

Estimation of Topological Dimension

D. R. Hundley*

M. J. Kirby†

Abstract

We present two extensions of the algorithm by Broomhead *et al* [2] which is based on the idea that singular values that scale linearly with the radius of the data ball can be exploited to develop algorithms for computing topological dimension and for detecting whether data models based on manifolds are appropriate. We present a geometric scaling property and dimensionality criterion that permit the automated application of the algorithm as well as a significant reduction in computational expense. For irregularly distributed data this approach can provide a detailed analysis of the structure of the data including an estimated dimension distribution function. We present our approach on several data sets.

Keywords: Intrinsic dimension, topological dimension, local PCA, local KL.

1 Introduction

The estimation of topological, or intrinsic, dimension is an important first step in data analysis, as it can provide useful insights into the local geometric structure - in fact, it can be used to determine if there are “interesting” local regions that may need extra attention.

We approach the topic from a geometric point of view, in that much of our data comes from high dimensional systems of differential equations, and so much of it can be viewed as noise-free. However, the methods can be applied to data sets that are mildly contaminated with noise, and we address this issue in the examples.

In contrast, much of the literature on intrinsic dimension comes from a statistical perspective (see for a sampling, [7, 6, 1, 12, 14]), and while this too provides one viewpoint, we stress that geometric structure can be very important- there is an inherent trade-off between noise reduction versus maintenance of small scale topological structure, which may in fact be necessary to preserve some desired properties of the manifold in the small dimensional representation. In particular, we don’t see our method as a competitor to statistical techniques, but as an alternative that may provide addi-

tional insight.

The method of singular value curves was originally described in [2], and subsequently utilized in [3], and it is based on the fact that smooth manifolds are locally linear and are well approximated by a truncated Taylor expansion. We extend this treatment in two ways. The first extension is to remove (to some extent) the dependency of the algorithm to the density of the data on the manifold, thus making the algorithm amenable to automation. The second extension is to define the SV curves not point-wise, but in terms of the density. This will allow us to use and interpret the SV curves when the data does not have a well defined manifold. Finally, we show the utility of these algorithms with several examples.

2 Local Dimensionality Estimation

The main ingredient to our approach is the notion of a k -dimensional manifold realized in an n -dimensional ambient space. By definition, such an object may be locally represented in terms of k coordinates

$$\psi_\alpha : \mathcal{M} \rightarrow \mathcal{N}, v = \psi_\alpha(u)$$

where $u \in \mathcal{M} \subset \mathbb{R}^k$ and $v \in \mathcal{N} \subset \mathbb{R}^n$, and α is an index of the local region. In terms of local coordinates, there exists a k -dimensional coordinate system in which the apparently n -dimensional data may be represented. Such a parameterization is afforded by the best linear approximation to the function f which may be expressed in terms of its Jacobian

$$df = Df(\bar{x})(x - \bar{x})$$

where $(Df(\bar{x}))_{ij} = \partial f_i / \partial x_j$. Fukunaga proposed that the characterization of a local region of the data in terms of a best linear approximation might be accomplished numerically by computing the Karhunen-Loève (KL) eigenvalues and eigenvectors (or principal components) in a spherical region about the point of interest [7, 6]. In theory, the locally dominant eigenvectors should span the same linear space as the column space of the Jacobian. As developed in the next section, this observation leads naturally to an approach for computing local bases and, *ipso facto*, the topological, or local intrinsic, dimension. In terms of noisy data, the method will attempt an encapsulation of the data.

*Mathematics Department, Whitman College, Walla Walla, WA 99362

†Mathematics Department, Colorado State University, Fort Collins, CO 80521

3 The Energy Criterion

The local description of a given data matrix $F = [f^{(1)}, \dots, f^{(P)}]$ is initiated by creating a partition in terms of local regions, or ϵ -balls, i.e., spheres of radius ϵ centered at the points $\{c_i\}_{i=1}^{n_c}$ where n_c is the number of balls in the decomposition. Each ball may be defined as

$$B_\epsilon(c_i) = \{f^{(j)} \in F : \|f^{(j)} - c_i\| < \epsilon\}$$

The number of data points contained in $B_\epsilon(c_i)$ will be denoted by P_i .

For a given radius, or a fixed number of points P_i , the eigenvalues and eigenvectors of the ensemble averaged covariance matrix may be computed to provide the decomposition

$$\frac{1}{P_i} B_\epsilon(c_i)^T B_\epsilon(c_i) = \Phi \Lambda \Phi^T$$

where now we view the local ball notation to indicate a matrix whose columns are the points in the ball. Throughout this paper the balls consist of data that is assumed to be centered, i.e., $\tilde{f}^{(\mu)} = (f^{(\mu)} - c_i)^T$; we drop the primed notation for simplicity. Furthermore, define r_i as the rank of the i 'th ball. For later algorithms it will also be convenient to assume that the data is ordered, i.e., the data is relabeled in terms of increasing magnitude, i.e., $\|f^{(1)}\| \leq \|f^{(2)}\| \leq \dots \leq \|f^{(P)}\|$.

Fukunaga proposed that the KL-dimension d_δ of the i 'th ball could be defined as the number of normalized eigenvalues $\tilde{\lambda}_j = \lambda_j/V_i$ that exceed a fixed percentage of the total variance $V_i = \sum_{j=1}^{r_i} \lambda_j$. Specifically, define the KL-dimension using δ at center c :

$$d_\delta = \#\{\tilde{\lambda}_i > \delta\}$$

where δ is the minimum fraction of energy an eigenvalue must retain for its eigenvector to be counted as spanning the tangent space. Typically $\delta = 0.01$, although this choice is *ad hoc*. In practice, this number d_δ is typically well-defined when the size of the ball is small enough, yet contains sufficient data to obtain good statistics.

In summary, Fukunaga's algorithm for determining the *ad-hoc* dimension d_δ proceeds as follows:

Energy Criterion Algorithm

- Select the energy level δ .
- Let ϵ be given by the data cluster centered at c_i .
- Compute the eigenvalues of the local data set $B_\epsilon(c_i)$.
- Determine the KL dimension d_δ at center c_i .

4 The Singular Value Curves

In contrast to Fukunaga's algorithm which is based on an *ad hoc* energy criterion as described in Section 3, the algorithm proposed by Broomhead *et al* [2] uses the scaling of the singular values for objectively determining the local dimension. This approach has the advantage that the linear nature of the tangent space is being directly exploited in the process of its identification.

Given data which resides in a local spherical region $B_\epsilon(c_i) \subset \mathbb{R}^n$, the goal of the scaling criterion is to determine the dependence of the singular values on the radius ϵ . We define the (*local*) *SV curve* for the j 'th singular value as $\sigma_j(\epsilon)$ where ϵ is varied on the interval $(0, \epsilon_{\max}]$; the scaling of which are revealed by the solution of the eigenvector problem

$$\frac{1}{P_\epsilon} B_\epsilon(c_i)^T B_\epsilon(c_i) \phi_j(\epsilon) = \sigma_j^2(\epsilon) \phi_j(\epsilon)$$

on the interval $(0, \epsilon_{\max}]$ where $\{\sigma_j, j = 1, \dots, n\}$ are the n singular values for the matrix of data whose distance to the center c_i is less than ϵ , and P_ϵ is the number of points in the B_ϵ . The discrete local SV curves τ_i are then represented by the pairs

$$\tau_i = \{(\epsilon, \sigma_j(\epsilon)) : 0 \leq \epsilon \leq \epsilon_{\max}\}$$

Consider the case of data which is a sampling of a low dimensional manifold, U with dimension k . In the local analysis discussed above, the data matrix consists of rows corresponding to $(f^{(\mu)} - c_i)^T$, where μ is the index of the data, and c_i is the i 'th center. Therefore, for sufficiently small ϵ , the rows of $B_\epsilon(c_i)$ are approximately tangent vectors to U at c_i . As $\epsilon \rightarrow 0$, the rank of $B_\epsilon(c_i)$ is the dimension of U at c_i . As ϵ is increased, the first k singular values will scale linearly until saturation or the effects of curvature become noticeable. The remaining singular values will grow as ϵ^2 or faster.

Singular Value (SV) Curves Algorithm

Let $i = 1 : n_c$, where n_c is the number of clusters.

1. Let $j = n + 1 : P_i$, where P_i is the number of points in cluster i .
2. Denote by $B(1 : j, :)$ the matrix formed by taking the smallest j center-subtracted data points as the j rows.
3. Compute the singular values of $\frac{1}{j} B(1 : j, :)$.
4. The local dimension is the number of SV curves that scale linearly with the radius of B .

There are two critical observations to make at this point. The first is that the Fukunaga and Olsen's algorithm [7] is a special case of the SV Curves- While

the former takes an eigenvalue/eigenvector computation at a single point, the SV curves algorithm takes measurements along many points. While the former algorithm asks for the dominating eigenvalues at one point, the SV asks for the relative scaling of the eigenvalues (or more precisely, the singular values) taken at many points. Thus, if desired, one can make a detailed analysis of the structure of the surface.

One difficulty of this algorithm as originally stated is the last step, as the slope of the linear functions will be dependent on the density of the data. We will resolve this problem by putting the curves into a kind of canonical form. Intuitively, we will force every (non-zero) singular value of the ball of data, at maximum radius, to be unity. Numerically, this is effected via whitening the data [6]. As a reminder of this method, if X is a matrix of data, and $X = U\Sigma V^T$ is its (reduced) singular value decomposition, then the whitening transformation is given by

$$Y = \Sigma^{-1}U^T X$$

Note that this has the built-in effect of filtering the dimensions corresponding to very small singular values.

An important feature of the modified algorithm is the fact that all the singular values now have the value $\sigma_i = 1$ when the full local data set (corresponding to $\epsilon = \epsilon_{\max}$) is used. In what follows we take $R \equiv \epsilon_{\max}$. In view of the fact that R is the maximum distance from a point in X (assuming it has been mean subtracted) to the center and that at R the singular values all must be unity we may make the following observation:

Geometric Scaling Property: The slopes of all linear scaled singular values must be $\frac{1}{R}$, and the quadratic coefficient must be $\frac{1}{R^2}$.

In other words, for eigenvectors which span the tangent space, their associated local SV curves must have the form

$$(4.1) \quad \sigma_j(\epsilon) = \frac{\epsilon}{R}$$

Similarly, the region of lowest order curvature of the function has local SV curves of the form

$$(4.2) \quad \sigma_j(\epsilon) = \frac{\epsilon^2}{R^2}$$

This standardization greatly facilitates the interpretation of the local SV curves and has an important algorithmic consequence which we now discuss.

The normalized line with $y_1(x) = x/R$ and a quadratic $y_2(x) = x^2/R^2$ have greatest deviation at the midpoint $x = R/2$. Given $y_1(R/2) = 1/2$ and $y_2(R/2) = 1/4$ it follows that if $\sigma_i > 3/8$ the line model

is more likely while if $\sigma_i < 3/8$ the quadratic (or higher order) model is more likely. Thus, in determining which dimensions are scaling linearly, we need make only a single computation, at a distance of $R/2$ from center, where the distinction between linear and higher order scaling is the greatest. The decision is thus:

Geometric Scaling Dimensionality Criterion:

If $\sigma_i(\epsilon) > \frac{3}{8}$ where $\epsilon = R/2$ then eigenvector i belongs to the span of the tangent plane. The local dimension is the number singular values for which this condition is satisfied.

With this criteria, we can automate the process of dimension estimation when the data represents a well defined manifold, and we call this the *modified SV curve algorithm*. We will now extend the interpretation of the algorithm in the case that the data is not so well represented.

5 The Analytic SV Curves

Consider the following definition of the Singular Value curves:

Let $p^{(i)}(x)$ denote the probability density function of the projection of x to the i^{th} coordinate vector. The corresponding singular value curve, $\sigma^{(i)}(t)$ is defined as:

$$(5.3) \quad \sigma^{(i)}(t) = \sqrt{\frac{\int_{-t}^t x^2 p^{(i)}(x) dx}{\int_{-t}^t p^{(i)}(x) dx}}$$

where the coordinate vectors are taken as the KL eigenvectors. The fact that the two methods are equivalent can be seen by considering the continuous version of Karhunen-Lo  ve¹.

Armed with this definition, one may prove the following:

1. Let x be uniformly distributed on $(-a, a)$. If $y = x^n$, then the SV curves of y scale linearly with the radius, t .
2. Let x^n have a uniform distribution. Then the SV curves of x^n scale proportionally to t^n , where t is the radius. Note that this is the case where the data set is a manifold where the data has been uniformly distributed over the surface.

The added value of Definition 5.3 can be seen in the fact that one can now experiment with different distributions and look at the effect on the scalings. For example, one can consider the signature scalings when the projection

¹For more details on this relationship, see [10].

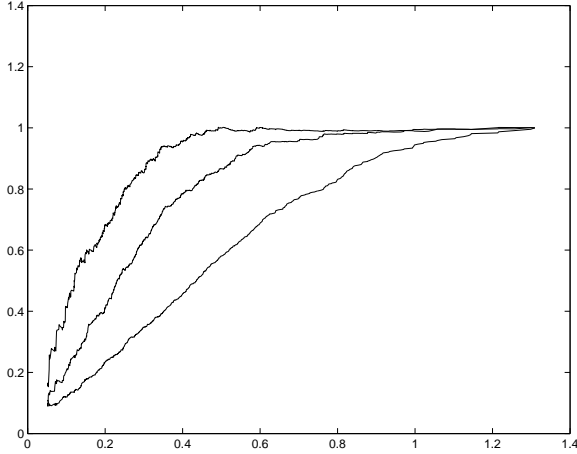


Figure 1: The (modified) SV curves for the multivariate normal data. Here we see a “signature scaling” that occurs for such data. The independent variable is the radius of the ball of data, the dependent variables are the singular values.

gives a normal distribution of data, which we will consider this further in the next section.

6 Examples

So far we have presented two extensions of the SV curves. Taken together, we can either perform a straight dimension estimation with possible automation, or do a detailed local analysis of the geometry of the surface in question. First, we will examine some template data sets in order to better interpret what we see in the more complicated cases. To begin, we will examine the case of Gaussian noise, then the surface of a sphere.

6.1 Gaussian Noise The data set in this example consisted of 1000 points in \mathbb{R}^3 , normally distributed with zero mean, and standard deviations of 0.4, 0.2, 0.1 in each coordinate direction respectively.

Figure 1 shows the modified SV curves (modified in the sense that the data has been whitened). Here we see the scaling that is typical of a normal distribution- the curves are scaling at a fractional power less than one, which is something we should not see on a noise-free manifold². We will keep these curves in mind as we will add normally distributed noise to a manifold below.

6.2 Surface of a Sphere Our second template example is given by data on the surface of a sphere, where

²In this case, a local Taylor expansion should locally well approximate the manifold, so that the scalings are in terms of positive integer powers

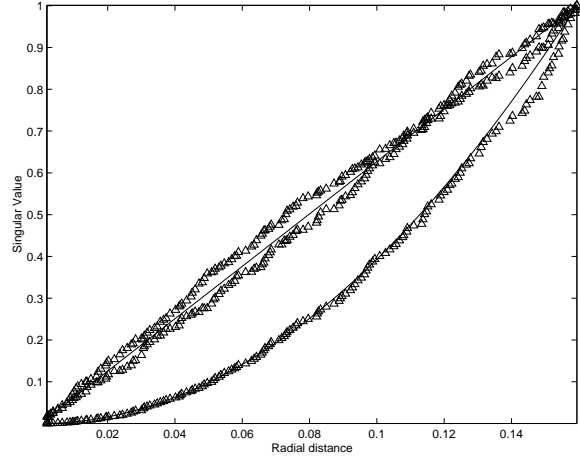


Figure 2: The modified SV curves on the unit sphere with non-uniformly distributed points. The predicted scalings are shown as solid lines, the computed values are shown as triangles. The horizontal axis is measuring the radial length of the ball of data, and the vertical axis is measuring the singular values. Indeed, this gives the expected result that the sphere is locally two dimensional, corresponding to the two linear scalings.

the data is centered at an arbitrary point on the surface.

If the points are distributed uniformly across the surface, then the dependence of the singular values as a function of the radial distance ϵ can be shown analytically to be:

$$\sigma_1(\epsilon) = \sigma_2(\epsilon) = \frac{1}{2}\epsilon + O(\epsilon^2)$$

$$\sigma_3(\epsilon) = \frac{1}{\sqrt{12}}\epsilon^2 + O(\epsilon^3)$$

In contrast, if the data is not uniformly distributed across the surface of the sphere, the slopes of the linear functions will change. This can make the determination of whether a function is scaling as a linear function of the radius very difficult. To obtain a canonical form for the linear/quadratic scalings, we whiten the data first.

In Figure 2, we show the result of computing the modified SV curves on the unit sphere with non-uniform data, and see that the predicted values (shown as solid lines, see Equations 4.1 and 4.2) are very close to the computed values (shown as triangles).

6.3 The Lorenz Attractor This example comes from an aged solution to the Lorenz equations,

$$\begin{aligned}\dot{x} &= -\frac{8}{3}x + yz \\ \dot{y} &= 10(y - z) \\ \dot{z} &= 28y - z - xy\end{aligned}$$

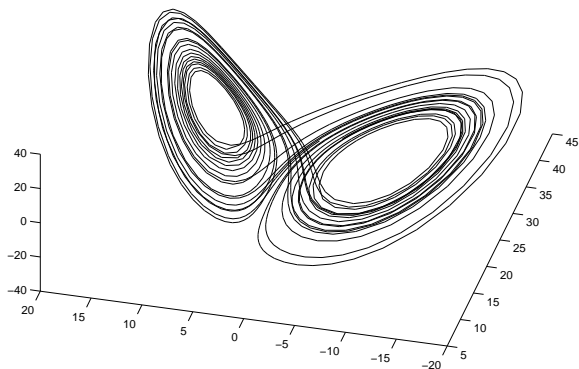


Figure 3: An aged solution to the Lorenz equations. Dimensionality estimation for this data can be difficult in the proximity of where the solutions seem to merge (in the center).

We chose this example because it is somewhat typical of a difficult problem in dimensionality estimation. The solution curve is shown in Figure 3, where we see that the solutions seem to merge (they do not, this is a fractal). However, there are many regions which can be embedded to \mathbb{R}^2 , while some cannot³.

The SV curves for a typical two dimensional cluster of data are given in Figure 4. Here we see the scalings predicted, even though the set is not a manifold. Figure 5 shows the result of adding normal noise (with standard deviation 0.1). The third SV curve is now showing the signature Gaussian scaling we discussed in Example 6.1. We show this for illustration purposes, since in practice this third singular value would have been filtered out.

We also show a typical set of SV curves for a local set that intersects the middle of the Lorenz data. Here, the curves initially show two linear scalings, but as the ball of data intersects more of the center, a third linear scaling appears.

6.4 Time Sequence of Images This example originally appeared in J. Bruske and G. Sommer [5]. The data consists of a movie in which a robot arm was revolving a cylinder about a fixed axis. Figure 7 shows a sample snapshot from the film. There were 360 grayscale snapshots (taken at each 1 degree progression), with a resolution of 256×256 pixels (so the dimension of the input space is 65,536). The noise is approximately Gaussian with a standard deviation of 1.75 gray values per pixel.

³Especially if one wishes to construct the inverse of the projection to \mathbb{R}^2 .

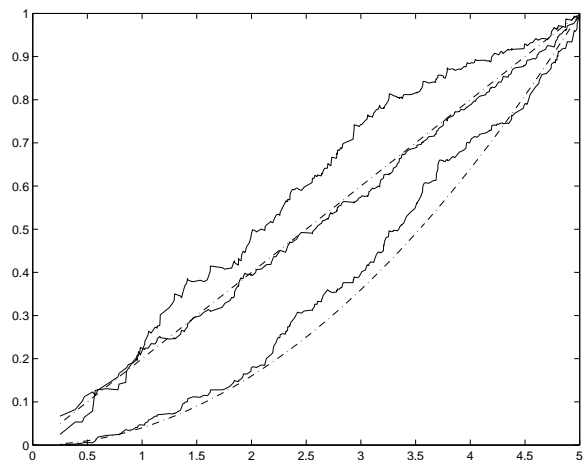


Figure 4: The SV curves for a typical two dimensional cluster from the Lorenz data. The dotted lines show the predicted linear and quadratic scalings for the modified SV curves (See Equations 4.1 and 4.2).

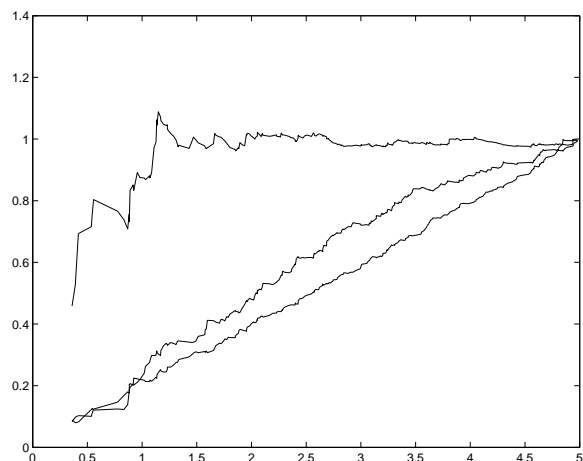


Figure 5: The same data as that shown in Figure 4, but with added normal noise (standard deviation 0.1). Here the third SV curve is shown with the signature Gaussian scaling. In the algorithm, this third curve would have been removed- it is shown here for illustrative purposes. From this picture, we might suspect that we have a local two dimensional set contaminated with noise.

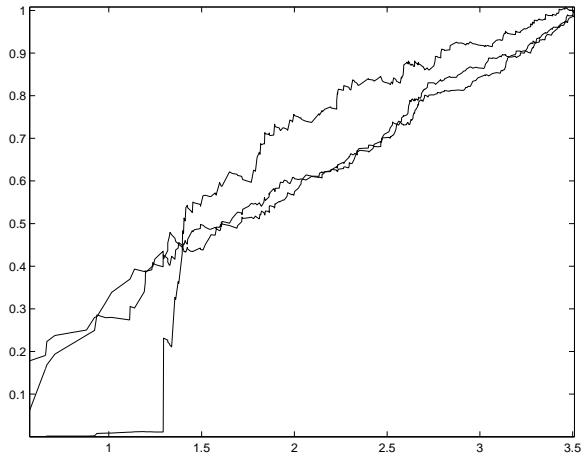


Figure 6: A typical set of SV curves for data that intersects the middle of the Lorenz set. Initially, we have the two linear scalings, but as the ball of data begins to encapsulate the center, three linear scalings appear.

We first performed a global reduction of the data using KL. The KL spectrum suggests that the data is three dimensional given three dimensions retain 97.5% of the total energy.

There are now two ways to proceed. In practice, we would perform the global reduction to three dimensions before constructing the SV curves, but for illustration purposes, we show the result of applying the SV algorithm on the full 65,536 dimensional space.

We took an arbitrary point to be the center of our cluster, then found the 50 closest points. Since we know that our dimension is no more than three, we began the SVD curve algorithm with only 4 points. A typical set of SV curves are shown in Figure 8. There is one primary curve that exhibits the linear scaling, and one curve that has the quadratic scaling. The remaining singular values are appearing as more and more data are included, and remain below the third singular value.

We now present the results of using the modified SV algorithm on the reduced dimensional set. This reduction is important, because the modified SV algorithm will be inverting the matrix of singular values. The result is shown in Figure 9, where again we see that the numerical scaling and the analytic scaling are matching quite well. From this we conclude that the topological dimension is one.

6.5 The Kuramoto-Sivashinsky (KS) Equation

Here is a typical example taken from a PDE simulation. The PDE under examination here is the Kuramoto-Sivashinsky equation, and is used to model turbulence

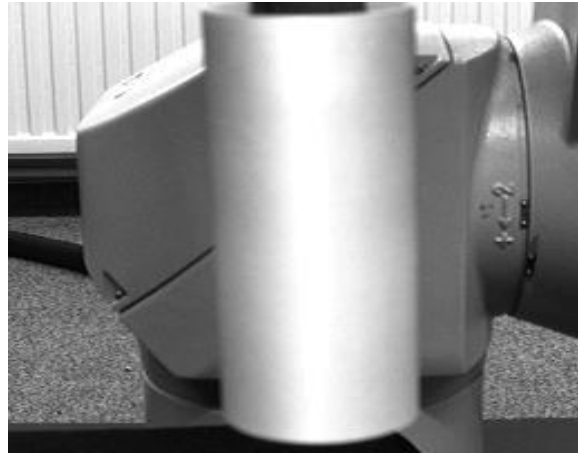


Figure 7: A sample snapshot from the movie of a cylinder turned by a robot arm.

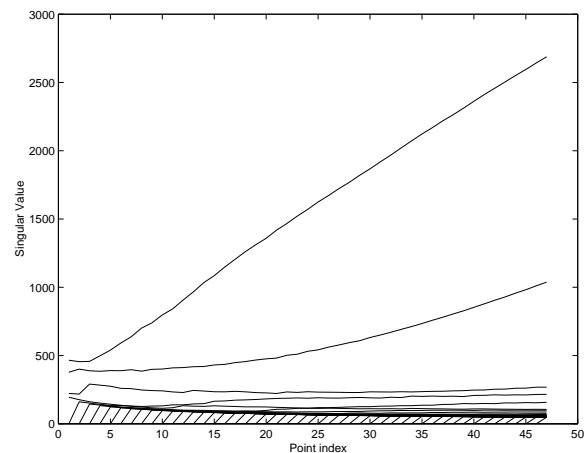


Figure 8: The singular value curves for an arbitrary center of the movie data. On the top are the original SV curves, where we see one linear scaling, one quadratic scaling, and the remaining singular values are remaining quite low.

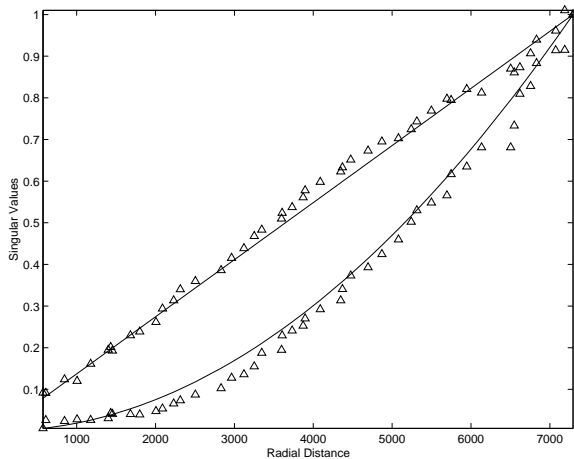


Figure 9: The result of the modified SV algorithm, performed on the reduced dimensional movie data. The local filtering has removed all but the first two singular values, and we can clearly see the linear vs. quadratic scaling.

in waves. Given a high dimensional simulation of the solution, our goal is to model the solution on the attracting set in low dimensional space. Thus, we must find the local dimensionality first. The KS Equation is given by:

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

with $\alpha = 87$. A simulation was performed using the Fourier-Galerkin method, retaining 10 complex modes (so the simulation was running in \mathbb{R}^{20}) [8, 9]. Global KL shows that the attracting set is a thin torus embedded in 20 dimensional space, and the data in the primary 6 dimensions are shown in Figure 10, along with the global (normalized) spectrum in Figure 11.

We used a clustering technique⁴ to place the centers for our local data sets on the surface of the torus, and proceeded to calculate the local dimension. The result of this technique is shown in Figure 12, where we see the placement of the clusters as solid dots (the curve represents the outline of the data). The next step involves computing the modified SV curves, and the result is shown in Figure 13, where we see the signature linear scalings of the first two SV curves, confirming our initial analysis and allowing us to proceed to the local modeling step. The full analysis and further examples are available in [8].

⁴The technique took into account both spatial location and velocity. More details can be found in [8]

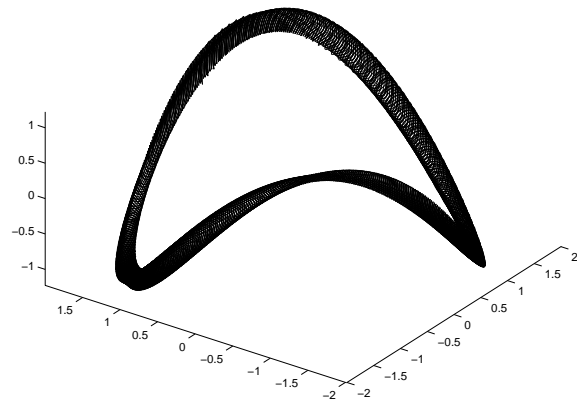


Figure 10: The aged solution to the KS Equation, $\alpha = 87$. The solution is shown in the primary 3 dimensional coordinate system. Note the appearance of a “thin torus”.

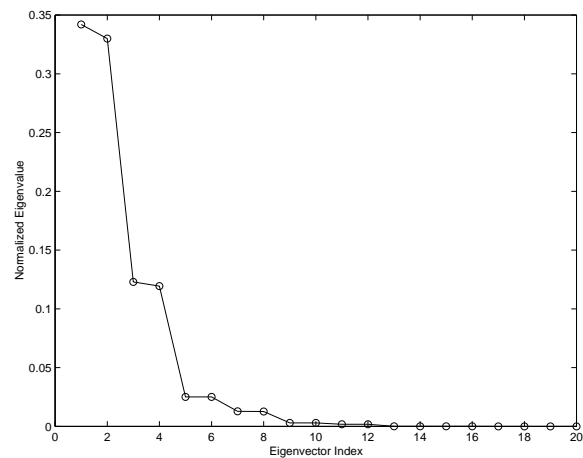


Figure 11: The normalized spectrum of the global covariance matrix for the aged solution to the KS equation. The values suggest that the torus seen in the previous figure has a high (global) dimension.

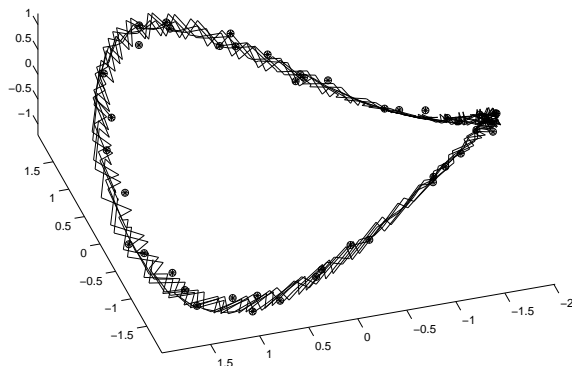


Figure 12: The locations of the cluster centers (in the KL coordinate system) for the solution to the KS Equation. The centers all lie on the surface, and are shown as filled dots, while an outline of the original data is drawn by the solid line.

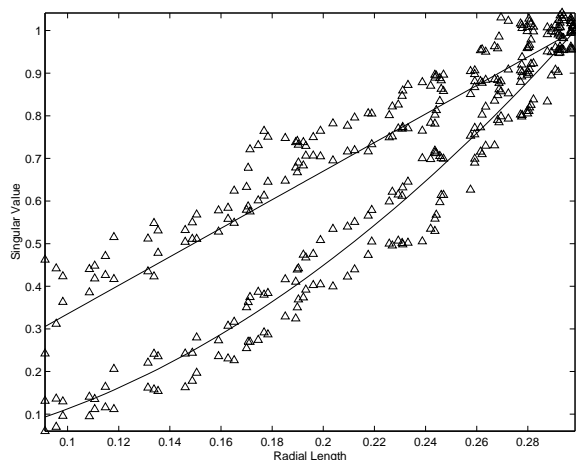


Figure 13: The result of the modified SV algorithm on the aged solutions to the KS Equation. Numerical values are shown with triangles, and the analytic curves are shown as solid lines. Here we clearly see the two dimensional structure from the two linear scalings.

7 Summary

In this paper, we have presented two extensions to the local SVD technique for determining the topological dimension of data representing a smooth manifold with low noise. The results provide for the option of automating the choice of dimension based on the scalings of the SV curves, or a detailed analysis of the surface structure. Indeed, such an analysis can be used to detect whether models based on manifolds are appropriate.

Our algorithm is based on the Taylor approximation to a surface and hence is ideally suited for manifolds. As such, it will not perform well for data that does not have a reasonable local linear approximation such as data in clouds of noise; for such problems methods based purely on the statistics of the data should be used, see, e.g., [13, 11].

Our primary interest is in determining local coordinate systems for building dynamical models using *neural charts* [8] or the Whitney Reduction Network [4]. Thus local dimensionality estimation techniques which also provide local coordinates systems are especially attractive. Given our models sit in high-dimensional spaces and associated massive data sets have been accumulated only automated procedures are practical.

Acknowledgements:

The authors would like to thank Dave Broomhead, Greg King and Mark Muldoon for useful discussions concerning the development of the preceding work, and to Gerald Sommer for graciously providing his data.

This work has been partially supported by NSF-DMS 9505863, NSF-INT 9513880.

References

- [1] David Banks and Robert Olszewski. Estimating local dimensionality. In *Proceedings of the Statistical Computing Section of the American Statistical Society*. ASA, 1997.
- [2] D. S. Broomhead, R. Jones, and G. P. King. Topological dimension and local coordinates from time series data. *J. Phys. A: Math. Gen.*, 20:L563–L569, 1987.
- [3] D.S. Broomhead, R. Indik, A.C. Newell, and D.A. Rand. Local adaptive Galerkin bases for large-dimensional dynamical systems. *Nonlinearity*, 4:159–197, 1991.
- [4] D.S. Broomhead and M.J. Kirby. The whitney reduction network: A method for computing autoassociative graphs. *Neural Computation*, 13:2595–2616, 2001.
- [5] J. Bruske and G. Summer. Intrinsic dimensionality estimation with optimally topology preserving maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5):572–575, 1998.

- [6] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Boston, MA, 1990.
- [7] K. Fukunaga and D.R. Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on Computers*, C-20(2):176–183, 1971.
- [8] Douglas Robert Hundley. *Local Nonlinear Modeling via Neural Charts*. PhD dissertation, Colorado State University, Department of Mathematics, June-August 1998.
- [9] M. Kirby. Minimal dynamical systems from partial differential equations using sobolev eigenfunctions. *Physica D*, 57:466–475, 1992.
- [10] Michael Kirby. *Geometric Data Analysis: An empirical approach to dimensionality reduction and the study of patterns*. John Wiley and Son, New York, 2001.
- [11] Thomas P. Minka. Automatic choice of dimensionality for pca. In Todd K. Leen, Thomas G. Dietterich, and Volker Tresp, editors, *Advances in Neural Information Processing Systems 13*, pages 598–604. MIT Press, 2001.
- [12] K.W. Pettis, T.A. Bailey, A.K. Jain, and R.C. Dubes. An intrinsic dimensionality estimator from near-neighbor information. *IEEE transactions on pattern analysis and machine intelligence*, PAMI-1:25–37, 1979.
- [13] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *J. Royal Statistical Society B*, 61(3), 1999.
- [14] Peter J. Verveer and Robert P.W. Duin. An evaluation of intrinsic dimensionality estimators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):81–86, 1995.