Nonlinear Dynamics (2005) 41: 47–67

Dimensionality Reduction Using Secant-Based Projection Methods: The Induced Dynamics in Projected Systems

D. S. BROOMHEAD[1,*] and M. J. KIRBY[2]

[1]*Department of Mathematics, UMIST, P.O. Box 88, Manchester M60 1QD, U.K.;* [2]*Department of Mathematics, Colorado State University, Fort Collins, CO 80523, U.S.A.;* **Author for correspondence (e-mail: david.broomhead@manchester.ac.uk; fax: +44-161-200-3669)*

(Received: 28 April 2004; accepted: 12 May 2004)

Abstract. In previous papers we have developed an approach to the data reduction problem which is based on a well-known, constructive proof of Whitney's embedding theorem [Broomhead, D. S. and Kirby, M., *SIAM Journal of Applied Mathematics* **60**(6), 2000, 2114–2142; Broomhead, D. S. and Kirby, M., *Neural Computation* **13**, 2001, 2595–2616]. This approach involves picking projections of the high-dimensional system which are optimised in the sense that they are easy to invert. This is done by considering the effect of the projections on the set of unit secants constructed from the data. In the present paper we discuss the implications of this idea in the case that the high-dimensional data is generated by a dynamical system. We ask if the existence of an easily invertible projection leads to practical methods for the construction of an equivalent, low-dimensional dynamical system. The paper consists of a review of the secant-based projection method and simple methods for finding good representations of the (nonlinear) inverse of the projections. We then discuss two variants of a way to find the dynamical system induced by a projection which lead to quite distinct numerical approximations. One of these is developed further as we describe various ways in which knowledge of the full dynamical system can be incorporated into the approximate projected system. The ideas of the paper are illustrated in some more or less simple examples, which range from a simple system of nonlinear ODEs which have an attracting limit cycle, to low-dimensional solutions of the Kuramoto–Sivashinsky equation which need many Galerkin modes for their description.

Key words: reduced dynamical systems, secant projections

1. Introduction

We are interested in dynamical systems which are defined on high-dimensional spaces, but which have low-dimensional attractors. As this volume will attest, there has been considerable activity directed towards finding reduced descriptions of such systems. The motivation for this research is due, at least in part, to the observation that many apparently complicated physical phenomena exhibit a tendency towards *self-organization* (see, e.g., [1, 2] and also [3] for an interesting biological perspective). For example, the existence of *coherent structures* is thought to be a direct consequence of such organizing tendencies [4]. Mathematically, we may examine the situation from a geometric viewpoint and imagine that the data representing a given process resides in (or traverses) a high-dimensional vector space. The effect of the self-organization is to limit the volume of space occupied by the data of the process. This being the case, it is possible that the restricted region of state space might be approximated by a subspace or a submanifold within which intrinsically low-dimensional descriptions may be sought – and found.

The existence of inertial manifolds for dissipative partial differential equations has given a sound mathematical foundation to the dynamical system reduction problem [5]. Moreover, this work suggests that nonlinear extensions to reduction methods may provide the key for extracting optimal *inertial forms*. Analytical approaches for the construction of approximate inertial manifolds have been proposed

see, e.g., [6–10]. Such approaches are distinguished by the fact that they employ analytical approximations such as gaps in the distribution of eigenvalues of linear operators for determining the number of parameterizing modes.

This paper, in contrast, deals with what we refer to as a *semi-analytical* approach in that we directly incorporate information about the geometry of the solution sets into the model. To illustrate this idea, consider the example of the Navier–Stokes system of nonlinear partial differential equations. These equations possess the blueprint for fluid motion, yet are intractable and have no known analytical solution [11]. It is through the acquisition of data associated with a specific parameter, via simulation or experiment, that the geometry of the solution space is revealed. However, general techniques of numerical simulation for such problems – finite difference methods and spectral methods, for example – do not exploit this knowledge of phase space; in fact, they are designed to model the whole space at every time-step. Thus, even low-dimensional phenomena, e.g., periodic solutions, may require the evolution of millions of equations.

It is not widely known, but research on the semi-analytical, or empirical–analytical, reduction of dynamical systems using Empirical Orthogonal Eigenfunctions (EOFs) was apparently initiated in [12]. The lack of citations of this paper may be due in part to the negative conclusions it drew; hardly surprising given the then limited availability of the computer resources essential to such investigations. Despite this early pessimism, a significant amount of successful effort has now been expended in the empirical computation of reduced systems. The series of seminal papers by Sirovich nicely illustrate these ideas [13–15] and still serve as an excellent introduction to the field. Additional papers in this vein include [16–18]. The optimal linear methods employed in these investigations were referred to as the Karhunen–Loève (KL) Decomposition – originally proposed in [19] – or, alternatively, as the Proper Orthogonal Decomposition – described in [20, 21]. Note also the independent work of Lorenz who proposed the technique of EOFs that motivated Sellers [22]. These methods are all well-known to be related to the singular value decomposition where the left or right singular vectors are computed as the eigenvectors of the appropriate covariance matrix,[1] see, e.g., [23]. Given the nature of this volume, we stop short here, and will not attempt a broad review of the literature, but will focus on outlining the work that directly influenced this current investigation.

Following the early successes in the construction of reduced dynamical systems, many interesting questions arose. For example, there is the deceptively simple question concerning how such reduced models characterize limit cycles. Ignoring issues concerning transients, the matter is rather straightforward if the limit cycle resides in a two-dimensional linear subspace, i.e., a plane. On the other hand, if the limit cycle has energy in many different modes as observed in the Kuramoto–Sivashinsky equation [24], then linear approaches for the reduction of the dynamics are essentially doomed to failure as no linear transformation can transport the dynamics to a plane. Such one-dimensional manifolds can, however, be transported nonlinearly and globally to planes [25]. Nonlinear transformations are very effective but equally hard to calculate, especially for more complicated dynamics [26, 27]. Thus, one may appeal to local theory and construct an atlas of charts for representing dynamics [28, 29].

The approach we take in this investigation is based on Whitney's theorem which states that m-dimensional manifolds may be parameterized by $(2m + 1)$-dimensional linear subspaces. Thus, this paper proposes a new approach for constructing *optimal* global models of dynamical systems that are evolved in the parameterisation subspace for the attracting set. Further, we demonstrate that a reasonable level of sophistication may be obtained by the incorporation of dynamical features of the original model

[1] It is worth noting that computation via such products can reduce the numerical accuracy [32].

using a set of constraint equations. Like all of the reduced order models, we propose that this geometric view, which exploits the intrinsic dimension of the data set, has many significant advantages which center on the fact that low-dimensional representations are considerably easier to investigate than their high-dimensional counterparts.

2. Whitney's Theorem and Empirical Reduction

Our main tool will be global projection onto a linear subspace. Our motivation for taking such a simple approach is that there are a variety of results which ensure the existence of faithful – in various senses – projections of low-dimensional sets. In the case that the low-dimensional set is (a subset of) a compact submanifold, say $\mathcal{M}$ ($\dim \mathcal{M} = m$), of some ambient space, the Whitney embedding theorem [30] (see also [31]) provides the necessary foundation by asserting the existence of an embedding of $\mathcal{M}$ in $\mathbb{R}^{2m+1}$.

The basis of the proof of Whitney's theorem is a recursive argument about the $(n-1)$-dimensional set of projections $\pi_{\hat{v}} : \mathbb{R}^n \to \mathbb{R}^{n-1}$ for suitably large n. Here the projections are parameterised by the unit vector $\hat{v}$ which can be thought of as spanning the kernel of $\pi_{\hat{v}}$ (see Figure 1). The set of possible $\hat{v}$s which correspond to distinct projections is the projective $(n-1)$-space, P_{n-1}. The proof demonstrates that – given $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$ (this can always be arranged, given a large enough n) – there is an open dense subset of P_{n-1} for which the corresponding projections of $\mathcal{M}$ are invertible. This property holds – and, therefore, successive projections can be found – while $n > 2m + 1$. The theorem ensures that the inverse of any such projection is as smooth as the manifold itself.

This result has implications about the way that $\mathcal{M}$ can be represented in the ambient space, $\mathbb{R}^n$. Imagine decomposing $\mathbb{R}^n$ into two parts by writing a general point in $\mathbb{R}^n$ as (x, y) were $x \in \mathbb{R}^{n-1}$ and $y \in \ker \pi_{\hat{v}}$. In the case that $\pi_{\hat{v}}$ is an embedding of $\mathcal{M}$, then for each $x \in \pi_{\hat{v}}\mathcal{M}$ there is unique value y such that $(x, y) \in \mathcal{M}$. In fact, the manifold can be thought of as the graph of a smooth function $g : \pi_{\hat{v}}\mathcal{M} \subset \mathbb{R}^{n-1} \to \mathbb{R}$

$$\mathcal{M} = \{(x, g(x)) \,|\, x \in \pi_{\hat{v}}\mathcal{M} \subset \mathbb{R}^{n-1}\} \tag{1}$$

This simple argument provides the foundation of the approach we shall take: it will be assumed that there is data restricted to a low-dimensional submanifold of a higher-dimensional ambient space; a projection of this which is smoothly invertible will be sought; and the function g will be found by

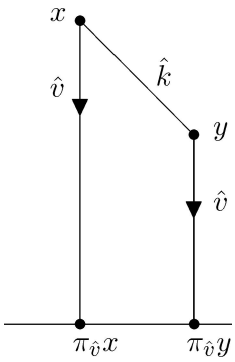

Figure 1. The projection of the data along direction $\hat{v}$. Here $\hat{k}$ represents the unit secant associated with the points x and y.

some suitable fitting method. In principle, the projection process can be done recursively, reducing the dimensionality by one at each step. However, as we shall describe in Section 3, it is generally more convenient to find the projection in one step. The argument can clearly be generalised to this case. We denote the chosen projection as $\pi_d : \mathbb{R}^n \to \mathbb{R}^d$ for some $d < n$ and the corresponding inverse as $\pi_d^{-1} := (id, g)$ where $g : \pi_d \mathcal{M} \to \mathbb{R}^{(n-d)}$.

The Whitney Embedding Theorem gives conditions for the existence of a large (open dense) set of projections which are embeddings of $\mathcal{M}$. Our approach is to search this set for projections which are particularly easy to invert. In this semi-analytic approach, finding the inverse of a projection is a matter of fitting the mapping from projected data to unprojected data. The ease with which this can be done can be related to how ill-conditioned the mapping is. Since we are dealing with smooth manifolds, we know that the inverse function – and, therefore, the function $g : \pi_d \mathcal{M} \to \mathbb{R}^{(n-d)}$ – is Lipschitz; that is, there exists a finite constant K_{π_d} such that

$$\|(x, g(x)) - (y, g(y))\| \leq K_{\pi_d} \|x - y\|$$

for all $x, y \in \pi_d \mathcal{M} \subset \mathbb{R}^d$. This can readily be rewritten as an inequality condition on the corresponding projection

$$\|\pi_d x - \pi_d y\| \geq \kappa_{\pi_d} \|x - y\| \tag{2}$$

for all $x, y \in \mathcal{M}$. For our purposes $\kappa_{\pi_d} = K_{\pi_d}^{-1} \leq 1$ will be taken as a measure of how well-conditioned the inverse of the projection is; the larger that κ_{π_d} is, the more slowly varying is the inverse, and the easier it is to fit the function g.

3. Secant-Based Projection Methods

The set of bad projections is the set of projections which are not injective. (The set of non-immersive projections is also bad, but can be thought of as the boundary of this set.) For simplicity, let us return to the case of finding a good projection $\pi_{\hat{v}} : \mathbb{R}^n \to \mathbb{R}^{n-1}$. The set of non-injective projections consists of those which are parameterised by unit vectors, $\hat{v}$, which are members of Σ, the set of unit secants of the manifold

$$\Sigma = \left\{ \frac{x - y}{\|x - y\|} : \forall x, y \in \mathcal{M},\ x \neq y \right\}$$

If we introduce the equivalence relation $\hat{v} \sim -\hat{v}$ then we can consider that $\Sigma \subset P_{n-1}$. The central plank of Whitney's theorem is a proof that – given the hypotheses of the theorem – Σ is actually nowhere-dense in P_{n-1}. In fact, taken with the immersive part of the proof, the closure of this set, $\bar{\Sigma}$, is shown to be nowhere dense in P_{n-1}.

Let us choose a projection $\pi_{\hat{v}}$ where $\hat{v} \notin \bar{\Sigma}$, and consider its effect on the unit secants of $\mathcal{M}$. Since $\bar{\Sigma}$ is compact we can define the following quantity for any $\hat{v} \notin \bar{\Sigma}$

$$\kappa_{\hat{v}} = \min_{\hat{k} \in \bar{\Sigma}} \|\pi_{\hat{v}} \hat{k}\| > 0 \tag{3}$$

Given the value of $\kappa_{\hat{v}}$ for a particular projection, we have

$$\|\pi_{\hat{v}} x - \pi_{\hat{v}} y\| \geq \kappa_{\hat{v}} \|x - y\| \tag{4}$$

for all $x \neq y \in \mathcal{M}$. This is (a special case of) the inequality given in Equation (2); by considering the effect of the projection on the unit secants of $\mathcal{M}$ we have a direct approach to estimating the Lipschitz constant of the inverse of the projection. In practise, we will look for orthogonal projections from $\mathbb{R}^n$ to $\mathbb{R}^d$. In this case the kernel of each projection is an $(n-d)$-dimensional linear subspace of $\mathbb{R}^n$. The generalisation to this case is then a matter of looking for a good – or optimal – kernel. In the following we describe two approaches to this problem.

3.1. A Maximin Algorithm

The first approach is a search for *good* projections rather than the unique optimum. This relies on satisfying a point-wise criterion

$$\|\pi_d \hat{k}\|_2 \geq \kappa_{\pi_d} \tag{5}$$

for some suitable choice of κ_{π_d}. (For this purpose we use, $\|\cdot\|_2$, the usual Euclidean norm.) Satisfaction of this criterion is necessary for the projection to be admissible in the sense of Whitney's theorem and has the added feature of ensuring an inverse mapping that is Lipschitz with constant $K_{\pi_d} \leq 1/\kappa_{\pi_d}$. Thus, for example, picking $\kappa_{\pi_d} = 0.1$ makes $K_{\pi_d} \leq 10$. This point-wise criterion has, therefore, considerable appeal from the point of view of our general approach. It allows us to impose a constraint on the steepness of the inverse map that we have to fit when reconstructing the dynamics in the full, high-dimensional state space of the system.

Adaptive Basis Algorithm

- Compute the initial basis by applying KL to the data set of unit secants.
- Determine the smallest dimension d such that $\|\pi_d \hat{k}\|_2 \geq \kappa_{\pi_d}$ for all $\hat{k} \in \Sigma$.
- Find the set S of all *bad* secants defined by

  $$S = \{\hat{k} \in \Sigma \,:\, \|\pi_{d-1} \hat{k}\|_2 < \kappa_{\pi_d}\}.$$

- Update the covariance matrix via Equation (6) above and find new basis: repeat.

We have proposed in [33, 34] an adaptive basis algorithm that seeks to optimise the pointwise criterion given by Equation (5). The foundation of this algorithm is the KL expansion of the set of unit secants. Assume that we have N secants which we think of as n-dimensional vectors. Formally, the KL method uses the diagonalisation of the covariance matrix of these N n-dimensional vectors to find a d-dimensional orthogonal basis such that the projection of all N vectors onto this basis minimises the mean square projection error. Actually, our algorithm simply uses this least squares solution as a starting point. The norms of all of the projected unit secants are computed and those secants whose lengths are reduced to less than κ_{π_d} are treated as a bad set. A new covariance matrix is computed which weights the secants of this bad set more strongly; in this way the method iteratively improves the projection by rotating its kernel to increase the projected norm of the secants that failed the test. This iteration proceeds according to

$$\Theta^{k+1} = \left(1 - \frac{\alpha}{N}\right)\Theta^k + \frac{\alpha}{|S|} S S^{\mathrm{T}} \tag{6}$$

where Θ^k is the weighted covariance matrix at the kth step, and S is a matrix whose columns are the unit secants which failed the test at the kth step. The notation $|S|$ indicates the number of columns of S and α is a positive parameter which scales the weighting given to the bad secants in the new covariance matrix. Note that the test for bad secants occurs at the maximum rank of the projection for which Equation (5) is violated.

3.2. Optimization over Grassman Manifolds

Here we[2] propose a new approach for finding optimal projection matrices that employs some relatively recent developments in geometric optimization theory on Stiefel and Grassman manifolds [35].

Orthogonal projectors of rank d are built on matrices whose columns are a set of orthonormal basis vectors, $U = [u^{(1)}|u^{(2)}|\cdots|u^{(d)}]$, so that $\pi_d = UU^{\mathrm{T}}$. We have described above how a point-wise cost function may be associated with the quality of the projection that produces our reduced system. Here we propose an alternative approach that employs a smooth cost function, $F(U)$, where we seek an optimum U which is constrained such that $U^{\mathrm{T}}U = I_{d\times d}$. Using the terminology of Edelman et al. [35] we are looking for a minimum of F over the set of *orthonormal* matrices U. This set of constraints actually forms a surface named the Stiefel manifold that consists of all the $n \times d$ orthonormal matrices.

In fact, we are interested in a subspace of dimension $d = 2m + 1$ (or less), rather than a particular orthonormal basis. The actual mathematical object of interest is therefore the Grassman manifold obtained from the Stiefel manifold by equivalencing all the orthonormal matrices whose columns span the same linear space. (The Grassman manifold, let us call it $G(n, d)$, can be thought of as the set of d-dimensional linear subspaces of $\mathbb{R}^n$. Using this notation, the projective space, P_{n-1}, introduced in Section 2 is $G(n, n-1)$.) For the purposes of parameterizing our manifold, $\mathcal{M}$, we may employ the *homogeneity assumption* that

$$F(U) = F(UQ)$$

where Q is any $d \times d$ orthogonal matrix.[3] This builds in the fact that we are interested in the subspace and are not directly concerned with the coordinate system up to a rotation.

Thus we propose to minimize the smooth cost function

$$F(U) = \frac{1}{|\Sigma|}\sum_{\hat{k}\in\Sigma}\frac{1}{\|UU^{\mathrm{T}}\hat{k}\|_2} \tag{7}$$

Like the point-wise optimization criterion given by Equation (5), small projections of unit secant norms are heavily penalized.

It has been shown in Edelman et al. [35] that geodesics on the Grassman manifold have a closed analytic form that is convenient for computation, i.e., the geodesic for orthonormal matrices is

$$Y(t) = (Y(0)V\ U)\begin{pmatrix}\cos\Sigma t\\ \sin\Sigma t\end{pmatrix}V^{\mathrm{T}}$$

[2] The work of this section was done in collaboration with J.P. Huke and will be described in more detail in a forthcoming paper.

[3] Recall that orthogonal matrices, unlike orthonormal matrices, are assumed to be square so $U^{\mathrm{T}}U = I$ implies $UU^{\mathrm{T}} = I$.

where

$$\dot{Y}(0) = U\Sigma V^{\mathrm{T}}$$

and $U\Sigma V^{\mathrm{T}}$ is the compact singular value decomposition of the horizontal component of (in this case) the gradient of $F(U)$. We refer the reader to [35] for details of this derivation.

We employ the conjugate gradient algorithm described in [35] in our calculations. However, to illustrate the elegance of these ideas we have included a sample code for implementing a simple steepest decent with line search algorithm using Matlab. Note that although the conjugate gradient method is faster, the steepest descent version of the algorithm performed respectably.

Matlab Code for Steepest Descent on Grassmanians

```
for i = 2:number_iterations
    [FY] = deriv_FY(Y,S);
    %returns the derivatives of the cost function.
    T = -(eye(size(Y,1)) - Y*Y')*FY;
    %computation of the tangent vector.
    [UH, DD, VH] = svd(T,0);
    best_eps = fminbnd('fncerr',0.0001,1)
    %Matlab line search code.
    %line search via call to cost function subroutine.
    DH = best_eps*diag(DD);
    DC = diag(cos(DH));
    DS = diag(sin(DH));
    Y = Y*VH*DC*VH'+UH*DS*VH';
    %evaluation of basis on geodesic.
end
```

3.3. Example

We conclude this section with a simple example. Consider the following map of the circle $\mathsf{P} : \mathcal{S}^1 \to \mathbb{R}^3$ where $\theta \mapsto (\sin\theta, \cos\theta, \sin 2\theta)$. The *pringle* set $P = \mathsf{P}(\mathcal{S}^1) \subset \mathbb{R}^3$, which is shown in the upper left of Figure 2, is an embedding of the circle and will be $\mathcal{M}$ in this example. If we imagine the projection of this set into a plane, then since $2 < 2m + 1 = 3$, we note that $\bar{\Sigma}$ is not nowhere dense in P_2 (see the bottom of Figure 2). There are, however, projections which are not in $\bar{\Sigma}$ and are thus invertible. Some of the possibilities are shown in the upper part of Figure 2. In the upper middle of the figure, we have the optimal projection in the sense of Equation (5) with the largest possible choice of κ_{π_d}, while to the right of this the minimum projected secant norm is clearly zero. This example originally appeared in [33] and additional pathologies of improperly picked projections are illustrated there, including the loss of immersivity.

4. The Induced Dynamics in the Projected System

4.1. Basic Formalities

Let us assume that we are given data which represents a dynamical process in some high-dimensional state space. Given that this data can be considered to lie in a low-dimensional submanifold of the

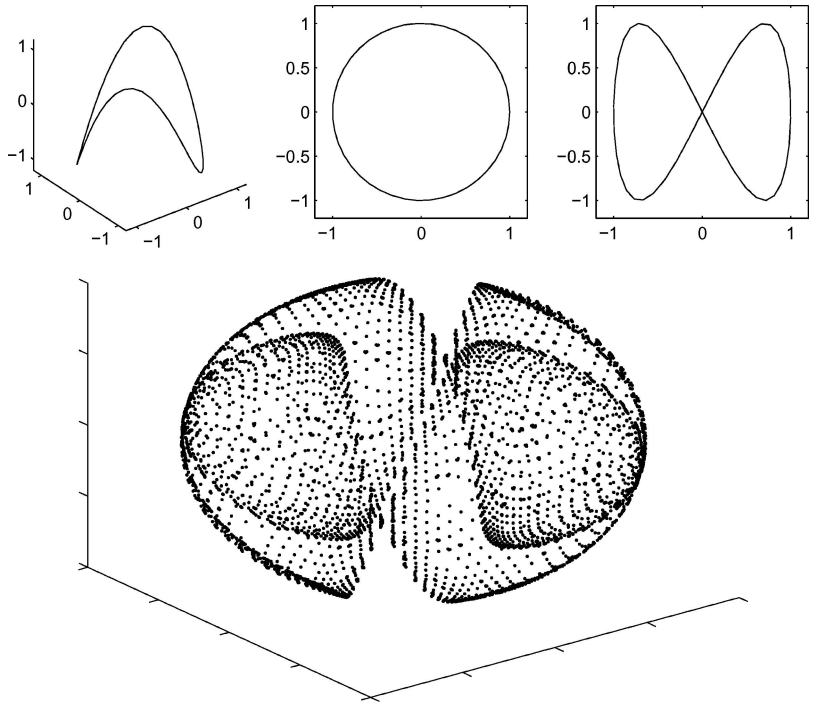

Figure 2. Examples of projections of $P = \mathsf{P}(\mathcal{S}^1)$ onto two-dimensional linear subspaces of $\mathbb{R}^3$. *Left:* An example of a one-dimensional manifold embedded in $\mathbb{R}^3$, i.e., the image of $\mathcal{S}^1$ under the action of P. *Middle:* The invertible projection onto the vectors that maximize the minimum projected secant norm, i.e., along the z-axis. *Right:* The non-invertible projection onto the best two vectors in the KL or POD sense, i.e., along the y-axis. *Bottom:* A sample of the set of unit secants.

state space, we want to describe here how it may be used to estimate an equivalent low-dimensional dynamical system. We shall describe two variants of a basic approach to this, which – for our purposes – differ in the way that numerical approximation enters the result.

Assume that we have direct access to, let us say, an ODE which is the governing equation for a process:

$$\dot{u} = F(u(t))$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ describes analytically the vector field generating trajectories defined on the high-dimensional space. Assume also that we have a (good) projection

$$\pi_d : \mathbb{R}^n \to \mathbb{R}^d$$

which has been derived from the data, or may be known *a priori*. Then, there is a well-established method by which an equivalent low-dimensional dynamical system can be obtained. This is based on using the projection to decompose a state $u \in \mathbb{R}^n$ (for the rest of this section we shall drop the subscript from π_d)

$$u = \pi u + (I_{n\times n} - \pi)u$$

and to separate the ODE to give

$$\frac{d\pi u}{dt} = \pi \circ F(\pi u + (I_{n\times n} - \pi)u) \tag{8}$$

and

$$\frac{d(I_{n\times n} - \pi)u}{dt} = (I_{n\times n} - \pi) \circ F(\pi u + (I_{n\times n} - \pi)u).$$

Given the assumption that the data lies in an m-manifold $\mathcal{M}$ and that, suitably restricted, the projection is invertible:

$$\pi^{-1} : \pi\mathcal{M} \subset \mathbb{R}^d \to \mathcal{M} \subset \mathbb{R}^n$$

it is possible to close Equation (8). In this slightly generalised situation the smooth function corresponding to that used in Equation (1) is

$$g : \pi\mathcal{M} \subset \mathbb{R}^d \to (I_{n\times n} - \pi)\mathcal{M}$$

such that $\pi^{-1}x = (x, gx)$. Using this in Equation (8) gives

$$\frac{d\pi u}{dt} = \pi \circ F(\pi u + g \circ \pi u). \tag{9}$$

In practice, the application of this formula requires that there is an explicit expression for the function g. This is likely to be obtained by fitting the relationship between the projected and the unprojected data using, for example, a radial basis function expansion [23, 36]. When integrating the resulting ODE in $\mathbb{R}^d$, we see that everything is known explicitly except the function g which is given approximately by the data fitting procedure. It is implicit in this formulation that each time the vector field of the projected dynamical system is evaluated it is by means of an evaluation of F where the argument is obtained by lifting to $\mathbb{R}^n$ using g.

4.2. A Modified Approach

In this section we propose a modification to the procedure described in the previous section. In Equation (9) the vector field induced by the projection has the form

$$\pi \circ F \circ \pi^{-1} \tag{10}$$

Here the explicit appearance of the inverse projection indicates the lifting into $\mathbb{R}^n$. As a practical issue, this involves doing calculations in $\mathbb{R}^n$ on each occasion that the numerical algorithm for integration has to evaluate the vector field. Here we suggest an alternative approach which makes this lifting step unnecessary by evolving the trajectories completely within the projected space. To this end, we introduce the mapping f which has the property

$$\pi \circ F = f \circ \pi$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ is the vector field transported to the projected space. Mathematically, since we are assuming that the projection is smoothly invertible, we recognise that Equation (10) could be taken as a definition of f. However, by writing f in this way we see that Equation (9) can now be recast in a form which makes no explicit reference to $\mathbb{R}^n$

$$\frac{d\pi u}{dt} = f(\pi u). \tag{11}$$

Of course, the problem now is that, although we know F, we do not know f explicitly. It turns out in this case that we need to find a fitted approximation for f; this can be done because we do know f explicitly at the projected data points. Assuming that we have data $\{u_k \in \mathbb{R}^n : k = 1, \ldots, K\}$ on the attractor, we know $F(u_k)$. Given the good projection π then

$$f(\pi u_k) = \pi \circ F(u_k).$$

We are, therefore, in a position to construct an interpolation of f in $\mathbb{R}^d$, using, for example, radial basis functions [23, 36].

In the examples given later, the class of models we actually use consists of radial basis functions augmented with global linear terms

$$\hat{f}(x) = Lx + \sum_{j=1}^{J} w_j \phi(\|x - c_j\|) \tag{12}$$

Here the set of points $\{c_j \in \mathbb{R}^d : j = 1, \ldots J\}$ is a set of suitably chosen fixed *centres* which give the basis functions for the expansion by translation of the argument of a suitable, real-valued function ϕ. The coefficients of the $d \times d$ matrix L and the $w_j \in \mathbb{R}^d$ are disposable parameters, chosen to give the least squares solution of the equations

$$\hat{f}(\pi u_k) = \pi \circ F(u_k)$$

where the index k, which runs from 1 to K, labels the data on the attractor.

This problem may be summarised as follows

$$y = A_1 w \tag{13}$$

Here y is a $K \times d$ matrix whose rows are the vectors $\pi \circ F(u_k)$. The structure of A_1 takes account of the fact that we have a mixture of a polynomial and radial basis function fit; A_1 is a $K \times (d + J)$ matrix which has composite rows consisting of the concatenation of a vector πu_k with a vector $(\phi(\|\pi u_k - c_1\|, \ldots, \phi(\|\pi u_k - c_J\|))$. Similarly, w is a $(d + J) \times d$ matrix which has the coefficients of L^{T} as its top $d \times d$ block and the rest consisting of J rows which are the vectors w_j. Standard numerical linear algebra routines can be used to solve this problem.

4.3. Including More Information

In principle, we have much more information about the projected vector field than simply the values $f(\pi u_k)$. Is it possible to incorporate this into the fitting process? For example, the time-ordered positions

of the projected points should be consistent with the projected vector field; we might like to require that the integration of the projected vector field would have a trajectory that passed through – or, at least, close to – the projected data points in the correct order. Can we find a fit, $\hat{f}$, that takes account of this requirement?

If the model data is initially generated via a numerical integration scheme (or may be approximated as such) then we may impose this assumption in the reduced space. For example, in [37] it was assumed that the data could be produced by a Runge-Kutta algorithm. Here it is more convenient to restrict our attention to multi-step methods that use the projected vector field in a linear manner

$$p_{k+1} = \sum_{t=0}^{T-1} \alpha_t f(p_{k-t}).$$

Here the α_t are known parameters which characterise the particular algorithm being used. Replacing f with the radial basis function interpolation $\hat{f}$, and inserting the known projected points $\{\pi u_k\}$ we obtain

$$\pi u_{k+1} = L \sum_{t=0}^{T-1} \alpha_t \pi u_{k-t} + \sum_{j=1}^{J} w_j \sum_{t=0}^{T-1} \alpha_t \phi(\|\pi u_{k-t} - c_j\|) \tag{14}$$

which we interpret as a linear equation for the coefficients of L and the weight vectors w_j; the same quantities as appeared in Equation (13). Therefore we have a second linear least squares problem

$$\tilde{y} = A_2 w \tag{15}$$

where, as before, w is a $(d + J) \times d$ matrix which has the coefficients of L^{T} as its top $d \times d$ block and the rest consisting of J rows which are the vectors w_j. In this case, $\tilde{y}$ is a $K \times d$ matrix whose rows are the vectors πu_{k+1}. The structure of the matrix, A_2 can be inferred from Equation (14). (As with the case of A_1, its form takes account of the mixture of basis functions used for the fit.)

A further, potentially useful, source of information about the projected vector field is the derivative $DF(x)$. The importance of this is that it can provide local information in the neighbourhood of the attractor. If, for example, the attractor of F were to be a limit cycle we would see a circle in the reduced space and hope to be able to interpolate the vector field, f, based on points on the circle. In principle, the $\hat{f}$ so constructed would have an invariant circle; there would be nothing, however, to ensure that this would be attracting, and hence the integration in the reduced space would not necessarily be stable. Incorporating knowledge of DF could resolve this problem.

The fitting of the derivatives in this example is equivalent to another least squares problem. In this case we require the derivative of the radial basis function model $\hat{f}$

$$D\hat{f}(x) = L + \sum_{j=1}^{J} w_j D\phi(\|x - c_j\|) \tag{16}$$

(where the derivatives can be calculated explicitly before doing any numerical work). We can now set up a new linear least squares problem:

$$D\hat{f}(\pi u_k) = \pi \circ DF(u_k)$$

where the right-hand side can be found from knowledge of F and the data points. Again, the index k runs from 1 to K labelling the data on the attractor, and we can write the whole problem in the familiar form:

$$\bar{y} = A_3 w \tag{17}$$

In summary we have proposed that three interpolation problems be solved simultaneously, namely a radial basis function fitting of
– the low-dimensional vector field f,
– the numerical integrator, and
– the derivative of the vector field.
These conditions may be simultaneously satisfied by solving a set of equations of the form

$$A_i w = a_i$$

where the A_i are the matrices appearing in Equations (13), (15) and (17), and $a_1 = y$, $a_2 = \tilde{y}$ and $a_3 = \bar{y}$.

To set up the simultaneous solution of these problems we introduce a weighted objective function defined as follows:

$$E^2(w) = \sum_i \beta_i E_i^2(w)$$

where

$$E_i^2(w) = \|a_i - A_i w\|^2$$

and the β_i are parameters representing the importance we attach to the different pieces of information being incorporated in the final model. Formally we may derive normal equations from this, the solution of which may be written

$$w = \left(\sum_i \beta_i A_i^{\mathrm{T}} A_i \right)^{-1} \sum_i \beta_i A_i^{\mathrm{T}} a_i .$$

Let us define w_i to be the weights that would be obtained by solving the individual normal equations

$$A_i^{\mathrm{T}} A_i w_i = A_i^{\mathrm{T}} a_i .$$

Introducing these above we get

$$w = \left(\sum_i \beta_i A_i^{\mathrm{T}} A_i \right)^{-1} \sum_i \beta_i A_i^{\mathrm{T}} A_i w_i .$$

This result suggests that we can solve the individual fitting problems separately to obtain the w_i and then compute w by interpolating between these values using weightings based on the β_i. If we define

$$\mathbb{P}_i = \left(\sum_i \beta_i A_i^{\mathrm{T}} A_i \right)^{-1} \sum_i \beta_i A_i^{\mathrm{T}} A_i$$

then we may write the solution for the weights which satisfy the fitting problems simultaneously as

$$w = \sum_i \mathbb{P}_i w_i.$$

Note that

$$\sum_i \mathbb{P}_i = 1.$$

5. The Dynamical Pringle Example

This section contains a very simple example which illustrates many of the issues discussed earlier. It is based on a skew product system of ODEs which has a limit cycle attractor in the form of the pringle set considered earlier. We show how the base dynamics can be obtained by projection and the application of a suitable fitting procedure to the data.

The system of equations we shall consider is:

$$\frac{dx}{dt} = y \tag{18}$$

$$\frac{dy}{dt} = -x - (x^2 + y^2 - 1)y \tag{19}$$

$$\frac{dz}{dt} = -\lambda z + 2(\lambda xy + \omega(x^2 - y^2)) \tag{20}$$

where $\lambda > 0$ and ω are parameters. This system is in the form of a skew product where the base dynamics are in the (x, y)-plane. It is easy to prove that this system has an attracting limit cycle which is the pringle set P. The basin of attraction of this is the whole of $\mathbb{R}^3 \setminus \{0\}$. In Figure 3, a numerically integrated

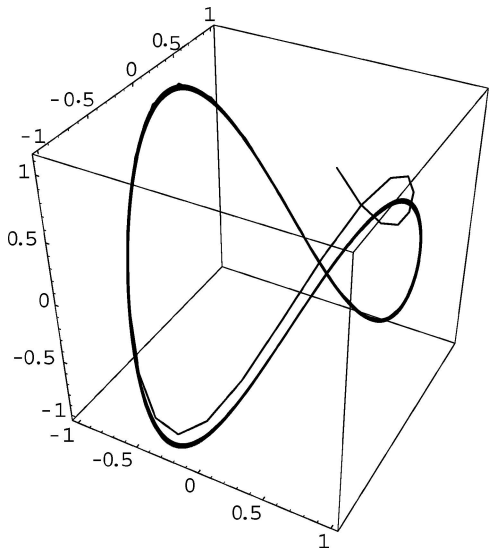

Figure 3. The convergence of a trajectory to the pringle attractor.

trajectory of this system is shown. The transient behaviour, showing convergence of the trajectory to P is evident.

5.1. Numerical Results

For these computations, Equations (18)–(20) were integrated numerically (using the *Mathematica* NDSolve function). The differential equation was integrated initially for 1500 time units to ensure that sufficient convergence to the attractor. The data consisted of 700 points in $\mathbb{R}^3$ sampled uniformly in time (sampling interval 0.01 time units); this time interval corresponds to just over one period of the pringle limit cycle. In addition, numerical values of the Jacobian at each sample point were computed.

In Example 3.3, we established that the secant method finds the optimum projection of data distributed on the pringle set (see also [33]). For the purposes of this calculation, therefore, this optimum projection was used (it corresponds to taking the first two components of each of the data points). The values of the projected vector field were computed at each of the data points and these used in a least squares radial basis function estimation of the projected vector field. The radial basis function fit used a model of the form shown in Equation (12); the linear part L is a 2×2 matrix. The centres were placed on a regular lattice in the plane which extended beyond the region occupied by the projected data.

In Figure 4, we see the results of integrating a simple radial basis function fit – using the model given in Equation (12) with cubic functions ($\phi(r) = r^3$) – of the projected vector field. In this case the centres were distributed on a 7×7 lattice as shown in the figure and the fitted model gives a good representation of f, the vector field induced by projecting Equations (18)–(20). The fitted vector field has an attracting

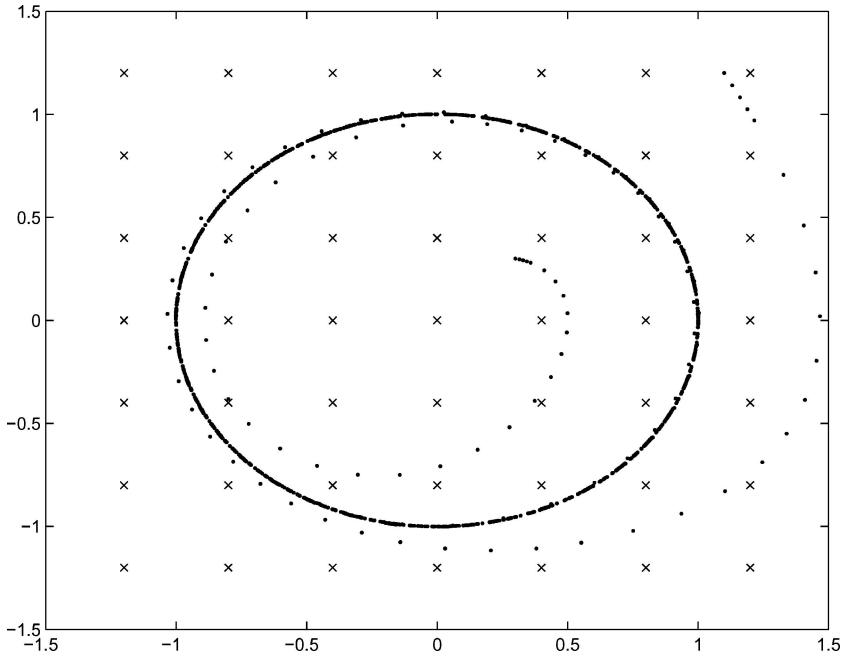

Figure 4. Numerical integrations of the fitted projected vector field corresponding to Equations (18)–(20). In this case the radial basis function model fitted consisted of cubic functions ($\phi(r) = r^3$) and a 7×7 square lattice of centres (positions marked by crosses in the figure), as well as a 2×2 linear part L. The solid dots indicate two integrated trajectories with initial conditions inside and outside the projected limit cycle. We see that the vector field has an attracting limit cycle which is indistinguishable from the projected pringle data.

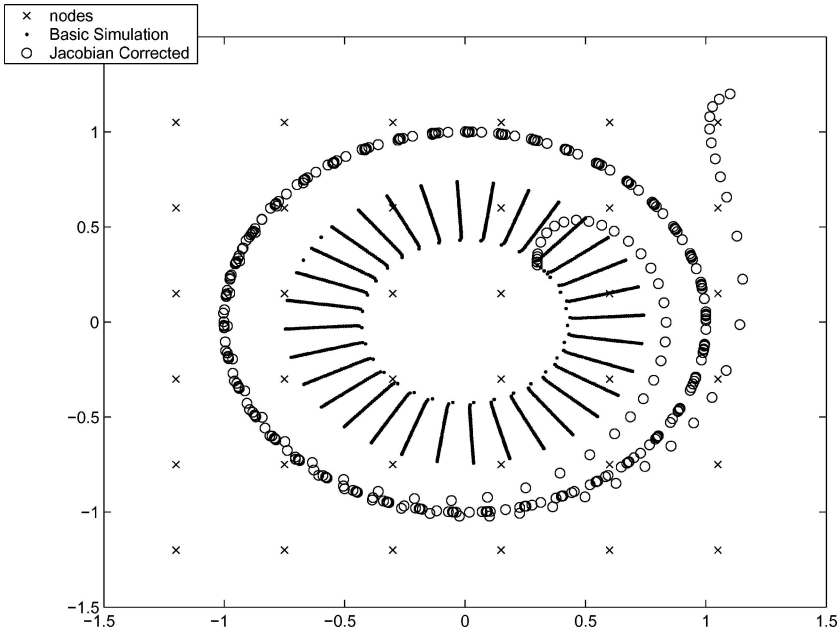

Figure 5. Numerical integrations of the fitted projected vector field corresponding to Equations (18)–(20). In this case the radial basis function model fitted consisted of cubic functions ($\phi(r) = r^3$) and a 6×6 square lattice of centres (positions marked by crosses in the figure), as well as a 2×2 linear part L. The solid dots indicate an integrated trajectory with initial conditions inside the projected limit cycle. It is clear that the convergence to the limit cycle is very weak here – and may even be due to numerical artifact. The open circles correspond to integrated trajectories for a vector field obtained by including information about the Jacobian (see text) in the radial basis function fit. In this case we see that the vector field has an attracting limit cycle which is indistinguishable from the projected pringle data.

limit cycle solution which clearly has a substantial basin of attraction and is very close to (on the scale of the figure, indistinguishable from) the projected pringle data.

Figure 5 shows the results of two kinds of computation. In the first case the standard fit of the projected vector field was carried out but this time using a 6×6 square lattice of centres. Here the fit fails to capture adequately the fact that the vector field should have an attracting limit cycle. However, when the Jacobian data is allowed to have a small weighted contribution to the fit, the resulting vector field again has an attracting limit cycle. Like that shown in Figure 4, the attracting limit cycle solution has a substantial basin of attraction and is, on the scale of the figure, indistinguishable from the projected pringle data.

6. A Kuramoto–Sivasinsky Limit Cycle Example

The Kuramoto–Sivashinsky (K–S) equation

$$u_t + 4u_{xxxx} + \alpha\left(u_{xx} + \frac{1}{2}(u_x)^2\right) = 0 \tag{21}$$

is an appealing partial differential equation for the study low order models as it has been shown to possess finite dimensional dynamics [38–40]. Furthermore, the parameter α essentially serves as a dimensionality control, in the sense that the larger the α the larger the dimension of the phase-space of the solution.

We generate data for this problem in the usual manner, i.e., by decomposing the velocity field using the expansion

$$u(x,t) = \sum_{-\infty}^{\infty} a_n(t) e^{inx}. \tag{22}$$

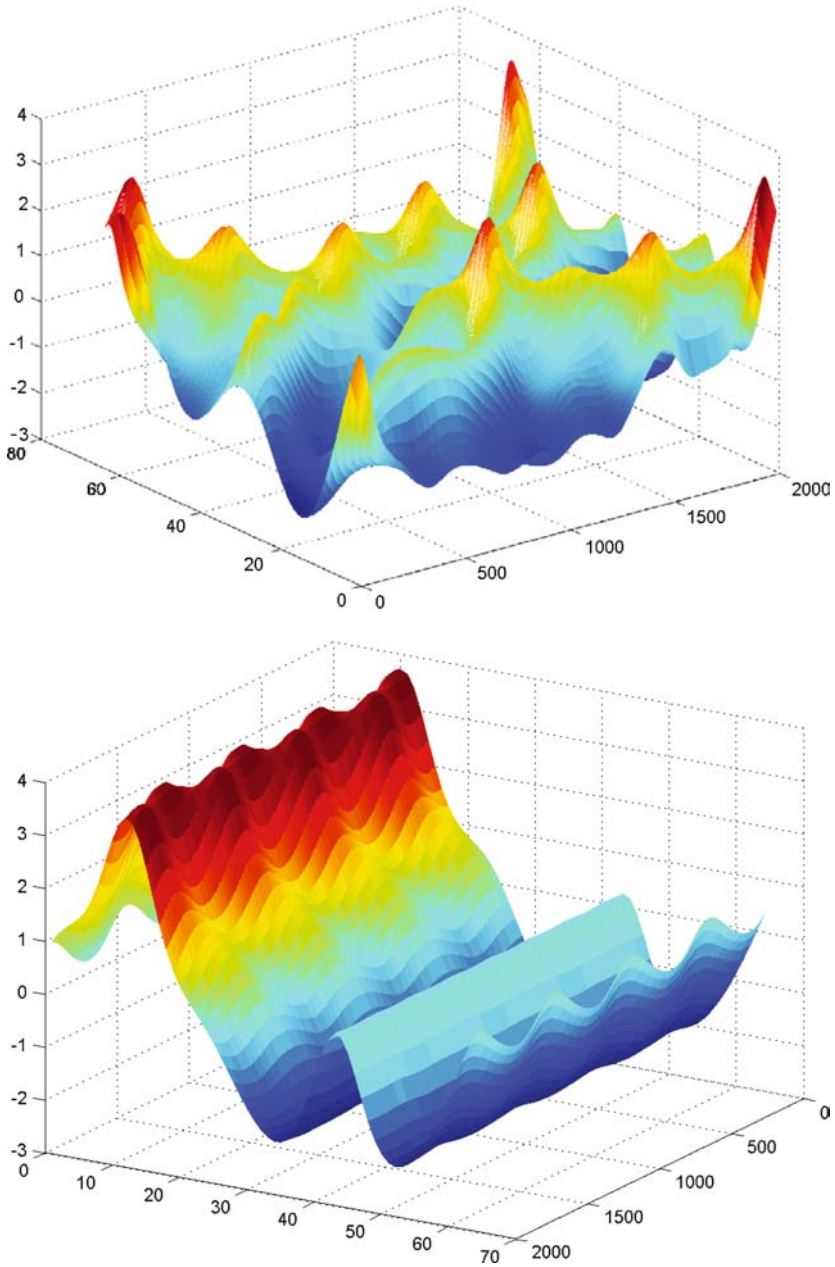

Figure 6. Numerical integrations of Equations (25)–(27) in physical coordinates $u(x,t)$ for $\alpha = 84.25$. The transients corresponding to $t = 0, \ldots, 0.04$ are shown on the left sampled every 0.0002 units. The stable limit cycle is shown on the right for $t = 0.4, \ldots, 0.42$. In graphical units time goes from $0, \ldots, 2000$ and the interval 2π is divided into 64 intervals.

The Fourier coefficients a_n are found by employing the standard orthogonality relationship $\int_0^{2\pi} e^{ikx} e^{-ijx} dx = 2\pi\delta_{j,k}$. Carrying out this procedure gives

$$\sum_{n=-\infty}^{\infty} (\dot{a}_n(t) + (4n^4 - \alpha n^2) a_n(t)) e^{inx} - \frac{\alpha}{2} \left(\sum_{n=-\infty}^{\infty} n a_n(t) e^{inx} \right)^2 = 0 \tag{23}$$

Applying the orthogonality condition and truncating the expansion results in

$$\dot{a}_l(t) = (\alpha l^2 - 4l^2) a_l(t) + \frac{\alpha}{2} \sum_{n=-N+l}^{N} (l-n) n a_{l-n} a_n \tag{24}$$

where $-N \leq l \leq N$. Making use of the reality condition $a_l = \bar{a}_{-l}$ and the fact that the a_0 term decouples from the system gives

$$\dot{a}_l = l^2(\alpha - 4l^2) a_l + \frac{\alpha}{2} \sum_{n=1}^{l-1} (l-n) n a_{l-n} a_n - \frac{\alpha}{2} \sum_{n=1}^{N-l} (l+n) n a_{l+n} \bar{a}_n + \frac{\alpha}{2} \sum_{n=l+1}^{N} (l-n) n \bar{a}_{n-l} a_n \tag{25}$$

where $2 \leq l \leq N-1$ and for $l = 1$

$$\dot{a}_1 = (\alpha - 4) a_1 + \frac{\alpha}{2} \sum_{n=2}^{N} (1-n) n \bar{a}_{n-1} a_n - \frac{\alpha}{2} \sum_{n=1}^{N-1} (1+n) n \bar{a}_n a_{n+1} \tag{26}$$

and $l = N$

$$\dot{a}_N = N^2(\alpha - 4N^2) a_N + \frac{\alpha}{2} \sum_{n=1}^{N} (N-n) n a_{N-n} a_n \tag{27}$$

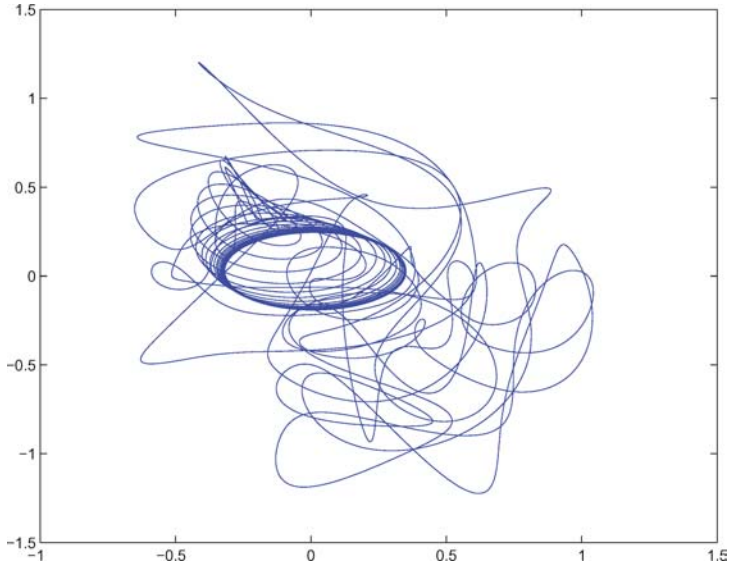

Figure 7. The two-dimensional optimized projection of the Fourier coefficients associated with the numerical integrations displayed in Figure 6. This integration corresponds to $t = 0, \ldots, 0.5$.

6.1. NUMERICAL RESULTS

The 10-dimensional complex Fourier–Galerkin approximation to the Kuramoto–Sivashinsky equations (25)–(27) was integrated in this study using Matlab's variable order integration routine ODE113. The initial conditions $a_1(0) = a_2(0) = 1 + i$ were employed and the solution integrated for $t = 0, \ldots, 0.5$ which resulted in 50000 samples in $\mathbb{C}^{10}$ or, equivalently, $\mathbb{R}^{20}$. The routines for computing the 20×20 Jacobian matrix were also implemented in Matlab in terms of the 20-dimensional real system generated by the complex coefficients. Results of this integration are show for both transient (left) and limit cycle (right) segments in physical coordinates in Figure 6.

6.1.1. *The Projected Dynamics*

We now propose to determine a reduced order model of the 20-dimensional dynamics of the Kuramoto–Sivashinsky equation displayed in Figure 6. If we restrict our attention to data that has converged

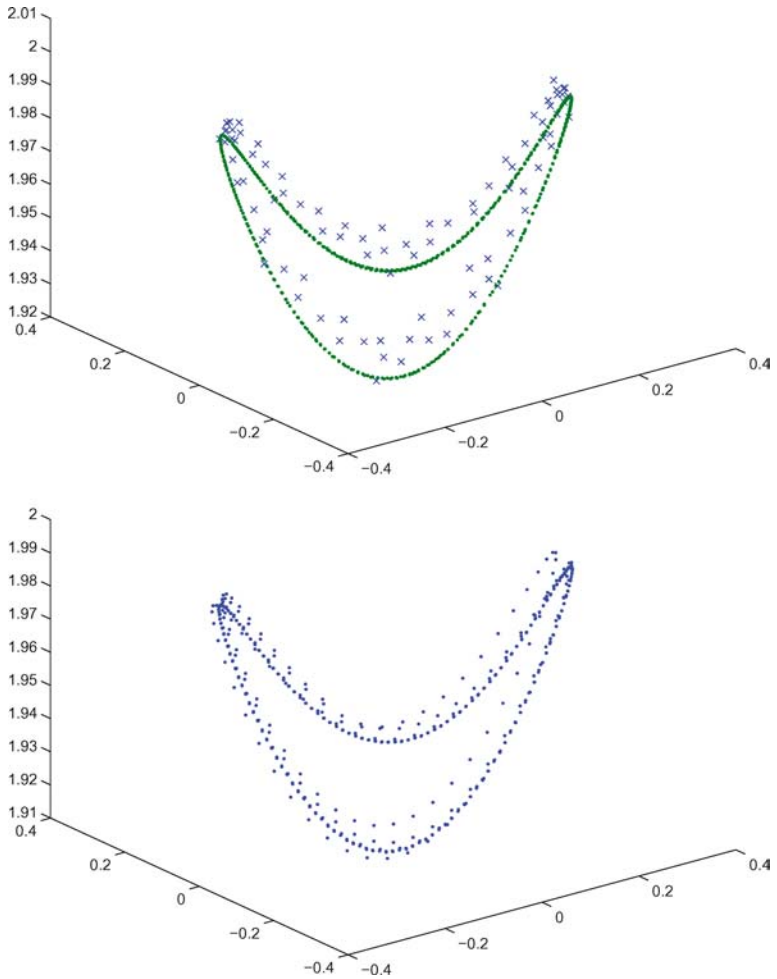

Figure 8. Top: the solid dots represent data on the limit cycle projected from $\mathbb{R}^{20}$ to $\mathbb{R}^3$. The '$\times$' mark the locations of the 88 centers of the scattered basis functions covering both transient and cycle data. *Bottom:* The results of numerically integrating the three-dimensional reduced system using Gaussian radial basis functions. We see that the field is faithfully reproduced for an initial condition off the limit cycle.

to the limit cycle (as shown in right of Figure 6) we obtain a two-dimensional projection with a minimum projected secant norm $\kappa_{\pi_d} = 0.98$. Note this norm has been maximized using the conjugate gradient optimization algorithm over Grassman manifolds as discussed in Section 3.2. While this two-dimensional linear subspace captures the limit cycle the transients now intersect as shown in Figure 7.

Given the intersection of the projected transient trajectories in two dimensions we seek to model the radial basis function ODE in three dimensions. Here we obtain a minimum projected secant norm $\kappa_{\pi_d} = 0.1$ where the transient data has now been included. Certainly the inverse mapping, if needed, will be more difficult to fit on the transient data but a Lipschitz constant of 10 is still reasonable. We emphasize that this inverse is not required for the evolution of our reduced order model dynamical system.

One advantage of radial basis function models as that they are appropriate for mapping scattered data where the domain is a subset of a high-dimensional space. In this example the radial basis function mapping is $f : \mathbb{R}^3 \to \mathbb{R}^3$. Rather than use a potentially large three-dimensional lattice we have opted to place the centers of the functions on the limit cycle as well as at points in the transient data. See the top of Figure 8 where the solid dots represent actual data and the crosses show the placement of the radial basis function centres. In this example the conditioning of the problem was also improved by employing Gaussian radial basis functions $\phi(y) = \exp(-y^2)$ in contrast to the previous example where cubic functions were employed. At the bottom of Figure 8, we see the result of integrating the dynamical system starting with a point off of the limit cycle. Note that the transient here is well modelled in that the trajectory converges to the appropriate limit cycle.

7. Conclusion

We have presented an essentially new empirical approach for the dimensionality reduction of dynamical systems. Our approach is motivated by Whitney's theorem for representing data on (possibly subsets of) manifolds. We show a general algorithm for constructing a bilipschitz mapping such that the Lipschitz constant of the nonlinear inverse is optimally small. We present alternatives for computing the projection of the model based on maximin and smooth optimization criteria.

The power of the data reduction method presented here is that the resulting model is defined completely in terms of the parameterization space and lifting the data back to the ambient space is not required as in other approaches for computing approximate inertial manifolds. One novel aspect of the approach presented here is that we demonstrate how dynamical information such as the Jacobian of the original analytical system may be built into the reduced system. We demonstrate the effect on our pringle model system as well as a high-dimensional limit cycle associated with the Kuramoto–Sivashinsky equation. These examples serve to illustrate variations in the method including lattice based versus cluster based RBF models for the vector field.

Note that the reconstruction of the dynamical models back to their original ambient coordinate systems is certainly possible using methods such as radial basis functions and has been described elsewhere [23, 33, 34].

In future work we propose to explore larger systems with more complicated attracting sets. In instances where we undertake the reconstruction of the dynamics, it will be possible to envision combining our approach with the actual lifting of the trajectories to form a hybrid predictor corrector integrator. This paper is the result of the evolution of many algorithms for computing reduced order models and has several theoretical and algorithmic features that we feel make it an approach of potentially significant interest.

Acknowledgements

This research has been partially supported by the National Science Foundation award DMS 9973303 and DOD USAF Office of Scientific Research, Contract # FA9550-04-1-0094.

References

1. Kohonen, T., *Self-Organization and Associative Memory*, Springer, Berlin, 1984.
2. Linsker, R., 'Self-organization in a perceptual network', *Computer* **21**, 1988, 105–117.
3. Kelso, J. A. Scott, *Dynamic Patterns : The Self-Organization of Brain and Behavior*, MIT Press, Boston, MA, 1995.
4. Peitgen, H.-O., 'Hartmut Jürgens, and Dietmar Saupe', *Chaos and Fractals: New Frontiers of Science*, Springer, New York, 1992.
5. Constantin, P., Foias, C., Nicolaenko, B., and Temam, R., *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations*, Springer, Berlin, Heidelberg, New York, 1989.
6. Jauberteau, F., Rosier, C., and Temam, R., 'The nonlinear Galerkin method in computational fluid dynamics', *Applied Numerical Mathematics* **6**, 1989/1990, 361–370.
7. Jolly, M. S., Kevrekidis, I. G., and Titi, E. S., 'Approximate inertial manifolds for the Kuramoto-Sivashinsky equation: Analysis and computations', *Physica D* **44**, 1990, 38–60.
8. Marion, M. and Temam, R., 'Nonlinear Galerkin methods', *SIAM Journal of Numerical Analysis* **26**, 1990, 1139–1157.
9. Foias, C., Sell, G. R., and Temam, R., 'Inertial manifolds for nonlinear evolutionary equations', *Journal of Differential Equations* **73**, 1988, 309–353.
10. Sirovich, L., Knight, B. W., and Rodriguez, J. D., 'Optimal low-dimensional dynamical approximations', *Quarterly of Applied Mathematics* **XLVIII**, 1990, 535.
11. Landau, L. D. and Lifshitz, E. M., *Fluid Mechanics*, Pergamon Press, New York, 1959.
12. Sellers, W. D., 'A statistical-dynamic approach to numerical weather prediction', Science Report No. 2, Statistical Forecasting Project 2, MIT, Cambridge, MA, 1957.
13. Sirovich, L., 'Turbulence and the dynamics of coherent structures, Part I: Coherent structures', *Quarterly of Applied Mathematics* **XLV**(3), 1987, 561–571.
14. Sirovich, L., 'Turbulence and the dynamics of coherent structures, Part II: Symmetries and transformations', *Quarterly of Applied Mathematics* **XLV**(3), 1987, 573–582.
15. Sirovich, L., 'Turbulence and the dynamics of coherent structures, Part III: Dynamics and scaling', *Quarterly of Applied Mathematics* **XLV**(3), 1987, 583–590.
16. Aubry, N., Holmes, P., Lumley, J. L., and Stone, E., 'Models for coherent structures in the wall layer', *Journal of Fluid Mechanics* **192**, 1988, 115–172.
17. Rodriguez, J. D. and Sirovich, L., 'Low-dimensional dynamics for the complex Ginzburg–Landau equation', *Physica D* **43**, 1990, 77.
18. Berkooz, G., Philip, H., and Lumley, J. L., 'The proper orthogonal decomposition in the analysis of turbulent flow', *Annual Review on Fluid Mechanics* **25**, 1993, 539–575.
19. Karhunen, K., 'Über lineare Methoden in der Wahrscheinlichkeitsrechnung', *Annales Academiae Scientiarum Fennicae* **37**, 1946.
20. Loève, M., *Probability Theory*, von Nostrand, Princeton, NJ, 1955.
21. Lumley, J. L., *Stochastic Tools in Turbulence*, Academic Press, New York, 1970.
22. Lorenz, E., 'Empirical orthogonal eigenfunctions and statistical weather prediction', Science Report No. 1, Statistical Forecasting Project 1, MIT, Cambridge, MA, 1956.
23. Kirby, M., *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*, Wiley, New York, 2001.
24. Kirby, M. and Armbruster, D. 'Reconstructing phase-space from PDE simulations', *Zeitschrift für angewandte Mathematik und Physik* **43**, 1992, 999–1022.
25. Kirby, M. and Miranda, R., 'Nonlinear reduction of high-dimensional dynamical systems via neural networks', *Physics Review Letters* **72**(12), 1994, 1822–1825.
26. Kirby, M. and Miranda, R., 'The remodeling of chaotic dynamical systems', in *Intelligent Engineering Through Artificial Neural Networks*, S. H. Dagli, B. R. Fernandez, J. Ghosh, and R. T. Soundar Kumara (eds.), ASME, New York, 1994, Number 4, pp. 831–836.
27. Kirby, M. and Miranda, R., 'Empirical dynamical system reduction I: Global nonlinear transformations', in *Semi-analytic Methods for the Navier–Stokes Equations (Montreal, 1995)*, K. Coughlin (ed.), Vol. 20 of CRM Proceedings Lecture Notes, American Mathematical Society, Providence, RI, 1999, pp. 41–64.

28. Broomhead, D. S., Indik, R., Newell, A. C., and Rand, D. A., 'Local adaptive Galerkin bases for large-dimensional dynamical systems', *Nonlinearity* **4**, 1991, 159–197.
29. Hundley, D., Kirby, M., and Miranda, R., 'Empirical dynamical system reduction II: Neural charts', in *Semi-Analytic Methods for the Navier–Stokes Equations (Montreal, 1995)*, K. Coughlin (ed.), Vol. 20 of CRM Proceedings Lecture Notes, American Mathematical Society, Providence, RI, 1999, pp. 65–83.
30. Hirsch, M. W., *Differential Topology*, Graduate Texts in Mathematics 33. Springer, Berlin, Heidelberg, New York, 1976.
31. Guillemin, V. and Pollack, A., *Differential Topology*, Prentice Hall, Englewood Cliffs, NJ, 1974.
32. Trefethen, L. N. and Bau III, D., *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.
33. Broomhead, D. S. and Kirby, M., 'A new approach for dimensionality reduction: Theory and algorithms', *SIAM Journal of Applied Mathematics* **60**(6), 2000, 2114–2142.
34. Broomhead, D. S. and Kirby, M., 'The Whitney reduction network: A method for computing autoassociative graphs', *Neural Computation* **13**, 2001, 2595–2616.
35. Edelman, A., Arias, T. A., and Smith, S. T., 'The geometry of algorithms with orthogonality constraints', *SIAM Journal on Matrix Analysis and Applications* **20**, 1998, 303–353.
36. Broomhead, D. S. and David, L., 'Multivariable functional interpolation and adaptive networks', *Complex Systems* **2**, 1988, 321–355.
37. Krischer, K., Rico-Martinez, R., Kevrikidis, I. G., Rotermund, H. H., Ertl, G., and Hudson, J. L., 'Model identification of a spatiotemporally varying catalytic reaction', *AIChE Journal* **39**(1), 1993, 89–98.
38. Hyman, J. M., Nicolaenko, B., and Zaleski, S., 'Order and complexity in the Kuramoto–Sivashinsky model of weakly turbulent interfaces', *Physica D* **23**, 1986, 265–292.
39. Hyman, J. M. and Nicolaenko, B., 'The Kuramoto–Sivashinsky equation: A bridge between PDE's and dynamical systems', Technical Report LA-UR 85-1556, Los Alamos National Laboratory, 1985.
40. Kevrikidis, I. G., Nicolaenko, B., and Scovel, C., 'Back in the saddle again: A computer assisted study of the Kuramoto–Sivashinsky equation', *SIAM Journal of Applied Mathematics* **50**, 1990, 760–790.