SIAM J. NUMER. ANAL.
Vol. 41, No. 5, pp. 1893–1925

A NEW LOOK AT PROPER ORTHOGONAL DECOMPOSITION*

MURUHAN RATHINAM† AND LINDA R. PETZOLD‡

Abstract. We investigate some basic properties of the proper orthogonal decomposition (POD) method as it is applied to data compression and model reduction of finite dimensional nonlinear systems. First we provide an analysis of the errors involved in solving a nonlinear ODE initial value problem using a POD reduced order model. Then we study the effects of small perturbations in the ensemble of data from which the POD reduced order model is constructed on the reduced order model. We explain why in some applications this sensitivity is a concern while in others it is not. We also provide an analysis of computational complexity of solving an ODE initial value problem and study the computational savings obtained by using a POD reduced order model. We provide several examples to illustrate our theoretical results.

Key words. proper orthogonal decomposition, model reduction, dynamical systems, numerical methods

AMS subject classification. 37M99

DOI. 10.1137/S0036142901389049

1. Introduction.

1.1. Background on proper orthogonal decomposition. Proper orthogonal decomposition (POD), also known as Karhunen–Loève decomposition or principal component analysis, provides a technique for analyzing multidimensional data. This method essentially provides an orthonormal basis for representing the given data in a certain least squares optimal sense. The POD method may be applied to infinite dimensional data such as fluid flow patterns as well. Truncation of the optimal basis provides a way to find optimal lower dimensional approximations of the given data.

In addition to being optimal in a least squares sense, POD has the property that it uses a modal decomposition that is completely data dependent and does not assume any prior knowledge of the process that generates the data. This property is advantageous in situations where a priori knowledge of the underlying process is insufficient to warrant a certain choice of basis. It also helps in exploring patterns in data that may reveal some insight into the underlying process that generates it.

Combined with the Galerkin projection procedure, POD provides a powerful method for generating lower dimensional models of dynamical systems that have a very large or even infinite dimensional phase space. The fact that this approach always looks for linear (or affine) subspaces instead of curved submanifolds makes it computationally tractable. However, it must be noted that POD does not neglect the nonlinearities of the original vector-field. This is so because if the original dynamical system is nonlinear, then the resulting POD reduced order model will also typically be nonlinear.

These properties of POD are the reason for its wide application in data analysis, data compression, and model reduction in various fields of engineering and science.

*Received by the editors May 7, 2001; accepted for publication (in revised form) March 6, 2003; published electronically November 19, 2003. This research was supported in part by grants EPRI WO-8333-06, NSF/KDI ATM-9873133, and NSF ACI-0086061.
http://www.siam.org/journals/sinum/41-5/38904.html
†Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250 (muruhan@math.umbc.edu).
‡Computational Science and Engineering, University of California Santa Barbara, Santa Barbara, CA 93106 (petzold@engineering.ucsb.edu).

Applications of POD include image processing [22], data compression, signal analysis [2], modeling and control of chemical reaction systems [12, 25, 26], turbulence models [14], coherent structures in fluids [14], control of fluids [10], electrical power grids [21, 20, 18], and wind engineering to name a few.

Extensions and modifications to POD have been proposed by various researchers to accommodate properties of the applications at hand. For instance, instead of time averaging, arclength-based averaging has been found to be useful in capturing dynamics involving "intermittent" attractors in [12]. The predefined POD method has been studied in [8], where modes are selected not only on the basis of energy of the data but also on some prior knowledge of the system. Structure preserving model reduction based on POD for mechanical systems with Lagrangian structure has been developed in [15].

Systems with symmetry deserve special attention. Several authors have made important contributions. Expanding the data set using symmetry was proposed in [27, 28, 29], and later works have shown that it is an essential step in capturing the correct dynamics [3, 4]. Methods for combining reduction theory with POD have been developed in [23].

1.2. Contributions of this work. In this paper we study some basic questions about POD. We focus on finite dimensional systems and follow a deterministic approach. The contributions of this paper include a study of the errors involved in solving an initial value problem using a POD reduced order model of a dynamical system, the sensitivity of the results of POD to perturbations in the data that is used to form the reduced model, as well as computational efficiency gained in using POD in model reduction applications. Even though these are some fundamental questions relating to POD, we believe that they have not been given sufficient attention in the literature.

1.3. Outline of the paper. The rest of the paper is organized as follows. In section 2, we review the POD method as it is applied in data representation as well as in model reduction. In section 3, we present some mathematical preliminaries on the manifold of projection matrices and finite time solution norms of linear time invariant systems. The former is relevant in the sensitivity analysis, and the latter will be useful since throughout this paper we derive particular results for linear time invariant systems. In section 4, we provide an error analysis of the POD method of model reduction as applied to a general nonlinear system. An example is provided to illustrate the various factors affecting the errors. In section 5, we study the sensitivity of the POD projection matrix P (Proposition 5.4), the projected data $\tilde{y}$, and the reduced model solution $\hat{y}$ to perturbations in the data x that is used to form the reduced model. We also study the particular case $y = x$, where the particular data/solution y for which the reduced model is applied is the same as the ensemble of data x from which the reduced model is constructed. Two examples are provided to illustrate the sensitivity results, one focusing on the $y = x$ case. In section 6, we present an estimate of the computational complexity involved in integrating a system of ODEs with and without the use of POD reduced order models. We also provide two examples to illustrate the various factors affecting the computational savings. Finally, in section 7, we make concluding remarks.

2. Proper orthogonal decomposition (POD). POD provides a method for finding the best approximating subspace to a given set of data. Originally POD was used as a data representation technique. For model reduction of dynamical systems,

POD may be used on data points obtained from system trajectories obtained via experiments, numerical simulations, or analytical derivations. Additional information may be found in [14, 19, 17, 16].

2.1. POD in data representation. We shall assume that the *data points* lie in $\mathbb{R}^n$. In the case of a dynamical system this is the phase space. A *data set* is a collection $x^\alpha \in \mathbb{R}^n$, where $\alpha \in \mathcal{I}$. The *index set* $\mathcal{I}$ may be a finite set $\{1, \ldots, N\}$, or a time interval $[0, T]$, or more generally of the form $\mathcal{I} = [0, T] \times \{1, \ldots, N\}$. The latter corresponds to a collection of trajectories. For example, in an image coding problem $\mathcal{I}$ is a finite discrete set. In model reduction of dynamical systems, $\mathcal{I}$ could be of the more general form above. We define an inner product between sets of data x_1 and x_2 with the same index set $\mathcal{I}$ in the obvious way. For example, if x_1 and x_2 are each a collection of N trajectories in the common interval $[0, T]$, i.e., $x_i^\alpha : [0, T] \to \mathbb{R}^n$ for $\alpha = 1, \ldots, N$ and $i = 1, 2$, then

$$(x_1, x_2) = \sum_{\alpha=1}^{N} \int_0^T (x_1^\alpha(t))^T x_2^\alpha(t) dt.$$

The corresponding norm is denoted $\|.\|$.

Remark 2.1. Note that we are using the inner product in our data space ($\mathbb{R}^n$) to induce an inner product in the space of data sets with the same index set.

We shall explain the POD method using the index set $\mathcal{I} = [0, T] \times \{1, \ldots, N\}$. Given a data set x, POD seeks a subspace $S \subset \mathbb{R}^n$ so that the total square distance

$$\|x - \rho_S x\|^2 = \sum_{\alpha=1}^{N} \int_0^T \|x^\alpha(t) - \rho_S x^\alpha(t)\|^2 dt$$

is minimized. Here ρ_S is the orthogonal projection onto the subspace S and $\rho_S x$ is the projected data set. The solution to this problem requires the construction of the *correlation matrix* defined by

$$R = \sum_{\alpha=1}^{N} \int_0^T x^\alpha(t)(x^\alpha(t))^T dt.$$

Note that R is symmetric positive semidefinite. Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N \geq 0$ be the ordered eigenvalues of R. Then the minimum value of $\|x - \rho_S x\|^2$ over all $k(\leq n)$ dimensional subspaces S is given by $\sum_{j=k+1}^{n} \lambda_j$ [14]. In addition the minimizing S is the invariant subspace corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_k$.

Often it may be best to find an affine subspace as opposed to a linear subspace. This requires us first to find the mean value of the data points

$$\bar{x} = \frac{1}{NT} \sum_{\alpha=1}^{N} \int_0^T x^\alpha(t) dt$$

and then construct the *covariance matrix* $\bar{R}$ given by

$$\bar{R} = \sum_{\alpha=1}^{N} \int_0^T (x^\alpha(t) - \bar{x})(x^\alpha(t) - \bar{x})^T dt.$$

Let S_0 be the invariant subspace of the largest k eigenvalues of $\bar{R}$. Then the best approximating affine subspace S passes through $\bar{x}$ and is obtained by shifting S_0 by $\bar{x}$. Algebraically the projection onto the subspace S is given by

$$z = \rho(x - \bar{x}), \tag{2.1}$$

where $z \in \mathbb{R}^k$ are coordinates in the subspace S, $x \in \mathbb{R}^n$ are coordinates in the original coordinate system in $\mathbb{R}^n$, and the matrix ρ of the projection consists of row vectors ϕ_i^T $(i = 1, \ldots, k)$, where ϕ_i are the unit eigenvectors corresponding to the largest k eigenvalues of $\bar{R}$. Note that given any point $p \in S$ with coordinates $z \in \mathbb{R}^k$, the coordinates $x \in \mathbb{R}^n$ of the same point in the original coordinate system are given by

$$x = \rho^T z + \bar{x}.$$

The affine projection $\tilde{x} \subset S$ of a point $x \in \mathbb{R}^n$ in the original coordinates is given by

$$\tilde{x} = P(x - \bar{x}) + \bar{x},$$

where $P = \rho^T \rho \in \mathbb{R}^{n \times n}$ is the matrix of the (linear) projection expressed in the original coordinate system in $\mathbb{R}^n$.

Remark 2.2. Note that the reduced subspace is uniquely characterized by the pair $(\bar{x}, P)$. Different data sets may lead to the same pair $(\bar{x}, P)$, and the detailed information about the data x is lost.

2.2. POD in model reduction. The POD method may also be used in obtaining a lower dimensional model of a dynamical system. In this case, having found the approximating subspace for our system data, the next task is to construct a vector-field on this subspace that represents the reduced order model. The procedure we describe is known as Galerkin projection and has been widely used in reducing PDEs to ODEs by projecting onto appropriate basis functions that describe the spatial variations in the solution. The procedure is applicable to any subspace; the subspace need not be obtained from the POD method. See [14] for more details.

Suppose the original dynamical system in $\mathbb{R}^n$ is given by a vector-field f,

$$\dot{x} = f(x, t).$$

Let $S \subset \mathbb{R}^n$ be the best k dimensional approximating affine subspace with projection given by (2.1). A vector-field f_a in the subspace S is constructed by the following rule: for any point $p \in S$ compute the vector-field $f(p, t)$ and take the projection $\rho f(p, t)$ onto the subspace S to be the value of $f_a(p, t)$. If z are the subspace coordinates of p, then $f_a(z, t) = \rho f(\rho^T z + \bar{x}, t)$. Thus we obtain the following reduced model:

$$\dot{z} = f_a(z, t) = \rho f(\rho^T z + \bar{x}, t). \tag{2.2}$$

If we are solving an initial value problem with $x(0) = x_0$, then in the reduced model one has the initial condition $z(0) = z_0$, where

$$z_0 = \rho(x_0 - \bar{x}).$$

Hence the approximating solution $\hat{x}(t)$ in the original coordinates in $\mathbb{R}^n$ is given by

$$\hat{x}(t) = \rho^T z(t) + \bar{x}.$$

From the above it is easy to see that the approximating solution $\hat{x}(t)$ is the solution to the following initial value problem:

$$\dot{\hat{x}} = Pf(\hat{x}, t); \quad \hat{x}(0) = \hat{x}_0 = P(x_0 - \bar{x}) + \bar{x}. \tag{2.3}$$

Note that $\hat{x}_0$ is just the projection of x_0 onto the affine subspace S.

3. Mathematical preliminaries.

3.1. Manifold of projection matrices. Let $\mathcal{P} \subset \mathbb{R}^{n\times n}$ be the manifold of all rank $k(< n)$ (orthogonal) projection matrices. ($\mathcal{P}$ is known in the literature as the Grassmannian [6].) For a general introduction to manifolds, tangent spaces, and differentiation on manifolds, see [1, 6]. Since we will be dealing with variations E of projections $P \in \mathcal{P}$ in our POD sensitivity analysis, we need a characterization of the tangent space $T_P\mathcal{P}$ to $\mathcal{P}$ at a given point $P \in \mathcal{P}$. The variation $E \in T_P\mathcal{P}$ cannot be any arbitrary matrix. In fact, the dimension of $\mathcal{P}$ and hence that of $T_P\mathcal{P}$ for any P is $k(n-k)$.

Without loss of generality, we can induce an orthonormal change of coordinates in $\mathbb{R}^n$ such that a given projection P becomes the canonical projection (P_0) in these coordinates, i.e.,

$$P_0 = \begin{bmatrix} I_{k\times k} & 0_{k\times n-k} \\ 0_{n-k\times k} & 0_{n-k\times n-k} \end{bmatrix}.$$

In many places in our analysis we shall assume the use of these canonical coordinates.

Let $V = T_{P_0}\mathcal{P}$, i.e., the tangent space to $\mathcal{P}$ at the canonical projection P_0. Using the relations $P^2 = P$ and $P^T = P$ (symmetric) and letting $P = P_0$, it is easy to see that V consists of matrices of the form

$$\begin{bmatrix} 0_{k\times k} & X_{k\times n-k} \\ X^T_{n-k\times k} & 0_{n-k\times n-k} \end{bmatrix},$$

where X is arbitrary. We will use the Frobenius norm for projection matrices P and their variations E in our analysis. We consider the basis $\{E^{ij} \ : \ i = 1, \ldots, k; j = 1, \ldots, n-k\}$ for V, where

$$E^{ij} = \begin{bmatrix} 0 & X_{ij} \\ X^T_{ij} & 0 \end{bmatrix}$$

and X_{ij} is the $k \times (n-k)$ matrix with all zeros except for a 1 in the (i,j)th element. Clearly $\|E^{ij}\| = \sqrt{2}$.

Remark 3.1. It may be noted that E^{ij} corresponds to an infinitesimal rotation of the subspace S (onto which P_0 projects) in the plane of the coordinates x_i and x_{j+k}. Consider the family of subspaces $S(\theta)$ which are spanned by

$$\{e_1, \ldots, e_{i-1}, \cos\theta e_i + \sin\theta e_{j+k}, e_{i+1}, \ldots, e_k\},$$

where $e_1, \ldots, e_n$ are the canonical basis vectors in $\mathbb{R}^n$. Note that when $\theta = 0$, S corresponds to the image of P_0. Computing the corresponding family of projection matrices $P(\theta)$, we can see that $E^{ij} = \frac{dP}{d\theta}(\theta = 0)$.

3.2. Finite time response of a linear time invariant system with time varying input. Some of the analysis in this paper requires estimating the norm of the trajectory of a linear time invariant system in a finite interval in response to a forcing input term.

Consider the system

$$\dot{x} = Ax + u$$

(where $x \in \mathbb{R}^n$) with input $u(t) \in \mathbb{R}^n$ and initial condition $x(0) = x_0$ in the interval $[0, T]$. The solution is

$$x(t) = \int_0^t e^{A(t-\tau)} u(\tau) d\tau + e^{At} x_0.$$

This may be written in the form

$$x = F(T; A)u + G(T; A)x_0, \tag{3.1}$$

where $F(T; A) : \mathcal{L}_2([0, T], \mathbb{R}^n) \to \mathcal{L}_2([0, T], \mathbb{R}^n)$ and $G(T; A) : \mathbb{R}^n \to \mathcal{L}_2([0, T], \mathbb{R}^n)$ are linear operators. It is in general very difficult to obtain sharp estimates for the norms of $F(T; A)$ and $G(T; A)$, and in fact this basically reduces to the problem of estimating the norm of the matrix exponential. As such we shall not provide an estimate, but we remark that these norms grow exponentially with T at a rate that is determined by the largest real part of any eigenvalue of A and in addition depend on the nonnormality of A. See [11] for an estimate of matrix exponential. In our analysis we shall estimate $\|x\|$ as

$$\|x\| \leq \|F(T; A)\|\|u\| + \|G(T; A)\|\|x_0\|, \tag{3.2}$$

expressing the results in terms of $\|F(T; A)\|$ and $\|G(T; A)\|$.

4. Error analysis of the POD method of model reduction. Consider solving the initial value problem $\dot{x} = f(x, t)$, $x(0) = x_0$, using a POD reduced order model in the interval $[0, T]$. Then in effect we are solving the initial value problem (2.3). We shall derive an estimate for the error $e(t) = \hat{x}(t) - x(t)$. Denote the component of $e(t)$ orthogonal to the subspace S by $e_o(t)$ and the component parallel to S by $e_i(t)$. Thus $e_o(t)$ and $e_i(t)$ are orthogonal vectors. Hence by definition $Pe_o(t) = 0$ and $Pe_i(t) = e_i(t)$. It is important to observe that $e_o(t)$ comes from the first part of the method, i.e., the subspace approximation. It is the error between $x(t)$ and its projection onto the subspace S. If one is considering a data compression problem, then $e_o(t) = e(t)$. But since we form a reduced order model by projecting the vector-field onto S, we make further approximations resulting in the additional error $e_i(t)$.

Remark 4.1. Note that for any function $g : [0, T] \to \mathbb{R}^n$, $\|g(t)\|$ is a norm in $\mathbb{R}^n$ which shall be the 2-norm throughout this paper. The function norm will be denoted by $\|g\|$, and unless explicitly stated otherwise it will be assumed to be the 2-norm.

We can derive an error estimate for $e_i(t)$ in terms of $e_o(t)$. Differentiating $e_o(t) + e_i(t) = \hat{x}(t) - x(t)$ and substituting into the ODEs for $\hat{x}$ and x, we get

$$\dot{e}_o + \dot{e}_i = Pf(\hat{x}, t) - f(x, t).$$

Multiplying on the left by P and using $P^2 = P$, we obtain the initial value problem for $e_i(t)$:

$$\dot{e}_i = P(f(x(t) + e_o(t) + e_i, t) - f(x(t), t)); \quad e_i(0) = 0. \tag{4.1}$$

Note that $e_i(0) = 0$ since the starting point $\hat{x}_0$ is the projection of x_0 onto S. Thus the error e_i is governed by (4.1), where we may regard $x(t)$ and $e_o(t)$ as forcing terms. See Figure 4.1, where x is the true solution, $\tilde{x}$ the projected solution, and $\hat{x}$ the solution of the reduced model. The errors e_i and e_o are also shown.

In the case of a linear time invariant system $\dot{x} = Ax$, (4.1) takes a simple form:

$$\dot{e}_i = PAe_i + PAe_o(t); \quad e_i(0) = 0. \tag{4.2}$$

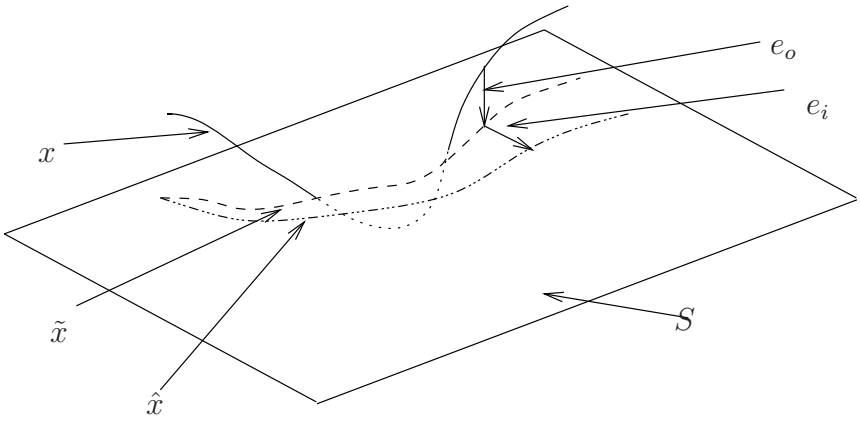

FIG. 4.1. *POD error.*

Applying the notation of (3.1), we get the estimate

$$\|e_i\|_2 \leq \|F(T;\hat{A})\|\|\tilde{A}\|\epsilon.$$

Hence the total error is

$$\|e\|_2 \leq \left(\|F(T;\hat{A})\|\|\tilde{A}\| + 1\right)\epsilon. \tag{4.3}$$

Here $\hat{A} = \rho A \rho^T$ and $\tilde{A} = \rho A \rho_c^T$, where ρ is the projection in subspace coordinates ($P = \rho^T \rho$), and ρ_c is the orthogonal complement to ρ. ϵ is the 2-norm of e_o (i.e., $\|e_o\|_2 = \epsilon$).

Before we state a proposition for the general nonlinear case, recall the definition of a *logarithmic norm related to a* 2*-norm* of a square matrix $A \in \mathbb{R}^{k\times k}$ denoted by $\mu(A)$:

$$\mu(A) = \lim_{h\to 0, h>0} \frac{\|I + hA\|_2 - 1}{h},$$

where I is the identity matrix [13].

PROPOSITION 4.2. *Consider solving the initial value problem $\dot{x} = f(x,t)$, $x(0) = x_0$, using the POD reduced order model in the interval $[0,T]$. Let $\rho \in \mathbb{R}^{k\times n}$ be the relevant projection matrix, and let S denote the affine subspace onto which POD projects. Write the solution (of the full model) $x(t)$ and the solution $\hat{x}(t)$ of the reduced model as*

$$x(t) = \rho^T u(t) + \rho_c^T v(t) + \bar{x}$$

and

$$\hat{x}(t) = \rho^T u(t) + \rho^T w(t) + \bar{x}$$

so that the errors $e_o(t)$ and $e_i(t)$ and the projected solution $\tilde{x}(t)$ are given by

$$e_o(t) = -\rho_c^T v(t),$$

$$e_i(t) = \rho^T w(t),$$

and

$$\tilde{x}(t) = \rho^T u(t) + \bar{x}.$$

Note that $u(t) \in \mathbb{R}^k$, $w(t) \in \mathbb{R}^k$, *and* $v(t) \in \mathbb{R}^{n-k}$. *Let* $\gamma \geq 0$ *be the Lipschitz constant of* $\rho f(x,t)$ *in the directions orthogonal to* S *in a region containing* $x(t)$ *and* $\tilde{x}(t)$. *To be precise, suppose*

$$\|\rho f(\tilde{x}(t) + \rho_c^T v, t) - \rho f(\tilde{x}(t), t)\| \leq \gamma \|v\|$$

for all $(v,t) \in D \subset \mathbb{R}^{n-k} \times [0,T]$, *where the region* D *is such that the associated region* $\tilde{D} = \{(\tilde{x}(t) + \rho_c^T v, t) \ : \ (v,t) \in D\} \subset \mathbb{R}^n \times [0,T]$ *contains* $(\tilde{x}(t), t)$ *and* $(x(t), t)$ *for all* $t \in [0,T]$. *Let* $\mu(\rho \frac{\partial f}{\partial x}(\bar{x} + \rho^T z, t)\rho^T) \leq \bar{\mu}$ *for* $(z,t) \in V \subset \mathbb{R}^k \times [0,T]$, *where the region* V *is such that it contains* $(u(t), t)$ *and* $(u(t) + w(t), t)$ *for all* $t \in [0,T]$ *and* μ *denotes the logarithmic norm related to the* 2*-norm. Let* $\epsilon = \|e_o\|_2$. *Then the error* e_i *in the* ∞*-norm satisfies*

$$\|e_i\|_\infty \leq \epsilon \frac{\gamma}{\sqrt{2\bar{\mu}}} \sqrt{e^{2\bar{\mu}T} - 1}, \tag{4.4}$$

and the 2*-norm of the total error satisfies*

$$\|e\|_2 \leq \epsilon \sqrt{1 + \frac{\gamma^2}{4\bar{\mu}^2}(e^{2\bar{\mu}T} - 1 - 2\bar{\mu}T)}. \tag{4.5}$$

Proof. We shall closely follow the ideas in [13, pp. 54–60]. Since

$$\dot{w}(t) = \rho f(\bar{x} + \rho^T u(t) + \rho^T w(t), t) - \rho f(\bar{x} + \rho^T u(t) + \rho_c^T v(t), t),$$

for $h > 0$ using Taylor expansion we have

$$\begin{aligned}
\|w(t+h)\| &= \|w(t) + h\rho f(\bar{x} + \rho^T u(t) + \rho^T w(t), t) - h\rho f(\bar{x} + \rho^T u(t) + \rho_c^T v(t), t)\| \\
&\quad + O(h^2) \\
&\leq \|w(t) + h\rho f(\bar{x} + \rho^T u(t) + \rho^T w(t), t) - h\rho f(\bar{x} + \rho^T u(t), t)\| \\
&\quad + h\|\rho f(\bar{x} + \rho^T u(t) + \rho_c^T v(t), t) - \rho f(\bar{x} + \rho^T u(t), t)\| + O(h^2).
\end{aligned}$$

Applying the mean value theorem to $\eta \mapsto \eta + h\rho f(\bar{x} + \rho^T \eta, t)$, we get

$$\begin{aligned}
&\|w(t) + h\rho f(\bar{x} + \rho^T u(t) + \rho^T w(t), t) - h\rho f(\bar{x} + \rho^T u(t), t)\| \\
&\qquad \leq \left(\max_{\eta \in [u(t), u(t)+w(t)]} \left\| I + h\rho \frac{\partial f}{\partial x}(\bar{x} + \rho^T \eta, t)\rho^T \right\| \right) \|w(t)\|,
\end{aligned}$$

where for any two vectors η_1, η_2 in $\mathbb{R}^k$, $[\eta_1, \eta_2]$ denotes the line segment joining the two. It follows that

$$\frac{\|w(t+h)\| - \|w(t)\|}{h} \leq \bar{\mu}\|w(t)\| + \gamma\|v(t)\| + O(h),$$

where the $O(h)$ term may be uniformly bounded independent of $w(t)$ [13]. Then it follows from Theorem 10.3 of [13] (also see Theorem 10.6 in [13]) that

$$\|e_i(t)\| = \|w(t)\| \leq \gamma \int_0^t e^{\bar{\mu}(t-\tau)} \|v(t)\| d\tau,$$

since $e_i(t) = \rho^T w(t)$. After applying the Cauchy–Schwarz inequality on the right side, we get

$$\|e_i(t)\| \le \frac{\gamma}{\sqrt{2\bar{\mu}}}\sqrt{e^{2\bar{\mu}t}-1}\sqrt{\int_0^t \|e_o(\tau)\|^2 d\tau}. \tag{4.6}$$

From this we readily obtain (4.4). Also bounding $\sqrt{\int_0^t \|e_o(\tau)\|^2 d\tau}$ by ϵ and integrating (4.6), we obtain an upper bound for $\|e_i\|_2$ which can be combined with $\|e_o\|_2$ to get (4.5). ∎

Remark 4.3. This analysis separates the two different errors and provides a bound for the total error in terms of the projection error ϵ of the true solution $x(t)$. The value of ϵ depends only on the true solution $x(t)$ and on the pair $(P, \bar{x})$ which determines the reduced order model (but not directly on f). If P and $\bar{x}$ were computed from the true solution $x(t)$ in the interval $[0, T]$ (this is somewhat an ideal situation), then $\epsilon = \sqrt{\sum_{j=k+1}^{n} \lambda_j}$, where λ_i are the eigenvalues of the covariance matrix. However, if the reduced model was computed from some other trajectories as often is the case in applications of model reduction methods, then ϵ would depend on how close $x(t)$ was to the trajectories used as data in addition to the quantity $\sum_{j=k+1}^{n} \lambda_j$ (typically the fractional error $\frac{\epsilon}{\|x\|_2}$ will be larger than $\frac{\sum_{j=k+1}^{n} \lambda_j}{\sum_{j=1}^{n} \lambda_j}$). For instance, in hybrid systems such as power systems where discrete events abruptly change some system parameters, data obtained from trajectories before the event results in a reduced order model with a large ϵ for simulations after the event [7].

Example 1. This example serves to illustrate the various factors that affect e_i given the same projection error e_o. We shall consider a linear time invariant system $\dot{x} = Ax$; $x(0) = x_0$. Assume A has distinct eigenvalues and that it possesses some fast decaying modes (eigenvalues with large negative real parts). Let $S \subset \mathbb{R}^n$ be the invariant subspace corresponding to the rest of the eigenvalues, where S is $k(< n)$ dimensional. If we have sufficiently many trajectories that have initial conditions symmetrically placed with respect to S, then the POD method will pick S as the subspace to project onto. We shall assume this to be the case. Performing an orthonormal change of coordinates if needed, we may assume that S corresponds to the last $n-k$ coordinates being zero. In these coordinates, the A matrix has the form

$$A = \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix},$$

where $A_1 \in \mathbb{R}^{k\times k}$, $A_{12} \in \mathbb{R}^{k\times(n-k)}$, and $A_2 \in \mathbb{R}^{(n-k)\times(n-k)}$. In fact the real Schur decomposition of A will put it in the above form. We shall say A is "block normal" if the off diagonal block $A_{12} = 0$.

Also note that

$$\rho = \begin{bmatrix} I_{k\times k} & 0_{k\times(n-k)} \end{bmatrix}$$

and

$$\rho_c = \begin{bmatrix} 0_{(n-k)\times k} & I_{(n-k)\times(n-k)} \end{bmatrix}.$$

Hence $\bar{\mu} = \mu(A_1)$ and $\gamma = \|A_{12}\|$.

For a given initial condition and time interval, the error e_o relates to the last three components of the solution and does not change if A_2 is unchanged. We can

independently change $\bar{\mu}$ and γ by changing A_1 and A_{12}, respectively. We kept A_2 unchanged, thus keeping $\epsilon = \|e_o\|_2$ unchanged, and studied the effects of changing A_1 (and hence $\bar{\mu}$) and A_{12} (and hence γ) independently on the POD error. First we chose

$$A_1 = \begin{bmatrix} -0.1000 & 0 & 0 \\ 0 & -0.1732 & 2.0 \\ 0 & -2.0 & -0.1732 \end{bmatrix},$$

$$A_{12} = \begin{bmatrix} 0.3893 & 0.5179 & -1.543 \\ 1.390 & 1.300 & 0.8841 \\ 0.06293 & -0.9078 & -1.184 \end{bmatrix},$$

and

$$A_2 = \begin{bmatrix} -1.0 & 0 & 0 \\ 0 & -1.226 & -0.7080 \\ 0 & 0.7080 & -1.226 \end{bmatrix}.$$

Note that the eigenvalues of A_2 have large negative real parts compared to the eigenvalues of A_1, and the eigenvalues of A are the union of these two sets. The corresponding $\bar{\mu} = -1$ and $\gamma = 2.4419$. We randomly chose an initial condition and computed $x(t)$, $\tilde{x}(t)$, and $\hat{x}(t)$ in the interval $[0, 5]$. Note that the reduced model has dimension 3 and that the last three components of both $\tilde{x}(t)$ and $\hat{x}(t)$ are zero. Similarly, the first three components of $e_i(t)$ and $e_o(t)$ are zero. See Figure 4.2, where only the nonzero components are plotted. The computed value of the projection error was $\epsilon = \|e_o\|_2 = 1.4575$. The sup-norm and the 2-norm of the error in the subspace S were also computed and found to be $\|e_i\|_\infty = 1.5589$ and $\|e_i\|_2 = 2.5733$. The bounds provided by the theory were $\|e_i\|_\infty \leq 6.3271$ and $\|e_i\|_2 \leq 10.7930$.

The second choice was to keep A_1 and A_2 the same but scale A_{12} down by a factor of 2. We kept the same initial condition and time interval. This results in the same $\bar{\mu}$, but $\gamma = 1.2209$. In fact, according to (4.2), the effect of scaling A_{12} affects the error e_i linearly, and we expect $e_i(t)$ to be scaled down by the same factor of 2. Figure 4.3 shows a plot of $e_i(t)$ for both cases. This highlights how the rate of change of S components of the vector-field in the directions orthogonal to S affect the error. In the extreme case when $A_{12} = 0$ (i.e., A is "block normal"), the components of the vector-field parallel to S are invariant in the directions perpendicular to S, and the error e_i is zero. Thus the error e_i is zero for a matrix that is "block normal" (with respect to a decomposition of the space based on "fast decay" and the rest of the eigenmodes) if the POD indeed captures the attracting subspace S correctly.

The error e_i is more influenced by $\bar{\mu}$ than γ (as long as $\gamma > 0$), and $\bar{\mu}$ is supposed to capture the growth or decay of solutions of the vector-field of the reduced model. Keeping A_2 and A_{12} the same, we changed A_1 so that

$$A_1 = \begin{bmatrix} -0.1000 & 0 & 1.0 \\ 0 & -0.1732 & 2.0 \\ 0 & -2.0 & -0.1732 \end{bmatrix}.$$

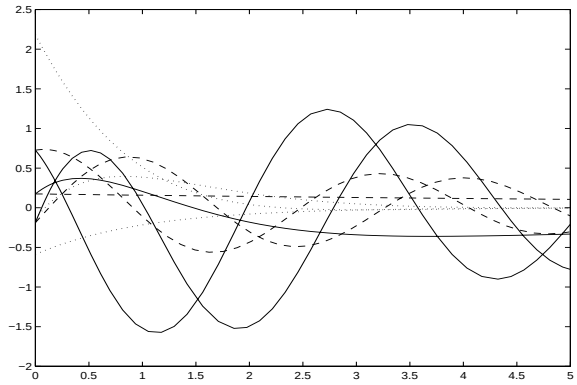

FIG. 4.2. *Example* 1 *on POD error. Solid: Projected solution* $\tilde{x}(t)$*. Dashed: Reduced model solution* $\hat{x}(t)$*. Dotted: Projection error* $e_o(t)$*. Only the three nonzero components* $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$*,* $\hat{x}_1, \hat{x}_2, \hat{x}_3$ *and* e_{o4}, e_{o5}, e_{o6} *are plotted.*

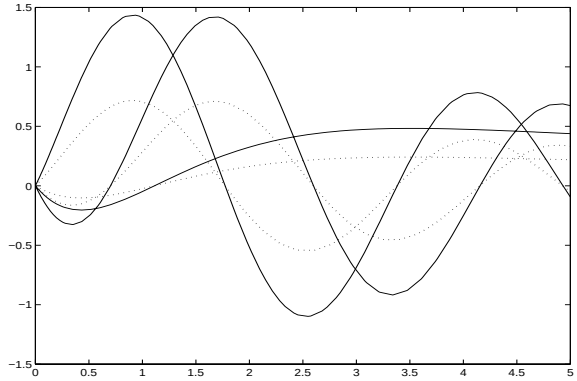

FIG. 4.3. *Example* 1 *on POD error. The effect of scaling* A_{12} *on the error* e_i *in the subspace* S*. Solid:* e_i *for unscaled* A_{12}*. Dotted:* e_i *for scaled down* A_{12}*. Only the three nonzero components are plotted.*

The corresponding $\bar{\mu} = 0.3647$. Note that this does not change the eigenvalues of A_1, but it does change its normality. This choice of A_1 is no longer normal, and even though the eigenvalues remain the same, the short term behavior of $e_i(t)$ is changed. In fact, $e_i(t)$ does not decay as much as in the normal case. This results in $\|e_i\|_\infty = 2.2088$ and $\|e_i\|_2 = 3.4565$. The bounds provided by the theory are $\|e_i\|_\infty \leq 25.4739$ and $\|e_i\|_2 \leq 28.3330$.

5. Sensitivity of POD to perturbations in data. Given a data set x, POD constructs a projection $P(x)$ onto a subspace which may then be used to approximate some other data set y. If POD is applied to model reduction to compute the approximation $\hat{y}$ to the true solution y of some ODE initial value problem, then the projection $P(x)$ will influence $\hat{y}$. Typically in POD applications the data set x comes from experimental measurement or numerical computations. Hence the data x has some error associated with it. Therefore, it is important to study the effect of these errors on the outcome of the POD model reduction procedure. In this section, we shall theoretically investigate the effect of infinitesimal perturbations of x on $P(x)$, $\tilde{y}$, and $\hat{y}$. We also look at the special case when $y = x$.

5.1. POD sensitivity factor. Let x be a data set, and let $P(x)$ be the corresponding POD projection. In this section, we analyze the sensitivities of the POD projection matrix $P(x)$, with respect to variations in the data x. Our analysis applies to any data set x taking values in $\mathbb{R}^n$, but for simplicity of exposition we assume x to be a single trajectory ($x : [0, T] \to \mathbb{R}^n$) whenever we need to be concrete.

Shifting the origin in data space if necessary, we may assume the mean data values $\bar{x} = 0$. In addition, we can find an orthonormal change of coordinates such that the covariance matrix of x is diagonal. We shall call this a *canonical coordinate system* for data set x. Assuming the use of these canonical coordinates, let

$$x = \sum_{\alpha=1}^{n} x_\alpha e_\alpha, \tag{5.1}$$

where e_α are the standard basis vectors in $\mathbb{R}^n$. Then it can be shown that the scalar data sets x_α are orthogonal. More specifically,

$$(x_\alpha, x_\beta) = \lambda_\alpha \delta_{\alpha,\beta}.$$

Here $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the covariance matrix of x. If, in addition, we permute the coordinates such that $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n \geq 0$ are the ordered eigenvalues, then we call this an *ordered canonical coordinate system*. Throughout the rest of the analysis, we shall assume that, after ordering, $\lambda_k > \lambda_{k+1}$ unless stated otherwise.

The POD projection matrix $P \in \mathcal{P} \subset \mathbb{R}^{n\times n}$ is defined as the minimizer of the function

$$e(P, x) = (Px - x, Px - x).$$

Differentiating with respect to P in the direction of E, we obtain

$$\frac{\partial e}{\partial P}(E) = 2(Px - x, Ex).$$

Thus stationary points P of e are given by the condition

$$(Px - x, Ex) = 0, \quad E \in T_P\mathcal{P}.$$

Performing an orthonormal change of coordinates if necessary, we may assume $P = P_0$ (canonical projection) is a stationary point. Then all variations $E \in V = T_{P_0}\mathcal{P}$. Requiring $(P_0 x - x, E^{ij}x) = 0$ for all E^{ij} gives us the conditions that $(x_i, x_{j+k}) = 0$ for all $1 \leq i \leq k$ and $1 \leq j \leq n - k$. This shows that all the stationary points of e are given by P that project onto any of the k dimensional invariant subspaces of the covariance matrix of x. A solution P to the above equation will be a strong local minimum if and only if the second derivative $\frac{\partial^2 e}{\partial P^2}$ is positive definite and this may be shown to be equivalent to $\lambda_k > \lambda_{k+1}$. Under this assumption, P is also a well-defined function of x locally.

LEMMA 5.1. *Without loss of generality, let $P = P_0$ be a stationary point in some canonical coordinate system (this may not be ordered). Let $E \in V$ and $\tilde{E} \in V$ be given by*

$$E = \begin{bmatrix} 0_{k\times k} & X_{k\times n-k} \\ X^T_{n-k\times k} & 0_{n-k\times n-k} \end{bmatrix}$$

and

$$\tilde{E} = \begin{bmatrix} 0_{k\times k} & \tilde{X}^T_{k\times n-k} \\ \tilde{X}_{n-k\times k} & 0_{n-k\times n-k} \end{bmatrix}.$$

Then the Hessian satisfies

$$\frac{\partial^2 e}{\partial P^2}(E)(\tilde{E}) = 2(X^T x_1, \tilde{X}^T x_1) - 2(Xx_2, \tilde{X}x_2). \tag{5.2}$$

Proof. Since e is a function on the manifold $\mathcal{P}$, one could introduce local coordinates on $\mathcal{P}$ to compute the Hessian of e at a stationary point. However, we shall use a coordinate independent method which allows us to work with matrices and keep the algebra simple. It may be shown that if $P(t,s)$ is a smooth mapping from $\mathbb{R}^2$ into $\mathcal{P}$ such that $P(0,0) = P_0$, $P_t(0,0) = E \in V$ and $P_s(0,0) = \tilde{E} \in V$ and if $\frac{\partial e}{\partial P}(P_0) = 0$, then the Hessian at $P = P_0$ is given by

$$\frac{\partial^2 e}{\partial P^2}(E)(\tilde{E}) = e_{ts}(0,0),$$

where P and e are regarded as functions of t and s and subscripts denote partial derivatives.

Differentiating $e = 2(Px - x, Px - x)$ with respect to t, we get

$$e_t = 2(P_t x, Px - x),$$

and differentiating again with respect to s, we get

$$e_{ts} = 2(P_{ts}x, Px - x) + 2(P_t x, P_s x). \tag{5.3}$$

Suppose

$$P_{ts}(0,0) = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix}.$$

The matrices W_1, W_2, W_3, and W_4 are not arbitrary but satisfy some relations. These are obtained by differentiating the relation $P^2 = P$ twice. In fact, we get

$$P_{ts}P + P_t P_s + P_s P_t + P P_{ts} = P_{ts},$$

and after substituting expressions for P, P_t, P_s, and P_{ts} (at $(t,s) = (0,0)$) in the above and using the fact that P_{ts} is symmetric, we obtain that $W_1 = -\tilde{X}X^T - X\tilde{X}^T, W_4 = \tilde{X}^T X + X^T \tilde{X}, W_3^T = W_2$, where W_2 is an arbitrary $k \times (n-k)$ matrix. It then follows that $P_{ts}(0,0) = F + W$, where

$$F = \begin{bmatrix} -\tilde{X}X^T - X\tilde{X}^T & 0 \\ 0 & \tilde{X}^T X + X^T \tilde{X} \end{bmatrix}$$

and

$$W = \begin{bmatrix} 0 & W_2 \\ W_2^T & 0 \end{bmatrix},$$

and hence $W \in V$. Hence from (5.3) we obtain

$$e_{ts}(0,0) = 2(Fx, P_0 x - x) + 2(Wx, P_0 x - x) + 2(Ex, \tilde{E}x).$$

Since $\frac{\partial e}{\partial P} = 0$ at $P(0,0) = P_0$ by assumption and $W \in V$, it follows that $(Wx, P_0x - x) = 0$. Let $x = (x_1, x_2)$, where $x_1(t) \in \mathbb{R}^k$ and $x_2(t) \in \mathbb{R}^{n-k}$. It is easy to see that

$$(Ex, \tilde{E}x) = (Xx_2, \tilde{X}x_2) + (X^T x_1, \tilde{X}^T x_1),$$

where the inner products on the right-hand side are in the appropriate function spaces. Since $(Fx, P_0x - x) = ((P_0 - 1)Fx, x)$, after computing $(P_0 - 1)F$ it can be shown that

$$(Fx, P_0x - x) = -(\tilde{X}^T X x_2 + X^T \tilde{X} x_2, x_2) = -2(Xx_2, \tilde{X}x_2).$$

Equation (5.2) follows from this. □

Remark 5.2. From (5.2) it may be shown that the Hessian $\frac{\partial^2 e}{\partial P^2}$ has E^{ij} as its eigenvectors with corresponding eigenvalues $2(\lambda_i - \lambda_{j+k})$ for $1 \le i \le k$ and $1 \le j \le n-k$ (note that we did not order the eigenvalues). Thus the stationary point $P = P_0$ is a strong minimum (maximum) if and only if the first k eigenvalues are strictly greater (smaller) than the rest. It is clear that if, after ordering, $\lambda_k > \lambda_{k+1}$, then there is a unique strong minimum and a unique strong maximum. The rest of the stationary points are saddle points.

The sensitivity of P to variations in data x is given by $\frac{dP}{dx}(\delta x)$, the directional derivative of P with respect to x in the direction δx, where $\delta x : [0,T] \to \mathbb{R}^n$ is assumed to be a unit-norm variation of x ($\|\delta x\| = 1$). It suffices to consider zero mean variations. This is because one may decompose any variation $\delta x \in \mathcal{L}_2([0,T] \to \mathbb{R}^n)$ into a constant function plus a zero mean function, and it is easy to see that the constant function part affects only the mean value $\bar{x}$ of the data while the zero mean function part affects only the projection P.

Remark 5.3. Variations of variables are denoted by prefix δ except for variations of P, which are denoted by E (or $\tilde{E}$, etc.). We will use the 2-norm for functions and the Frobenius norm for matrices P and E.

The norm $\|\frac{dP}{dx}\|$ is defined by

$$\left\| \frac{dP}{dx} \right\| = \sup_{\|\delta x\|=1} \left\| \frac{dP}{dx}(\delta x) \right\|$$

and measures the worst-case sensitivity of P to unit-norm variations of x. However, it makes more sense to consider the nondimensional quantity defined by

$$S_k(x) = \left\| \frac{dP}{dx} \right\| \|x - \bar{x}\|, \tag{5.4}$$

which we shall call the *POD sensitivity factor*. It is the worst-case ratio (in the limit of zero perturbation) of the perturbation of P to the fractional perturbation $\frac{\delta(x-\bar{x})}{\|x-\bar{x}\|}$. We use $x - \bar{x}$ instead of x because P depends only on $x - \bar{x}$. If we scale the data set x by a constant $c \in \mathbb{R}$, then both $\bar{x}$ and $x - \bar{x}$ also scale by c, but P remains unchanged ($P(cx) = P(x)$). The definition of S_k takes care of this scaling symmetry. In fact, we get $S_k(x) = S_k(cx)$. Note that the suffix k stands for the dimension of the reduced subspace $S \subset \mathbb{R}^n$ in which the projected data lives.

PROPOSITION 5.4. *Consider applying POD to a data set x to find the best approximating $k(< n)$ dimensional subspace. Let the ordered eigenvalues of the covariance matrix of the data x be given by $\lambda_1 \ge \cdots \ge \lambda_n$. Suppose $\lambda_k > \lambda_{k+1}$, which ensures*

that $P(x)$ is well defined. Then

$$S_k(x) = \max_{i\le k,\, j\le n-k} \sqrt{2}\frac{\sqrt{\lambda_i+\lambda_{j+k}}}{\lambda_i-\lambda_{j+k}}\sqrt{\lambda_1+\cdots+\lambda_n} \ge \sqrt{2}. \tag{5.5}$$

Furthermore, in the ordered canonical coordinates corresponding to data x, the unit-norm variation δx that causes the maximal variation in P is given by

$$\delta x = \frac{E^{\tilde{i},\tilde{j}}(x-\bar{x})}{\sqrt{\lambda_{\tilde{i}}+\lambda_{\tilde{j}+k}}}, \tag{5.6}$$

where $i=\tilde{i}$ and $j=\tilde{j}$ maximize the right-hand side of (5.5).

Proof. In this proof we will use ordered canonical coordinates. Differentiating $\frac{\partial e}{\partial P}(E)=0$ totally with respect to x in the δx direction, we get

$$\frac{\partial^2 e}{\partial P^2}(E)\left(\frac{dP}{dx}(\delta x)\right)+\frac{\partial^2 e}{\partial P\partial x}(E)(\delta x)=0 \quad \forall E\in V. \tag{5.7}$$

Hence $\frac{dP}{dx}(\delta x)$ is implicitly defined through the above equation.

The mixed partial $\frac{\partial^2 e}{\partial P\partial x}$ is given by

$$\begin{aligned}\frac{\partial^2 e}{\partial P\partial x}(E)(\delta x) &= 2(Ex,(P-1)\delta x)+2(E\delta x,(P-1)x)\\ &= 2((PE-E)x,\delta x)+2((EP-E)x,\delta x)\\ &= -2(Ex,\delta x),\end{aligned} \tag{5.8}$$

where in the first step we used the fact that $E^T=E$ and $P^T=P$, and in the second step we used the fact that $PE+EP=E$ (which comes from $P^2=P$). Note that if $\bar{x}\neq 0$, then $e=(P(x-\bar{x})-(x-\bar{x}),P(x-\bar{x})-(x-\bar{x}))$. Even though we assumed without loss of generality that $\bar{x}=0$, when we take variations of x we need to consider the corresponding variations of $\bar{x}$. However, since we care only about zero mean variations δx, for those the corresponding variation $\delta\bar{x}=0$. Hence we are justified in neglecting the term $\bar{x}$.

It is instructive to examine the finite dimensional space U of $\mathbb{R}^n$-valued functions defined by

$$U=\operatorname{span}\{E'(x-\bar{x}) \ : \ E'\in V\}. \tag{5.9}$$

From (5.8) (note that we assumed $\bar{x}=0$) it can be seen that if δx is orthogonal to U, then $\frac{\partial^2 e}{\partial P\partial x}=0$, and hence by (5.7) $\frac{dP}{dx}(\delta x)=0$. Since we are only interested in variations δx that introduce nonzero variations in P, we shall assume $\delta x\in U$. It can be shown that the map $E'\in V\to E'(x-\bar{x})\in U$ is an isomorphism. This is readily seen by evaluating this map on the basis E^{ij} and showing that $E^{ij}(x-\bar{x})=E^{ij}x$ form an independent set. In fact,

$$E^{ij}x = x_i e_{j+k}+x_{j+k}e_i,$$

and hence

$$\begin{aligned}(E^{ij}x,E^{lm}x) &= \lambda_i+\lambda_{j+k}, \quad l=i, j=m,\\ &= 0 \quad \text{otherwise.}\end{aligned} \tag{5.10}$$

Hence $\{E^{ij}x \ : \ i = 1, \ldots, k; j = 1, \ldots, n-k\}$ is an orthogonal set (and is clearly independent as well).

Substitute (5.2) and (5.8) into the implicit equation (5.7) for $\frac{dP}{dx}$, and let

$$\frac{dP}{dx}(\delta x) = \begin{bmatrix} 0 & \tilde{X} \\ \tilde{X}^T & 0 \end{bmatrix}$$

and $\delta x = (\delta x_1, \delta x_2)$, where $\delta x_1(t) \in \mathbb{R}^k$ and $\delta x_2(t) \in \mathbb{R}^{n-k}$. Then we obtain an equation for $\tilde{X}$:

$$(X^T x_1, \tilde{X}^T x_1) - (X x_2, \tilde{X} x_2) = (X x_2, \delta x_1) + (X^T x_1, \delta x_2) \quad \forall X \in \mathbb{R}^{k \times (n-k)}. \tag{5.11}$$

Let $\delta x = E' x \in U$, where

$$E' = \sum_{l,m} \alpha_{lm} \frac{E^{lm}}{\sqrt{\lambda_l + \lambda_{m+k}}},$$

and let $\tilde{X} = \sum_{l,m} \beta_{lm} X_{lm}$. Substituting these into (5.11) for $X = X_{ij}$, we get

$$\beta_{ij} = \alpha_{ij} \frac{\sqrt{\lambda_i + \lambda_{j+k}}}{\lambda_i - \lambda_{j+k}}.$$

Hence it follows that

$$\frac{dP}{dx}(\delta x) = \sum_{i,j} \alpha_{ij} \frac{\sqrt{\lambda_i + \lambda_{j+k}}}{\lambda_i - \lambda_{j+k}} E^{ij}. \tag{5.12}$$

The requirement that $\|\delta x\| = 1$ is equivalent to $\sum_{i,j} \alpha_{ij}^2 = 1$. Since $\|E^{ij}\| = \sqrt{2}$, it follows that

$$\left\| \frac{dP}{dx} \right\| = \max_{i \le k,\ j \le n-k} \sqrt{2} \frac{\sqrt{\lambda_i + \lambda_{j+k}}}{\lambda_i - \lambda_{j+k}}, \tag{5.13}$$

with the maximizing unit-norm variation δx given by (5.6). (Note that we need to replace x by $x - \bar{x}$, since we assumed for simplicity that $\bar{x} = 0$.) The equation in (5.5) follows from this. The inequality in (5.5) follows because

$$\max_{i \le k, j \le n-k} \frac{\sqrt{\lambda_i + \lambda_{j+k}}}{\lambda_i - \lambda_{j+k}} \sqrt{\lambda_1 + \cdots + \lambda_n} \ge \max_{i \le k, j \le n-k} \frac{\lambda_i + \lambda_{j+k}}{\lambda_i - \lambda_{j+k}} \ge 1. \qquad \square$$

The following corollary is obvious from the above proof.

COROLLARY 5.5. *Assuming $\lambda_k > \lambda_{k+1}$ as before and the use of ordered canonical coordinates, the linear map $\delta x \in U \mapsto E(x - \bar{x}) \in U$, where $E = \frac{dP}{dx}(\delta x)$, is self-adjoint and has as its eigenvectors the orthonormal basis of U given by $\{u_{ij}\}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n-k$, which are defined by*

$$u_{ij} = E^{ij}(x - \bar{x}) = \frac{x_i e_{j+k} + x_{j+k} e_i}{\sqrt{\lambda_i + \lambda_{j+k}}}.$$

The corresponding eigenvalues are $\frac{\lambda_i + \lambda_{j+k}}{\lambda_i - \lambda_{j+k}}$. Hence the induced 2-norm of this operator is $\frac{\lambda_k + \lambda_{k+1}}{\lambda_k - \lambda_{k+1}}$.

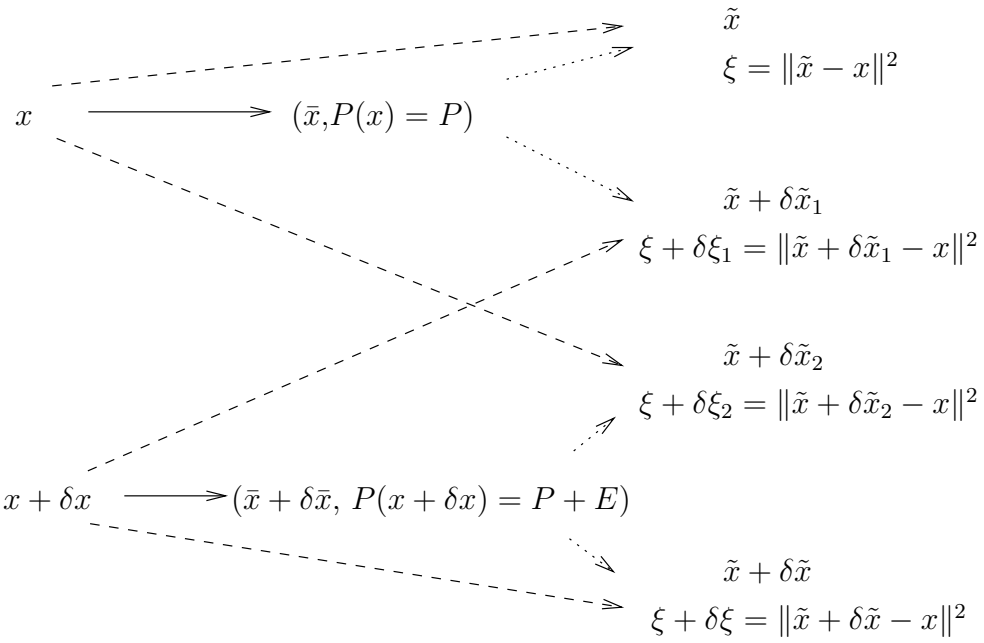

FIG. 5.1. *POD sensitivity for y near x: This shows the four different possibilities of using the reduced models computed from x and $y = x + \delta x$ to approximate these data sets. The solid arrows indicate the construction of a POD reduced model from a data set. A pair of dashed and dotted arrows together show the reduced model being applied to a data set to obtain a reduced and approximate data set. The approximate data obtained and its square error with respect to x are shown on the far right.*

Remark 5.6. Proposition 5.4 was concerned with the POD method of finding the best approximating affine subspace using the mean and the covariance matrix of the data x. Instead, if we considered the POD method of finding the best approximating linear subspace using the correlation matrix, then we get the same equations and the same final expression (5.5) for $S_k(x)$ (in the definition of the space U, $(x - \bar{x})$ needs to be replaced by x). However, x, the perturbation δx, and the worst-case perturbation of δx as well as functions in the space U are no longer necessarily zero mean.

5.2. Sensitivity of the projected data $\tilde{y} = P(x)(y - \bar{x}) + \bar{x}$ and the error $\|\tilde{y} - y\|^2$ when $y = x$ and/or $y = x + \delta x$. In some applications the POD reduced model $(\bar{x}, P(x))$ constructed from a data set x may be used to approximate x itself ($y = x$ situation) or some nearby data $y = x + \delta x$. For instance, consider coding a 512×512 grey scale image by dividing it into subimages of size 8×8 to provide an ensemble of $4096 (= 64 \times 64)$ points in the 64 dimensional subimage space. Suppose that by applying POD to this ensemble we find a subspace of dimension 6 that captures 99.9% of the energy. We could then apply the POD projection to the subimages and code the entire image using 4096×6 grey scale values. This is the $y = x$ situation. If we have a sequence of nearby images (such as in video), then we can use the same reduced model $(\bar{x}, P(x))$ for the nearby images $y = x + \delta x$.

From a theoretical point of view, several different sensitivities may be of interest. These are shown in Figure 5.1. The sensitivities $\delta\tilde{x}_1$ and $\delta\xi_1$ correspond to the situation where the same reduced model $(\bar{x}, P(x))$ (obtained from x) is applied to both x and to a nearby $y = x + \delta x$. The quantity $\delta\tilde{x}_1$ is the perturbation of the

approximate data, and $\delta\xi_1$ is the perturbation of the square error $\xi = \|\tilde{x} - x\|^2$. Thus

$$\delta\tilde{x}_1 = (P(x)(x + \delta x - \bar{x}) + \bar{x}) - \tilde{x}$$

and

$$\delta\xi_1 = \|(\tilde{x} + \delta\tilde{x}_1) - (x + \delta x)\|^2 - \xi,$$

where $\tilde{x} = P(x)(x - \bar{x}) + \bar{x}$. This is the most common kind of sensitivity one is interested in in practice. Note that the reason for considering the square of the error rather than the error $\|\tilde{x} - x\|$ itself is because the square root is not smooth when its argument is zero.

The sensitivities $\delta\tilde{x}_2$ and $\delta\xi_2$ correspond to the situation where two nearby reduced models $(\bar{x}, P(x))$ and $(\bar{x} + \delta\bar{x}, P(x + \delta x)$ are applied to the same data x. Thus

$$\delta\tilde{x}_2 = (P(x + \delta x)(x - (\bar{x} + \delta\bar{x})) + (\bar{x} + \delta\bar{x})) - \tilde{x}$$

and

$$\delta\xi_2 = \|(\tilde{x} + \delta\tilde{x}_2) - x\|^2 - \xi.$$

The sensitivities $\delta\tilde{x}$ and $\delta\xi$ correspond to the situation where two nearby reduced models $(\bar{x}, P(x))$ and $(\bar{x} + \delta\bar{x}, P(x + \delta x))$ are applied to the respective data sets x and $x + \delta x$ from which they were constructed. Thus

$$\delta\tilde{x} = (P(x + \delta x)((x + \delta x) - (\bar{x} + \delta\bar{x})) + (\bar{x} + \delta\bar{x})) - \tilde{x}$$

and

$$\delta\xi = \|(\tilde{x} + \delta\tilde{x}) - (x + \delta x)\|^2 - \xi.$$

We provide a useful and easy-to-prove lemma stated without proof.

LEMMA 5.7. *Let $L : H \to H$ be a linear operator in the Hilbert space H. Let $K \subset H$ be a closed linear subspace of H. Then we can write $H = K \oplus K^\perp$. Furthermore, suppose $L(K) \subset K$ and $L(K^\perp) \subset K^\perp$ and that the restrictions $L|_K$ and $L|_{K^\perp}$ are bounded operators. Then $\|L\|_2 = \max\{\|L|_K\|_2, \|L|_{K^\perp}\|_2\}$.*

PROPOSITION 5.8. *Consider applying POD to a data set x to find the best approximating $k(< n)$ dimensional subspace. Let the ordered eigenvalues of the covariance matrix of the data x be given by $\lambda_1 \geq \cdots \geq \lambda_n$. Suppose $\lambda_k > \lambda_{k+1}$, which ensures that $P(x)$ is well defined. Consider the sensitivities depicted in Figure* 5.1. *For unit-norm (infinitesimal) variations δx of x, the worst-case variations are given by*

$$\|\delta\tilde{x}_1\| = \|\delta x\|, \tag{5.14}$$

$$\|\delta\tilde{x}_2\| = \frac{\lambda_k + \lambda_{k+1}}{\lambda_k - \lambda_{k+1}} > 1, \tag{5.15}$$

$$\|\delta\tilde{x}\| = \frac{\lambda_k + \sqrt{\lambda_k\lambda_{k+1}}}{\lambda_k - \lambda_{k+1}} > 1, \tag{5.16}$$

$$|\delta\xi_1| = |\delta\xi| = 2\|\tilde{x} - x\| = 2\sqrt{\xi}, \tag{5.17}$$

$$\delta\xi_2 = 0. \tag{5.18}$$

(All norms are 2-norms.)

Proof. We shall use the ordered canonical coordinate system whenever necessary.

We define some relevant subspaces. As before we shall consider the data x to be a single trajectory $x : [0,T] \to \mathbb{R}^n$ for simplicity of exposition. However, the results will hold for more general types of data. We shall assume $x \in \mathcal{L}_2([0,T],\mathbb{R}^n)$, the space of all square integrable $\mathbb{R}^n$-valued functions in $[0,T]$. Let $Z \subset \mathcal{L}_2([0,T],\mathbb{R}^n)$ denote the (closed) subspace of all zero mean functions:

$$Z = \left\{ x \in \mathcal{L}_2([0,T],\mathbb{R}^n) \ : \ \int_0^T x dt = 0 \right\}.$$

Its orthogonal complement $Z^\perp$ is finite dimensional and consists of functions that are constant-valued (almost everywhere) in $[0,T]$. We shall further decompose Z into the orthogonal sum $Z = W \oplus Y$, where

$$W = \operatorname{span}\left\{ \frac{x_\alpha}{\sqrt{\lambda_\alpha}} e_\beta \ : \ \alpha = 1,\dots,\tilde{n},\ \beta = 1,\dots,n \right\}. \tag{5.19}$$

Here $\tilde{n}$ is the number of nonzero eigenvalues of the covariance matrix associated with trajectory x (thus W depends on x), and x_α are its components in the ordered canonical coordinate system. Since Y is the orthogonal complement of W in Z, it is closed. The $n\tilde{n}$ dimensional W is further decomposed into the orthogonal sum $W = U \oplus U_2 \oplus V_1 \oplus V_2$, where

$$\begin{aligned}
U &= \operatorname{span}\left\{ u_{ij} = \frac{x_i e_{j+k} + x_{j+k} e_i}{\sqrt{\lambda_i + \lambda_{j+k}}} \ : \ i = 1,\dots,k;\ j = 1,\dots,n-k \right\}, \\
U_2 &= \operatorname{span}\left\{ u^2_{ij} = \frac{\lambda_{j+k} x_i e_{j+k} - \lambda_i x_{j+k} e_i}{\sqrt{\lambda_i \lambda_{j+k}(\lambda_i + \lambda_{j+k})}} \ : \ i = 1,\dots,k;\ j = 1,\dots,\tilde{n}-k \right\}, \\
V_1 &= \operatorname{span}\left\{ \frac{x_i e_\alpha}{\sqrt{\lambda_i}} \ : \ i = 1,\dots,k;\ \alpha = 1,\dots,k \right\}, \\
V_2 &= \operatorname{span}\left\{ \frac{x_{j+k} e_{\beta+k}}{\sqrt{\lambda_{j+k}}} \ : \ j = 1,\dots,\tilde{n}-k;\ \beta = 1,\dots,n-k \right\}.
\end{aligned}$$

It should be noted that the spanning elements above form orthonormal bases for the respective subspaces. Furthermore, define $\tilde{U} = U \oplus U_2$ and $V = V_1 \oplus V_2$. Also note that U defined above is the same as in (5.9).

The perturbation $\delta\tilde{x}_1$ is given by

$$\delta\tilde{x}_1 = P(x)(x + \delta x - \bar{x}) + \bar{x} - P(x)(x - \bar{x}) - \bar{x}$$

and simplifies to $\delta\tilde{x}_1 = P(x)\delta x$. Hence the worst perturbation δx is in the image of $P(x)$, resulting in $\delta\tilde{x}_1 = \delta x$. Note that this holds for finite as well as infinitesimal perturbations.

The variation $\delta\tilde{x}_2$ is given by

$$\delta\tilde{x}_2 = E(x - \bar{x}) + (1 - P(x))\delta\bar{x},$$

where $E = \frac{dP}{dx}(\delta x)$. Note that $\delta x \in Z^\perp$ implies that $E = 0$ and hence that $\delta\tilde{x}_2 = (1-P)\delta\bar{x} \in Z^\perp$. Also note that $\delta x \in Z$ implies $\delta\tilde{x}_2 \in Z$. Furthermore, if $\delta x \in Z$ and $\delta x \perp U$, then $\delta\tilde{x}_2 = 0$. If $\delta x \in U$, then $\delta\tilde{x}_2 = E(x - \bar{x}) \in U \subset Z$. From Corollary

5.5 and Lemma 5.7 it is clear that the worst-case variation $\|\delta\tilde{x}_2\| = \frac{\lambda_k+\lambda_{k+1}}{\lambda_k-\lambda_{k+1}} > 1$ and that it corresponds to $\delta x = \frac{x_k e_{k+1}+x_{k+1}e_k}{\sqrt{\lambda_k+\lambda_{k+1}}}$.

The variation $\delta\tilde{x}$ is given by

$$\delta\tilde{x} = E(x-\bar{x}) + P\delta x - P\delta\bar{x} + \delta\bar{x}.$$

Denote by L the operator that maps δx to $\delta\tilde{x}$. The following are easy to establish: $L(Z) \subset Z$, $L(Z^\perp) \subset Z^\perp$, $L(W) \subset W$, $L(Y) \subset Y$, $L(V) \subset V$, and $L(\tilde{U}) \subset \tilde{U}$. If $\delta x \in Z^\perp$, then $\delta\tilde{x} = \delta x = \delta\bar{x}$, so $\|L|_{Z^\perp}\| = 1$. If $\delta x \in Z$ and $\delta x \perp \tilde{U}$, it follows that $\delta\tilde{x} = P\delta x$. Hence by Lemma 5.7

$$\|L\| = \max\{\|L|_{\tilde{U}}\|, 1\}.$$

It can be verified that the following orthonormal basis of $\tilde{U}$ are eigenvectors of the finite dimensional operator $L|_{\tilde{U}} : \tilde{U} \to \tilde{U}$:

$$\left\{ \frac{x_i e_{j+k}}{\sqrt{2\lambda_i}} + \frac{x_{j+k}e_i}{\sqrt{2\lambda_{j+k}}}, \frac{x_i e_{j+k}}{\sqrt{2\lambda_i}} - \frac{x_{j+k}e_i}{\sqrt{2\lambda_{j+k}}}, \frac{x_i e_{\tilde{j}+k}}{\sqrt{2\lambda_i}} \ : i = 1,\dots,k;\ j = 1,\dots,\tilde{n}-k; \right.$$
$$\left. \tilde{j} = \tilde{n}-k+1,\dots,n-k \right\}.$$

The eigenvectors $\{\frac{x_i e_{\tilde{j}+k}}{\sqrt{2\lambda_i}}\}$ all have eigenvalue 1. The eigenvectors $\{\frac{x_i e_{j+k}}{\sqrt{2\lambda_i}} + \frac{x_{j+k}e_i}{\sqrt{2\lambda_{j+k}}}\}$ have corresponding eigenvalues $\frac{\lambda_i+\sqrt{\lambda_i\lambda_{j+k}}}{\lambda_i-\lambda_{j+k}} > 1$. The eigenvectors $\{\frac{x_i e_{j+k}}{\sqrt{2\lambda_i}} - \frac{x_{j+k}e_i}{\sqrt{2\lambda_{j+k}}}\}$ have corresponding (positive) eigenvalues $\frac{\lambda_i-\sqrt{\lambda_i\lambda_{j+k}}}{\lambda_i-\lambda_{j+k}} < 1$. Hence the norm $\|L|_{\tilde{U}}\|$ is given by the largest eigenvalue $\frac{\lambda_k+\sqrt{\lambda_k\lambda_{k+1}}}{\lambda_k-\lambda_{k+1}} > 1$. Hence $\|L\| = \frac{\lambda_k+\sqrt{\lambda_k\lambda_{k+1}}}{\lambda_k-\lambda_{k+1}}$.

The (infinitesimal) variation $\delta\xi_1$ is given by

$$\delta\xi_1 = 2(\tilde{x}-x, \delta\tilde{x}_1 - \delta x) = (\tilde{x}-x, (P-1)\delta x),$$

and hence the worst case is when $\delta x = \pm\frac{\tilde{x}-x}{\|\tilde{x}-x\|}$ and results in $|\delta\xi_1| = 2\|\tilde{x}-x\|$.

The variation $\delta\xi$ is given by

$$\begin{aligned}\delta\xi &= 2(\tilde{x}-x, \delta\tilde{x}-\delta x)\\ &= 2(\tilde{x}-x, E(x-\bar{x})) + 2(\tilde{x}-x, (1-P)\delta\bar{x}) + 2(\tilde{x}-x, (1-P)\delta x)\\ &= 2(\tilde{x}-x, (1-P)\delta x).\end{aligned}$$

As before, the worst-case variation is given by $\delta x = \pm\frac{\tilde{x}-x}{\|\tilde{x}-x\|}$ and results in $|\delta\xi| = 2\|\tilde{x}-x\|$.

The variation $\delta\xi_2$ is given by

$$\begin{aligned}\delta\xi_2 &= 2(\tilde{x}-x, \delta\tilde{x}_2)\\ &= 2(\tilde{x}-x, E(x-\bar{x})) + 2(\tilde{x}-x, (1-P)\delta\bar{x}).\end{aligned}$$

Since $\tilde{x}-x \in V_2$, $E(x-\bar{x}) \in U$, and $(1-P)\delta\bar{x} \in Z^\perp$, it follows that $\delta\xi_2 = 0$. ∎

Remark 5.9. The above proposition shows that the projected data $\tilde{x}$ may become extremely sensitive to perturbations in the data set when $\lambda_k \approx \lambda_{k+1}$. However, the (square of the) error itself does not show this sensitivity. This is related to the fact that when $\lambda_k = \lambda_{k+1}$ there are infinitely many choices for $P(x)$ and thus for $\tilde{x}$, and

these different choices for $\tilde{x}$ may be quite different from each other. However, they all have exactly the same error $\sqrt{\lambda_{k+1} + \cdots + \lambda_n}$. It should also be noted that our sensitivity results hold for infinitesimal variations, and finite perturbations are likely not to be as sensitive as the first derivative may suggest.

Example 2. This example illustrates a situation where the reduced model solution is very sensitive to perturbations of the trajectory x used as POD data. We consider a dissipative ODE example which has a periodic orbit which is a global attractor. Consider the ODE

$$\dot{x} = A(x - f(t)) + f'(t), \quad x \in \mathbb{R}^n,$$

where $f : \mathbb{R} \to \mathbb{R}^n$ is smooth. Observe that for any choice of A and f, $x = f(t)$ is a trajectory of this system. We exploit this fact to independently choose $f(t)$ and A to create an interesting example which highlights some of the potential problems with the POD procedure.

We chose $f(t)$ to be periodic and A to be a constant matrix with all of its eigenvalues in the complex left half plane. Thus $x = f(t)$ will be a global attractor of this system. Specifically we chose $f(t)$ to be of the form

$$f(t) = \left(\sqrt{a_1}\sin\left(\frac{2\pi t}{25}\right),\ \sqrt{a_2}\cos\left(\frac{2\pi t}{25}\right),\ \sqrt{a_3}\sin\left(\frac{4\pi t}{25}\right),\ \sqrt{a_4}\cos\left(\frac{4\pi t}{25}\right)\right)^T,$$

where a_i are real nonnegative constants. This trajectory has period 25. If we use this trajectory in the interval $[0, 50]$ (two periods) as POD data, we will get a reduced model with $\bar{x} = 0$ and a diagonal covariance matrix R with $R_{ii} = 25a_i$. This is because the component functions of $f(t)$ in the interval $[0, 50]$ form an orthogonal set. If we choose $a_4 = 0$ (or very small), then the POD procedure based on this trajectory will give a reduced model ODE by projecting onto the first three components in $\mathbb{R}^4$. This projection will preserve all (or almost all) of the energy of the POD data trajectory. Now consider a matrix A that has all of its eigenvalues in the complex left half plane, but its submatrix consisting of the first three rows and columns (i.e., the projection of A onto the first three components in $\mathbb{R}^4$) has an eigenvalue in the complex right half plane. Such a choice of A will lead to a reduced model which is unstable. If $a_4 = 0$, the global attractor $x = f(t)$ will still be a trajectory of the reduced model but it will not be an attractor. Thus the qualitative behavior of the reduced model will be quite different even though the POD procedure is based on a global attractor of a dissipative system.

In order to find such an A, we first chose A to be diagonalizable with eigenvalues

$$\lambda(A) = \{-0.7 + 0.4i, -0.7 - 0.4i, -0.2, -0.1\}.$$

Then by trial and error, applying random similarity transformations, we found an A with the above canonical form such that its submatrix consisting of the first three rows and columns had an eigenvalue of about 1.8 in the complex right half plane.

If we choose $a_4 = 0$, then with $k = 3$ we do not expect high sensitivity to perturbations in the data. However, we get an interesting example where doing POD on a lower dimensional global attractor still leads to a reduced model which is unstable and qualitatively different. Since we were interested in studying the effects of perturbations in the POD data on the final outcome of a POD reduced model solution, instead of choosing $a_4 = 0$, we chose the following values for a_i:

$$a_1 = 5,\ a_2 = 0.5,\ a_3 = 0.011,\ a_4 = 0.01,$$

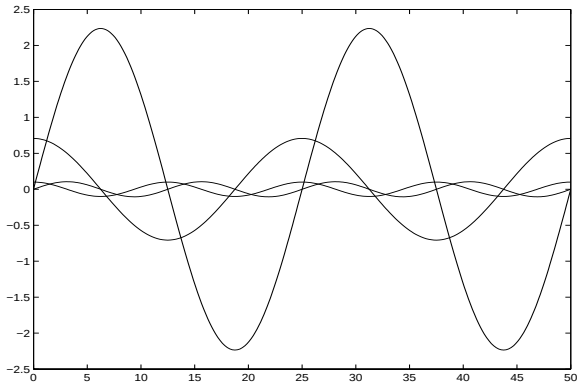

FIG. 5.2. *Example* 2: *True solution* $x(t)$.

where $a_3 \approx a_4$. With this choice, a reduced model of dimension $k = 3$ will capture most of the energy, but we will expect high sensitivity to small perturbations in the POD data. We chose the initial value problem with $x(0) = f(0)$ and time interval $[0, 50]$. Thus the solution trajectory is $x = f(t)$ and consists of two periods. In order to incoorporate the effects of numerical errors, we computed the solution using the MATLAB solver `ode45` and then computed the POD reduced model of dimension $k = 3$ numerically. We also numerically computed the projected trajectory $\tilde{x}$ as well as the solution $\hat{x}$ of the reduced ODE model. We found that the POD procedure preserved 99.8% of the energy and that the sensitivity factor was $S_k = 480$. We found $\|x\| = 11.75$ and $\|\tilde{x}\| = 11.74$. The reduced model solution was highly unstable and $\|\hat{x}\| = 1.25 \times 10^{38}$. The eigenvalues of the reduced model matrix were $\{1.80, -0.281 + 0.217i, -0.281 - 0.217i\}$.

Then we perturbed the trajectory x by δx in the direction given by (5.6) that creates the worst perturbation in P. We chose $\|\delta x\| = 0.1$. We then computed the POD reduced model corresponding to $x + \delta x$ and also computed the perturbed projected trajectory $\tilde{x} + \delta\tilde{x}$ as well as the perturbed reduced model solution $\hat{x} + \delta\hat{x}$. Figure 5.2 shows a plot of the numerically computed true solution $x(t)$ (i.e., the full model solution). Figures 5.3 and 5.4 show how a small perturbation in x leads to a larger perturbation in $\tilde{x}$, and Figure 5.5 shows an even larger perturbation in $\hat{x}$. We also observed that while the unperturbed reduced model projected almost onto the first three components in $\mathbb{R}^4$, the perturbed reduced model was projecting onto a subspace that consisted of the span of $\{e_1, e_2\}$ and a combination of e_3 and e_4, and this subspace was rotated from the span of $\{e_1, e_2, e_3\}$ by an angle of about 41° (e_i being the standard basis vectors in $\mathbb{R}^4$). It was also observed that the eigenvalues of the perturbed reduced model matrix were $\{2.89, -0.337, -0.166\}$, which correspond to a larger instability and a qualitatively different nonoscillatory behavior from that of the unperturbed reduced model.

This example illustrates two potential inadequacies of the POD method. One is that even capturing 100% of the energy of a globally attracting low dimensional trajectory may still lead to a POD reduced model with the wrong dynamics. Second, it also illustrates how POD sensitivity to the data trajectory may lead to qualitatively different reduced models.

The first problem is related to two factors. One is that a single trajectory (even a global attractor) or a set of trajectories alone does not carry all the information

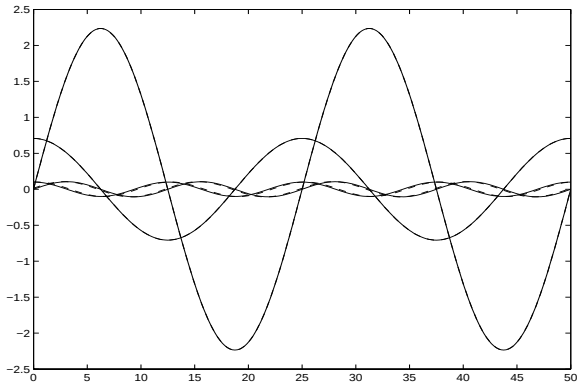

FIG. 5.3. *Example* 2. *Perturbation of* x*: Solid:* x*; dashed:* $x+\delta x$*. The perturbation is so small* ($\frac{\|\delta x\|}{\|x\|} = 0.0085$) *that the two trajectories are barely distinguishable. Note that all four components are plotted for both* x *and* $x+\delta x$*. The perturbation is only noticeable in the two smaller components.*

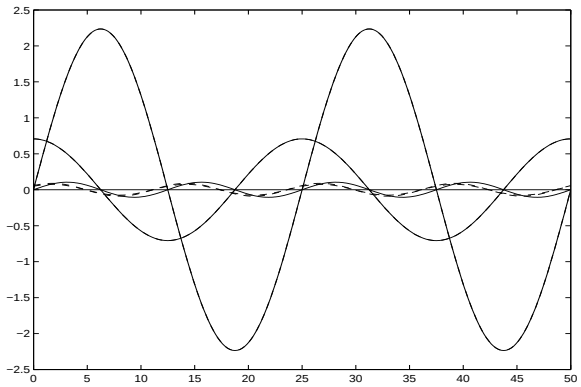

FIG. 5.4. *Example* 2. *Perturbation of* $\tilde{x}$*: Solid:* $\tilde{x}$*; dashed:* $\tilde{x}+\delta\tilde{x}$*. The perturbation is larger than that of* x *but still not noticeable for the two large components.*

about the dynamics. Second, the projection of a vector-field onto a given subspace does not preserve its stability characteristics. Our example has these characteristics and in addition has a high sensitivity factor.

5.3. Effect of POD sensitivity in data representation and model reduction. Consider the reduced model $(\bar{x}, P(x))$ obtained from a data set x. Suppose we apply this reduced model to represent another data set y and obtain $\tilde{y} = P(x)(y-\bar{x})+\bar{x}$. The previous subsection was concerned with the special situation where $y = x$. In general situations, the data set y is different from x. The variation $\delta\tilde{y}$ due to a variation δx is given by

$$\delta\tilde{y} = E(y-\bar{x}) + \delta\bar{x} - P\delta\bar{x},$$

where E is the corresponding variation of P. Since $\|E\| \le S_k \frac{\|\delta(x-\bar{x})\|}{\|x-\bar{x}\|}$ (assuming $\|x-\bar{x}\| \neq 0$), we obtain

$$\|\delta\tilde{y}\| \le S_k \frac{\|y-\bar{x}\|}{\|x-\bar{x}\|}\|\delta(x-\bar{x})\| + \|\delta\bar{x}\|.$$

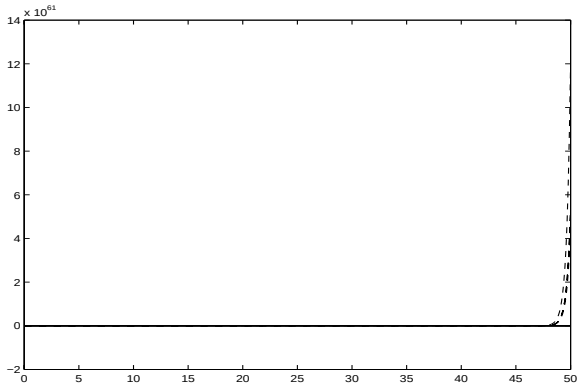

FIG. 5.5. *Example* 2. *Perturbation of* $\hat{x}$*: Solid:* $\hat{x}$*; dashed:* $\hat{x} + \delta\hat{x}$*. Note that* $\hat{x}$ *appears to be zero since it is of the order* 10^{38}*, while* $\hat{x} + \delta\hat{x}$ *is of the order* 10^{61}*.*

Assuming further that $\|y\| \neq 0$, we obtain the following fractional sensitivity relation:

$$\frac{\|\delta\tilde{y}\|}{\|y\|} \leq S_k \frac{\|y - \bar{x}\|}{\|y\|} \frac{\|\delta(x - \bar{x})\|}{\|x - \bar{x}\|} + \frac{\|\delta\bar{x}\|}{\|y\|}. \tag{5.20}$$

Now let us consider the case where we use the reduced model $(\bar{x}, P(x))$ to compute the solution of an ODE initial value problem for a linear time invariant system $\dot{y} = Ay$, with initial condition $y(0) = y_0$ in the interval $[0, T]$. Let y denote the true solution and $\hat{y}$ denote the reduced model solution. Then $\hat{y}$ satisfies the initial value problem $\dot{\hat{y}} = PA\hat{y}$, $\hat{y}(0) = P(y_0 - \bar{x}) + \bar{x}$. Taking variations, we get

$$\delta\dot{\hat{y}} = PA\delta\hat{y} + EA\hat{y},$$

with initial condition $\delta\hat{y}(0) = E(y_0 - \bar{x}) - P\delta\bar{x} + \delta\bar{x}$. Hence applying the estimate (3.2), we get

$$\|\delta\hat{y}\| \leq \|F(T; PA)\|\|E\|\|A\|\|\hat{y}\| + \|G(T; PA)\|\|E\|\|y_0 - \bar{x}\| + \|G(T; PA)\|\|\delta\bar{x}\|,$$

where F and G are defined by (3.1). Since $\|E\| \leq S_k \frac{\|\delta(x-\bar{x})\|}{\|x-\bar{x}\|}$, it follows that

$$\begin{aligned} \frac{\|\delta\hat{y}\|}{\|\hat{y}\|} \leq & \left(\|F(T; PA)\|\|A\| + \|G(T; PA)\| \frac{\|y_0 - \bar{x}\|}{\|\hat{y}\|} \right) S_k \frac{\|\delta(x - \bar{x})\|}{\|x - \bar{x}\|} \\ & + \|G(T; PA)\| \frac{\|\delta\bar{x}\|}{\|x - \bar{x}\|}. \end{aligned} \tag{5.21}$$

Example 3. We considered the same initial value problem of Example 2 in the same interval. However, instead of using the true solution as POD data, we used the set of eight trajectories x obtained by solving the system in the same interval with the symmetrically placed initial conditions $x(0) = e_i$ and $x(0) = -e_i$ for $i = 1, \ldots, 4$, where $e_i \in \mathbb{R}^4$ are the standard basis vectors as POD data, and computed the rank 3 projection matrix $P(x)$. The sensitivity factor was $S_k = 10.140$. We then perturbed this data set x in the direction given by (5.6) (this gives the worst perturbation in P) by an amount $\|\delta x\| = 0.5$. The norm of the data set was $\|x\| = 54.070$, and $\|x - \bar{x}\| = 52.90$. Thus we had the fractional change $\|\delta x\|/\|x - \bar{x}\| = 0.0095$. We also

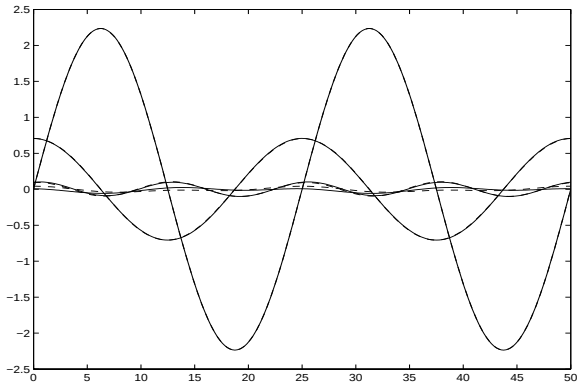

FIG. 5.6. *Example* 3. *Perturbation of* $\tilde{y}$*: Solid:* $\tilde{y}$*; dashed:* $\tilde{y} + \delta\tilde{y}$.

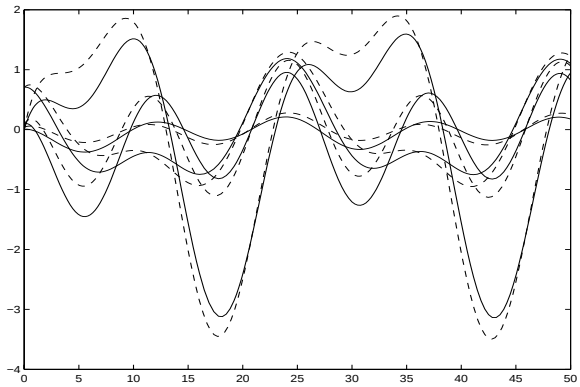

FIG. 5.7. *Example* 3. *Perturbation of* $\hat{y}$*: Solid:* $\hat{y}$*; dashed:* $\hat{y} + \delta\hat{y}$.

computed the projections $P(x)$ and $P(x + \delta x)$ corresponding to the data sets x and $x + \delta x$.

Denote by y the true solution of the initial value problem. (This is the same as $x(t)$ of Example 2 in Figure 5.2.) We applied both reduced models $P(x)$ and $P(x+\delta x)$ to compute $\tilde{y} = P(x)(y - \bar{x}) + \bar{x}$ and $\tilde{y} + \delta\tilde{y} = P(x + \delta x)(y - \bar{x} - \delta\bar{x}) + \bar{x} + \delta\bar{x}$, the projected solutions. Figure 5.6 shows the two different projected solutions.

We then computed the reduced model solutions $\hat{y}$ and $\hat{y}+\delta\hat{y}$ corresponding to $P(x)$ and $P(x + \delta x)$, respectively. These are plotted in Figure 5.7. This again illustrates how a small perturbation in the POD data set may cause a large perturbation in the reduced model solution.

Remark 5.10. In this section we basically saw how POD results may be very sensitive to slight perturbations in the data when $\lambda_k \approx \lambda_{k+1}$. However, one needs to be careful in interpreting these results. This raises the question of whether one should consider the sensitivity factor S_k (in addition to the projection error $\sqrt{\lambda_{k+1} + \cdots + \lambda_n}$) as an important factor in choosing an appropriate dimension k for the reduced model. Sometimes the distribution of eigenvalues may be such that seeking higher accuracy may lead to a high sensitivity factor S_k. The importance of S_k depends on the nature of the application. It must also be noted that our sensitivity analysis holds only for infinitesimal perturbations; the sensitivity for finite perturbations is likely to be

different. For instance, infinitesimal analysis predicts that in the limit $\lambda_k \to \lambda_{k+1}$ the sensitivity of P with respect to x grows indefinitely. However, since projection matrices live on a compact set ($\|P\| = 1$), the finite perturbations of P cannot grow indefinitely. As mentioned in Remark 5.9, when $y \approx x$, for data compression type problems we have already argued that the sensitivity factor is not a serious issue. However, in reduced order models for ODEs, Examples 2 and 3 show how a small perturbation in the POD data may lead to very large perturbations in the reduced model solutions.

In addition, we would like to note that in iterative methods based on POD such as the DIRM method [20, 21], the convergence of the iterations will depend on the sensitivity factor. This has been theoretically and numerically demonstrated in [20].

6. Computational complexity of POD reduced order models. Since the POD method of model reduction results in a smaller dimensional system of ODEs, one might expect computational savings when integrating the resulting system. However, this may not be so in practice. In this section we shall take a close look at the complexity of computation for integrating a system of ODEs (initial value problems) and evaluate the savings if any in using the POD method.

To be precise, by complexity we shall mean the asymptotic behavior of the number of floating point operations (flops—addition, multiplication, and elementary functions each count as one flop) involved per integration step as the system size n (k for reduced models) becomes very large. Table 6.1 shows the complexity of various basic operations that are used in integration of ODEs. All matrices are $n \times n$ and vectors are n dimensional. Banded matrices are assumed to have $b+1$ nonzero entries symmetrically placed around the diagonal. See [11] for details on complexity of linear algebraic operations. We slightly abuse the notation and denote by $f(n)$ the number of flops involved in computing a nonlinear vector-field $f(x,t)$, where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$. Computing the reduced order vector-field $\rho f(\rho^T z + \bar{x}, t)$ could potentially be more expensive than $f(n)$. If this is naively treated as a composition of functions, then the complexity is $f(n) + 4nk$ (two matrix-vector multiplications should be included). Depending on the form of f, one may not be able to improve on this. However, often the analytical formula $\rho f(\rho^T z + \bar{x}, t)$ may be simplified, especially if f is a polynomial in x. We shall denote the complexity of this term by $\hat{f}(k,n)$. This may be bounded as follows:

$$\hat{f}(k,n) \leq f(n) + 4nk.$$

For Jacobian evaluations we have assumed the use of centered finite differences. See [9] for efficient numerical evaluation of banded Jacobians by finite difference approximation. Throughout the rest of this section we will assume that $f(n)$ and $\hat{f}(k,n)$ are of order greater than or equal to n and k, respectively. Under this assumption we can ignore the subtractions and divisions involved in computing the finite differences. If analytical Jacobians are used, the corresponding complexity is likely to be similar [5]. It must be noted that even if the original Jacobian is banded, the reduced model Jacobian is not likely to be.

First we shall consider a linear time invariant system $\dot{x} = Ax$. Table 6.2 shows the asymptotic complexities for various cases. The explicit method considered is forward Euler, and the implicit method considered is backward Euler. The explicit case ($x_n = x_{n-1} + h_n A x_{n-1}$) involves basically a matrix-vector product. The implicit case involves solving the equation $(I - h_n A)x_n = x_{n-1}$ at each time step. We assumed that this is done by Gaussian elimination, first doing an LU decomposition and then two

TABLE 6.1
Complexity of some basic operations. Dense and banded refer to the full model Jacobians. Jacobian evaluation assumes centered finite differencing.

	Dense	Banded
Matrix-vector product Ab	$2n^2$	$2bn$
LU decomposition	$\frac{2n^3}{3}$	$\frac{b^2 n}{2}$
Triangular linear system solve	n^2	bn
Nonlinear function evaluation	$f(n)$	$f(n)$
Nonlinear function evaluation (reduced model)	$\hat{f}(k,n)$	$\hat{f}(k,n)$
Nonlinear Jacobian evaluation	$2nf(n)$	$2(b+1)f(n)$
Nonlinear Jacobian evaluation (reduced model)	$2k\hat{f}(k,n)$	$2k\hat{f}(k,n)$

TABLE 6.2
Asymptotic complexity for linear systems.

	Full model explicit	Reduced model explicit	Full model implicit	Reduced model implicit
Dense	$2n^2$	$2k^2$	$\frac{n^3}{15}$	$\frac{k^3}{15}$
Banded	$2bn$	$2k^2$	$(\frac{b^2}{20}+3b+2)n$	$\frac{k^3}{15}$

triangular system solves (one forward and one backward). Usually the LU decomposition needs to be computed only whenever the time step h_n changes. Throughout this section we shall assume that on average, the LU decomposition needs to be computed only once every 10 time steps. Thus we obtain a complexity of $\frac{n^3}{15}+2n^2+\frac{n^2}{10}+\frac{n}{10} \sim \frac{n^3}{15}$ for a dense matrix A and $(\frac{b^2}{20}+3b+2)n$ for a banded matrix A (the quantity b is held constant). If a POD reduced order model of dimension $k(< n)$ is used, then we replace n by k in most of the expressions except for that for banded A (the reduced model matrix $\rho A \rho^T$ is not likely to be banded, and we shall assume it to be dense). It must be noted that in several examples when $n \to \infty$, the adequate size k of a reduced model remains constant after an initial growth. This is especially true in discretized PDE systems, since a finite number of empirical modes are adequate to capture any given percentage of the energy. As a result the asymptotic formulae for $k \to \infty$ are often not applicable.

For nonlinear systems $\dot{x} = f(x,t)$, the explicit method (forward Euler) involves evaluating $x_n = x_{n-1} + h_n f(x_{n-1}, t_{n-1})$; hence the complexity is $f(n) + 2n \sim f(n)$. If the reduced model is used, then one needs to evaluate $\rho f(\rho^T z_{n-1} + \bar{x}, t_{n-1})$, and the corresponding complexity is $\hat{f}(k,n) + 2k \sim \hat{f}(k,n)$.

For the implicit case one needs to solve the nonlinear system of equations

$$F(x_n) \triangleq x_n - x_{n-1} - h_n f(x_n, t_n) = 0$$

for x_n. This is done typically by Newton iteration

$$\left[I - h_n \frac{\partial f}{\partial x}\right] \delta_n^{(m)} = x_{n-1} - x_n^{(m-1)} + h_n f(x_n^{(m-1)}, t_n), \tag{6.1}$$

where $\delta_n^{(m)} = x_n^{(m)} - x_n^{(m-1)}$ is the correction. Ideally

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial x}(x_n^{(m-1)}, t_n);$$

TABLE 6.3
Asymptotic complexity for nonlinear systems.

	Full model explicit	Reduced model explicit	Full model implicit	Reduced model implicit
Dense	$f(n)$	$\hat{f}(k,n)$	$\frac{nf(n)}{5} + \frac{n^3}{15}$	$\frac{k\hat{f}(k,n)}{5} + \frac{k^3}{15}$
Banded	$f(n)$	$\hat{f}(k,n)$	$\frac{bf(n)}{5} + \frac{b^2n}{20}$	$\frac{k\hat{f}(k,n)}{5} + \frac{k^3}{15}$

hence the Jacobian should be evaluated at every Newton iteration inside a given time step. In practice it has been observed that one could get away with keeping the Jacobian unchanged not only during the Newton iteration but also for a few time steps, without severely compromising the accuracy. We assumed that the Jacobian update and the LU decomposition typically need to be done only once every 10 time steps. Note that the right-hand side of (6.1), however, needs to be computed at every Newton iteration. The evaluation of the right-hand side alone requires an asymptotic complexity of $f(n)$ at every Newton iteration for the full model. For the model reduced system one needs to evaluate $\rho f(\rho^T z_n + \bar{x}, t_n)$, which involves an asymptotic complexity of $\hat{f}(k,n)$. It is reasonable to assume that the number of Newton iterations per time step is on average a number independent of system size n (or k for the reduced model). Often in practice this could be about 2. Thus asymptotically the complexity of the Jacobian evaluations dominates over the complexity of evaluating the right-hand side of (6.1). Under these assumptions we get the asymptotic complexities shown in Table 6.3.

For nonlinear systems with explicit solver the asymptotic savings depend solely on the complexity of the nonlinear function evaluations $f(n)$ and $\hat{f}(k,n)$. Hence savings can be expected only if $\rho f(\rho^T z + \bar{x}, t)$ can be analytically simplified.

For nonlinear systems with implicit solver the complexity has two components: one from the nonlinear function evaluations and the other from the linear algebra. Depending on the complexity of f (or $\hat{f}$ for reduced models), one of these terms may be dominant. Asymptotic savings achieved depend on several factors including complexity of $f(n)$ and $\hat{f}(k,n)$ as well as the assumptions on asymptotic behavior of k as $n \to \infty$.

In our complexity analysis we have made several assumptions which are not always valid in practice. Even though most of these assumptions are reasonable, the combined error in our estimate of computational savings can sometimes be wrong by more than a factor of 10. For instance, we looked at complexity per time step. This is useful only if both the full model and the reduced model took more or less the same number of time steps. In the two examples here, the number of time steps did not vary by more than a factor of 2, except for the case of the explicit solver applied to the nonlinear PDE example with reduced model dimension $k = 6$. The asymptotic formulae may not be very applicable for the reduced models because of their smaller size. In addition we ignored computations associated with adaptive stepsize control, which seem to be a significant percentage, for the banded Jacobian case.

It must also be noted that we assumed that we care only about the solution value at the final time (or perhaps only at a few different points in the time interval), and this allowed us to ignore the cost of computing $\hat{x} = \rho^T z + \bar{x}$. The next two examples illustrate how various complex factors can affect the computational savings achieved by the use of POD reduced order models.

Example 4 (RC circuit—dense Jacobian). We considered the example of an

electric circuit with resistors and capacitors. By connecting each node to every other node, we obtain a dense Jacobian. One of the nodes is considered the ground (has zero voltage). Such a circuit with n nodes other than ground is described by a first order system of n linear ODEs. The current i_{jk} from the jth node to the k node is given by

$$i_{jk} = g_{jk}(v_j - v_k) + c_{jk}(\dot{v}_j - \dot{v}_k),$$

where g_{jk} and c_{jk} are the appropriate conductances and the capacitances, and $v \in \mathbb{R}^n$ is the vector of node voltages. We may add a nonlinearity to the resistors to obtain

$$i_{jk} = g_{jk}(v_j - v_k) + h_{jk}(v_j - v_k)^3 + c_{jk}(\dot{v}_j - \dot{v}_k).$$

These equations when combined with Kirchoff's current law give rise to an equation $\dot{v} = f(v)$.

We chose the parameter values somewhat ad hoc so that the resulting linearized system had a reasonable range of eigenvalues. We chose a system with $n = 500$. A random initial condition was chosen, and the time interval was chosen to be $[0, 2]$. Both linear and nonlinear versions were simulated, with both explicit and implicit solvers `ode45` and `ode15s` from MATLAB. Reduced order models were computed from the resulting trajectories and applied to both the linear and nonlinear systems. A reduced order model of size $k = 50$ was used, even though a size of $k = 2$ would have preserved more than 99% of the energy. The reason for using $k = 50$ is that $k = 2$ is too small for the asymptotic formulae to be valid. The number of floating point operations as counted in MATLAB are shown in Table 6.4.

The asymptotic complexity of function evaluations for this example are given by

$$f(n) = Cn^2,$$

where $C = 13$ for the nonlinear (cubic) case and $C = 2$ for the linear case. The complexity for the nonlinear reduced model is

$$\hat{f}(k, n) = Cn^2 + 4nk,$$

when $\rho f(\rho^T z + \bar{x})$ is not analytically simplified. Since $f(x)$ is cubic for the nonlinear case, it is possible to analytically simplify $\rho f(\rho^T z + \bar{x})$. This will improve the savings achieved by the reduced order model. In fact assuming that we get a cubic polynomial (in z) with all the possible monomials (dense cubic), we can estimate the complexity of $\hat{f}$. Table 6.5 compares the complexities of $\hat{f}$ after analytical simplification (assuming a dense cubic) for different values of k with that of $f(n)$. It is clear that for values of $k = 15$ or $k = 8$ we can expect significant savings.

We can see from Table 6.6 that the savings predicted by our theory is within an order of magnitude of the actual savings. It must be noted that our theory is only valid asymptotically as n and k get arbitrarily large and that our theory was based on forward and backward Euler methods, while the example used a Runge–Kutta method for the explicit case and a numerical differentiation formula (NDF) for the implicit case [24]. The discrepancies are due to several other factors as well. One is that the reduced model size $k = 50$ is not large enough to use the asymptotic formula. This is an important factor in the linear implicit case but not in the explicit case. Another reason is that the number of steps taken were different for the full model and the reduced model. Furthermore, there are computations associated with

TABLE 6.4
Computational cost in 10^3 flops: RC circuit. Full model versus unsimplified reduced model with $k = 50$.

	Full model	Reduced model (unsimplified, $k = 50$)
Linear explicit	$37,632$	645
Nonlinear explicit	$238,382$	$305,041$
Linear implicit	$646,460$	$2,323$
Nonlinear implicit	$2,609,800$	$400,177$

TABLE 6.5
Cost (in flops) of f and simplified $\hat{f}$: RC circuit.

Full model $n = 500$	Reduced model $k = 50$	Reduced model $k = 15$	Reduced model $k = 8$
3250000	2342500	24450	2624

TABLE 6.6
Computational savings ratio: RC circuit, unsimplified reduced model with $k = 50$.

	Observed savings	Asymptotic theoretical savings
Linear explicit	58	100
Nonlinear explicit	0.78	1
Linear implicit	278	1000
Nonlinear implicit	6.5	10

adaptive stepsize control which were not accounted for by our theory. The latter was a significant factor in the explicit case. We found that the computations associated with adaptive stepsize control grew linearly with system dimension. Furthermore, we found that the Jacobian evaluations were done much less often than the LU decomposition, contrary to our initial assumption. In fact there was only one Jacobian evaluation for the whole simulation in all of the implicit solvers.

Example 5 (reaction-diffusion PDE in one dimension—banded Jacobian). We considered the one dimensional reaction diffusion equation

$$x_t = 0.1x_{ss} - cx^3$$

in the spatial interval $s \in [0, 6]$ with zero boundary conditions. We discretized the spatial dimension on a uniform grid of n interior points using centered differences for both first and second derivatives. This yields a system of ODEs: $\dot{x} = f(x)$ with $x \in \mathbb{R}^n$. We chose two values $c = 0$ and $c = 1$. The first gives rise to a linear system and the second to a nonlinear system, both being dissipative. In both cases the Jacobian is a tridiagonal matrix (hence is banded with $b = 2$). We chose a system of size $n = 499$ and used a reduced order model of size $k = 50$. The following smooth initial condition was chosen:

$$x(0, s) = \exp\left(-\frac{(s-3)^2}{9}\right)\sin^2(\frac{\pi s}{2}), \quad s \in [0, 6].$$

The time interval of simulation was $[0, 5]$. We simulated the systems with two different MATLAB solvers: `ode23` (explicit) and `ode15s` (implicit). The cost of the computation is shown in Table 6.7.

The asymptotic complexity of function evaluations for this example is given by

$$f(n) = Cn,$$

TABLE 6.7
Computational cost in 10^3 flops: One dimensional PDE example, full model and unsimplified reduced model with $k = 50$.

	Full model	Reduced model (unsimplified, $k = 50$)
Linear explicit	149,150	69,527
Nonlinear explicit	182,310	1,342,400
Linear implicit	1,732	3,508
Nonlinear implicit	2,089	30,458

TABLE 6.8
Computational savings factor: One dimensional PDE example, $k = 50$, unsimplified $\hat{f}$.

	Observed savings	Asymptotic theoretical savings
Linear explicit	2.1	0.6
Nonlinear explicit	0.14	0.048
Linear implicit	0.49	0.49
Nonlinear implicit	0.069	0.002

where the constant $C = 2(b + 1) = 6$ for the linear case and $C = 10$ for the nonlinear case. The complexity of the nonlinear reduced model function evaluations is

$$\hat{f}(k, n) = Cn + 4nk,$$

when $\rho f(\rho^T z + \bar{x})$ is not analytically simplified.

The computational savings achieved and the theoretical asymptotic values are compared in Table 6.8. With the exception of the nonlinear implicit case, the asymptotic theory is within an order of magnitude of the observed values. For the same reasons as in the dense Jacobian case, one cannot expect the theory to be very accurate. In addition, in the banded Jacobian case there are other factors. Since the costs associated with overhead such as stepsize control as well as the rest of the computational costs grow linearly with system size for the unreduced systems, it is no longer valid to neglect the cost of such overhead even asymptotically. Hence the theory underestimates the cost for the unreduced case. This basically explains why the savings were better than predicted by theory. Another reason for the observed discrepancies is that MATLAB (version 5) does not take advantage of the banded structure of the Jacobian. It uses only the sparsity pattern. As a result, the cost of LU decomposition and triangular system solves is greater than that predicted by the theory.

It must be noted that one reason why there are no computational savings is due to the fact that the cost of function evaluation is significantly worse for the reduced model in the nonlinear case: $\hat{f} \approx 10n + 4nk = 210n \gg 10n \approx f(n)$. This was with no expression simplification applied to $\rho f(\rho^T z + \bar{x})$. If we simplify the expression and treat it as a dense cubic, the cost of evaluation (for the $k = 50$ case) is $\hat{f} = 2342500 \gg 104790 = 217n$, which is even worse. This is because the cubic nonlinearity in f is diagonal in the original x coordinates, while the simplified expression for $\hat{f}(z)$ is (typically) a dense cubic. The cost of evaluating a dense cubic $\mathbb{R}^k \to \mathbb{R}^k$ is of order $\frac{k^4}{3}$ and can be prohibitively large if k is not sufficiently small. However, if we use a smaller reduced order model of size $k = 6$, which in this example preserves all of the energy of the solution trajectory up to eight digit accuracy, we indeed get significant computational savings. For $k = 6$, the cost of function evaluation (assuming a dense cubic) is only ≈ 996 flops. Table 6.9 shows the savings factor for $k = 6$ when the simplified expression for nonlinear $\hat{f}$ is used and compares this with the theoretical asymptotic values. The theoretical values are within an order of magnitude

TABLE 6.9
Computational savings factor: One dimensional PDE example, $k = 6$, simplified nonlinear $\hat{f}$.

	Observed savings	Asymptotic theoretical savings
Nonlinear explicit	860	105
Nonlinear implicit	10.9	1.75

of the observed values. The discrepancies are again due to various factors.

This example illustrates how a multitude of factors affects the computational costs of using a POD reduced model in place of the full original model.

7. Conclusions. We investigated some basic properties of the POD method in finite dimensions. We provided an analysis of the errors involved in computing the solution of a nonlinear ODE initial value problem using a POD reduced order model. In addition to providing quantitatively reasonable error estimates, the analysis also explains the various factors that affect the error.

We also provided a sensitivity analysis of the POD method. We introduced the POD sensitivity factor which was a nondimensional measure of the sensitivity of the resulting projection with respect to perturbations in the data. We studied the effect of data perturbation on the projected data as well as the reduced model solution of the POD method. The POD sensitivity factor is relevant in some applications of POD while it is not in some other applications. We provided a discussion of this issue.

In addition to the error and sensitivity analysis, we also provided an analysis of computational complexity in using the POD reduced model in computing the solution to an ODE initial value problem. Our analysis showed that the computational savings achieved by POD depend on several factors and that the complexity of the nonlinear function evaluations can significantly affect the savings that might be gained by the use of a POD reduced model. Our examples suggest that combining expression simplification with reduced order models (for the class of polynomial vector-fields) may achieve significant savings if the reduced models are small enough.

Acknowledgment. We would like to thank the reviewers for their constructive comments.

REFERENCES

[1] R. ABRAHAM, J. E. MARSDEN, AND T. RATIU, *Manifolds, Tensor Analysis and Applications*, 2nd ed., Springer-Verlag, New York, 1988.
[2] V. ALGAZI AND D. SAKRISON, *On the optimality of Karhunen-Loève expansion*, IEEE Trans. Inform. Theory, 15 (1969), pp. 319–321.
[3] N. AUBRY, W. LIAN, AND E. TITI, *Preserving symmetries in the proper orthogonal decomposition*, SIAM J. Sci. Comput., 14 (1993), pp. 483–505.
[4] G. BERKOOZ AND E. TITI, *Galerkin projections and the proper orthogonal decomposition for equivariant equations*, Phys. Lett. A, 174 (1993), pp. 94–102.
[5] C. BISCHOF, A. CARLE, P. HOVLAND, P. KHADEMI, AND A. MAUER, *Adifor* 2.0 *Users' Guide (Revision D)*, Technical report CRPC-95516-S, Center for Research on Parallel Computation, Houston, TX, 1995.
[6] W. BOOTHBY, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed., Academic Press, New York, 1994.
[7] Y. CAO AND L. R. PETZOLD, *A Note on Model Reduction for Analysis of Cascading Failures in Power Systems*, manuscript.
[8] E. A. CHRISTENSEN, M. BRØNS, AND J. N. SØRENSEN, *Evaluation of proper orthogonal decomposition–based decomposition techniques applied to parameter-dependent nonturbulent flows*, SIAM J. Sci. Comput., 21 (2000), pp. 1419–1434.

[9] A. Curtis, M. Powell, and J. Reid, *On the estimation of sparse Jacobian matrices*, J. Inst. Math. Appl., 13 (1974), pp. 117–120.
[10] S. Glavaski, J. E. Marsden, and R. Murray, *Model reduction, centering, and the Karhunen-Loève expansion*, in Proceedings of the IEEE Conference on Decision and Control, Tampa, FL, 1998, pp. 2071–2076.
[11] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1996.
[12] M. Graham and I. Kevrekidis, *Alternative approaches to the Karhunen-Loève decomposition for model reduction and data analysis*, Computers and Chemical Engineering, 20 (1996), pp. 495–506.
[13] E. Hairer, S. Norsett, and G. Wanner, *Solving Ordinary Differential Equations* I: *Nonstiff Problems*, Springer-Verlag, New York, 1980.
[14] P. Holmes, J. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, UK, 1996.
[15] S. Lall, P. Krysl, and J. E. Marsden, *Structure-preserving model reduction for mechanical systems*, Phys. D, 184 (2003), pp. 304–318.
[16] S. Lall, J. E. Marsden, and S. Glavaski, *Empirical model reduction of controlled nonlinear systems*, in Proceedings of the IFAC World Congress, Beijing, 1999, pp. 473–478.
[17] B. Moore, *Principal component analysis in linear systems: Controllability, observability, and model reduction*, IEEE Trans. Automat. Control, 26 (1981), pp. 17–31.
[18] P. Parrilo, F. Paganini, G. Verghese, B. Lesieutre, and J. E. Marsden, *Model reduction for analysis of cascading failures in power systems*, in Proceedings of the American Control Conference, San Diego, CA, 1999, pp. 4028–4212.
[19] K. Pearson, *On lines and planes of closest fit to a system of points in space*, Philosophical Magazine, 2 (1833), pp. 609–629.
[20] M. Rathinam and L. R. Petzold, *Dynamic iteration using reduced order models: A method for simulation of large scale modular systems*, SIAM J. Numer. Anal., 40 (2002), pp. 1446–1474.
[21] M. Rathinam and L. Petzold, *An iterative method for simulation of large scale modular systems using reduced order models*, in Proceedings of the IEEE Conference on Decision and Control, Sydney, Australia, 2000, pp. 4630–4635.
[22] A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.
[23] C. Rowley and J. E. Marsden, *Reconstruction equations and the Karhunen-Loève expansion for systems with symmetry*, Phys. D, 142 (2000), pp. 1–19.
[24] L. F. Shampine and M. W. Reichelt, *The MATLAB ODE suite*, SIAM J. Sci. Comput., 18 (1997), pp. 1–22.
[25] S. Shvartsman and I. Kevrekidis, *Low-dimensional approximation and control of periodic solutions in spatially extended systems*, Phys. Rev. E(3), 58 (1998), pp. 361–368.
[26] S. Shvartsman, C. Theodoropoulos, R. Rico-Martinez, I. Kevrekidis, E. Titi, and T. Mountziaris, *Order reduction for nonlinear dynamic models of distributed reacting systems*, J. Process Control, 10 (2000), pp. 177–184.
[27] L. Sirovich, *Turbulence and the dynamics of coherent structures*, I, Quart. Appl. Math., 45 (1987), pp. 561–571.
[28] L. Sirovich, *Turbulence and the dynamics of coherent structures*, II, Quart. Appl. Math., 45 (1987), pp. 573–582.
[29] L. Sirovich, *Turbulence and the dynamics of coherent structures*, III, Quart. Appl. Math., 45 (1987), pp. 583–590.