International Journal of Bifurcation and Chaos, Vol. 17, No. 4 (2007) 1199–1219

ATTRACTOR MODELING AND EMPIRICAL NONLINEAR MODEL REDUCTION OF DISSIPATIVE DYNAMICAL SYSTEMS

ERIK BOLLT
Department of Mathematics & Computer Science,
Department of Physics, Clarkson University,
Potsdam, NY 13699-5815, USA

Received November 8, 2005; Revised May 26, 2006

In a broad sense, model reduction means producing a low-dimensional dynamical system that replicates either approximately, or more strictly, exactly and topologically, the output of a dynamical system. Model reduction has an important role in the study of dynamical systems and also with engineering problems. In many cases, there exists a good low-dimensional model for even very high-dimensional systems, even infinite dimensional systems in the case of a PDE with a low-dimensional attractor. The theory of global attractors approaches these issues analytically, and focuses on finding (depending on the question at hand), a slow-manifold, inertial manifold, or center manifold, on which a restricted dynamical system represents the interesting behavior of the dynamical system; the main issue depends on defining a stable invariant manifold in which the dynamical system is invariant. These approaches are analytical in nature, however, and are therefore not always appropriate for dynamical systems known only empirically through a dataset. Empirically, the collection of tools available are much more restricted, and are essentially linear in nature. Usually variants of Galerkin's method, project the dynamical system onto a function linear subspace spanned by modes of some chosen spanning set. Even the popular Karhunen–Loeve decomposition, or POD, method is exactly such a method. As such, it is forced to either make severe errors in the case that the invariant space is intrinsically a highly nonlinear manifold, or bypass low-dimensionality by retaining many modes in order to capture the manifold. In this work, we present a method of modeling a low-dimensional nonlinear manifold known only through the dataset. The manifold is modeled as a discrete graph structure. Intrinsic manifold coordinates will be found specifically through the ISOMAP algorithm recently developed in the Machine Learning community originally for purposes of image recognition.

Keywords: Attractor; model reduction; POD; KL analysis; ISOMAP; singular perturbation.

1. Introduction

A fundamental problem in dynamical systems is to reduce a high dimensional problem to a simpler low dimensional problem, when such a reduction exists. This is what is meant when using the phrase looking for "hidden order within chaos," [Prigogine, 1984; Bollt, 2005] , which implies that there is some form of dissipation in the system. This is the cornerstone of several major techniques in the field for dissipative systems with global attractors, including singular-perturbation theory [Carr, 1981], the method of multiple-scales [Kevorkian & Cole, 1996], and the inertial manifold theory [Teman, 1997; Robinson, 2001]. Each of these techniques strive to find a lower-dimensional equation, restricted to some stable invariant manifold, and whose dynamics is the same (conjugate) as the long term behavior of the original system. The idea of an attractor

in chaos theory, and especially a chaotic attractor, focuses on the concept that a low dimensional process exists in a seemingly high dimensional and complex process. The methods usually require a closed form of the model to analytically reduce the system dynamics onto a stable invariant manifold.

Data-driven and empirical techniques are also important for the obvious reason that they are designed to deal with real world problems where only measurement from a laboratory realization of a dynamical system is available. Galerkin's method and finite element methods are very well regarded and well developed [Hughes, 2000], but these are linear methods of analysis of nonlinear evolution equations, and they require in advance an analytic form of the model. The method of approximate inertial manifolds [Jolly *et al.*, 2001], also does an excellent job of finding a lower dimensional manifold for restriction of the dynamics, but this too requires an analytic form of the model.

Most notably, based on the Taken's embedding method [Takens, 1980] there are techniques of the time-series embedding literature [Abarbanel, 1996; Kantz, 1997], which are concerned first with finding a good embedding, meaning the right embedding dimension such that data on the embedding manifold is properly unfolded (such as false-nearest neighbors [Kennel, 1992]), and then a good delay, such that the data is well distinguished from each other, in a mutual information theoretic sense [Fraser & Swinney, 1986]. The techniques in this area have been successful for prediction [Farmer & Sirowich, 1987; Weigenbend & Gershenfeld 1993; Abarbanel, 1996; Kantz, 1997], control [Ott *et al.*, 1994] as well as characterization [Kantz, 1997; Eckmann & Ruelle, 1985] of datasets from real measurements; they allow for local modeling of the dynamical system on whatever might be the attractor manifold containing the attractor set. However, such methods do not model an invariant manifold directly. Rather, a different local model, a coordinate chart [Conlon, 2001], is made for each neighborhood, but no effort is generally made to connect the local models into a global framework, or atlas.

The main concern of this paper will be to introduce a method to construct an empirical model of the global invariant manifold. When a dynamical system has a stable invariant manifold, onto which empirical data is attracted, that manifold can be detected and modeled as a discrete graph structure. In this sense, our goal could be described as dimension reduction.

2. Linear Versus Nonlinear Model Reduction

2.1. *Linear model reduction*

Consider a spatiotemporal pattern, such as the solution of an evolution equation

$$u_t(x,t) = Au(x,t) + f(u), \tag{1}$$

sampled on a grid in x, and in,

$$t : \{u^n(x)\} = \{\mathbf{u}(x,t_n)\}_{n=1,M}. \tag{2}$$

The form Eq. (1) also represents a large spatially extended ODE, taking u to be a time varying vector valued function, and x to now be a discrete value identifying a lattice index position. A favorite method of approach is formally called Galerkin's method [Hughes, 2000], and it relies on formal substitution into a finite expansion of a finite basis set of functions, $\Phi_n \in L^2$,

$$u(x,t) = \sum_n a_n(t)\psi_n(x), \tag{3}$$

into the PDE, which when the basis set is orthonormal, exercising the inner product condition,

$$(u,v) = \int u(x)v(x)d\mu(x), \tag{4}$$

and an assumed orthonormality condition of the basis functions,

$$(\Phi_i, \Phi_j) = \delta_{i,j}, \tag{5}$$

results in a coupled set of time varying Fourier coefficients. The evolution equations of the $a_n(t)$ are then a coupled set of ODEs. The question becomes then how to represent the full dynamics in the Banach space in terms of projection onto a finite basis set. This is the problem of inertial manifold theory [Teman, 1997; Robinson, 2001]. In general, it is expected that a small number of basis functions do not result in a sufficient representation of the dynamics on the true nonlinear invariant manifold.

A popular method of model reduction of modeling a given empirical dataset is a special case of a Galerkin's method, in terms of an "optimal basis" called KL analysis [Karhunen, 1946; Loeve, 1955; Lumley, 1970; Holmes *et al.*, 1996; Sirovich, 1989]. The KL (Karhun–Loeve) modes (a form of POD analysis — principal orthogonal decomposition, or PCA — Principal Component Analysis), is

fundamentally a linear analysis using eigenfunctions $\Psi_n(x)$ of the time-averaged covariance matrix,

$$K(x,x') = \langle \mathbf{u}(x,t_n)\mathbf{u}(x',t_n)\rangle, \tag{6}$$

which may be arrived at by a singular value decomposition [Golub & Van Loan, 1996]. Then u may be expanded in the resulting orthoganol basis,

$$u(x,t) = \sum_n a_n(t)\psi_n(x), \tag{7}$$

and this is the optimal basis in the sense of projection:

$$\max_{\psi\in L^2(D)} \frac{\langle |(u,\psi)|\rangle}{\|\psi\|}, \tag{8}$$

[Holmes *et al.*, 1996]. These functions are orthogonal in time, meaning in terms of time-averaging,

$$\langle a_n(t)a_m(t)\rangle = \lambda_n\delta_{nm}, \tag{9}$$

in terms of eigenvalues of,

$$K : \lambda_n = \frac{(\psi_n, K\psi_n)}{\|\psi_n\|}. \tag{10}$$

Thus, the time-varying Fourier coefficients $a_n(t)$ are decorrelated *in time average*. A computationally important approach [Sirovich, 1989], to solve this eigenvalue problem involves successive computation to maximize mean square energy. Formal substitution of a finite expansion of empirical modes $u(x,t) = \sum_n a_n(t)\psi_n(x)$ into the PDE, and then projection onto each basis element $\psi_m(x)$ produces an ODE which is expected to be a maximal energy model of the PDE. POD does a good empirical job of capturing a high energy model of the true dynamics. Improvements, such as Balanced POD [Rowley, 2005] based on balanced truncation further improves upon this picture.

A fundamental topological problem with the use of KL modes as a dimension reduction technique for nonlinear dynamical systems is that KL analysis is fundamentally a linear analysis. Given a dataset of high dimensional and randomly distributed data points, principle component analysis gives the principle axis of the time-averaged covariance matrix. That is, it treats that data as an ellipsoidal cloud, and it yields the major and minor axes. It does not have the ability to cope with a truly nonlinearly curved invariant manifold in a properly nonlinear way.

2.2. *Nonlinear reduced models from dynamical systems*

In general, the classic issue of concern here is that of the presence of a significant spectral gap, or said differently, a system with two or more significantly different time scales. A system with a stable invariant manifold is well charicatured in the context of a singularly perturbed system [Carr, 1981; Fenichel, 1979],

$$\begin{aligned} \dot{x} &= F(x,y), \\ \epsilon\dot{y} &= G(x,y). \end{aligned} \tag{11}$$

where,

$$\begin{gathered} x\in\Re^m, \quad y\in\Re^n, \quad F:\Re^m\times\Re^n\to\Re^m, \\ \text{and} \quad G:\Re^m\times\Re^n\to\Re^n. \end{gathered} \tag{12}$$

It is easy to see that for $0<\epsilon\ll 1$, that the $y(t)$-equation runs fast, relative to the slow dynamics of the first equation for evolution of $x(t)$. Such systems are called singularly perturbed, since if $\epsilon = 0$ we get a differential-algebraic equation

$$\begin{aligned} \dot{x} &= F(x,y), \\ G(x,y) &= 0. \end{aligned} \tag{13}$$

The second ODE becomes an algebraic constraint. Under sufficient smoothness assumptions on the functions F and G so that implicit function theorem [Tikhonov *et al.*, 1985; Carr, 1981; Fenichel, 1979] can be applied, there is a function, or $\epsilon = 0$ slow-manifold,

$$y = h_\epsilon(x)|_{\epsilon=0}, \tag{14}$$

such that,

$$G(x, h_\epsilon(x)|_{\epsilon=0}) = 0. \tag{15}$$

The singular perturbation theory concerns itself with continuation, and persistence of stability of this manifold $h_\epsilon(x)$ within $O(\epsilon)$ of $h_\epsilon(x)|_{\epsilon=0}$, for $0<\epsilon\ll 1$ and even for larger ϵ. For the rest of this paper, we will refer to a stable invariant manifold in an $\Re^{m+n}$ space, generally as the graph of an expression,

$$\begin{aligned} H:\ &\Re^m\to\Re^n \\ &x\mapsto y = H(x). \end{aligned} \tag{16}$$

if it exists. In such a case, the reduced model becomes,

$$\dot{x} = F(x, H(x)), \tag{17}$$

which is equivalent to Eq. (11), subject to substitution of Eq. (16), and yields the manifold equation,

$$\epsilon DH|_x \cdot F(x,H(x)) = G(x,H(x)). \tag{18}$$

For the purposes of this paper, we presume that we cannot approach the model reduction in an analytic form of the equations, Eq. (11), either because we only know the dynamical system through a dataset, or because the problem is otherwise difficult to put into such a form. However, the goal is to find an empirical representation of what is the manifold, $y = H(x)$ which leads to a manifold reduced equation, Eq. (17).

If the data is mainly confined to a low dimensional subspace, that is, $H(x)$ is approximately flat, then linear methods work pretty well to discover a flat, nearly invariant and stable subspace and estimate its dimensionality. More generally, however, if the data lies on (or near) a highly curved low dimensional submanifold, then linear methods overly simplify the topological picture.

3. Manifold Learning in the Machine Learning Community

Manifold learning can be described as an established area, and a quickly evolving area, in the machine learning community, for wide ranging practical problems of detecting low dimensional structures in very high dimensional datasets. It is also an idea for problems of handwritten character recognition [LeCun *et al.*, 1995], object recognition [Schlkopf & Smola, 2001], facial recognition [Schlkopf & Smola, 2001], and other classification and feature extraction problems.

Popular methods of reducing the dimensionality of a dataset include SOM-self-organizing maps [Kohonen, 1988], GTM-generative topographic mapping [Bishop *et al.*, 1998], and autoencoder neural networks, [DeMers & Cottrell, 1993], attempt to generalize PCA-principal component analysis to build a single global low-dimensional nonlinear model, as do similarly SVM-support vector machine and kernel methods [Schlkopf & Smola, 2001; Vapnik, 1998]. However, such methods can be difficult to apply to real datasets depending on dataset size, manifold complications and dimensionality. In brief, they rely on greedy optimization criteria which can lead to problems with unwanted local minima that result in unrevealing suboptimal results. On the other hand, there are local methods [Bregler & Omohundro, 1995; Hinton *et al.*, 1995] which build a set of local models that are usually linear, and therefore only valid in a limit range, and can have smoothness problems. Interestingly, the local methods in the machine learning literature parallel what can be found in the (now) traditional time-series embedding literature of the dynamical systems community [Abarbanel, 1996; Abarbanel *et al.*, 1993; Kantz, 1997; Farmer & Sirowich, 1987; Eckmann & Ruelle, 1985], although the two communities have developed their techniques as apparently parallel but independent efforts.

Recently, the Isometric mapping-ISOMAP method has been developed [Bernsetin *et al.*, 2000; Roweis & Saul, 2000; Tenenbaum *et al.*, 2000], that approximates the manifold by an undirected graph whose geodesics are meant to coincide with those of the true nonlinear manifold. Other recent methods approximate a manifold's global nonlinear structure by a discrete graph, notably the LLE-local linear embedding method [Saul & Roweis, 2000] which preserves linear structure on the manifold. We will focus here on the ISOMAP method.

For description of the ISOMAP method, assume a dataset consisting of N data points in q-dimensional Euclidean space $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \Re^q$. We wish to construct a corresponding dataset, $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N \subset \Re^p$ appropriately embedded within an invariant manifold, and hopefully $p \ll q$. In the following, we will describe $\mathbf{X}$ as represented in the variables of the ambient space $\Re^q$, and we will describe $\mathbf{Y}$ as represented in the intrinsic variables of the manifold (meaning that the manifold locally has neighborhoods like, homeomorphic to, $\Re^p$).

In review, to embed a point x from q-dimensional Euclidean space, into intrinsic variables y of a p-dimensional manifold, means that we need to represent the manifold in terms of a parameterization,

$$\Phi : \mathbf{Y} \mapsto X, \tag{19}$$

where,

$$\begin{aligned} x &= \Phi(y) \\ &= \langle \phi_1(y_1, y_2, \ldots, y_p), \phi_2(y_1, y_2, \ldots, y_p), \ldots, \\ &\quad \phi_q(y_1, y_2, \ldots, y_p) \rangle, \end{aligned} \tag{20}$$

For example,

- The familiar $p =$ one-dimensional circle is parameterized by two functions,

$$x = \langle \phi_1(y), \phi_2(y) \rangle = \langle \cos(y), \sin(y) \rangle, \quad y \in \Re, \tag{21}$$

to represent the $q = 2$ coordinates of the ambient space.

- Similarly,

$$x = \langle y_1, y_2, y_1^2 + y_2^2 \rangle, \tag{22}$$

and $y_1, y_2 \in \Re$, is a $p =$ two-dimensional parameterization of each point on the manifold in $q =$ three-dimensional ambient space.

These, y-variables are what we call the intrinsic variables, and can be described as directions to any point on the manifold relative to a base point on the manifold. Thus the job of approximating the manifold requires two related parts,

- **Marks on the manifold**. The dataset in intrinsic variables, $\mathbf{X}$, which is assumed to lie on the lower dimensional manifold $\mathcal{M}$ embedded in $\Re^q$, serves this purpose. For dynamical systems purposes, this manifold is presumed to embed an attractor set $\mathcal{A}$.
- **Relative positions on the manifold**. The parameterization describes vector positions in the manifold of each data relative to each other data point.

3.1. *Review of ISOMAP*

ISOMAP is a manifold "learning" algorithm that extends the classical MDS-multidimensional scaling method [Cox & Cox, 1994] by using approximations of geodesic distances with shortest paths of a discrete graph approximation of the manifold, instead of directly applying MDS to the ambient Euclidean space.

There are several main steps in developing the ISOMAP [Tenenbaum *et al.*, 2000] embedding, meaning to represent the parameters Y in Eq. (20).

(1) **Build a neighbors graph to approximate the embedding manifold**. A graph $G = (V, E)$ consists of the set of vertices $V = \{v_i\}$ which we assign to match the data points, $X = \{\mathbf{x}_i\}_{i=1}^N \to \{v_i\}_{i=1}^N$, and edges which are ordered pairs of edges present in the graph, $E = \{v_i, v_j\}$. One can choose either *epsilon*-neighborhoods, or neighborhoods of k-nearest neighbors. To build a k-near neighbors graph, construct the graph consisting of edges $\{v_i, v_j\}$ corresponding to the k-closest data points x_j to x_i, for each i, with respect to the Euclidean distance function of the ambient space $d_X(\cdot, \cdot)$. Let $\mathcal{N}_M$ be a matrix encoding the weighted graph of intrinsic manifold distances corresponding to the graph G; for each vertex present $\{v_i, v_j\}$ in G, we associate the *neighbors* distances $\mathcal{N}_M(i, j) \approx d_X(x_i, x_j)$, and for each edge not present, we associate the distance $\mathcal{N}_M(i, j) = \infty$, to forbid geodesics of the graph from jumping between branches of the underlying manifold.

(2) **Compute geodesics of the graph**, to approximate geodesics of the manifold. There are popular methods to compute shortest paths of the graph, including Floyd's algorithm for small to medium sized datasets or Dijtsra's algorithms for small to large datasets. Thus compute an approximate geodesic distance matrix $D_M(i, j)$ consisting of shortest weighted path lengths from $\mathcal{N}_M$ between each i to j, to approximate in manifold geodesic distances.

(3) **Approximate in Manifold Distance by k-Near Neighbor Distance.** The D_M distance matrix of graph geodesics from the previous step is taken to approximate the true geodesic distances of the manifold between x_i, and x_j, $d_M(x_i, x_j)$. This approximation improves as data density increases. If k is chosen too large, or data density is too low, then consequently, some neighbors could be on separate branches of the manifold, and the approximation is poor, resulting in illegal shortcuts and a poor representation of the manifold.

(4) **Perform an MDS on D_M**. MDS requires only the D_M matrix of in manifold distances as input, which was computed from X input data above, but X is not required further, to form projective variables Y of the intrinsic variables. Therefore the results in variables $\tilde{X}$ defined below, and Y, will also represent intrinsic manifold variables.

3.2. *Review of MDS*

We review the classical MDS algorithm as follows, [Cox & Cox, 1994a]. Given D_M, which approximates in manifold geodesic distances for our purposes, the goal is to form a matrix of projected d-dimensional data Y to optimize the residual error,

$$E = \|\tau(D_M) - \tau(D_Y)\|_{L^2}, \tag{23}$$

denoting the matrix norm, $\|A\| \equiv \sqrt{\sum_{i,j} A_{I,j}^2}$, (this is therefore not a so-called natural norm [Golub & Van Loan, 1996]), $\tau(\cdot)$ is a centered distance function, and the matrix $[D_Y]_{i,j}$ describes the geodesic distances between y_i and y_j in the projective space Y of intrinsic variables. To compute Y, we list the

following string of facts which are central to the theory of MDS [Cox & Cox, 1994]. First define.

$$\tau(D_M) = -\frac{1}{2} H D_M^2 H, \tag{24}$$

where H is a centering matrix,

$$H = \frac{1}{N} I - \mathbf{1} \cdot \mathbf{1}^t, \tag{25}$$

$\mathbf{1}$ is a $N \times 1$ matrix of ones, and I is the $N \times N$ identity. Likewise, define, but we do not need to compute,

$$\tau(D_Y) = -\frac{1}{2} H D_Y^2 H. \tag{26}$$

Then, again for exposition but not for computation, we note that,

$$\begin{aligned} \min \|\tau(D_M) - \tau(D_y)\| &= \min \left\| -\frac{1}{2} H (D_M^2 - D_Y^2) H \right\| \\ &= \min \|\tilde{X}^t \tilde{X} - Y^t Y\|^2 \\ &= \min \operatorname{trace} (\tilde{X}^t \tilde{X} - Y^t Y)^2 \end{aligned} \tag{27}$$

The last equalities follow from a theorem [Cox & Cox, 1994], highlighted by Eq. (32), that yields for any squared matrix D_M^2, there exists points $\tilde{x}_i$, $\tilde{x}_j$ such that,

$$d_{i,j}^2 = (\tilde{x}_i - \tilde{x}_j)^t \cdot (\tilde{x}_i - \tilde{x}_j), \tag{28}$$

from which,

$$\tau(D_M)^2 = \tilde{X}^t \tilde{X}, \tag{29}$$

and likewise for $\tau(D_Y)^2$. The coordinates $\tilde{x}$ are to be understood as centered in such a way that pairwise Euclidean distances are D_M. A key advantage of the MDS algorithm over the more common POD algorithm is that all matrix manipulations to compute an output Y require only the centered distance matrix $\tau(D_M)$, which represents geodesic distances on the manifold. Whereas by contrast POD works directly with the original X input data by attempting to linearize it by a singular value decomposition-SVD of the X. Therefore $\tilde{X}$ is allowed to be in the manifold appropriate to the geodesic distances D_M, and $\tilde{X}$ is thus distinguished from the original input data X. If needed, $\tilde{X}$ can be found by Eq. (32), since we never have any need to distinguish between Y or $\tilde{Y}$, and we will always write the former.

Since $\tau(D_M)$ is symmetric and positive semi-definite, the computation of MDS uses the spectral decomposition,

$$\tau(D_M) = V \Sigma V^t, \tag{30}$$

where $\Sigma = \operatorname{diag}(\lambda_i)$ is the matrix of eigenvalues, and V is the orthogonal matrix of eigenvectors,

$$V^2 = I, \text{ and } \tau(D_M) V = \Sigma V. \tag{31}$$

Comparing representations for $\tau(D_M)$, Eqs. (24) and (29), to the spectral decomposition Eq. (30) gives a representation of the variables $\tilde{x_i}$ mentioned by Eq. (28),

$$\tilde{X} = V \Sigma^{\frac{1}{2}}, \tag{32}$$

where the square matrix of non-negative eigenvalues has a simple square root $\Sigma = \operatorname{diag}(\sqrt{\lambda_i})$. The MDS solution is then,

$$Y \equiv Y_{\text{MDS}} = V_p \Sigma_p^{\frac{1}{2}}, \tag{33}$$

where $\Sigma_p^{\frac{1}{2}}$ and V_p use the top p (significant) eigenvalues and eigenvectors of $\tau(D_M)$.

Note that the MDS is essentially a linear analysis, and the main nonlinear step to restrict to the nonlinear manifold is the formation of the neighbors graph based on k-nearest neighbors, the idea being that with enough data density, geodesics of the graph will not take any shortcuts not allowed by paths in the manifold.

There is a simple relationship between a PCA projection, at the heart of POD used in KL analysis, and the MDS projection. Whereas POD forms a rank-p projection of maximal variance, MDS forms the rank-p projection that optimizes the dissimilarity, meaning intra-point distances. Specifically, the variables of the corresponding projections relate according to,

$$Y_{\text{PCA}} = \Sigma_{\text{PCA}}^{\frac{1}{2}} Y_{\text{MDS}}, \tag{34}$$

where $\Sigma_{\text{MDS}} = \Sigma_{\text{PCA}} = \Sigma_p$ as used in Eq. (33). Also, there is a relationship of the basis vectors,

$$V_{\text{PCA}} = \tilde{X} V_{\text{MDS}}, \tag{35}$$

where similarly $V_{\text{MDS}} = V_p$ from Eq. (33). The two algorithms yield essentially the same thing when the distance matrix is Euclidean distance, but since we take D_M to be discretely approximated in manifold distance in ISOMAP, the results are different, as are the steps of computation. The most important difference in the algorithmic steps between MDS and PCA is that MDS does not explicitly use X in its computations, and therefore since variables to be found are in some unknown nonlinear manifold, this is a good dependency to avoid.

Most interesting to us for future exploration is how to couple recent results of convergence of the graph approximation of the manifold, [Bernsetin *et al.*, 2000], to our particular setting. The theory in [Bernsetin *et al.*, 2000] concerns considerations of sampling a random variable on the submanifold, which can be considered in the setting of a dynamical system on a manifold through understanding the invariant measure of the dynamical system, and convergence rate of the initial distribution function.

4. Examples

4.1. *A Duffing on paraboloid singularly perturbed system*

First, we consider a benchmark problem, consisting of a singularly perturbed problem, in the form of Eq. (11), of relaxation of a Duffing oscillator onto a stable nonlinear manifold consisting of a paraboloid.

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= \sin(x_3) - ax_2 - x_1^3 + x_1, \\ \dot{x}_3 &= 1, \\ \epsilon\dot{y} &= y - \alpha(x_1^2 + x_2^2). \end{aligned} \tag{36}$$

If we choose, $a = 0.02, b = 3, \alpha = 1$, and $\epsilon = 0.001$, we get the chaotic dataset shown wrapped onto a paraboloid in Fig. 1.

Clearly, any dataset from a sampling of the flow,

$$\{z_i\}_i \equiv \{(x_1(t_i), x_2(t_i), y(t_i)\}_i, \tag{37}$$

is poorly modeled as lying in any planar or linear subspace, but nonetheless, it lies on a two-dimensional nonlinear submanifold. It is no surprise that applying the KL-method to this dataset cannot properly reveal the true two-dimensionality of this process. Furthermore, this example could have been augmented by creating many fast variables which decay to a two-dimensional paraboloid in as high-dimensional an ambient space as we like; we chose only one fast variable for sake of artistic simplicity of displaying the attractor as a three-dimensional rendering. We know of no previous method to properly model the dynamics on the parabolic invariant manifold.

There are approximately 65,000 data points comprising of the sampling of the flow shown in Fig. 1(a). In principle, one could perform ISOMAP directly on this dataset, $\{z_i\}$, but it is computationally too expensive on this large dataset, and empirically redundant. In practice, we find a subsampling of the dataset $\{z_{i_k}\}$ of 1000 points, shown in Fig. 1(b), to be quite sufficient. A subsampling of a high-dimensional dataset for better computational efficiency has been called "landmark ISOMAP." The critical issue is that the subsampling must have similar statistics (the same distribution) as the larger dataset, and the subsampling should be sufficiently dense so that the approximations of the manifold by the discrete graph structure is good. We see in Fig. 2(a) a clear indication to justify that the expected benchmark result that the embedding manifold should be two-dimensional.

The intrinsic manifold coordinates $y = (y_1, y_2)$ shown in Fig. 2(b) serve sufficiently for revealing the underlying Duffing flow, which is the empirical version of Eq. (17), which we rewrite here calling the intrinsic variables-y,

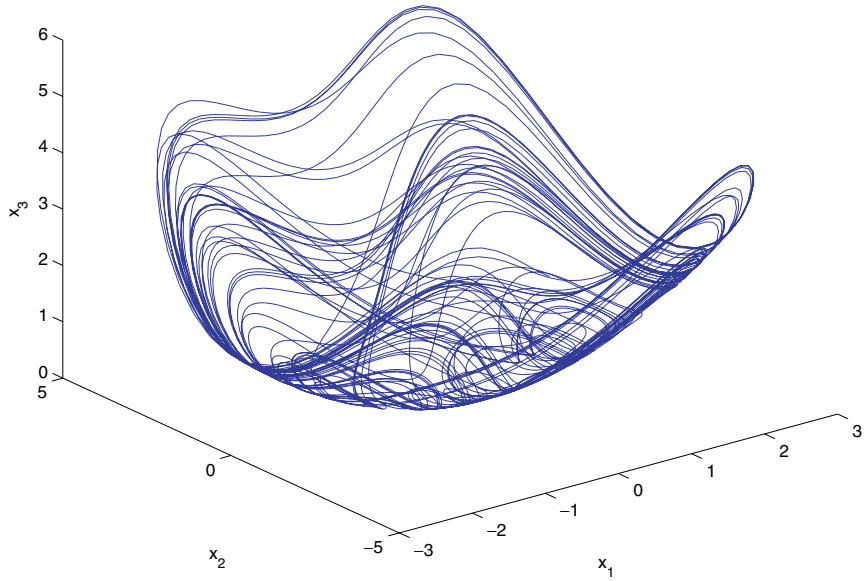
$$\dot{y} = f(y) = F(y, H(y)). \tag{38}$$

The red curve shown in Fig. 3(b) shows just such an empirical curve of the approximated flow of $\dot{y} = f(y)$. This curve can now be used for any number of other purposes, such as through a nonlinear parameter estimation by synchronization, or by the least squares Kalman-type methods, or alternatively, the red curve could be used for predication and/or control. The point is, that each of these activities is now performed in a reduced dimensional space, which could in principle be much lower dimensionally.

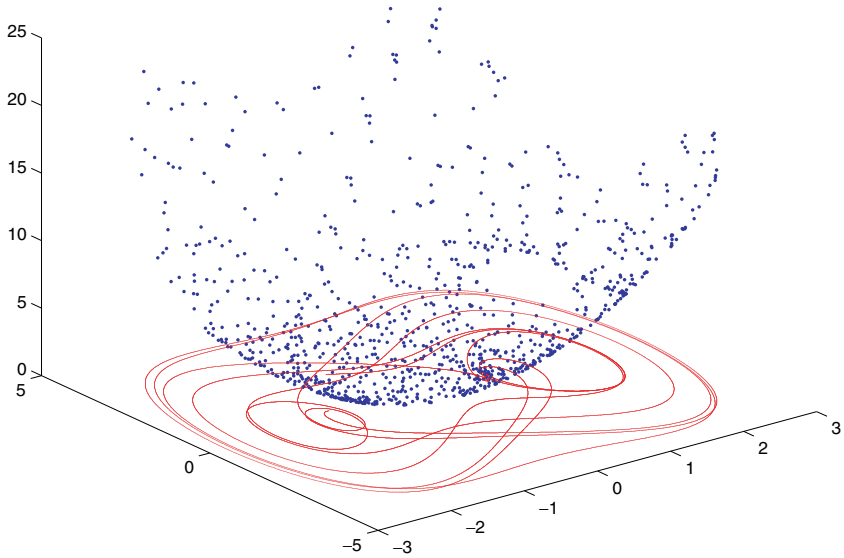
In this particular case, where the embedding manifold is particularly smooth and low dimensional, we demonstrate a further operation to reduce the size of the dataset being sent to ISOMAP, and to nonetheless ensure that there is sufficient data density to justify the discrete graph to manifold approximation. The problem is one of regridding, to reduce the number of data points where they are particularly dense, and to use the underlying manifold smoothness to increase data density where it is sparse. A traditional method of regridding uses multivariate splines. For example, a bivariate B-spline which could be used in this example would have the form [Eubank, 1999; Messer, 1991; Nyschka, 1995],

$$f(x_1, x_2) = \sum_i \sum_j B_{i,k}(x_1) B_{j,l}(x_2) a_{i,j}, \tag{39}$$

which can be quite successful for a smooth slow manifold with a fast normal contraction, so the dataset is essentially on the manifold. If the data is not expected to be as close to the manifold, then



(a)



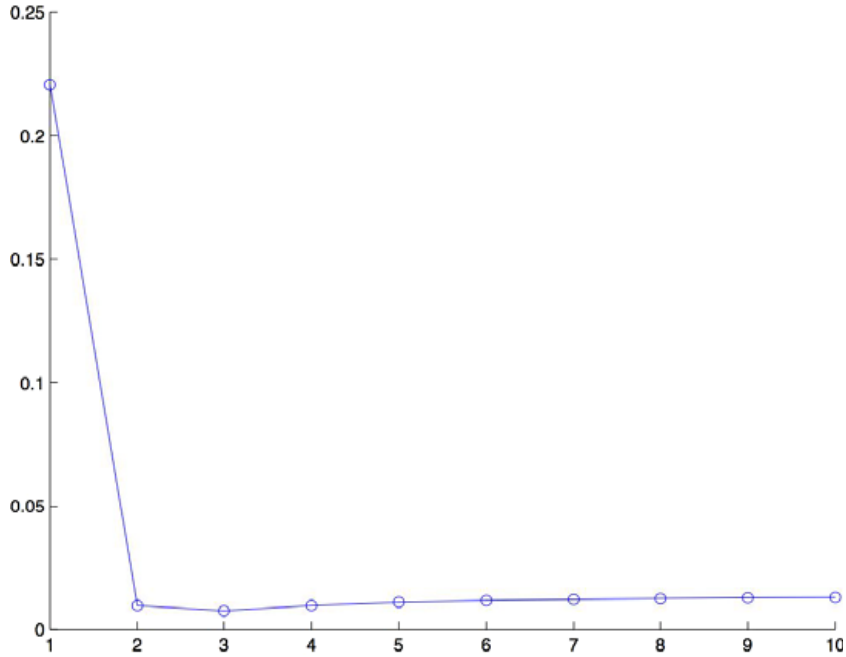
(b)

Fig. 1. (a) Data of a singularly perturbed relaxation of a Duffing oscillator onto a stable nonlinear manifold, by Eqs. (36). (b) A subsampling of the data, the blue dots, is a sufficient landmark set for more efficient processing by ISOMAP.

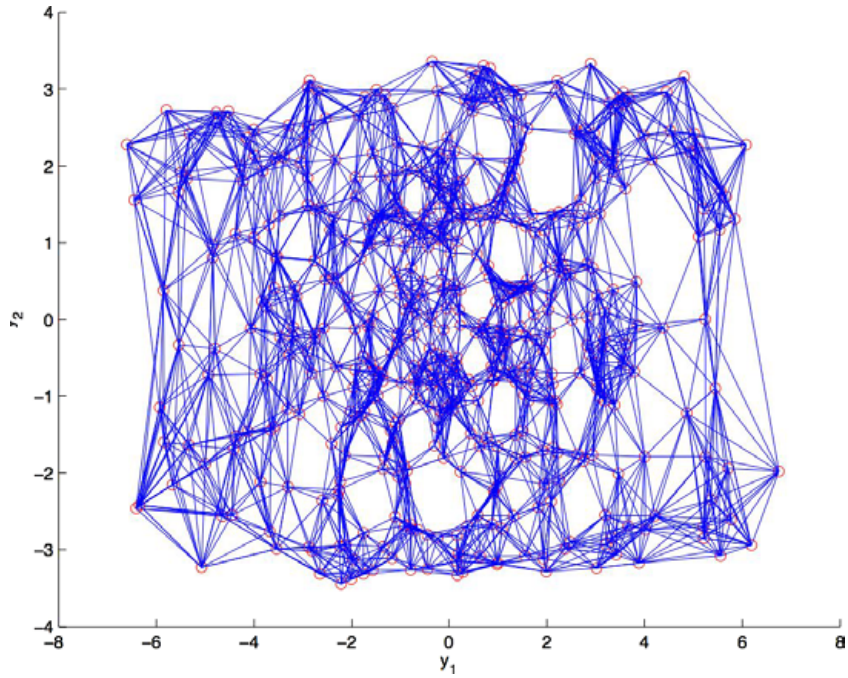
a degree of smoothing can be inferred by a multivariate version of a smoothing spline, which in the bivariate case can be written as the minimizer of the functional,

$$p\sum_i |x_{3,i} - fc_i|^2 + (1-p)\int (|D_{1,1}f|^2 + 2|D_{1,2}f|^2 + |D_{2,2}f|^2), \tag{40}$$

which gives rise to thin-plate splines. Notice that this functional is a balance between least squares smoothing in the case $p = 0$ and an exact fitting spline when $p = 1$, with a balance between data fidelity in the first term, and curvature relationships in the second. We have used the Matlab Spline Toolbox [Matlab, 2005] of both the multivariate spline, and the plate spline, and in this case the results are essentially the same, due to the strong normal hyperbolicity of the slow manifold. In Fig. 3(a) we see the singularly perturbed Duffing oscillator Eq. (36) and the results of a bivariate spline applied to the Duffing data. The spline has been evaluated

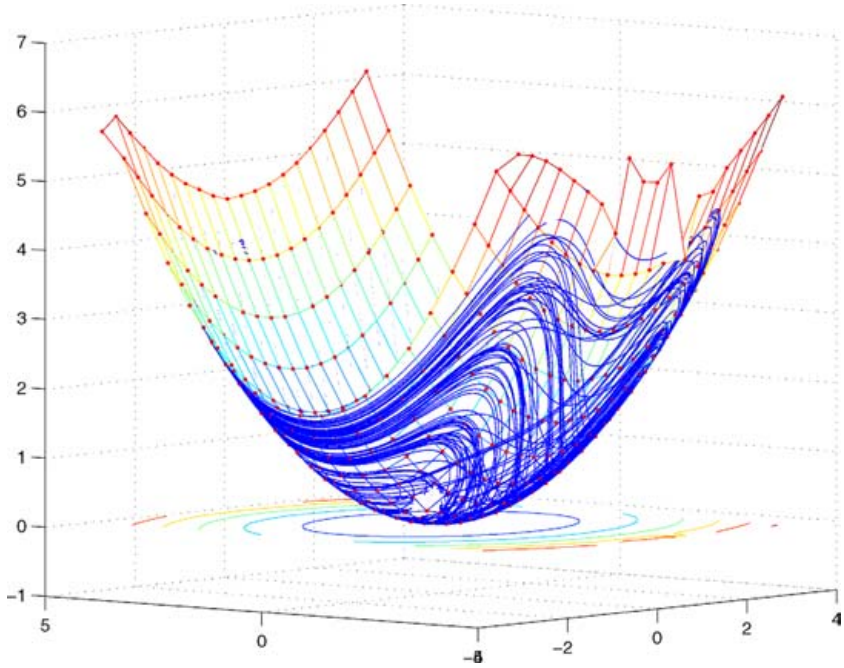


(a)

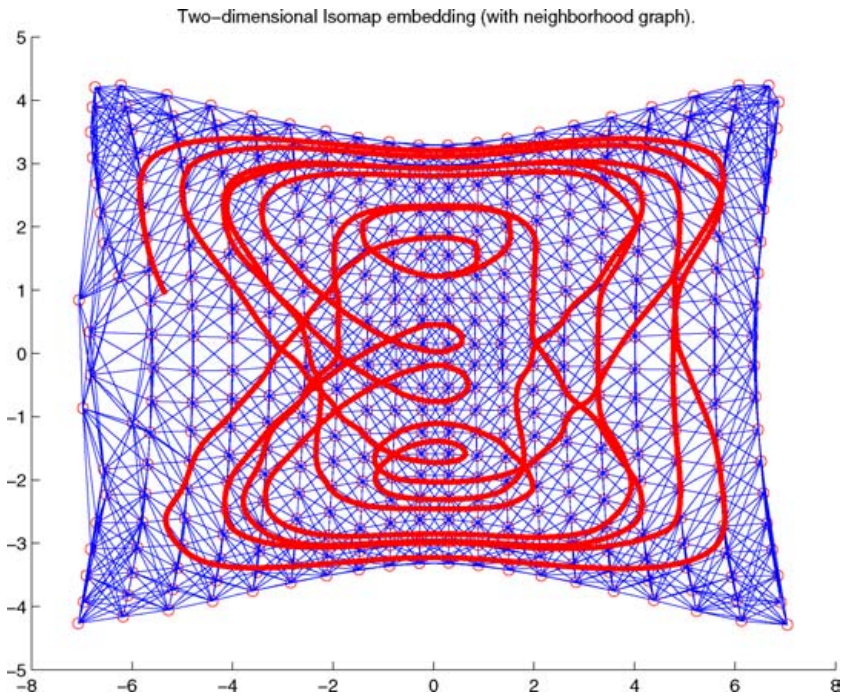


(b)

Fig. 2. (a) Dimensionality found by the ISOMAP algorithm applied to the dataset shown in Fig. 1(a). The horizontal axis is the test dimension, d, and the vertical axis is error. The algorithm indicates quite clearly that the dataset justifies a two-dimensional embedding. (b) The two-dimensional nonlinear embedding of the dataset onto the (paraboloid) manifold. The variables shown are the intrinsic variables, $y = (y_1, y_2)$, as in Eq. (20). This is the discrete graph model of the manifold in intrinsic variables.



(a)



(b)

Fig. 3. (a) The Duffing singularly perturbed flow data from Eq. (36) in blue, and a 20×20 regridding derived by bivariate spline, used for a reduced dataset for ISOMAP. (b) The uniform grid, but now in intrinsic variables, and its underlying neighborhood graph. Also shown in red is the embedded flow data in intrinsic variables.

on a 20×20 uniform grid of $n = 400$ points, resulting in the square-looking grid shown on the paraboloid. The $n = 400$ corners of these squares have been passed to the ISOMAP algorithm, as an alternative and smaller landmark set than using the Duffing data, and constructed to be reliably more uniform than a subsampling of the flow. The resulting embedding is shown in Fig. 3(b), which shows

the discrete graph corresponding to grid shown on the right used to approximate the manifold, but arranged in the intrinsic manifold distances. Likewise, the red curve shows the embedded flow. A remark to point out at this stage is the apparent and obvious result that the grid in Fig. 3(a) which is uniform in x_1, x_2 ambiant variables, is not uniform in the intrinsic manifold variables. This is, of course, expected since the relationship of the slow manifold, $x_3 = H(x_1, x_2)$ gives distances in the intrinsic variables infinitesimally by,

$$\begin{aligned} dL &= \sqrt{dx_1^2 + dx_2^2 + dx_3^2} \\ &= \sqrt{dx_1^2 + dx_2^2 + (D_1H(x_1, x_2)Dx_1 + D_2H(x_1, x_2)Dx_2)^2}, \end{aligned} \tag{41}$$

4.2. *Lorenz equations*

For our second example, we take the famous Lorenz equations [Lorenz, 1963],

$$\begin{aligned} \dot{x_1} &= \sigma(x_2 - x_1), \\ \dot{x_2} &= \rho x_1 - x_2 - x_1 x_3, \\ \dot{x_3} &= x_1 x_2 - \beta x_3, \end{aligned} \tag{42}$$

where we choose as usual, $\sigma = 10, r = 28, b = 8/3$. There is no apparent invariant manifold for these equations, but there is a famous butterfly shaped attractor, with a fractal dimension D_F slightly larger than 2 (it is beside the point here to specify which fractal dimension). In fact, it is known that the Lorenz attractor is better described as a branched manifold, [Williams, 1979; Birman & Williams, 1983]. Therefore, this makes a good example dataset to test an algorithm which insists on treating the data as if there is a manifold.

In Fig. 4, we see results of directly fitting the Lorenz data. As validated by Fig. 4(b), it is no surprise that a two-dimensional manifold fits the attractor well, but not perfectly. The manifold found, shown in Fig. 4(c) appears as two joined annuli, joined at an edge, where in fact it is known that the folding part of the Lorenz chaos occurs; the branched part of the branched manifold has been flattened. This benchmark serves as an instructive example of the sort of topological errors which might occur in higher dimensional attractors. Numerically however, for predictive purposes, this sort of modeling error is not always a problem.

In Fig. 5, we show a second way of processing the Lorenz data, using the splines regridding method mentioned in the previous example. In Fig. 5(a) we show the Lorenz attractor data, together with a splined uniform grid running approximately through the data. Using this regridded data results in the same data in intrinsic variables on the approximated manifold shown in Fig. 5(b). It is in agreement with what is known about the Lorenz attractor, and branched manifolds, and our own previous result, that now the manifold approximation now appears as two planar manifolds with two apparent tears. Recent work has analytically discussed aspects of the Lorenz equations which do indeed give rise to a singular perturbation form [Ramdani *et al.*, 2000].

4.3. *Chua's circuit equations*

An important system in the theory of chaos in nonlinear electronics elements has been the famous Chua's circuit, [Matsumoto, 1984; Chua *et al.*, 1986] which depending on the elements present has been modeled by a differential equation with either a cubic nonlinearity, or a piecewise linear function. In either case, these systems are well known for their rich collection of attractors [Chua, 1992; Tsuneda, 2005] and bifurcations between them as the parameters are varied. Here we take a single example for each of the two types of nonlinearities as presentation of the types of result one can expect when performing an ISOMAP embedding model of the nonlinear attractor.

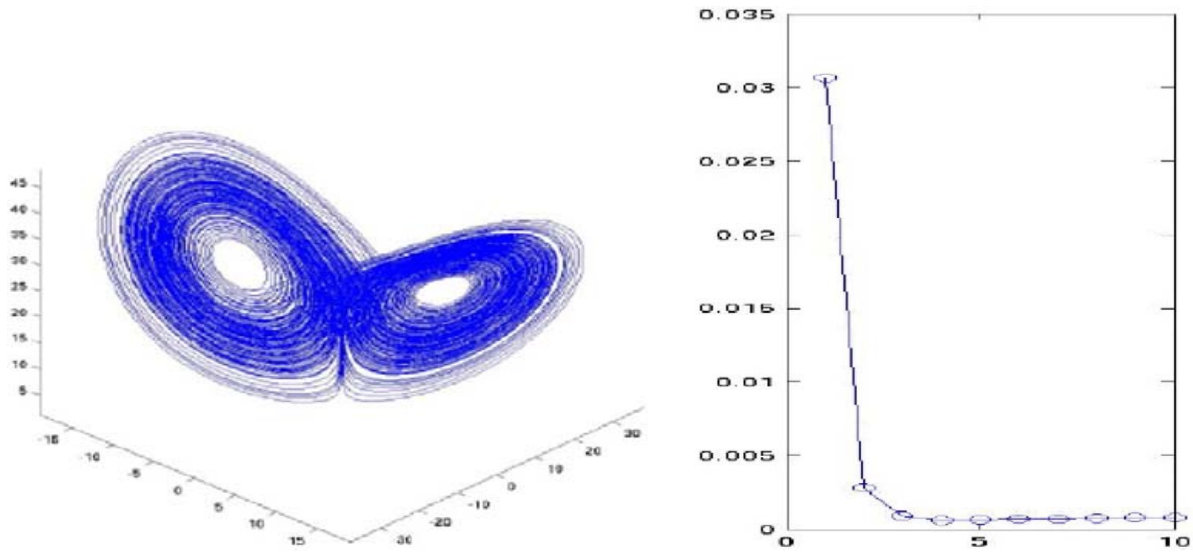
4.3.1. *Chua's equations with piecewise linear nonlinearity*

Consider the equations [Tsuneda, 2005],

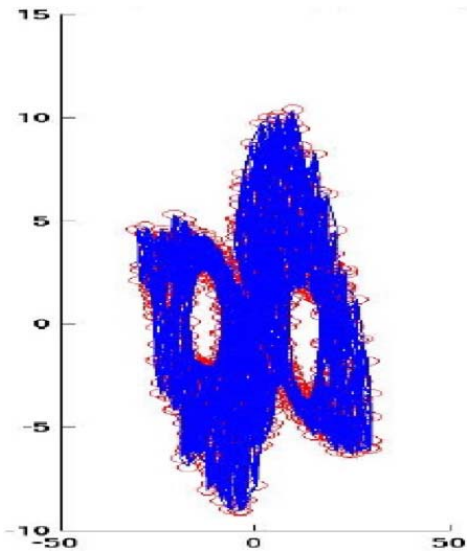
$$\begin{aligned} \dot{x} &= k\alpha(y - x - f_L) \\ \dot{y} &= k(x - y + z) \\ \dot{z} &= k(-\beta y - \gamma z) \end{aligned} \tag{43}$$

with parameters,

$$\begin{aligned} \alpha &= 3 \\ \beta &= 30 \\ \gamma &= -0.86 \\ m_0 &= -3 \\ m_1 &= 0.4 \\ d &= 3.0 \end{aligned} \tag{44}$$



(a) (b)



(c)

Fig. 4. (a) Solution data from the Lorenz equations, Eq. (42) in ambiant variables. (b) Embedding error as a function of dimension. (c) Lorenz data in intrinsic variables, in two dimensions. The model manifold of what is known to be a branched manifold admits well the known location of the branched section which is near the joint shown.

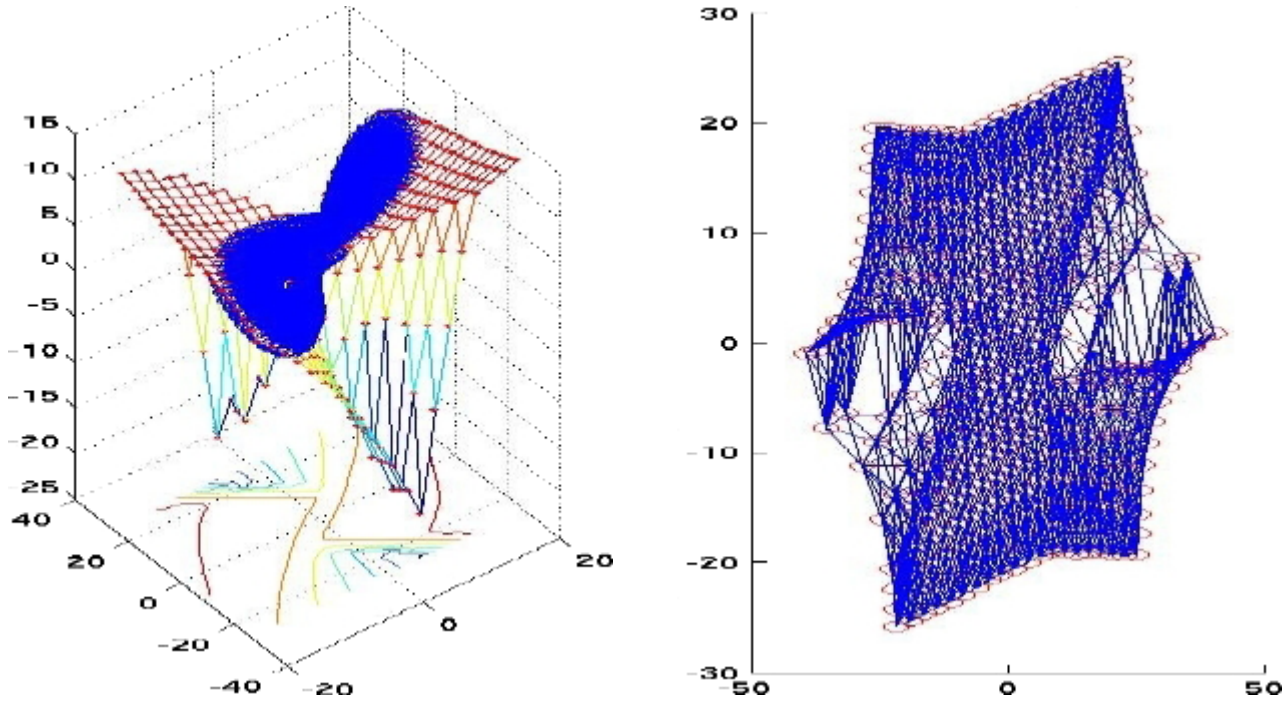
and nonlinear in the piecewise linear form,

$$a = \frac{-35(d^2-1)^2(m_0-m_1)}{16d^7}$$

$$b = \frac{(45d^4-50d^2+21)(m_0-m_1)}{16d^5} + m_1 \qquad (45)$$

$$f_L = m_1 x + \frac{1}{2}(m_0-m_1)(|x+1|-|x-1|).$$

In Fig. 6, we see an example of one of the many possible Chua attractors, this one from the parameters as specified above. Notice the twisting of the attractor. In Fig. 7 we see the results of an ISOMAP embedding of this data. The dimensional analysis is strongly suggestive of two dimensions, which is a correct description projectively, and this is shown also in Fig. 7. Thus in the projective intrinsic coordinates, we see the simple rotation aspect of the attractor, and we get intrinsic coordinates descriptive of this aspect. The example is instructive however, since projectively, there is the loss of information especially due to the twisting aspect of the attractor, but this appears as the nodules



(a) (b)

Fig. 5. (a) Data from the Lorenz equations, Eq. (42) in ambient variables, together with a uniform regridding using a stiff bivariate cubic smoothing spline. This uniform data is processed by ISOMAP. (b) The resulting manifold has apparent rips, which is expected by inspection of the data, considering the intrinsic distances Eq. (41), and the known way in which the two butterfly lobes are joined.

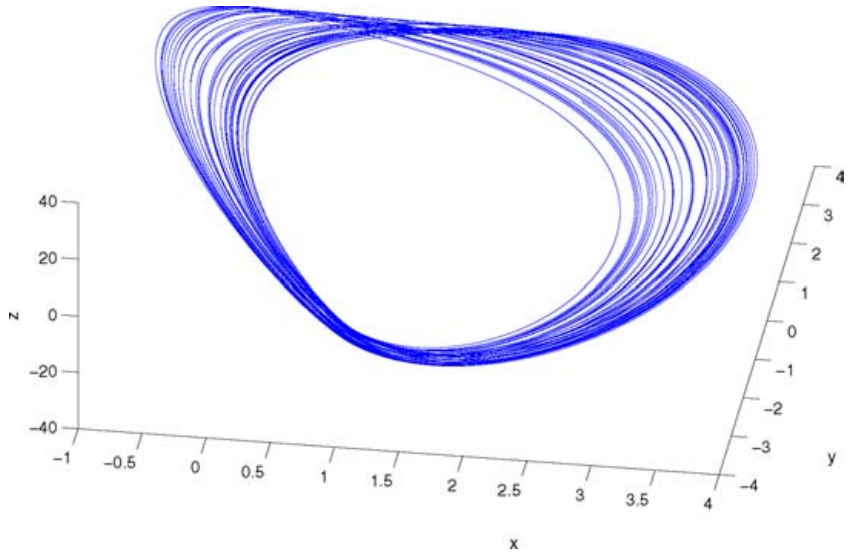
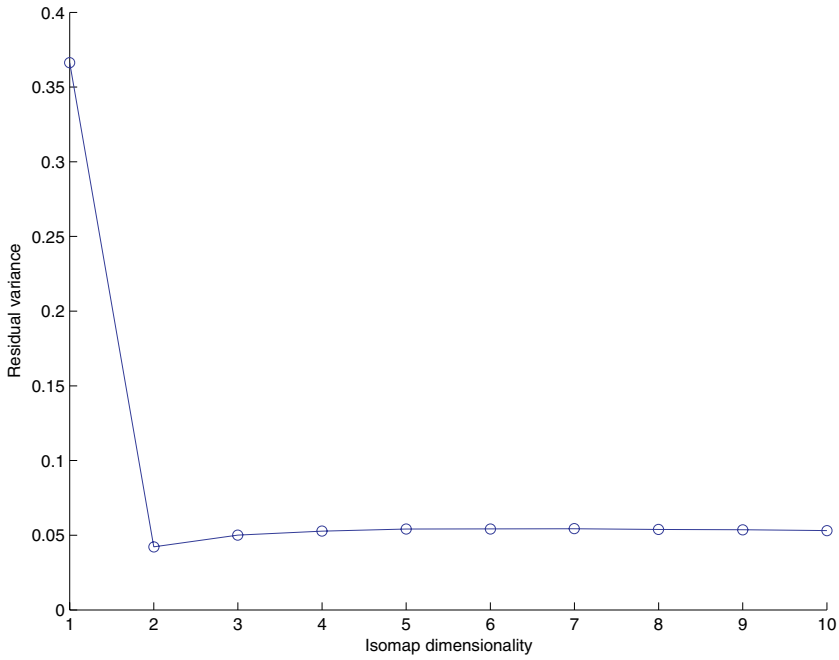
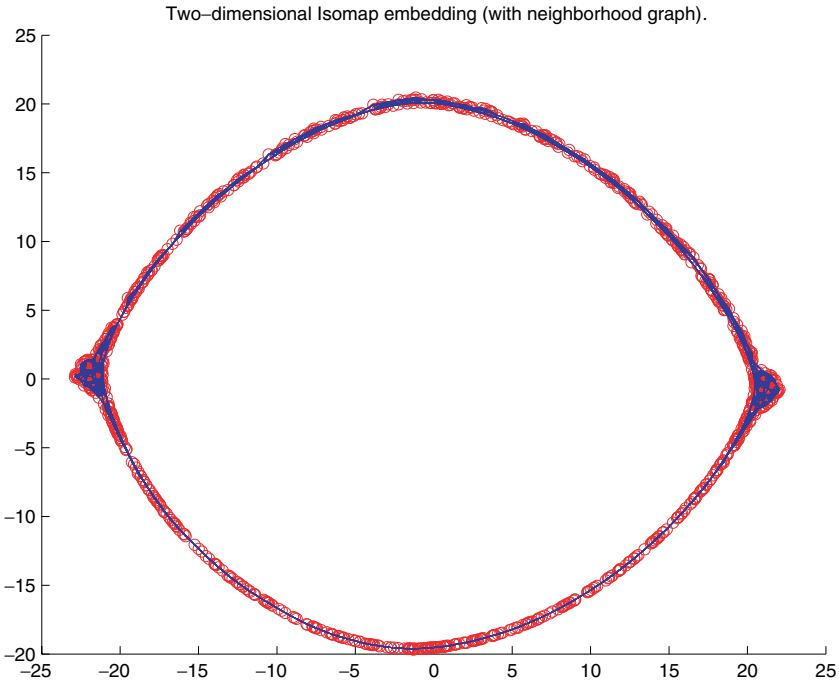


Fig. 6. A Chua attractor due to piecewise linear nonlinearity, according to Eqs. (43)–(45).

seen at the ends. To prevent such a projective loss, a mix of aspects of a false nearest neighbors type technique [Kennel, 1992] and ISOMAP would be necessary, and this is not explored here. In this case, the original coordinates would be required, since it is well known, by Poincaré–Bendixon theorem [Perko, 2006], that at least three independent coordinates are required for chaos. For projections of very high dimensional problems, such as the KS equations in the next subsection, projections which are still chaotic can be quite useful.



(a)



(b)

Fig. 7. ISOMAP embedding of piecewise linear Chua's flow shown in Fig. 6. (a) Dimension calculation strongly indicates a two-dimensional embedding dimension. (b) The resulting intrinsic coordinates suggest the simple rotation where the twist fold appears as the nodules at the ends.

4.3.2. *Chua's equations with cubic nonlinearity*

Consider the equations [Tsuneda, 2005],

$$\begin{aligned} \dot{x} &= k\alpha(y - x - f_C) \\ \dot{y} &= k(x - y + z) \\ \dot{z} &= k(-\beta y - \gamma z) \end{aligned} \tag{46}$$

again with parameters as in Eqs. (44), and now nonlinear in the cubic form,

$$\begin{aligned} a &= \frac{-35(d^2-1)^2(m_0 - m_1)}{16d^7} \\ b &= \frac{(45d^4 - 50d^2 + 21)(m_0 - m_1)}{16d^5} + m_1 \\ f_C &= ax^3 + bx. \end{aligned} \tag{47}$$

In Fig. 8, we see the resulting attractor. The ISOMAP embedding is again clearly two-dimensional, as shown by Fig. 9(a), but this time, the projection to intrinsic variables on the manifold corresponds to a domain which appears as an annulus, as seen in Fig. 9(b). Inspecting the three-dimensional picture in Fig. 8 this makes strong sense.

4.4. *Kuramoto–Shivasinky equations*

The Kuramoto–Shivasinky systems developed as a description of flame front flutter [Kuramoto & Tsuzuki, 1976] has become somewhat of a paradigm of spatiotemporal dynamics, and also one used in particular, in the study of low-dimensional behavior in high dimensional systems such as by using a KL modeling [Zoldi & Greenside, 1997] and one for which the inertial manifold theory can be carried far forward analytically [Foias *et al.*, 1988; Robinson, 2001], and numerically with the approximate inertial manifold theory [Jolly *et al.*, 2001]. As is typical with inertial manifold theory, there is an apparently very large gap between the minimally provable dimensionality of the inertial manifold, and the apparent dimension of the observed attractor.

We will take as our last example, the KS equations,

$$u_t = (u^2)_x - u_{xx} - \nu u_{xxxx}, x \in [0, 2\pi], \tag{48}$$

or periodically extended, $u(x,t) = u(x + 2\pi, t)$. It is an example of an evolution equation of the form Eq. (1), which can be written formally as an ODE in a Banach space as follows [Cvitanovic, 2003]. Let,

$$u(x,t) = \sum_{k=-\infty}^{\infty} b_k(t) e^{ikx}. \tag{49}$$

Assuming a real u forces $b_k = \overline{b}_k$. Substitution yields infinitely many ODEs for the time varying Fourier coefficients,

$$\dot{b}_k = (k^2 - \nu k^4) b_k + ik \sum_{k=-\infty}^{\infty} b_m b_{k-m}. \tag{50}$$

Restricting to pure imaginary solutions yields, $b_k = ia_k$ for real a_k gives,

$$\dot{a}_k = (k^2 - \nu k^4) a_k + ik \sum_{k=-\infty}^{\infty} a_m a_{k-m}, \tag{51}$$

and restricting to odd solutions, $u(x,t) = -u(-x,t)$ gives $a_{-k} = a_k$. Finally, for computational reasons, it is always necessary to truncate at the Nth term,

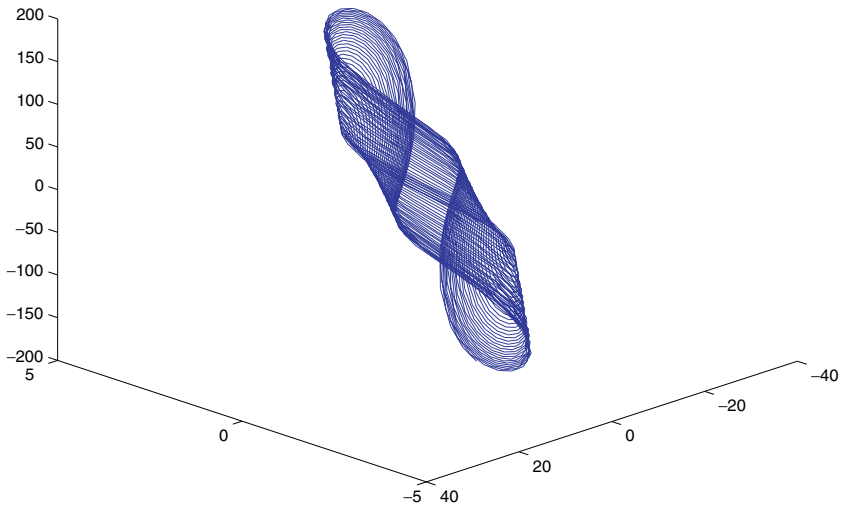
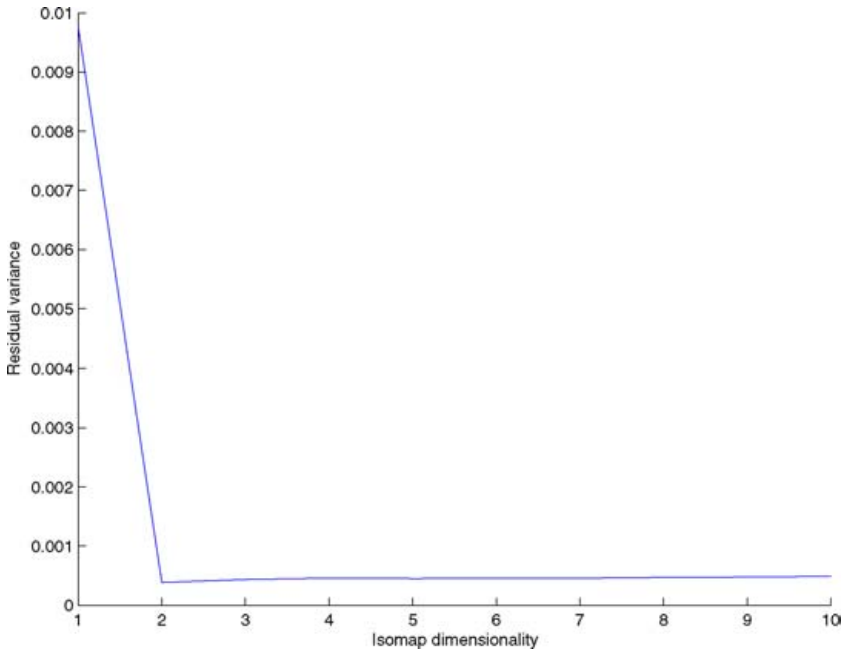
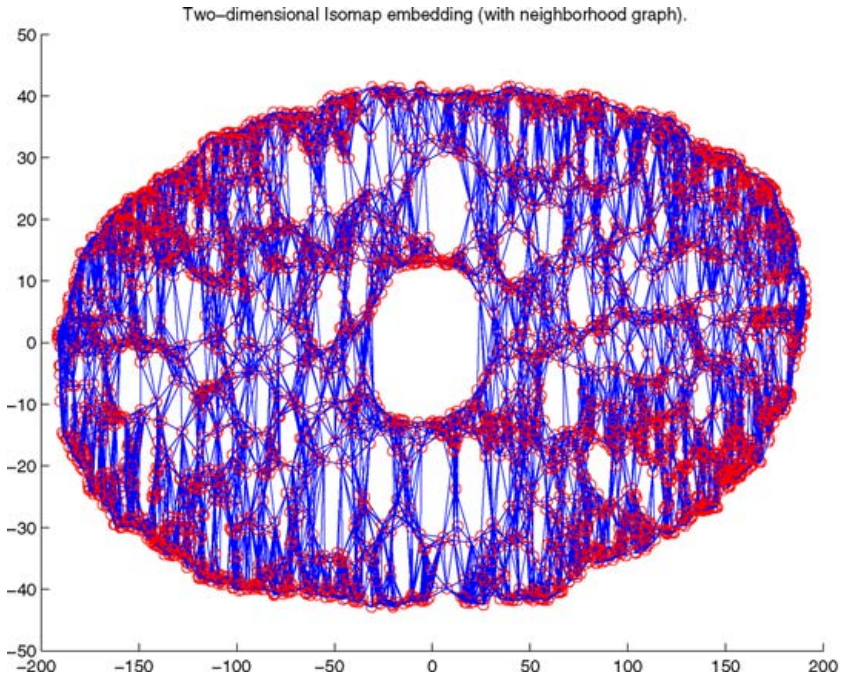


Fig. 8. A Chua attractor due to cubic nonlinearity, according to Eqs. (46) and (47).



(a)



(b)

Fig. 9. ISOMAP embedding of cubic nonlinearity Chua's flow shown in Fig. 8. (a) Dimension calculation strongly indicates a two-dimensional embedding dimension. (b) The resulting intrinsic coordinates suggest the simple rotation where the twist fold appears as the nodules at the ends.

which here means defining $a_k = 0$ if $k > N$. In Fig. 10, we show comparison between the solutions of the ODEs Eq. (51) for a 16 mode and 24 mode model of the KL attractor. We can see obvious differences, and the question is therefore whether either truncation can capture the true dynamics, and in fact the inertial manifold theory [Foias *et al.*, 1988; Robinson, 2001] indicates that possibly hundreds of thousands of modes may be necessary. The approximate inertial manifold theory

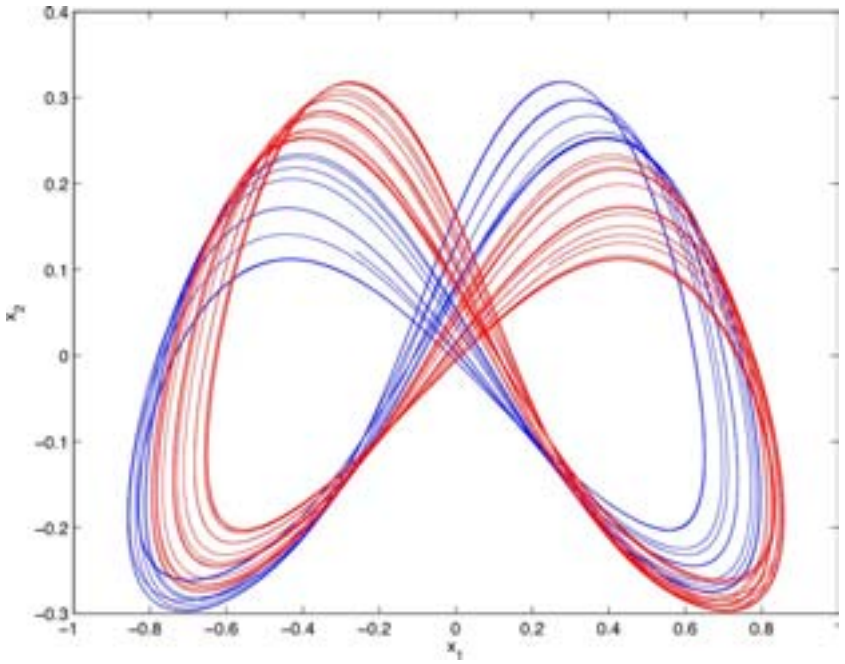


Fig. 10. Projective views of their respective a_1 and a_2 coordinates of KS ODE equations Eq. (51), using $N = 16$ and 24 mode models respectively in blue and red.

[Jolly *et al.*, 1990] suggests that somewhat less, but still large number of modes could be necessary to even approximately model the equations.

We will take only an observational approach, and assume that we have used a sufficiently large number of modes in our truncation so as to ensure that the data is a good topological model of the attractor. We will then use the data together with the manifold modeling procedures which we have described to observe very simple dynamics. For the sake of pictures, we show results from the 16 mode model, although we observed the same with the 24 and 32 mode models. We see from Figs. 11 and 12 that strongly indicate a three-dimensional

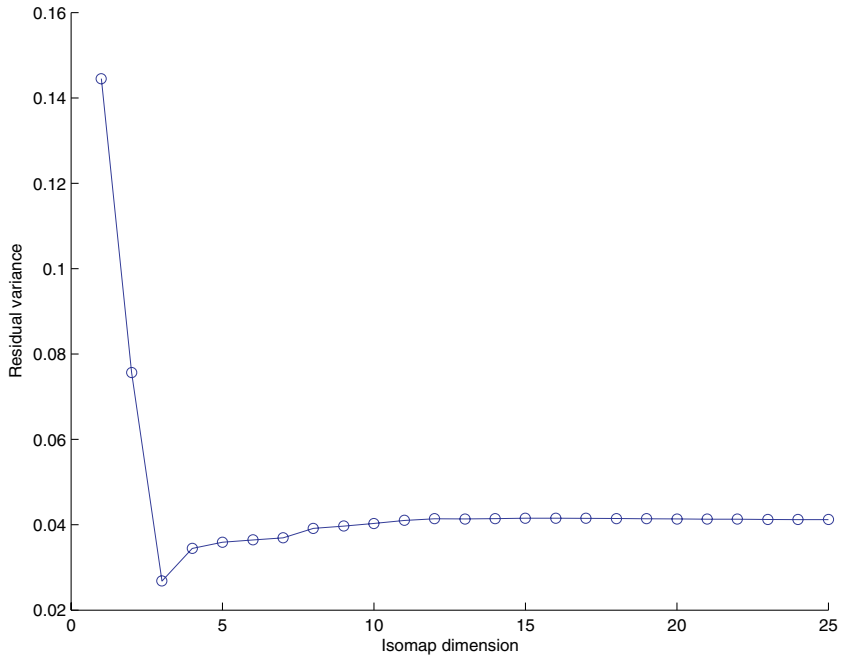
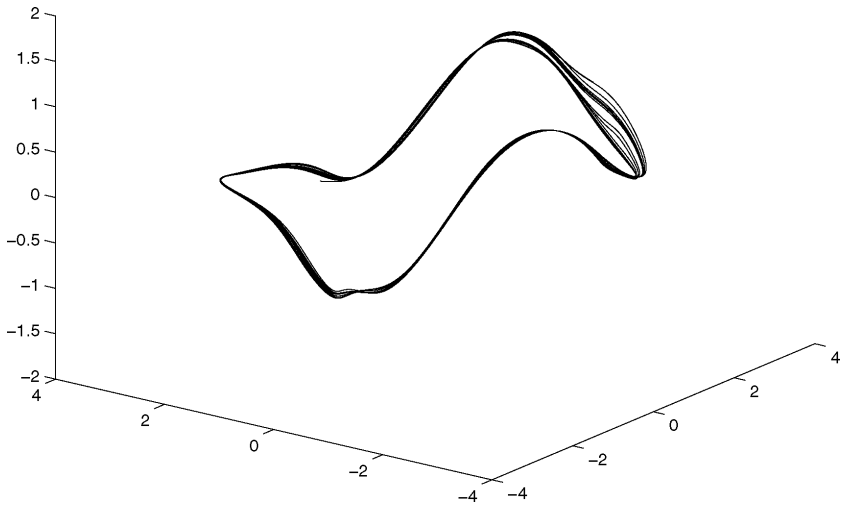
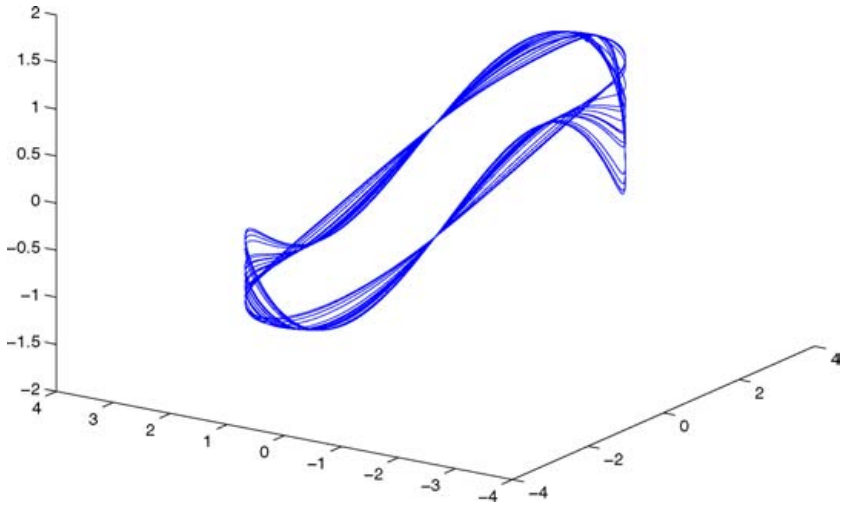


Fig. 11. Data on (near) the attractor of the KS ODE equations Eq. (51) strongly indicates a three-dimensional manifold. See Fig. 12(b) showing the three-dimensional embedding of the data into intrinsic variables from ISOMAP.



(a)



(b)

Fig. 12. (a) Projection of the data of the KS ODE equations Eq. (51) onto three a_1, a_2, a_3. (b) Results of the ISOMAP algorithm embedding the data in three intrinsic variables.

embedding is best. An apparent twist in the bands shown also justifies the projection.

4.5. *Generalized synchronization example*

Synchronized arrays of dynamical systems have attracted considerable attention in the past decade and a half. See for example one particularly nice review [Boccaletti *et al.*, 2002], amongst many that are available. Identical synchronization can be characterized by a stable invariant "synchronization" manifold. A more general version of synchronization, known as generalized synchronization can be defined as a coupled system having a stable normally hyperbolic invariant manifold [Josic, 2000]. Methods based on conditional Lyapunov exponents are popularly used to indicate synchronization, but except for special specific examples, such as the benchmark example used here, usually no direct observation or modeling of the generalized synchronization manifold is seen. The claim here is that the data from such a system reveals that invariant manifold, and it can be effectively approximated by the methods described here.

We take here as an example, a system presented in [Josic, 2000], which consists of a Lorenz coupled

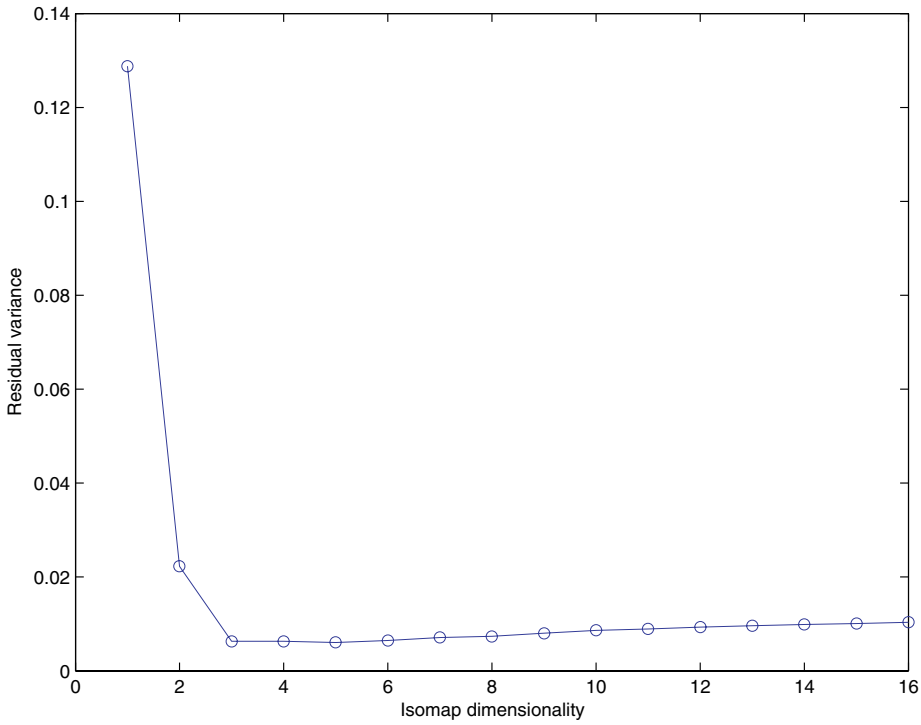


Fig. 13. The Lorenz driving Rossler generalized synchrony system Eq. (52) is indicated by an apparent three-dimensional invariant manifold, which is in this case known to be the slow manifold Eq. (53).

to Rossler equations,

$$\begin{aligned}
\dot{x}_1 &= \sigma(y_1 - x_1),\\
\dot{y}_1 &= \rho x_1 - y_1 - x_1 z_1\\
\dot{z}_1 &= x_1 y_1 - \beta z_1\\
\dot{x}_2 &= -y_2 - z_2 - c(x_2 - (x_1^2 + y_1^2))\\
\dot{y}_2 &= x_2 + \alpha y_2 - c(y_2 - (y_1^2 + z_1^2))\\
\dot{z}_2 &= b + z_2 x_2 - d z_2 - c(z_2 - (x_1^2 + z_1^2)),
\end{aligned} \tag{52}$$

and all parameters are chosen to be the famous chaotic values, $\sigma = 10, \rho = 28, \beta = 8/3, \alpha = 0.2, b = 0.2, d = 8.0$. We take $d = 10$ since it is a large enough value for synchronization. It was shown in [Josic, 2000] that a large class of symmetrically and asymmetrically coupled system can be cast by a simple change of variables into a formal singularly perturbed form written in Eq. (11), which according to the geometric singular perturbation theory [Fenichel, 1979] gives rise to the stable manifold defining generalized synchrony. For our purposes, we wish only to show that generally, such systems reveal themselves by their flow which restricts themselves to that manifold. See Fig. 13 which indicates a three-dimensional attractor. It is known for this benchmark example [Josic, 2000], that there is an attractor embedded in the manifold,

$$\begin{aligned}
x_2 &= x_1^2 + y_1^2,\\
y_2 &= y_1^2 + z_1^2,\\
z_2 &= x_1^2 + z_1^2.
\end{aligned} \tag{53}$$

5. Conclusion

Modeling high dimensional processes by a low dimensional process is a fundamental problem in dynamical systems. We have discussed how traditional linear methods have built in failings in the situation that asymptotic behavior of the dynamical system is restricted to a nonlinear manifold, specifically in the setting of singular perturbation theory. We have presented a way in which recent methods of modeling an invariant manifold, known only through a dataset near that manifold, can be effectively learned and modeled as a discrete graph structure. The methods locally based on MDS have certain strong parallels with what was developed in the time-series embedding community in the past decade(s) [Weigenbend & Gershenfeld, 1993; Fraser & Swinney, 1986; Kennel, 1992; Eckmann & Ruelle, 1985], which in its best implementations used SVD locally. However, the global aspects of the new

methods go much further, effectively modeling the global manifold, and therefore global intrinsic variables, by a discrete graph.

A good model of the dynamics restricted to the manifold means, as we have described, presenting the flow as in intrinsic variables, or in manifold coordinates, rather than ambient variables of the larger space. Such description is therefore a first step to one of several useful processes in applied dynamical systems theory, such as prediction of time series, global modeling of differential equations restricted to the manifold (filtering for example) and even control. Our current work is focusing on these applications. It is also our plan in future work to explore how to link analysis of convergence rates of initial distributions of ensembles of initial conditions of a singularly perturbed system, towards the invariant measure on an invariant manifold together with known recent results of sampling theorems of random data, from probability densities on a manifold with convergence of the modeling of the graph approximations of the manifold.

It is also of interest to better understand how more traditional (statistical) methods of identifying low dimensional behavior, such as measuring (box) fractal dimensions of the dataset, or measuring conditional Lyapunov exponents (for indicating generalized synchronization), work in harmony with the above introduction of constructive methods of modeling invariant manifolds.

Acknowledgment

The author has been supported by the National Science Foundation under DMS0404778.

References

Abarbanel, H., Brown, R., Sirowich, J. & Tsimring, L. [1993] "The analysis of observed chaotic data in physical systems," *Rev. Mod. Phys.* **65**, 1331–1392.

Abarbanel, H. [1996] *Analysis of Observed Chaotic Data* (Springer, NY).

Bernstein, M., de Silva, V., Langford, J. C. & Tenenbaum, J. B. [2000] "Graph approximations to geodesics in embedded manifolds," Technical report, Stanford and Carnegie Mellon.

Birman, J. & Williams, R. F. [1983] "Knotted periodic orbits in dynamical systems-I: Lorenz's equations," *Topology* **22**, 47–82.

Bishop, C., Svensen, M. & Williams, C. [1998] "GTM: The generative topographic mapping," *Neural Comput.* **10**.

Boccaletti, S., Kurths, J., Osipov, G., Valladares, D. L. & Zhou, C. S. [2002] "The synchronization of chaotic systems," *Phys. Rep.* **366**, 1–101.

Bollt, E. [2005] *Order from Chaos, Encyclopedia of Nonlinear Science*, ed. Scott, A. (Routledge, NY).

Bregler, C. & Omohundro, S. [1995] *Nonlinear Image Interpolation Using Manifold Learning*, NIPS 7 (MIT Press).

Carr, J. [1981] *Applications of Center Manifold Theory* (Springer-Verlag, NY).

Chua, L. O., Komuro, M. & Matsumoto, T. [1986] "The double scroll family," *IEEE Trans. Circuits Syst.* **CAS-33**, 1073–1118.

Chua, L. O. [1992] "A zoo of strange attractors from the canonical Chua's circuits," *Proc. 35th Midwest Symp. Circuits and Systems*, Vol. 2, pp. 916–926.

Conlon, L., [2001] *Differentiable Manifolds* (Birkhauser).

Cox, T. F. & Cox, M. A. [1994] *Multidimensional Scaling* (Chapman and Hall, London).

Cvitanovic, P. A., Dahlqvist, P., Mainieri, R., Tanner, G., Vattay, G., Whelan, N. & Wirzba, A. [2003] *Classical and Quantum Chaos*, *Part I*: *Deterministic Chaos*, unpublished web book, www.nbi.dk/Chaosbook, version 10.1.6.

DeMers, D. & Cottrell, G. [1993] *Nonlinear Dimensionality Reduction*, NIPS 5 (Morgan Kauffman).

Eckmann, J.-P. & Ruelle, D. [1985] "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**, 617–619.

Eubank, R. L. [1999] *Nonparametric and Spline Smoothing* (Marcel Dekker, NY).

Farmer, J. & Sirowich, J. [1987] "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845.

Fenichel, N. [1979] "Geometric singular perturbation theory for ordinary differential equations," *J. Diff. Eqs.* **31**, 53–98.

Foias, C., Nicolaenko, B., Sell, G. R. & Teman, R. [1988] "Inertial manifolds for the Kuramato–Sivasinsky equation and an estimate of their lowest dimension," *J. Math. Pures. Appl.* **67**, 197–226.

Fraser, A. M. & Swinney, H. L. [1986] "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A* **33**, 11–34.

Golub, G. H. & Van Loan, C. F. [1996] *Matrix Computations* (Johns Hopkins University Press).

Hinton, G., Revow, M. & Dayan, P. [1995] *Recognizing Handwritten Digits Using Mixtures of Linear Models*, NIPS 7 (MIT Press).

Holmes, P., Lumley, J. L. & Berkooz, G. [1996] *Turbulence*, *Coherent Structures*, *Dynamical Systems*, *and Symmetry* (Cambridge Press, NY).

Hughes, T. J. R. [2000] *The Finite Element Method*: *Linear Static and Dynamic Finite Element Analysis* (Dover Publications).

Jolly, M. S., Kevrekidis, I. G. & Titi, E. S. [1990] "Approximate inertial manifolds for the

Kuramato-Sivashinsky equation: Analysis and computations," *Physica D* **44**, 38–60.

Jolly, M. S., Rosa, R. & Temam, R. [2001] "Accurate computations on inertial manifolds," *SIAM J. Sci. Comput.* **22**, 2216–2238.

Josic, K. [2000] "Synchronization of chaotic systems and invariant manifolds," *Nonlinearity* **13**, 1321–1336.

Kantz, H. & Schrieber, T. [1997] *Nonlinear Time Series Analysis* (Cambridge University Press, NY).

Karhunen, K. [1946] "Zur spektraltheorie stochasticher," *Prozesse Ann. Acad. Sci. Fennicae* **37**.

Kennel, M. B., Brown, R. & Abarbanel, H. D. I. [1992] "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A* **45**, 3403.

Kevorkian, J. K. & Cole, J. D. [1996] *Multiple Scale and Singular Perturbation Methods*, Applied Mathematical Sciences (Springer).

Kohonen, T. [1988] *Self-Organization and Associative Memory* (Springer, Berlin).

Kuramoto & Tsuzuki, Y. T. [1976] "Persistent propagation of concentration waves in dissipative media far from equilibrium," *Prof. Theor. Phys.* **55**, 365–369.

LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cartes, C., Denker, J., Drucker, H., Guyon, I., Miiller, U., Sckinger, E., Simard, P. & Vapnik, V. [1995] "Comparison of learning algorithms for handwritten digit recognition," *Proc. Int. Conf. Artificial Neural Networks*, pp. 53–60.

Loeve, M. M. [1955] *Probability Theory* (Van Nostrand, Princeton, NJ).

Lorenz, E. N. [1963] "Deterministic nonperiodic flows," *J. Atmosph. Sci.* **20**, 130–141.

Lumley, J. L. [1970] *Stochastic Tools in Turbulence* (Academic, NY).

Matlab [2005] *The Spline Toolbox*, The MathWorks 7.0.4.352 (R14).

Matsumoto, T. [1984] "A chaotic attractor from Chua's circuit," *IEEE Trans. Circuits Syst.* **CAS-31**, 1055–1058.

Messer, K. [1991] "A comparison of a spline estimate to its equivalent kernel estimate," *Ann. Statist.* **19**, 817–829.

Nyschka, D. [1995] "Splines as local smoothers," *Ann. Statist.* **23**, 1175–1197.

Ott, E., Sauer, T. & Yorke, J. A. [1994] *Coping with Chaos*: *Analysis of Chaotic Data and the Exploitation of Chaotic Systems* (John Wiley, NY).

Perko, L. [2006] *Differential Equations and Dynamical Systems*, 3rd edition (Springer).

Prigogine, I. [1984] *Order Out of Chaos*, *Man's New Dialogue with Nature* (Bantam Books, NY).

Ramdani, S., Rossetto, B., Chua, L. O. & Lozi, R. [2000] "Slow manifolds of some chaotic systems with applications to laser systems," *Int. J. Bifurcation and Chaos* **10**, 2729–2744.

Robinson, J. C. [2001] *Infinite-Dimensional Dynamical Systems*: *An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors* (Cambridge University Texts).

Roweis, S. T. & Saul, L. K. [2000] "Nonlinear dimensionality reduction by locally linear embedding," *Science* **260**, 2323–2326.

Rowley, C. W. [2005] "Model reduction for fluids, using balanced proper orthogonal decomposition," *Int. J. Bifurcation and Chaos* **15**, 997–1013.

Saul, L. K. & Roweis, S. T. [2000] "An introduction to locally linear embedding," Technical report, AT&T Labs and Gatsby Computational Neuroscience Unit, UCL.

Schlkopf, B. & Smola, A. J. [2001] *Learning with Kernels*: *Support Vector Machines*, *Regularization*, *Optimization*, *and Beyond*, Adaptive Computation and Machine Learning (The MIT Press).

Sirovich, L. [1989] "Chaotic dynamics of coherent structures," *Physica D* **37**, 126–145.

Takens, F. [1980] *Dynamical Systems and Turbulence*, eds. Rand, D. & Young, L.-S., Lecture Notes in Mathematics, Vol. 898 (Springer, Berlin), p. 366.

Teman, R. [1997] *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, Applied Mathematical Sciences, 2nd edition (Springer).

Tenenbaum, J. B., de Silva, V. & Langford, J. C. [2000] "A global geometric framework for nonlinear dimensionality reduction," *Science* **260**, 2319–2323.

Tikhonov, A. N., Vasileva, A. B. & Sveshnikov, A. G. [1985] *Differential Equations* (Springer).

Tsuneda, A. [2005] "A gallery of attractors from smooth Chua's equation," *Int. J. Bifurcation and Chaos* **15**, 1–40.

Vapnik, V. [1998] *Statistical Learning Theory* (Wiley-Interscience).

Weigenbend, A. S. & Gershenfeld, N. A. [1993] *Times Series Prediction*: *Forecasting the Future and Understanding the Past* (Addison-Wesley, Reading, MA).

Williams, R. F. [1979] "The structure of Lorenz attractors," *Publications Mathmatiques de l'IHS*, Vol. 50, 73–99.

Zoldi, S. M. & Greenside, H. S. [1997] "Karhunen–Loeve decomposition of extensive chaos," *Phys. Rev. Lett.* **78**, 1687–1690.