the average over a finite number of realizations of Q with initial conditions $\hat{u}_F^0$ taken at uncorrelated times from a long time series generated by an integration of the system of Eq. (9) assuming ergodicity of the flow. Introducing the projection of the error at time τ onto the patterns

$$\varepsilon_i^\tau=[p_i,\hat{u}_{\mathrm{PIP}}^\tau-\hat{u}_F^\tau]=z_i^\tau-\tilde{z}_i^\tau \tag{28}$$

the error function can be split into two parts:

$$\chi=\chi_1+\chi_2 \tag{29}$$

with

$$\chi_1(P)=1-\frac{1}{\mathrm{Var}}\sum_i \lambda_i^{\mathrm{PIP}} \tag{30}$$

and

$$\chi_2(P;\tau_{\max})=\frac{1}{\mathrm{Var}\ \tau_{\max}}\int_0^{\tau_{\max}}\sum_i \langle(\varepsilon_i^\tau)^2\rangle d\tau. \tag{31}$$

χ_1 simply measures the mean squared error due to the projection of $\hat{u}_F$ onto PIP space and does not depend on the dynamical behavior of the reduced system; χ_2 represents the error in PIP space and refers to the dynamics of the PIP model. The maximum integration time $\tau_{\max}$ remains as a free parameter of the method. It may be taken as a characteristic time scale of the system based on some physical knowledge or reasoning or obtained from mathematical techniques (e.g., a dominant period in the temporal power spectrum of the system or the inverse of the largest Lyapunov exponent in the case of a chaotic system).

In order to actually evaluate the error function χ the temporal integral in Eq. (31) has to be approximated by a finite sum,

$$J=\frac{1}{\tau_{\max}}\int_0^{\tau_{\max}}\sum_i (\varepsilon_i^\tau)^2 d\tau=\sum_{k=1}^K \sum_i w_k(\varepsilon_i^{\tau_k})^2, \tag{32}$$

where τ_k are equally spaced (for simplicity) mesh points in the interval $[0,\tau_{\max}]$ $[\tau_k=(k/K)\tau_{\max}]$ and the weights w_k are given by some quadrature rule. When using, e.g., the Simpson quadrature rule the weights are

$$w_{2k-1}=\frac{4}{3K},\quad k=1,\ldots,\frac{K}{2}, \tag{33}$$

$$w_{2k}=\frac{2}{3K},\quad k=1,\ldots,\frac{K}{2}-1, \tag{34}$$

$$w_K=\frac{1}{3K}. \tag{35}$$

K then has to be an even integer.

The variational principle poses a high-dimensional nonlinear minimization task (with the NL elements of P as variables) that has to be solved numerically by iterative techniques. The gradient of χ with respect to the components of P can be calculated efficiently using an adjoint method originating from the theory of optimal control. See the Appendices for details on the calculation techniques of the numerical minimization procedure.

The methodology originally proposed by Hasselmann [8] and used by Kwasniok [9], Achatz *et al.* [10] and Achatz and Schmitz [11] may be viewed as a first-order approximation of that illustrated here. In these former studies (and also in the papers [12–15]) the optimality criterion defining the patterns only refers to the local time derivative of the reduced system whereas now the dynamical behavior over a finite time interval is taken into account. This more recent approach intuitively can be expected to be more powerful (although computationally more expensive) than the former one. For example, think of a periodic system. Then the newer method with $\tau_{\max}$ set at the oscillation period will be more efficient at finding a reduced model reproducing the correct amplitude and frequency of the oscillations than the former method, which just looks at the local time derivative. Especially the problem of ill posedness should be less severe with

the newer technique. Imagine two sets of patterns, which differ only slightly from each other. Then the errors in the time derivative of the reduced models associated with these pattern sets are usually also nearly identical whereas an error function based on the time evolution over a finite time can be expected to provide a better discrimination between the two pattern sets.

E. Karhunen-Loève modes as a limiting case

In the limit $\tau_{\max}\to 0$ the term χ_2 vanishes. One is then left with the problem of finding an L-dimensional subspace such that the mean squared projection error χ_1 is minimal. As is well known the solution to this minimization problem is given by the KL decomposition. The KL modes e_ϱ are given by the eigenvectors of the eigenvalue problem

$$\Gamma M E = E\,\mathrm{diag}(\lambda_1^{\mathrm{KL}},\ldots,\lambda_N^{\mathrm{KL}}), \tag{36}$$

where E is the $(N\times N)$ matrix with the Fourier components of the KL modes as its columns

$$e_\varrho = \sum_\mu E_{\mu\varrho} f_\mu . \tag{37}$$

The matrix ΓM as a product of symmetric, positive definite matrices can be diagonalized in a real basis and has only real and positive eigenvalues. The eigenvalues are in descending order $(\lambda_\varrho^{\mathrm{KL}} > \lambda_{\varrho+1}^{\mathrm{KL}})$. The KL modes form an orthonormal set:

$$[e_\varrho, e_\sigma] = \sum_{\mu,\nu} M_{\mu\nu} E_{\mu\varrho} E_{\nu\sigma} = \delta_{\varrho\sigma} . \tag{38}$$

The amplitudes of the KL eigenfunctions $w_\varrho = [e_\varrho, \hat{u}_F]$ are pairwise uncorrelated and the second moment of each is given by the corresponding eigenvalue

$$\langle w_\varrho w_\sigma \rangle = \sum_{\mu,\nu} (M\Gamma M)_{\mu\nu} E_{\mu\varrho} E_{\nu\sigma} = \delta_{\varrho\sigma} \lambda_\varrho^{\mathrm{KL}} . \tag{39}$$

The sum of all eigenvalues is equal to the spatially integrated variance of the system:

$$\mathrm{Var} = \sum_\varrho \lambda_\varrho^{\mathrm{KL}} . \tag{40}$$

The global minimum of χ_1 is attained when taking the eigenfunctions corresponding to the L largest eigenvalues. The value of the error function is then

$$\chi_1 = 1 - \frac{1}{\mathrm{Var}} \sum_i \lambda_i^{\mathrm{KL}} . \tag{41}$$

The PIP approach may be viewed as a nonlinear extension of the KL method. For short integration times $\tau_{\max}$ the error function χ is dominated by the term χ_1; the optimal patterns are then very close to the KL modes. With increasing $\tau_{\max}$ more and more information on the dynamical behavior of the truncated model is included.

III. THE KURAMOTO-SIVASHINSKY EQUATION

As a model system to illustrate the method outlined in Sec. II we consider the rescaled Kuramoto-Sivashinsky (KS) equation in one space dimension

$$\frac{\partial u}{\partial t} + 4\frac{\partial^4 u}{\partial x^4} + \alpha\left[\frac{\partial^2 u}{\partial x^2} + \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2\right] = 0 \tag{42}$$

subject to periodic boundary conditions

$$u(x,t) = u(x+2\pi,t) \quad \forall t \tag{43}$$

at $\alpha = 84.25$ where a limit cycle solution is observed. This case has already been treated in the literature in the context of reduced models using Sobolev eigenfunctions as basis functions [7].

The following class of scalar products is introduced:

$$[g,h]_\beta = \frac{1}{2\pi}\int_0^{2\pi} \frac{\partial^\beta g}{\partial x^\beta}\frac{\partial^\beta h}{\partial x^\beta}\,dx = (-1)^\beta\left[\frac{\partial^{2\beta} g}{\partial x^{2\beta}}, h\right]_0 ,$$
$$\beta = 0,1,2,\ldots . \tag{44}$$

For each scalar product and a given set of patterns p_i a linear projection operator $\mathcal{Z}_\beta^{\mathcal{P}}$ can be defined according to Eq. (13). For all β the projector is self-adjoint with respect to the corresponding scalar product

$$[g, \mathcal{Z}_\beta^{\mathcal{P}}(h)]_\beta = [\mathcal{Z}_\beta^{\mathcal{P}}(g), h]_\beta \quad \forall g,h \in \mathcal{H} \tag{45}$$

and idempotent

$$\mathcal{Z}_\beta^{\mathcal{P}}(\mathcal{Z}_\beta^{\mathcal{P}}(g)) = \mathcal{Z}_\beta^{\mathcal{P}}(g) \quad \forall g \in \mathcal{H}. \tag{46}$$

Note that the Burgers type nonlinearity in the KS equation conserves the integral quantity $I_k = (1/2\pi)\int_0^{2\pi}(\partial u/\partial x)^k dx$ for all $k \in \mathcal{N}$. I_1 is the spatial average of $\partial u/\partial x$ and vanishes for all times. $I_2 = [u,u]_1$ is similar to a kinetic energy; for $k \geqslant 3$ I_k does not represent a particular physical quantity. When separating u into the time mean and the anomalies the term that is nonlinear in the anomalies conserves the quantity $\hat{I}_k = (1/2\pi)\int_0^{2\pi}(\partial \hat{u}/\partial x)^k dx$ for all $k \in \mathcal{N}$. One has $\hat{I}_1 = 0$ for all times. $\hat{I}_2 = [\hat{u},\hat{u}]_1$ may be termed a turbulent kinetic energy.

A. Spectral basis

The appropriate set of basis functions for the present example is

$$\{f_\mu ; \mu = 1,\ldots,N\} = \{\sqrt{2}\cos kx, \sqrt{2}\sin kx; k = 1,\ldots,k_{\max}\}. \tag{47}$$

The modes are orthonormal with respect to $[\cdot,\cdot]_0$:

$$[f_\mu, f_\nu]_0 = \delta_{\mu\nu} . \tag{48}$$

When truncating at wave number $k_{\max}$ the number of modes is $N = 2k_{\max}$. The wave number zero mode is omitted in order to remove the mean drift of Eq. (42) as is usually done. Note that the dynamical system generated by a Galerkin procedure using a Fourier basis as given in Eq. (47) is identical for all scalar products out of the class introduced in Eq. (44)

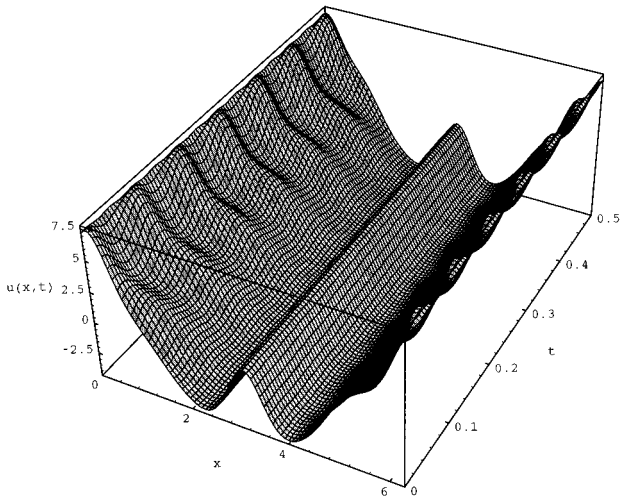

FIG. 1. Solution to the Kuramoto-Sivashinsky equation at $\alpha = 84.25$.

as can be immediately seen considering the corresponding adjoint modes f_μ^*. Test calculations revealed that a truncated model using the Fourier modes up to wave number $k_{\max} = 15$ (30 degrees of freedom) is sufficient to capture the long-term dynamics of the KS equation monitored by first and second moments as well as temporal Fourier spectra at the selected parameter value $\alpha = 84.25$. The model is integrated in time using a de-aliased pseudospectral transform method [16] for the evaluation of the nonlinear term based on a discrete Fourier transform on a grid consisting of 48 equally spaced mesh points (employing the fast Fourier transform algorithm). Given the extreme stiffness and sensitivity to the numerical accuracy of ODE models derived from the KS equation a variable-order, variable-step backward differentiation formula is used as ODE integrator with the error tolerance set at 10^{-12} per time unit. Figure 1 illustrates the time evolution of the solution. The system exhibits periodic behavior; the variability is spatially localized at the edges of the interval (cf. also Fig. 4). Already from visual inspection of Fig. 1 it is intuitively clear that the system is in fact very low dimensional and that the dynamical description based on Fourier modes requiring 30 degrees of freedom is far from optimal.

B. Principal interaction patterns

According to the general outline in Sec. II u_F is separated into the time mean $\langle u_F \rangle$ and the anomalies $\hat{u}_F$. $\hat{u}_F$ is expanded into a series of PIPs. Projection of the KS equation onto the PIP basis yields the system of modal equations

$$\dot{z}_i = \frac{1}{2} \sum_{j,k} a_{ijk} z_j z_k + \sum_j b_{ij} z_j + c_i, \tag{49}$$

where the interaction coefficients are given by

$$a_{ijk} = (-1)^{\beta+1} \alpha \left[\frac{\partial^{2\beta} p_i}{\partial x^{2\beta}}, \frac{\partial p_j}{\partial x} \frac{\partial p_k}{\partial x} \right]_0 = a_{ikj}, \tag{50}$$

$$b_{ij} = (-1)^{\beta+1} \left[\frac{\partial^{2\beta} p_i}{\partial x^{2\beta}}, 4 \frac{\partial^4 p_j}{\partial x^4} + \alpha \frac{\partial^2 p_j}{\partial x^2} + \alpha \frac{\partial \langle u_F \rangle}{\partial x} \frac{\partial p_j}{\partial x} \right]_0, \tag{51}$$

$$c_i = (-1)^{\beta+1} \left[\frac{\partial^{2\beta} p_i}{\partial x^{2\beta}}, 4 \frac{\partial^4 \langle u_F \rangle}{\partial x^4} + \alpha \frac{\partial^2 \langle u_F \rangle}{\partial x^2} + \frac{\alpha}{2} \left(\frac{\partial \langle u_F \rangle}{\partial x} \right)^2 \right]_0. \tag{52}$$

The interaction coefficients can be evaluated numerically using the pseudospectral transform method. In general for some choice of patterns the tensors of interaction coefficients a, b, c depend on the scalar product used in the projection. Only if the PIP space is identical with the space spanned by the Fourier modes up to a certain wave number do all scalar products of Eq. (44) generate the same dynamical system. Especially the full system of Sec. III A is the same for all β. The difference is due to the fact that in the special case of a Fourier subspace the projection operator $\mathcal{Z}_\beta^{\mathcal{P}}$ and the operator of spatial differentiation $\partial/\partial x$ commute for all β whereas in the general case they do not.

For completeness we remark that the projection procedure using the scalar product $[\cdot,\cdot]_\beta$ is equivalent to formulating the KS equation for $\partial^\beta u/\partial x^\beta$ rather than u [by β-fold spatial differentiation of Eq. (42)], expanding $\partial^\beta \hat{u}_F/\partial x^\beta$ rather than $\hat{u}_F$ into a series of PIPs and using the scalar product $[\cdot,\cdot]_0$ in the projection.

In view of the conserved integral quantities of the KS equation one may ask if these conservation properties are adopted by truncated models. The corresponding quantities in the subspace $\mathcal{P}$ are $\hat{I}_k^{\mathcal{P}} = (1/2\pi) \int_0^{2\pi} (\partial \hat{u}_{\mathrm{PIP}}/\partial x)^k dx$. $\hat{I}_1^{\mathcal{P}}$ by construction vanishes for all times and is thus trivially conserved by all PIP models. For $k \geqslant 2$ the conservation properties depend on the scalar product used in the projection. In the case $\beta = 1$ the nonlinear terms in a PIP model for arbitrary sets of patterns conserve the truncated turbulent kinetic energy $\hat{I}_2^{\mathcal{P}} = [\hat{u}_{\mathrm{PIP}}, \hat{u}_{\mathrm{PIP}}]_1$ but in general not $\hat{I}_k^{\mathcal{P}}$ for $k \geqslant 3$. This conservation statement can be shown by considering the reduced model in the form of Eq. (15) and using integration by parts as well as the properties of the projector given in Eqs. (45) and (46). The nonlinear interaction coefficients then satisfy the relationship

$$a_{ijk} + a_{jik} + a_{kij} = 0, \tag{53}$$

as can be readily verified from Eq. (50) using integration by parts. For all other scalar products $\hat{I}_k^{\mathcal{P}}$ is for all $k \geqslant 2$ generally not conserved by the nonlinear terms apart from the case of a Fourier subspace where the models coincide for all β. Especially the nonlinear terms of the complete system of Sec. III A [when written as an anomaly model like Eq. (9)] conserve the truncated turbulent kinetic energy $\hat{I}_2^{\mathcal{F}} = [\hat{u}_F, \hat{u}_F]_1$ but not $\hat{I}_k^{\mathcal{F}} = (1/2\pi) \int_0^{2\pi} (\partial \hat{u}_F/\partial x)^k dx$ for $k \geqslant 3$.

IV. RESULTS AND DISCUSSION

PIP models have been extracted from the KS equation based on four realizations of the error function Q corresponding to initial conditions separated by 0.2 system units

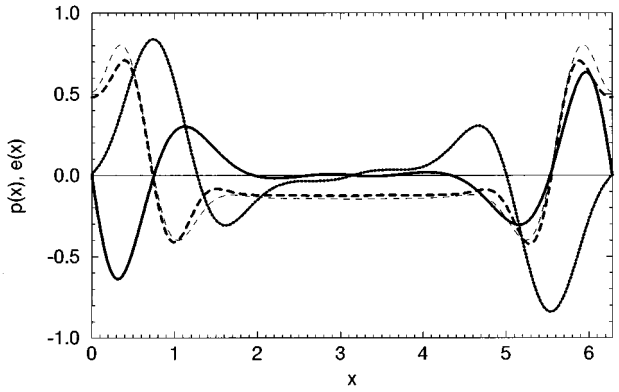

FIG. 2. First (solid), second (dotted), and third (dashed) PIP (thick lines) and KL mode (thin lines).

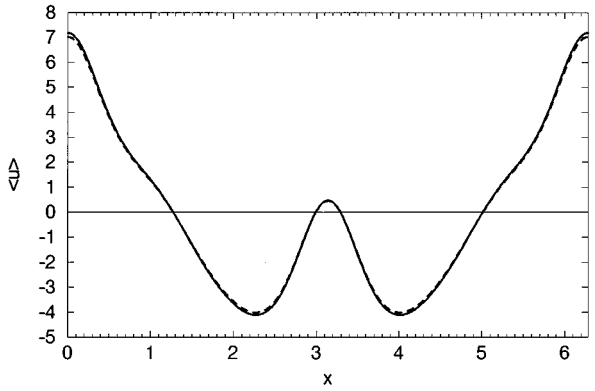

FIG. 3. Mean state obtained from a simulation with 30 Fourier modes (complete system, solid), 20 Fourier modes (identical with solid), 3 PIPs (identical with solid), 4 KL modes (dashed), and 5 KL modes (identical with solid).

(su). This small ensemble size here is due to the simple periodic structure of the present system; in the case of a chaotic system a much higher number of realizations is necessary dependent on the dimension of the underlying attractor [17]. The integration time $\tau_{\max}$ is taken to be 0.02 su, which is about a quarter of the oscillation period of the system. With this value of $\tau_{\max}$ one already gets the same results as with $\tau_{\max}$ equal to the oscillation period, which is the canonical choice for $\tau_{\max}$ in the case of a periodic system. The sampling interval of the data is 0.005 su corresponding to $K=4$ mesh points for the approximation of the temporal integral in the error function. Calculations have been done for the scalar products $[\cdot,\cdot]_0$, $[\cdot,\cdot]_1$, and $[\cdot,\cdot]_2$. Starting out with two PIPs and using progressively more modes, four, three and five PIPs turned out to be the minimum number of patterns to capture all essentials of the long-term behavior of the KS equation at the selected parameter value $\alpha=84.25$ for $[\cdot,\cdot]_0$, $[\cdot,\cdot]_1$, and $[\cdot,\cdot]_2$, respectively. The good performance of the scalar product $[\cdot,\cdot]_1$ may be due to the fact that it stands out against the others as to conservation of turbulent kinetic energy $\hat{I}_2^{\mathcal{P}}$ by the nonlinear terms. In the following the presentation of the results is restricted to the case $\beta=1$.

Figure 2 gives the spatial structure of the three PIPs extracted from the system. For comparison also the first three KL modes (also with respect to $[\cdot,\cdot]_1$) are shown. Table I gives the fraction of variance captured by these three PIPs and KL modes, respectively. The first and second PIP and KL eigenfunction are virtually indistinguishable from one another; they together account for 99.68% of the variance of the system and are indispensable for any dynamical description. In the third mode significant differences occur, especially in the zones of high variability at the edges of the interval. An expansion of the third PIP in terms of the KL modes reveals significant contributions up to the tenth KL mode.

TABLE I. Fraction of variance captured by three PIPs and KL modes.

i	$\lambda_i^{\mathrm{KL}}/\mathrm{Var}$	$\Sigma_{j=1}^{i}\lambda_j^{\mathrm{KL}}/\mathrm{Var}$	$\lambda_i^{\mathrm{PIP}}/\mathrm{Var}$	$\Sigma_{j=1}^{i}\lambda_j^{\mathrm{PIP}}/\mathrm{Var}$
1	0.8045	0.8045	0.8045	0.8045
2	0.1923	0.9968	0.1923	0.9968
3	0.0023	0.9991	0.0021	0.9989

When integrating the reduced model based on the three KL modes shown in Fig. 2 forward in time the amplitudes grow without bounds; i.e., this model does not possess a stable attractor and thus completely fails to capture the long-term dynamics of the complete system whereas the model based on the three PIPs has a stable attractor and reproduces all principal dynamical and statistical properties (as discussed in detail below). This is due to the fact that the property of dissipativity inherent in the KS equation is not adopted by the truncated KL model since the dissipation term mainly acts at the small spatial scales (high wave numbers) whereas the leading KL modes are dominated by large spatial scales. The third PIP has significantly larger contributions from higher wave numbers than the third KL mode. Hence the dissipation is captured better by the PIP space than by the KL space leading to the superior performance of the PIP model.

Now the ability of the three-dimensional PIP model [for convenience hereafter referred to as PIP(3) model] to reproduce the long-term behavior of the complete system is investigated in detail. For comparison also the performance of reduced models based on four KL modes [KL(4) model], five KL modes [KL(5) model], and the Fourier modes up to wave number 10 [F(20) model] is shown. In the F(20) model only the anomalies are truncated at wave number 10 but the full mean state $\langle u_F \rangle$ is used to allow for a direct comparison to the other models.

In Fig. 3 the mean state obtained from the complete system and from the reduced systems is given. The PIP(3) model reproduces the mean state perfectly as well as the KL(5) model and the F(20) model. The KL(4) model has a slight but significant error.

Figure 4 illustrates the variance $\langle u^2 \rangle - \langle u \rangle^2$ as a function of the space coordinate. The accordance for the PIP(3) model is perfect; the same holds for the KL(5) model. The KL(4) and the F(20) model both have large errors.

In order to check whether the reduced models are able to capture the behavior of the complete system in the time domain Fig. 5 shows the temporal Fourier spectrum of $u(x_0)$, where x_0 is a representative position in space taken as the

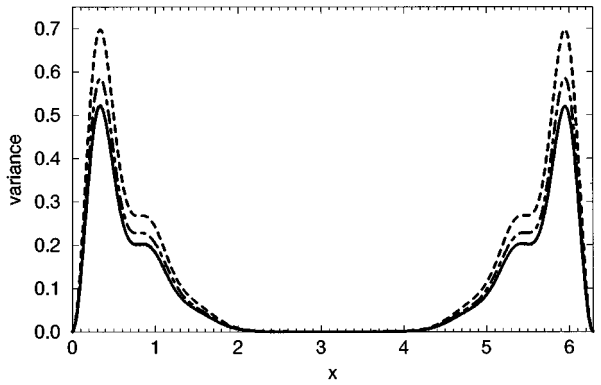

FIG. 4. Variance obtained from a simulation with 30 Fourier modes (complete system, solid), 20 Fourier modes (dot-dashed), 3 PIPs (dotted), 4 KL modes (dashed), and 5 KL modes (identical with solid).

position of maximum variance at the left edge of the interval: $x_0 \approx 0.3326$. All models reproduce the periodicity. One recognizes the basic frequency of the limit cycle; the higher harmonics are not shown. The Fourier spectrum is reproduced perfectly in the PIP(3) simulation as well as with the KL(5) model. The KL(4) simulation has a large shift in frequency; in the F(20) model the frequency is slightly too low.

Last, we compare the geometrical structure of the limit cycle in phase space in the complete system and in the reduced systems. For this purpose, Fig. 6 shows the projection of the limit cycle onto the plane spanned by the first two PIPs or KL modes, respectively, since these are virtually identical. The PIP(3) and the KL(5) model actually yield perfect agreement; in the KL(4) simulation the amplitude of the limit cycle is too large in accordance with Fig. 4. The F(20) model cannot be compared to the other models here since the first two PIPs are not fully contained in the subspace spanned by the Fourier modes up to wave number 10.

A plot of the solution obtained from a simulation with the PIP(3) model (not shown) is indistinguishable from Fig. 1. The same holds for the KL(5) model. With the KL(4) and the F(20) model the errors in amplitude and frequency of the oscillations are clearly visible.

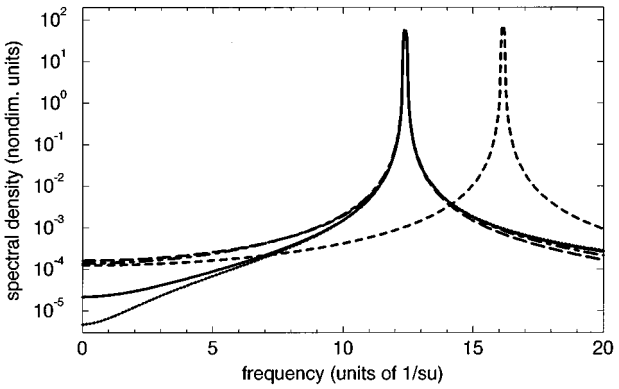

FIG. 5. Fourier spectrum of $u(x_0)$ obtained from a simulation with 30 Fourier modes (complete system, solid), 20 Fourier modes (dot-dashed), 3 PIPs (dotted), 4 KL modes (dashed), and 5 KL modes (long-dashed).

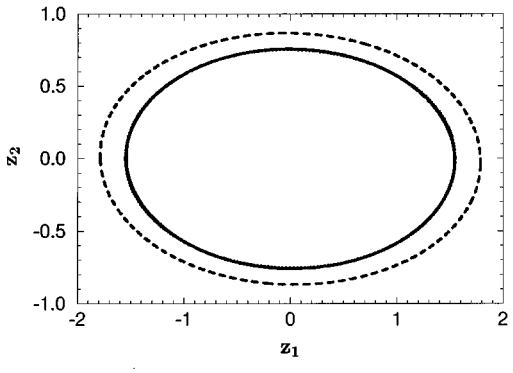

FIG. 6. Projection of the limit cycle obtained from a simulation with 30 Fourier modes (complete system, solid), 3 PIPs (dotted), 4 KL modes (dashed), and 5 KL modes (identical with solid) onto the plane spanned by the first two PIPs.

We also considered reduced models based on Sobolev eigenfunctions as introduced by Kirby [7]. A model using four patterns slightly improves on the KL(4) simulation; but in order to obtain a performance that is as good as that of the PIP(3) model five modes are necessary as with KL modes.

The minimum number of patterns involved in the models based on KL modes and Sobolev eigenfunctions, respectively, is much lower here than is reported in [7]. This may be due to several reasons: first, the reduced model in [7] is formulated for the complete state rather than the anomalies. Given the fact that the system under consideration here is characterized by relatively small fluctuations around a large amplitude mean state (cf. Figs. 3 and 4) a reduced model formulated *a priori* as an anomaly model may be expected to be more efficient since the dynamical interactions between the mean state and the anomaly patterns are fully preserved and the anomaly field is captured better with an expansion concentrating on the anomalies. Secondly, only the scalar product $[\cdot,\cdot]_0$ has been considered by Kirby. Moreover, in the former study the calculation of the KL modes and Sobolev eigenfunctions is based on a Fourier-Galerkin approximation truncated at wave number 10, which is not fully converged; this also may influence the results.

We also extracted PIP models from the KS equation with the algorithm illustrated in [9] and found a PIP model with four degrees of freedom that performs nearly as good as the PIP(3) model described above but did not succeed in describing the dynamics with three patterns.

V. CONCLUSIONS

An algorithm for constructing minimal systems of ODEs modeling the dynamics of nonlinear PDEs has been illustrated. Characteristic spatial structures are obtained from a nonlinear minimization procedure based on a dynamical optimality criterion and used as basis functions in a Galerkin approximation. The method is applied to a limit cycle solution of the KS equation. As to the number of modes required to capture the principal dynamical and statistical properties of the PDE a dynamical description based on PIPs provides a considerable improvement on more conventional techniques using Sobolev eigenfunctions or KL modes as basis functions, a PIP model with three patterns being as good as a

model based on five KL modes or Sobolev eigenfunctions, and is far more efficient than a model based on Fourier modes. Moreover, the present algorithm improves on a previously published PIP algorithm [9]. The methodology of the present paper has also been successfully applied to the complex Ginzburg-Landau equation in a chaotic regime [17].

Presumably, the reduced systems obtained from the method illustrated here are the minimal systems attainable from a PDE when using a linear Galerkin projection. A possibility to arrive at an even further reduced dynamical description may be offered by combining an optimization procedure of the type introduced here with a nonlinear Galerkin scheme, i.e., by projecting the PDE onto an optimized nonlinear approximate inertial manifold rather than an optimized linear subspace. This may be worth pursuing further in future research but clearly lies outside the scope of the present study.

APPENDIX A: MINIMIZATION OF THE ERROR FUNCTION

The minimization of the error function with respect to the patterns is performed numerically using a quasi-Newton algorithm with Broyden-Fletcher-Goldfarb-Shanno update of the approximated Hessian matrix [18,19]. The algorithm requires exact evaluation of the error function and its gradient for arbitrary sets of patterns.

The fact that the constraints of Eqs. (20) and (21) do not restrict the solution of the minimization problem but are only imposed to remove the ambiguity in the representation of the subspace $\mathcal{P}$ can be exploited to facilitate the calculations in the following manner. Imagine an arbitrary $(N\times L)$ matrix P with linearly independent columns representing a set of linearly independent patterns $\{\overline{p_1},\ldots,\overline{p_L}\}$. A function $\mathcal{W}$ is defined that maps P onto its normal form; i.e., onto the (uniquely defined) equivalent $(N\times L)$ matrix $P=\mathcal{W}(P)$ corresponding to a set of patterns $\{p_1,\ldots,p_L\}$ satisfying the constraints. The map $\mathcal{W}$ consists of two parts. First, the patterns $\overline{p_i}$ are orthonormalized using the standard Gram-Schmidt procedure. One ends up with a matrix $P^\perp$ corresponding to a set of orthonormal patterns $\{p_1^\perp,\ldots,p_L^\perp\}$, which are related to the patterns $\overline{p_i}$ by $p_i^\perp=\Sigma_j(T_1)_{ji}\overline{p_j}$, where T_1 is an upper triangular $(L\times L)$ matrix. Let $C^\perp$ be the covariance matrix of the amplitudes $z_i^\perp$ of the patterns $p_i^\perp$: $C_{ij}^\perp=\langle z_i^\perp z_j^\perp\rangle=\Sigma_{\mu,\nu}(M\Gamma M)_{\mu\nu}P_{\mu i}^\perp P_{\nu j}^\perp$. $C^\perp$ can be diagonalized by a real, orthonormal $(L\times L)$ matrix T_2: $T_2^t C^\perp T_2=\mathrm{diag}(\lambda_1^{\mathrm{PIP}},\ldots,\lambda_L^{\mathrm{PIP}})$. Then the matrix $P=P^\perp T_2=PT_1T_2$ represents a set of patterns $\{p_1,\ldots,p_L\}$ with $p_i=\Sigma_j(T_2)_{ji}p_j^\perp=\Sigma_j(T_1T_2)_{ji}\overline{p_j}$, which satisfies the constraints of both Eq. (20) and Eq. (21). Eventually some sign convention is applied to the patterns p_i. One then has $\chi_1(P)=\chi_1(P)$, $\chi_2(P;\tau_{\max})=\chi_2(P;\tau_{\max})$, and $\chi(P;\tau_{\max})=\chi(P;\tau_{\max})$. The elements of P are used as variables in the numerical minimization procedure (not applying any constraints); at each step of the minimization $P=\mathcal{W}(P)$ is calculated as described above. Then χ_1 is calculated according to Eq. (30) and the tensors of interaction coefficients a, b, c, and χ_2 are evaluated using the patterns p_i.

Classical perturbation theory for symmetric matrices applied to $C^\perp$ yields for the first-order variation of χ_1 with respect to the elements of P:

$$\frac{\partial\chi_1}{\partial\overline{P_{\mu r}}}=-\frac{1}{\mathrm{Var}}\sum_{i,j,k}\frac{\partial C_{jk}^\perp}{\partial\overline{P_{\mu r}}}(T_2)_{ji}(T_2)_{ki}. \tag{A1}$$

Forming $\partial C_{jk}^\perp/\partial\overline{P_{\mu r}}$ from the Gram-Schmidt procedure is straightforward using only elementary analysis and is therefore not given in detail here. The gradient of χ_2 is

$$\frac{\partial\chi_2}{\partial\overline{P_{\mu r}}}=\sum_{\nu,i}\frac{\partial\chi_2}{\partial P_{\nu i}}\frac{\partial P_{\nu i}}{\partial\overline{P_{\mu r}}} \tag{A2}$$

with

$$\frac{\partial\chi_2}{\partial P_{\mu r}}=\frac{1}{\mathrm{Var}}\left\langle\frac{\partial J}{\partial P_{\mu r}}\right\rangle=\frac{1}{\mathrm{Var}}\left\{\left\langle\frac{\partial J}{\partial z_r^0}\hat{u}_\mu^0\right\rangle-2\sum_{k=1}^{K}w_k\langle\varepsilon_r^{\tau_k}\hat{u}_\mu^{\tau_k}\rangle+\sum_{i,j}\left[\left\langle\frac{\partial J}{\partial a_{rij}}\right\rangle\Lambda_{\mu ij}+2\left\langle\frac{\partial J}{\partial a_{irj}}\right\rangle\Theta_{i\mu j}\right]\right.$$
$$\left.+\sum_i\left[\left\langle\frac{\partial J}{\partial b_{ri}}\right\rangle\Upsilon_{\mu i}+\left\langle\frac{\partial J}{\partial b_{ir}}\right\rangle\Xi_{i\mu}\right]+\left\langle\frac{\partial J}{\partial c_r}\right\rangle\Pi_\mu\right\}, \tag{A3}$$

$$\Lambda_{\mu ij}=(-1)^{\beta+1}\alpha\left[\frac{\partial^{2\beta}f_\mu}{\partial x^{2\beta}},\frac{\partial p_i}{\partial x}\frac{\partial p_j}{\partial x}\right]_0, \tag{A4}$$

$$\Theta_{i\mu j}=(-1)^{\beta+1}\alpha\left[\frac{\partial^{2\beta}p_i}{\partial x^{2\beta}},\frac{\partial f_\mu}{\partial x}\frac{\partial p_j}{\partial x}\right]_0, \tag{A5}$$

$$\Upsilon_{\mu i}=(-1)^{\beta+1}\left[\frac{\partial^{2\beta}f_\mu}{\partial x^{2\beta}},4\frac{\partial^4 p_i}{\partial x^4}+\alpha\frac{\partial^2 p_i}{\partial x^2}+\alpha\frac{\partial\langle u_F\rangle}{\partial x}\frac{\partial p_i}{\partial x}\right]_0, \tag{A6}$$

$$\Xi_{i\mu}=(-1)^{\beta+1}\left[\frac{\partial^{2\beta}p_i}{\partial x^{2\beta}},4\frac{\partial^4 f_\mu}{\partial x^4}+\alpha\frac{\partial^2 f_\mu}{\partial x^2}+\alpha\frac{\partial\langle u_F\rangle}{\partial x}\frac{\partial f_\mu}{\partial x}\right]_0, \tag{A7}$$

$$\Pi_\mu=(-1)^{\beta+1}\left[\frac{\partial^{2\beta}f_\mu}{\partial x^{2\beta}},4\,\frac{\partial^4\langle u_F\rangle}{\partial x^4}+\alpha\,\frac{\partial^2\langle u_F\rangle}{\partial x^2}+\frac{\alpha}{2}\left(\frac{\partial\langle u_F\rangle}{\partial x}\right)^2\right]_0. \tag{A8}$$

$\partial P_{\nu i}/\partial\overline{P}_{\mu r}$ can be obtained from $\partial C^{\perp}_{ij}/\partial\overline{P}_{\mu r}$ by standard first-order perturbation theory for symmetric matrices applied to $C^{\perp}$. The formulas (A4)–(A8) are evaluated by means of the pseudospectral transform method. In the case of $\beta=1$ the relation

$$\Lambda_{\mu ij}+\Theta_{i\mu j}+\Theta_{j\mu i}=0 \tag{A9}$$

holds. The expressions $\partial J/\partial z_i^0$, $\partial J/\partial a_{ijk}$, $\partial J/\partial b_{ij}$, and $\partial J/\partial c_i$ are calculated efficiently using an adjoint technique (see Appendix B).

Hence by introduction of the map $\mathcal{W}$ the variational principle is formulated as an unconstrained minimization problem; the error function and its derivatives only have to be considered for orthonormal pattern sets and the algorithm automatically supplies the PIP model in its normal form defined in Sec. II C. This greatly reduces the computation time since the computational effort involved in the map $\mathcal{W}$ and its derivatives is negligible compared to that involved in the ensemble of integrations of the PIP model necessary to evaluate χ_2 and its derivatives.

APPENDIX B: THE ADJOINT METHOD

Adjoint techniques originating from the theory of optimal control are an efficient tool to iteratively solve variational problems. They allow for the economical calculation of the gradient of an error function of the type considered here with respect to initial, boundary, or parametric conditions of the system under consideration. Adjoint methods are widely used in the fields of meteorology, oceanography, and climate research for many types of problems including variational data assimilation, parameter fitting, determination of optimally growing perturbations, and sensitivity analysis (see [20] for an overview). The mathematical foundations of the adjoint formalism can be found, e.g., in [21–25]. In the following the adjoint method is briefly described in the form in which it is relevant in the present context.

Given an autonomous nonlinear system of first-order differential equations in L-dimensional phase space,

$$\dot z=G(z;\vartheta),\quad z=(z_1,\ldots,z_L) \tag{B1}$$

depending on a set of adjustable parameters $\vartheta=(\vartheta_1,\ldots,\vartheta_R)$, an error function

$$J(z^0,\vartheta)=\sum_{k=1}^{K}\sum_i w_k(\varepsilon_i^{\tau_k})^2 \tag{B2}$$

is defined where $\varepsilon^{\tau_k}=(\varepsilon_1^{\tau_k},\ldots,\varepsilon_L^{\tau_k})$ is the error between the state $z^{\tau_k}=(z_1^{\tau_k},\ldots,z_L^{\tau_k})$ obtained when integrating the system of Eq. (B1) forward in time from $t=0$ to $t=\tau_k$ with initial condition $z^0=(z_1^0,\ldots,z_L^0)$ and a given set of data $\tilde z^{\tau_k}=(\tilde z_1^{\tau_k},\ldots,\tilde z_L^{\tau_k})$:

$$\varepsilon_i^{\tau_k}=z_i^{\tau_k}-\tilde z_i^{\tau_k}. \tag{B3}$$

τ_k and w_k are defined as in Sec. II D. In the present context the system of Eq. (B1) has to be identified with the PIP model [Eqs. (16) and (49), respectively], the initial condition is $z_i^0=[p_i,\hat u_F^0]$ and the data $\tilde z_i^{\tau_k}$ are given by the projection of the state of the complete system at time τ_k onto the PIPs $\tilde z_i^{\tau_k}=[p_i,\hat u_F^{\tau_k}]$.

At this point the set of adjoint equations is introduced:

$$\dot y_i=-\sum_j\frac{\partial G_j}{\partial z_i}y_j,\quad i=1,\ldots,L, \tag{B4}$$

$$\dot\xi_i=-\sum_j\frac{\partial G_j}{\partial\vartheta_i}y_j,\quad i=1,\ldots,R. \tag{B5}$$

Equations (B1), (B4), and (B5) together form an autonomous system of $2L+R$ first-order ordinary differential equations, which is nonlinear in the variables z_i and linear in the variables y_i and ξ_i. An operator $\mathcal{S}^{\tau_k,\tau_{k-1}}$ is defined by $(y^{\tau_{k-1}},\xi^{\tau_{k-1}})=\mathcal{S}^{\tau_k,\tau_{k-1}}(y^{\tau_k},\xi^{\tau_k})$, where $(y^{\tau_{k-1}},\xi^{\tau_{k-1}})=(y_1^{\tau_{k-1}},\ldots,y_L^{\tau_{k-1}},\xi_1^{\tau_{k-1}},\ldots,\xi_R^{\tau_{k-1}})$ is the state obtained when integrating the adjoint system backward in time from $t=\tau_k$ to $t=\tau_{k-1}$ with initial condition $(y^{\tau_k},\xi^{\tau_k})=(y_1^{\tau_k},\ldots,y_L^{\tau_k},\xi_1^{\tau_k},\ldots,\xi_R^{\tau_k})$. $\mathcal{S}^{\tau_k,\tau_{k-1}}$ depends on the whole trajectory z^τ of the nonlinear system on the interval $[\tau_{k-1},\tau_k]$. Using the main results from the theory of optimal control and exploiting the linearity of the adjoint equations in y_i and ξ_i the following algorithm can be proven: setting $y_i^{\tau_K}=0$, $\xi_i^{\tau_K}=0$ and calculating successively

$$\begin{aligned}
&(y^{\tau_{K-1}},\xi^{\tau_{K-1}})=\mathcal{S}^{\tau_K,\tau_{K-1}}(y^{\tau_K}+w_K\varepsilon^{\tau_K},\xi^{\tau_K})\\
&\vdots\\
&(y^{\tau_{k-1}},\xi^{\tau_{k-1}})=\mathcal{S}^{\tau_k,\tau_{k-1}}(y^{\tau_k}+w_k\varepsilon^{\tau_k},\xi^{\tau_k})\\
&\vdots\\
&(y^0,\xi^0)=\mathcal{S}^{\tau_1,0}(y^{\tau_1}+w_1\varepsilon^{\tau_1},\xi^{\tau_1}),
\end{aligned} \tag{B6}$$

with $\mathcal{S}^{\tau_k,\tau_{k-1}}$ referring to the trajectory z^τ in the interval $[0,\tau_{\max}]$ defined by the initial condition z^0 one gets

$$y_i^0=\frac{1}{2}\frac{\partial J}{\partial z_i^0}, \tag{B7}$$

$$\xi_i^0=\frac{1}{2}\frac{\partial J}{\partial\vartheta_i}. \tag{B8}$$

Hence first one integrates the nonlinear system [Eq. (B1)] forward in time from $t=0$ to $t=\tau_{\max}$ with initial condition z^0 to obtain the errors ε^{τ_k} and calculate J. Then one backward integration of the adjoint system from $t=\tau_{\max}$ to $t=0$

according to Eq. (B6) yields all components of the gradient of J with respect to the initial condition and the system parameters. In the course of the forward integration of the nonlinear system the trajectory z^τ has to be stored at sufficiently many points in the interval $[0,\tau_{\max}]$ since it is needed for the backward integration of the adjoint system.

In the case of the PIP model derived from the Kuramoto-Sivashinsky equation [Eq. (49)] the vector of system parameters ϑ is formed by the elements of the interaction tensors a, b, and c taking into account the symmetry of the quadratic interaction coefficients ($a_{ijk}=a_{ikj}$):

$$\{\vartheta_i;i=1,\ldots,R\}=\{a_{ijk};i,j,k=1,\ldots,L;j\geqslant k\}\cup\{b_{ij};i,j=1,\ldots,L\}\cup\{c_i;i=1,\ldots,L\}. \quad \text{(B9)}$$

This yields $R=\frac{1}{2}L^3+\frac{3}{2}L^2+L$ independent parameters. The set of corresponding adjoint variables is

$$\{\xi_i;i=1,\ldots,R\}=\{\omega_{ijk};i,j,k=1,\ldots,L;j\geqslant k\}\cup\{\eta_{ij};i,j=1,\ldots,L\}\cup\{\varphi_i;i=1,\ldots,L\}. \quad \text{(B10)}$$

The adjoint equations read:

$$\dot{y}_i=-\sum_{j,k} a_{jik}z_k y_j-\sum_j b_{ji}y_j, \quad \text{(B11)}$$

$$\dot{\omega}_{ijk}=-\frac{1}{2}z_j z_k y_i, \quad \text{(B12)}$$

$$\dot{\eta}_{ij}=-z_j y_i, \quad \text{(B13)}$$

$$\dot{\varphi}_i=-y_i. \quad \text{(B14)}$$

[1] L. Sirovich and J. D. Rodriguez, Phys. Lett. A **120**, 211 (1987).
[2] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, J. Fluid Mech. **192**, 115 (1988).
[3] L. Sirovich, Physica D **37**, 126 (1989).
[4] J. D. Rodriguez and L. Sirovich, Physica D **43**, 77 (1990).
[5] F. M. Selten, J. Atmos. Sci. **50**, 861 (1993).
[6] F. M. Selten, J. Atmos. Sci. **52**, 915 (1995).
[7] M. Kirby, Physica D **57**, 466 (1992).
[8] K. Hasselmann, J. Geophys. Res. **93**, 11 015 (1988).
[9] F. Kwasniok, Physica D **92**, 28 (1996).
[10] U. Achatz, G. Schmitz, and K.-M. Greisiger, J. Atmos. Sci. **52**, 3201 (1995).
[11] U. Achatz and G. Schmitz (unpublished).
[12] C. Uhl, R. Friedrich, and H. Haken, Z. Phys. B **92**, 211 (1993).
[13] C. Uhl, R. Friedrich, and H. Haken, Phys. Rev. E **51**, 3890 (1995).
[14] V. K. Jirsa, R. Friedrich, and H. Haken, Physica D **89**, 100 (1995).
[15] R. Friedrich and C. Uhl, Physica D **98**, 171 (1996).
[16] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics* (Springer-Verlag, New York, 1988).
[17] F. Kwasniok (unpublished).
[18] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic Press, London, 1981).
[19] I. M. Navon and D. M. Legler, Mon. Wea. Rev. **115**, 1479 (1987).
[20] P. Courtier, J. Derber, R. Errico, J.-F. Louis, and T. Vukićević, Tellus **45A**, 342 (1993).
[21] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations* (Springer-Verlag, Berlin, 1971).
[22] G. I. Marchuk, *Methods of Numerical Mathematics* (Springer-Verlag, Berlin, 1975).
[23] G. I. Marchuk, *Adjoint Equations and Analysis of Complex Systems* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995).
[24] D. G. Cacuci, J. Math. Phys. **22**, 2794 (1981).
[25] F.-X. Le Dimet and O. Talagrand, Tellus **38A**, 97 (1986).