

Time-Scale Separation from Diffusion-Mapped Delay Coordinates*

T. Berry[†], J. R. Cressman[‡], Z. Gregurić-Ferenček[‡], and T. Sauer[†]

Abstract. It has long been known that the method of time-delay embedding can be used to reconstruct nonlinear dynamics from time series data. A less-appreciated fact is that the induced geometry of time-delay coordinates increasingly biases the reconstruction toward the stable directions as delays are added. This bias can be exploited, using the diffusion maps approach to dimension reduction, to extract dynamics on desired time scales from high-dimensional observed data. We demonstrate the technique on a wide range of examples, including data generated by a model of meandering spiral waves and video recordings of a liquid-crystal experiment.

Key words. dimension reduction, time-scale separation, delay coordinates, diffusion map

AMS subject classifications. 37M10, 37L30, 62G07

DOI. 10.1137/12088183X

1. Introduction. The method of time-delay embedding was first introduced by Takens, Ruelle, and others for the purpose of reconstructing dynamical attractors from data [32, 23, 1, 14]. With a series of generic delayed observations, it was shown that topological properties are preserved in reconstruction dimensions greater than $2d$, for a smooth attractor of dimension d [32], and for an attractor of fractal dimension d [26]. Considerable effort followed to develop methods of estimating attractor dimension, in order to carry out embeddings in a Euclidean space of minimal dimension.

More recently, the development of pervasive and cheap sensors has caused a shift in emphasis toward methods capable of handling large quantities of time-ordered data. For example, we could imagine a complex system with a relatively low-dimensional attractor, possibly chaotic, where the observations are represented by a high-dimensional multivariate time series. In such a case, it is of great interest to apply a data analysis technique to a video of an experiment, for instance, and to seek ways to selectively project that data onto various dynamical time scales of interest.

Snapshots of spatiotemporal patterns produced by electroconvection in a liquid crystal are shown in Figure 1. Simulations of complex spatiotemporal models, such as Rayleigh–Bénard convection, indicate that there may be a low-dimensional representation of the process [15]. However, conventional techniques of dimensionality reduction such as the Karhunen–Loeve decomposition have been unable to recover a low-dimensional process even for low driving

*Received by the editors June 22, 2012; accepted for publication (in revised form) by B. Sandstede February 20, 2013; published electronically April 30, 2013. This research was supported by NSF grants EFRI-1024713 and DMS-0811096.

<http://www.siam.org/journals/siads/12-2/88183.html>

[†]Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030 (tberry@gmu.edu, tsauer@gmu.edu).

[‡]School of Physics, Astronomy, and Computational Sciences and The Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA 22030 (jcressman@gmu.edu, zrvongr@gmail.com).

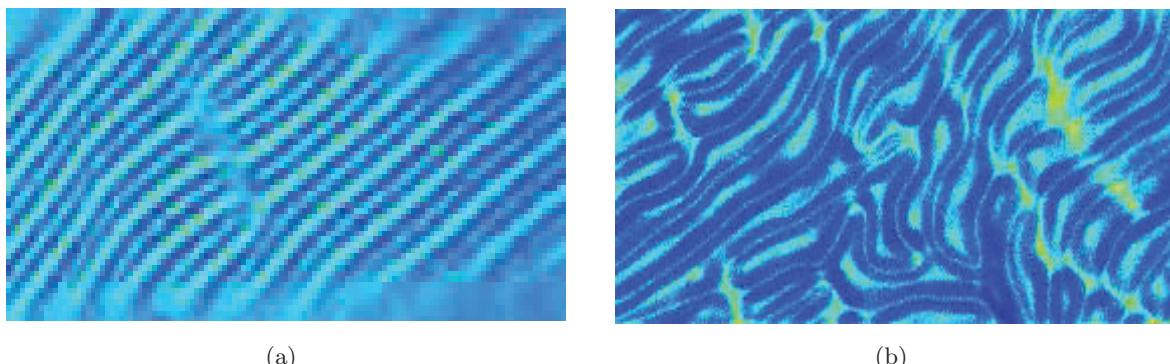


Figure 1. Electroconvection patterns display a variety of complex spatiotemporal behavior. Enlarged images from nematic liquid crustal experiment showing defects at (a) low voltage and (b) high voltage.

levels [15, 11].

Dealing with multivariate recordings in high dimensions requires some form of dimensionality reduction. In 1985, Broomhead and King [5] combined the ideas of time-delay embedding and singular value decomposition (SVD) to project high-dimensional reconstructions to lower dimension for analysis. This idea is in the tradition of Karhunen–Loeve decompositions of dynamical systems. Local versions of the same idea were used later to develop algorithms for time series prediction [25] and noise reduction. One purpose of this article is to provide a theoretical explanation for the power of time delays when combined with dimensionality reduction, as demonstrated in the following simplistic scenario.

Example 1.1. Consider the task of analyzing a video which contains a single black pixel which moves periodically in a circle, surrounded by grayscale static. The input to our method is the video, a set of time-ordered frames, each a two-dimensional array of pixel values. The deterministic dynamics of the underlying dynamics can be represented by a single variable: the angular position of the black pixel. However, such a representation of our data would contain only a tiny fraction of the variance of the data set (which is dominated by noise), and the reconstruction error would be large. Directly applying a standard dimensionality reduction technique like the SVD, which is greedy for variance, would ignore the black dot and instead focus on learning the moments and correlations in the noise. On the other hand, if delay coordinates are used, followed by a projection by SVD into the dominant directions, the desired dynamics are isolated from the noise. In Figure 2, taken from the associated video (88183_01.mov [[local](#)/[web](#) 3.33MB]), we show the results of applying the SVD to the images with and without delays. This example shows a key feature of delay coordinates: they project the data onto the most stable dynamical variables. We will describe this feature more quantitatively in section 2.

Although useful in Example 1.1, a severe limitation of SVD is that it is a linear projection. Generic linear projections of complicated sets retain their topological characteristics if the image dimension is large enough, but fine details of the geometry are in general lost. More refined dimensionality reduction techniques do less damage to the geometry. In particular, diffusion maps, developed by Coifman and his collaborators [7, 9, 6, 8], are able to partially

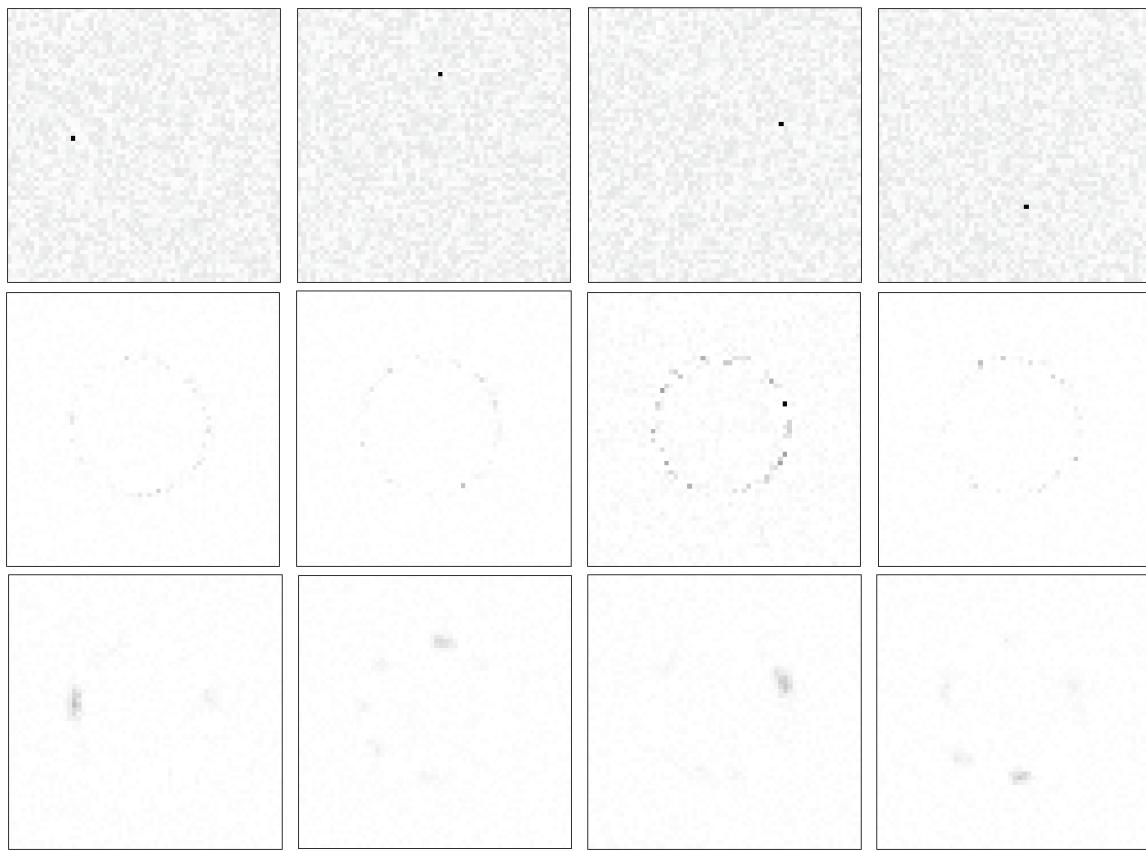


Figure 2. Top row: Four sample images from a video of the black pixel moving on a noisy circle. Middle row: Reconstruction of the four sample images using the first 32 SVD modes. Bottom row: Reconstruction with SVD applied to time-delay embedding, using the first 32 SVD modes. The snapshots are taken from the associated video (88183_01.mov [local/web 3.33MB]).

retain geometric properties of attractor manifolds under projection from high dimensions. In a second example, we explore replacing the SVD projection by a diffusion map.

Example 1.2. In this example we analyze a video with two periodically moving artificial objects, inspired by the liquid crystal data in Figure 1. The largest variance component of the video is made up of white stripes that move diagonally across the image, as shown in Figure 3(a). A smaller variance component is a localized “defect,” which moves periodically on a circle. As in Example 1.1, we consider the input “data” to be the unprocessed video itself. In Figure 3, we consider two versions of the video, where the stripes move slower (Figure 3(b)) or faster (Figure 3(c)) than the defect. The projection using SVD with time delays (green trace) picks out the higher variance component (the stripes) in both versions. On the other hand, pairing the delay-coordinate embedding with a diffusion map projection (black trace) projects to the slower mode in both versions: the stripes in Figure 3(b) and the defect in Figure 3(c). See the associated videos (88183_02.mov [local/web 5.03MB] and 88183_03.mov [local/web 3.29MB]) for a realtime view of the compared projections in Figures 3(b) and 3(c),

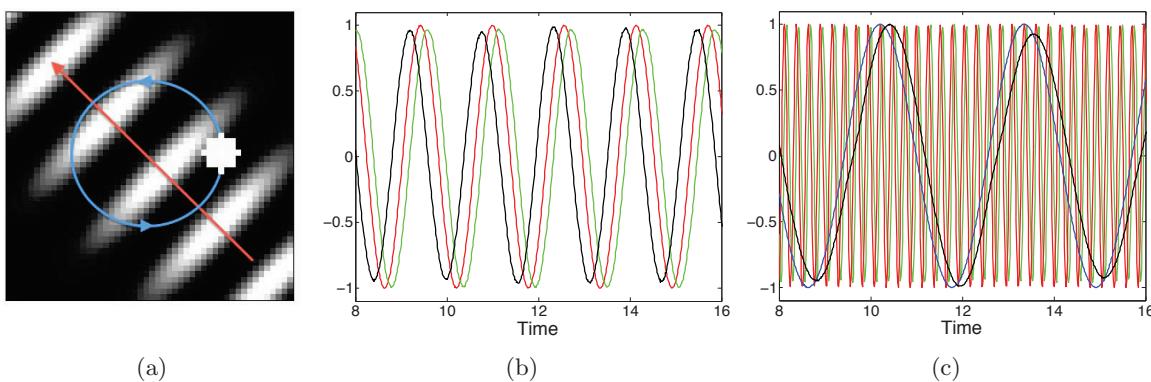


Figure 3. (a) A frame from the video with arrows added to indicate the dynamics; the stripes follow the red arrow, and the defect moves along the blue circle. (b) When the stripes move more slowly than the defect, both SVD (green trace) and DMDC (black trace) reconstruct the phase of the slow moving stripes (red). (c) When the stripes move more quickly, DMDC (black) finds the phase of the defect (blue) on its circular trajectory; however, SVD (green) still finds the phase of the stripes because they have higher variance. The associated videos (88183_02.mov [local/web 5.03MB] and 88183_03.mov [local/web 3.29MB]) demonstrate the projection onto the slow dynamics of scenarios (b) and (c), respectively.

respectively.

The goal of this article is to explain the mechanisms behind the above examples, namely (1) the contribution of delay coordinates to reconstructing intrinsic attractor geometry and (2) the use of diffusion maps to control the metric properties of the reconstructed attractor. We will use this knowledge to build a flexible computational framework called diffusion-mapped delay coordinates (DMDC) that can be used to analyze high-dimensional data such as videos for low-dimensional dynamics. In particular, we are interested in separating time scales in the data, by projecting onto the stablest dynamical directions (using weighted delay coordinates) and further slicing the dynamics there into Fourier-like components (using diffusion maps).

Nonlinear dimensionality reduction has been successfully applied to very high-dimensional observations such as images [33, 21]. However, these methods are typically designed for data that have no dynamics. Dynamical data, meaning data with a time ordering, is a unique problem with different goals than generic dimensionality reduction. In our view, the key to success of nonlinear dimensionality reduction techniques is the imposition of appropriate assumptions on the structure of the data. Whereas the goal of generic dimensionality reduction is typically to minimize the reconstruction error, or to maximize the variance of the low-dimensional representation, such a representation *may not contain dynamically interesting information*.

To adapt nonlinear dimensionality reduction for use on data measured from dynamical systems, we find it important to separate the extrinsic features of the data, such as the observation and embedding space, from the intrinsic dynamical features, such as Lyapunov exponents and associated invariant manifolds. These intrinsic dynamical features are independent of the observation space and can be represented by an intrinsic geometry on the state space, given by the Lyapunov metric, which is a Riemannian metric that characterizes the Oseledets splitting [22]. In section 2, we show that the classical technique of time-delay embeddings, with

appropriate weights, effectively recovers the Lyapunov metric in the most stable Lyapunov direction.

Within the most stable direction there may still be multiple time scales. A periodic or quasi-periodic trajectory such as Example 1.2 may have a zero Lyapunov exponent along the trajectory that harbors a fast mode and a slower mode. For other examples, see sections 4 and 5 below. We separate time scales within the most stable Lyapunov direction using the reconstructed eigenfunctions of the Laplace operator with respect to the Lyapunov metric. The Laplace operator is approximated by a diffusion map [7], as we discuss in section 3.

Previous work on time-scale separation from dynamical data includes the work [8] on stochastic systems governed by Brownian motion in a potential field. Ideally, time-scale separation finds a hierarchy of slow variables which are independent of the higher frequency oscillations of the dynamics. For example, the evolution of a dynamical system is independent of the fast oscillations of observation noise, as in Example 1.1. Moreover, if the system contains a slow manifold, then the evolution on the slow manifold will be independent of the faster dynamics. Time-scale separation involves finding the variables that are governed by these slow dynamics and ordering the variables according to the relevant time scale. Our goals of projecting onto slow dynamics have some overlap to those of [8, 17, 28, 35], although we assume no knowledge of the equations generating the data, or availability of surrogates such as microscopic models or legacy solvers. Furthermore, in analyzing video data of experiments, we are reduced to assuming that the observations are related to the true system state in an unknown way.

In section 2 below, we show that for data which is not necessarily observed in its latent state space, the standard technique of time-delay embeddings will partially recover the Lyapunov metric. This improves significantly on the results of Takens, which show that time-delay coordinates reconstruct the topology of the latent state space. This will allow the diffusion map to be applied to generic observations of a dynamical system.

In section 3 we explain how diffusion maps can be used for dimensionality reduction and for time-series analysis. As a method for dimensionality reduction, we show how a diffusion map can be interpreted as a kernel principal component analysis, resulting in an optimal low-dimensional embedding of a data set with respect to a conformal transformation of the geometry. In fact, the diffusion map constructs an operator \mathcal{L} which is the Laplacian operator with respect to the geometry introduced by the conformal transformation. Moreover, when the data has a time ordering, the diffusion map coordinates can also be interpreted as time series given by the projection of the dynamics onto eigenfunctions of \mathcal{L} . We show that for a dynamical system with an invariant measure, the conformal transformation will adapt the operator \mathcal{L} to this invariant measure.

In section 4 we show how projecting the dynamics onto the eigenfunctions of the operator \mathcal{L} can lead to time-scale separation. If the dynamics are elliptic when restricted to the most stable Lyapunov direction, then the operator \mathcal{L} can be written as a Laplacian plus lower order terms (with respect to a special Riemannian metric which depends on \mathcal{L}). In many cases the most stable Lyapunov direction will be one-dimensional, and the Laplacian will be unique up to a choice of measure. Then the eigenfunctions will simply be spatial sine and cosine functions which are rescaled to the invariant measure on the stable manifold. Projecting the dynamics onto this adapted Fourier basis on the manifold will regularize the lower order terms,

which is analogous to the smoothing of a low-pass filter. We interpret the projection onto each eigenfunction as a solution of a reduced dynamical system with a specific time scale which is related to the corresponding eigenvalue.

In section 5 we implement the proposed new technique and apply it to the high-dimensional observations of spatiotemporal dynamics. We apply DMDC to a computation model of a meandering spiral wave and also to video data of spatiotemporal patterns produced by electroconvection in a liquid crystal. We find that even for these complex dynamics, the method produces a time-scale separation, in addition to producing decorrelated latent variables and achieving significant noise reduction. We also include video demonstration of the decomposition of the spatiotemporal dynamics into separate time scales.

The appendix to this article includes a self-contained description of the DMDC algorithm and implementation details.

2. Geometry of time-delay coordinates. The theory of time-delay coordinates as introduced by Takens [32] shows that, by appending delayed values of a generic observation of a dynamical system, one achieves a diffeomorphic copy of the attractor in some Euclidean space. The fact that the embedding is given by a diffeomorphism of the attractor shows that the time-delay embedding is topology-preserving, although crucially it introduces a new geometry to the data set. In this section, we will see that the new geometry is related to the Lyapunov metric [24] restricted to the most stable Oseledets subspace. In the past, the increased dimension of the ambient Euclidean space was a limitation for time-delay embeddings, and significant effort was made to look for parsimonious time-delay embeddings using as few delays as possible. However, as we will see in section 3, diffusion maps are not adversely affected by the ambient dimension, which will allow us to take hundreds or even thousands of time delays and computationally study the effect of the delay embedding on the geometry of the attractor.

The context for this section will be the natural setting of the multiplicative ergodic theorem (MET) on a smooth manifold. Let \mathcal{M} be an n -dimensional smooth compact Riemannian manifold which is the attractor of a system denoted $\dot{x} = f(x)$, with invariant measure μ for the induced flow F_t . To accommodate discrete observations of the dynamics we will consider the flow F_τ for a fixed time step $\tau > 0$. According to Oseledets [22], there exist real numbers $\sigma_1 < \dots < \sigma_k$, with $k \leq n$, such that for μ -almost every x there is a splitting $T_x\mathcal{M} = \bigoplus_{i=1}^k E_i(x)$, where $\dim E_i = d_i$ and where $d_1 + \dots + d_k = n$. Each Oseledets space $E_i(x)$ is invariant under the dynamics, meaning that any nonzero vector $u_i \in E_i(x)$ has image $DF_{-j\tau}(x)u_i \in E_i(F_{-j\tau}(x))$. Moreover, for any $u_{i,x} \in E_i(x)$,

$$\lim_{j \rightarrow \infty} \frac{1}{j} \ln \|DF_{j\tau}(x)u_i\| = \sigma_i,$$

$$\lim_{j \rightarrow -\infty} \frac{1}{j} \ln \|DF_{j\tau}(x)u_i\| = -\sigma_i.$$

Assume a multivariate observation of dimension r , given by a smooth nonlinear $h \in C^\infty(\mathcal{M}, \mathbb{R}^r)$. For $\kappa, \tau > 0$ define the κ -weighted *delay coordinate map* $H : \mathcal{M} \rightarrow \mathbb{R}^{r(s+1)}$ by

$$(2.1) \quad H(x) = [h(x), e^{-\kappa}h(F_{-\tau}(x)), e^{-2\kappa}h(F_{-2\tau}(x)), \dots, e^{-s\kappa}h(F_{-s\tau}(x))]^T.$$

Under weak technical assumptions [26], by choosing $r(s+1) \geq n$, the delay coordinate map is an immersion for a prevalent choice of h . Additionally, for $r(s+1) > 2n$, the delay coordinate map H is an embedding of the manifold \mathcal{M} into $\mathbb{R}^{r(s+1)}$. There is no reason to assume that this embedding preserves the metric on the manifold, and thus geometry of the embedded manifold $H(\mathcal{M})$ is an extrinsic factor from our point of view.

For $\epsilon > 0$, the ϵ -Lyapunov metric $\langle u, v \rangle_\epsilon$ is defined by

$$\langle u_i, v_i \rangle_\epsilon = \sum_{j \in \mathbb{Z}} e^{-2(\sigma_i j + \epsilon |j|)} \langle DF_{j\tau}(x)u_i, DF_{j\tau}(x)v_i \rangle_{T_x \mathcal{M}}$$

for $u_i, v_i \in E_i(x)$. The Lyapunov metric is intrinsic to the dynamics because, when measured in this metric, the dynamics satisfy the uniform bounds

$$e^{-(\sigma_i + \epsilon)j} \|u_i\|_\epsilon \leq \|DF_{-j\tau}(x)u_i\|_\epsilon \leq e^{-(\sigma_i - \epsilon)j} \|u_i\|_\epsilon.$$

Thus we will consider the Riemannian metric of interest on \mathcal{M} to be the Lyapunov metric.

We now return to the extrinsic geometry and investigate the metric induced on the embedded manifold $H(\mathcal{M})$. Let $u = u_1 + \dots + u_k, v = v_1 + \dots + v_k \in T_x \mathcal{M}$, where $u_i, v_i \in E_i(x)$, and denote $\hat{u} = DH(u), \hat{v} = DH(v)$ in $T_{H(x)} H(\mathcal{M})$. Then the inner product $\langle \cdot, \cdot \rangle$ in the Euclidean reconstruction space $R^{r(s+1)}$ is

$$\begin{aligned} \langle \hat{u}, \hat{v} \rangle &= \sum_{j=0}^s e^{-2j\kappa} \langle Dh(F_{-j\tau}(x))DF_{-j\tau}(x)u, Dh(F_{-j\tau}(x))DF_{-j\tau}(x)v \rangle_{\mathbb{R}^r} \\ &= \sum_{i=1}^k \sum_{j=0}^s e^{-2j\kappa} \langle Dh(F_{-j\tau}(x))DF_{-j\tau}(x)u_i, Dh(F_{-j\tau}(x))DF_{-j\tau}(x)v_i \rangle_{\mathbb{R}^r}. \end{aligned}$$

The main theorem of this article shows that, with the right choice of κ , the metric $\langle \cdot, \cdot \rangle$ in the embedding space projects onto the most stable Oseledets subspace E_1 .

Theorem 2.1. *Let \mathcal{M} be a compact manifold, $u, v \in T_x \mathcal{M}$, and let $\hat{u} = DH(u)$ and $\hat{v} = DH(v)$ be the images under the time-delay embedding H in (2.1). Let $u_i = \pi_i(u)$ be the projection onto the i th Oseledets space, and assume that u_1 and v_1 are nonzero. Let $0 < \kappa < -\sigma_1$. Then for a prevalent choice of h and for all $i \neq 1$,*

$$\lim_{s \rightarrow \infty} \frac{\langle \hat{u}_i, \hat{v}_i \rangle}{\|\hat{u}\| \|\hat{v}\|} = 0$$

and therefore

$$\lim_{s \rightarrow \infty} \frac{\langle \hat{u}, \hat{v} \rangle - \langle \hat{u}_1, \hat{v}_1 \rangle}{\|\hat{u}\| \|\hat{v}\|} = 0.$$

Proof. First note that

$$\begin{aligned} |\langle \hat{u}_i, \hat{v}_i \rangle| &= \left| \sum_{j=0}^s e^{-2j\kappa} \langle Dh(F_{-j\tau}(x))DF_{-j\tau}(x)u_i, Dh(F_{-j\tau}(x))DF_{-j\tau}(x)v_i \rangle_{\mathbb{R}^r} \right| \\ &\leq \sum_{j=0}^s e^{-2j\kappa} \|Dh(F_{-j\tau}(x))DF_{-j\tau}(x)u_i\|_{\mathbb{R}^r} \|Dh(F_{-j\tau}(x))DF_{-j\tau}(x)v_i\|_{\mathbb{R}^r} \\ &\leq h_{\max}^2 \sum_{j=0}^s e^{-2j\kappa} \|DF_{-j\tau}(x)u_i\|_{\epsilon} \|DF_{-j\tau}(x)v_i\|_{\epsilon} \\ &\leq h_{\max}^2 \|u_i\|_{\epsilon} \|v_i\|_{\epsilon} \sum_{j=0}^s e^{-2j(\sigma_i + \kappa - \epsilon)} \end{aligned}$$

for all $\epsilon > 0$, where h_{\max} is the maximum of the matrix norm $\|Dh\|$ over the compact manifold \mathcal{M} . This bounds the growth rate in all Oseledets subspaces.

Next we need to show that the component of \hat{u} in the most stable direction dominates as s increases. Thus we will bound $\|\hat{u}\|$ from below, and we will focus on the component u_1 , which is assumed to be nonzero. Choose $k \geq 0$ such that $r(k+1) \geq n$; then the delay coordinate map is an immersion for a prevalent choice of h [26]. From this we need only the fact that for all $x \in \mathcal{M}$ and all $j \geq 0$ the rank of the $r(k+1) \times n$ matrix

$$A_j(x) = \begin{bmatrix} e^{-\kappa j} Dh(F_{-j\tau}(x))DF_{-j\tau}(x) \\ \vdots \\ e^{-\kappa(k+j)} Dh(F_{-(k+j)\tau}(x))DF_{-(k+j)\tau}(x) \end{bmatrix}$$

is n , implying that the kernel of the matrix is zero. For any nonzero vector $u_1 \in E_1(x)$ the vector $A_j(x)u_1$ is nonzero, and for some $j \leq l \leq j+k$ the r -vector

$$e^{-\kappa l} Dh(F_{-l\tau}(x))DF_{-l\tau}(x)u_1 \neq 0.$$

Thus for each $x \in \mathcal{M}$ we have $\max_{j \leq l \leq j+k} \|Dh(F_{-l\tau}(x))e_1\|_{\mathbb{R}^r} > 0$ for any unit vector $e_1 \in E_1(x)$. Since \mathcal{M} is compact we can define

$$\begin{aligned} h_{\min} &\equiv \min_{x \in \mathcal{M}} \min_{j \geq 0} \max_{j \leq l \leq j+k} \|Dh(F_{-l\tau}(x))e_1\|_{\mathbb{R}^r} \\ &= \min_{x \in \mathcal{M}, j \geq 0, y = F_{-j\tau}(x)} \max_{0 \leq l \leq k} \|Dh(F_{-l\tau}(y))e_1\|_{\mathbb{R}^r} \\ &\geq \min_{x \in \mathcal{M}} \max_{0 \leq l \leq k} \|Dh(F_{-l\tau}(x))e_1\|_{\mathbb{R}^r} > 0. \end{aligned}$$

This allows us to establish the lower bound

$$\|A_j(x)u_1\|_{\mathbb{R}^{r(k+1)}} \geq h_{\min} e^{-j(\kappa + \sigma_1 + \epsilon)} \|u_1\|_{\epsilon}$$

for all $x \in \mathcal{M}$. We can use this to obtain a lower bound on $\|\hat{u}\|$ by splitting the s terms into $\lfloor s/k \rfloor$ (the greatest integer less than s/k) blocks of size k so that

$$\begin{aligned} \|\hat{u}\|^2 &\geq \sum_{j=0}^s e^{-2j\kappa} \|Dh(F_{-j\tau}(x))DF_{-j\tau}(x)u_1\|_{\mathbb{R}^r}^2 \\ &\geq \sum_{l=0}^{\lfloor s/k \rfloor} \|A_{lk}(x)u_1\|^2 \\ &\geq h_{\min}^2 \|u_1\|_{\epsilon}^2 e^{-2k\lfloor s/k \rfloor(\sigma_1+\kappa+\epsilon)} \\ &\geq h_{\min}^2 \|u_1\|_{\epsilon}^2 e^{-2s(\sigma_1+\kappa+\epsilon)}. \end{aligned}$$

By combining the upper and lower bounds, we have

$$\frac{|\langle \hat{u}_i, \hat{v}_i \rangle|}{\|\hat{u}\| \cdot \|\hat{v}\|} \leq \left(\frac{h_{\max}^2}{h_{\min}^2} \cdot \frac{\|u_i\|_{\epsilon}}{\|u_1\|_{\epsilon}} \cdot \frac{\|v_i\|_{\epsilon}}{\|v_1\|_{\epsilon}} \right) \frac{|1 - e^{-2(s+1)(\sigma_i+\kappa-\epsilon)}|}{e^{-2s(\sigma_1+\kappa+\epsilon)}} \rightarrow 0$$

as $s \rightarrow \infty$ for $i \neq 1$ since by hypothesis $\sigma_1 + \kappa < 0$ and $\sigma_1 < \sigma_i$. To find the second limit note that $\hat{u} = \hat{u}_1 \oplus \hat{u}_1^\perp$, and by writing $\hat{u}_1^\perp = \sum_{i \neq 1} \tilde{u}_i$ with $\tilde{u}_i \in E_i$, we can apply the first limit on each component. ■

For large s we have $\langle \hat{u}, \hat{v} \rangle \approx \langle \hat{u}_1, \hat{v}_1 \rangle$, so the metric in the embedding space is negligible in all but the most stable Lyapunov direction. The proof fails when the hypothesis $0 < \kappa < -\sigma_1$ is not satisfied. When $\kappa \leq 0$, we cannot bound the matrix norm of $A(x)$, and in fact the construction does not yield a well-defined metric. When $\kappa \geq -\sigma_1$, the norm converges to a finite value in each Oseledets component, destroying the projection onto the stable component.

The constants h_{\min}, h_{\max} allow for local deviations from the long term behavior of the dynamical system, which is governed by the Lyapunov exponents. These constants are an extrinsic feature of the observed dynamics in the sense that they are accidental aspects of the observation function h . Of course, the dynamics would be uniform if we could reconstruct the intrinsic Lyapunov metric. Unfortunately, delays alone cannot reconstruct the Lyapunov metric, because we have shown that delay coordinates project the tangent spaces onto the most stable Oseledets subspace. The proof also illuminates the natural trade-off in the choice of κ . For κ near zero, the reconstruction projects strongly onto the stable component, but the dynamics may not be regular on this component. As we increase κ , the dynamics are increasingly regularized on the stable projection until we get to close to $-\sigma_1$, at which point the projection fails, according to Theorem 2.1. See Example 2.1 below and the associated videos (88183_04.mov [local/web 1.21MB] and 88183_05.mov [local/web 1.04MB]) for an illustration of this trade-off.

If the dynamics are reversible, the least stable Oseledets space in forward time is the most stable in reverse time. For a time series we can easily reverse time by inverting the time ordering of the data. Thus for $0 < \kappa < \max_i \{\sigma_i\}$ the algorithm can be applied to the reversed time series to project onto the least stable Oseledets space. This is equivalent to taking sequels instead of delays in the original time series. The bottom row in Figure 4 demonstrates the projection onto the least stable Oseledets space for the cat map.

This new interpretation of weighted time-delay coordinates reveals not only that they reconstruct the topology of the state space but also that they can regularize the dynamics,

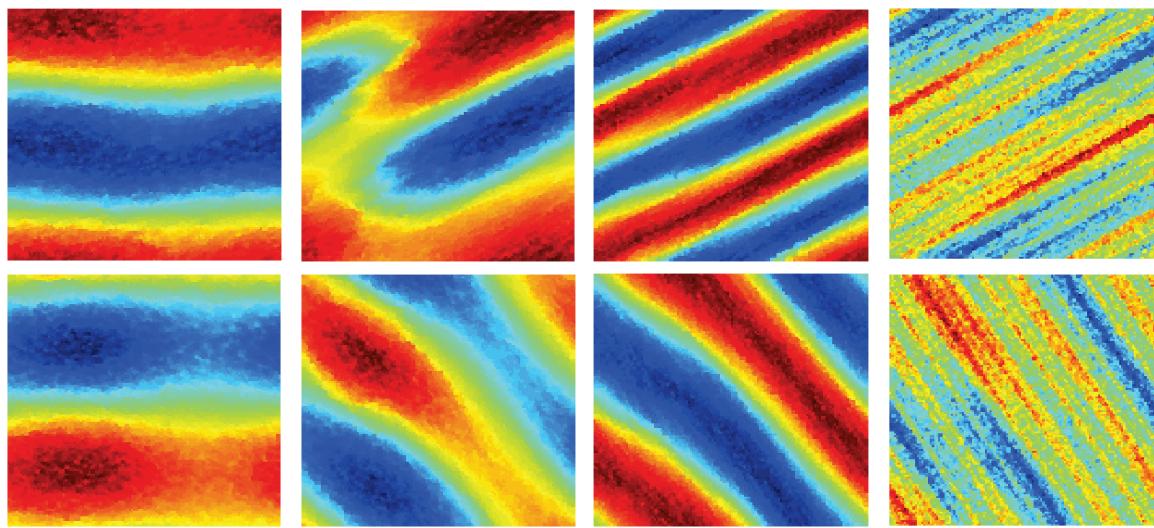


Figure 4. The top row shows the first nontrivial eigenfunctions of the Laplace–Beltrami operator in the (a, b) -plane, calculated from DMDC, for a time-delay embedding of the cat map with $s = 2048$ delays and $\kappa = 1.2, 0.8, 0.4, 0.01$ from left to right. When $\kappa = 1.2$, Theorem 2.1 does not apply, and the projection to the stable direction fails. When $0 < \kappa < -\sigma_1 \approx 0.962$, the projection successfully reconstructs the stable Oseledets direction in theory. As $\kappa \rightarrow 0$, the stable projection remains but regularization begins to fail. The bottom row shows DMDC applied to the time series in reverse order for the same values of κ . The entire range of $\kappa \in [.01, 1.2]$ for the top and bottom rows is visualized in the associated videos (88183-04.mov [local/web 1.21MB] and 88183-05.mov [local/web 1.04MB], respectively).

while projecting onto the most stable Lyapunov direction. Of course, if the other Lyapunov directions correspond to dynamical noise, this projection could be a useful feature. Moreover, even if the other Lyapunov directions correspond to interesting dynamics, these dynamics will be less stable, and therefore operate on a faster time scale, than the most stable directions. Most importantly, the projection onto the most stable manifold will often achieve a significant dimensionality reduction.

Example 2.1 (cat map). In this example we illustrate the effect on the geometry of the torus by weighted time-delay embedding via Arnold’s cat map, a discrete time map on the torus $S^1 \times S^1$ given by

$$\begin{aligned} a_{j+1} &= 2a_j + b_j \bmod 1, \\ b_{j+1} &= a_j + b_j \bmod 1. \end{aligned}$$

We choose a point $(a_0, b_0) \in [0, 1]^2$ randomly and apply the cat map for N iterations, producing the points $\{(a_j, b_j)\}_{j=0}^N$.

To test Theorem 2.1, we embed the torus into \mathbb{R}^3 via

$$\begin{aligned} x_j &= (2 + \sin 2\pi a_j) \sin 2\pi b_j, \\ y_j &= (2 + \sin 2\pi a_j) \cos 2\pi b_j, \\ z_j &= \cos 2\pi a_j, \end{aligned}$$

and observe the dynamics through the weighted delay coordinates,

$$w_j = [x_j, y_j, z_j, e^{-\kappa} x_{j-1}, e^{-\kappa} y_{j-1}, e^{-\kappa} z_{j-1}, \dots, e^{-s\kappa} x_{j-s}, e^{-s\kappa} y_{j-s}, e^{-s\kappa} z_{j-s}].$$

Here we have chosen $r = 3$ in the application of Theorem 2.1.

The system is ergodic on the torus with one positive Lyapunov exponent and one negative. We can visualize the geometry of the time-delay embedding by plotting the eigenfunctions of the Laplace–Beltrami operator on the embedded torus (see Figure 4). These eigenfunctions are computed using diffusion maps, described in section 3; they are shown here to display the effect of varying κ .

When $\kappa > -\sigma_1 = -\log \frac{1}{2}(3-\sqrt{5}) \approx 0.962$, we do not achieve the projection onto the stable Oseledets space, and thus the Riemannian metric of the torus in \mathbb{R}^3 (our observation space) is recovered. This is shown in Figure 4 when $\kappa = 1.2$ by the eigenfunction oscillating vertically in the (a, b) -plane, which corresponds to the large circle on the torus in the (x, y, z) -space. As κ decreases, the metric becomes increasing localized in the stable Lyapunov direction. Note that for $\kappa < -\sigma_1$ the eigenfunctions become constant in the unstable direction $(1, (\sqrt{5}-1)/2) \approx (1, 0.618)$ and oscillatory in the orthogonal stable direction. Thus the eigenfunctions of the Laplacian in the embedding space represent only the location of the point on the stable manifold.

3. Diffusion maps for delay coordinates. In the previous section we found that an appropriately weighted time-delay embedding can reconstruct the intrinsic geometry of a smooth attractor from a generic observable. In cases where the observable is high-dimensional, such as a video, we will want to pair the delay embedding with a dimensionality reduction technique that preserves the geometry as faithfully as possible. In this section we show that a carefully chosen diffusion map will give, in a specific sense, the best preservation of the delay geometry. Moreover, we will show that the diffusion map has a natural dynamical interpretation when applied to time series.

For a finite set of points on a manifold embedded in a high-dimensional Euclidean space, a diffusion map is a nonlinear map to a lower-dimensional space. In rough analogy to the principal components from an SVD, the components of a diffusion map [7, 9] are eigenvectors of a transition matrix for a random walk on the data set. Under appropriate normalizations, the transition matrix is a discrete approximation to the Laplace–Beltrami operator [20], which is by definition the divergence of the gradient on the manifold inherited from the embedding. Thus the components of the diffusion map will be approximations to eigenfunctions of this operator, and we will see that the diffusion map minimizes an energy functional that measures the distortion of the manifold’s geometry.

We have seen that a time-delay embedding introduces a natural geometry on a dynamical system, and we wish to preserve this geometry in the new (lower-dimensional) coordinates. Unfortunately, the sampling density in the embedding space also influences the geometry. The power of the diffusion maps technique is the ability to control the influence of the sampling density on the geometry in the new coordinates. We will see that for a certain normalization a diffusion map can match the invariant measure of a dynamical system.

In general there are many techniques for dimensionality reduction (see [21] for an overview). For the application to delay coordinates, a diffusion map is the natural choice because

(1) it preserves geometry and (2) it has a dynamical interpretation. First, we will see that for any dimension l a diffusion map into \mathbb{R}^l yields the minimum distortion of the metric (and for $l > 2n$ the map will generically be an embedding). Second, when the data is a time series, each coordinate function of the diffusion map can be interpreted as a time series. In fact, each coordinate function will correspond to a generalized low-pass filter of the time series. In section 4 we will show that this dual interpretation, as a geometry-preserving map and a low-pass filter, can be readily exploited for time-scale separation. In this section we review the main results of diffusion maps [7] and give a new interpretation of the construction as a conformal change of metric. We then interpret diffusion maps for generic observations of ergodic dynamical systems as a generalization of Fourier analysis.

3.1. Diffusion maps and sampling. Assume that N data points are sampled from an n -dimensional manifold \mathcal{M} which is embedded in \mathbb{R}^m . Consider a symmetric kernel $J_\epsilon(x, y) = h_\epsilon(|x - y|)$; in our applications, we use a Gaussian $h_\epsilon(z) = e^{-z^2/(2\epsilon)}$. The kernel is first used to interpolate the sampling density $p(x)$ of the data as

$$(3.1) \quad P(x) = \int_{\mathcal{M}} J_\epsilon(x, y)p(y)dy \approx \sum_{i=1}^N J_\epsilon(x, x_i),$$

where $\{x_i\}_{i=1}^N \in \mathcal{M} \subset \mathbb{R}^m$ are the discrete observations, which are assumed to be sampled from the true density $p(x)$. In the limit of large N the discrete approximation becomes equality.

The idea is to use a kernel density estimate to approximate the heat kernel $e^{-t\Delta}$, where Δ is the Laplace–Beltrami operator on \mathcal{M} inherited from \mathbb{R}^m . Here we use the convention that the eigenvalues of Δ are positive. Unfortunately, the sampling density will affect such a kernel estimate, but a subtle renormalization can recover the correct operator. The sampling bias parameter α will be introduced to control the influence that the sampling density will have on the geometry.

Define the discrete version of the kernel to be the $N \times N$ matrix J , where $J_{ij} = J_\epsilon(x_i, x_j)$. The discrete version of the interpolated measure above is the $N \times N$ matrix $P = \text{diag}(J\mathbf{1}_N)$, where $\mathbf{1}_N$ denotes the N -vector of ones. That is, P is the diagonal matrix where P_{ii} is the i th row sum of J .

An $N \times N$ matrix with nonnegative entries is *stochastic* if its rows add to 1. The eigenvalues of a stochastic matrix are nonnegative, and the largest eigenvalue is 1. In our case, the matrix $P^{-1}J$ is stochastic. However, it is still biased by the sampling density.

The parameter α controls the influence of the sampling. For fixed $\epsilon > 0$ and $\alpha \geq 0$, define

$$(3.2) \quad \begin{aligned} K(x, y) &= \frac{J_\epsilon(x, y)}{P(x)^\alpha P(y)^\alpha}, & K &= P^{-\alpha}JP^{-\alpha}, \\ Q(x) &= \int_{\mathcal{M}} K(x, y)p(y)dy, & Q &= \text{diag}(K\mathbf{1}_N), \\ F_{\epsilon, \alpha}(f)(x) &= \int_{\mathcal{M}} \frac{K(x, y)}{Q(x)} f(y)p(y)dy, & T &= Q^{-1}K, \end{aligned}$$

where we have, by abuse of notation, listed the discrete counterpart of each operator on the left as an $N \times N$ matrix on the right. Note that T is a stochastic matrix.

The key result of diffusion maps is that the matrix T can be used to approximate a differential operator \mathcal{L} that captures the geometry of the data set. Namely, it is shown in [7] that

$$\mathcal{L}\varphi \equiv \lim_{\epsilon \rightarrow 0} \frac{I - F_{\epsilon,\alpha}}{\epsilon} \varphi = \Delta\varphi + 2 \frac{\nabla(p^{1-\alpha})}{p^{1-\alpha}} \cdot \nabla\varphi.$$

Alternatively, if we write the sampling density as $p = e^{-U}$, then

$$(3.3) \quad \mathcal{L}\varphi = \Delta\varphi - 2(1-\alpha)\nabla U \cdot \nabla\varphi.$$

In the special case $\alpha = 1$, we recover the Laplace–Beltrami operator Δ ; in the case of general α , we recover a backwards Fokker–Planck operator \mathcal{L} for a diffusion process.

Now we can define the diffusion maps as in [7], so that they give a low-dimensional representation of the data. They are defined for each time scale t by the projection of the data onto the eigenfunctions of the operator $F_{\epsilon,\alpha}^{t/\epsilon} \approx T^{t/\epsilon}$. Without loss of generality, on the scale of 1 time unit, we can define the eigenfunctions of $T^{1/\epsilon}$ by $T^{1/\epsilon}\psi_l = \lambda_l^2\psi_l$, where $1 = \lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_L > 0$. Here we have used the fact that since T is a stochastic matrix, it has eigenvalues between 0 and 1, and therefore so does $T^{t/\epsilon}$ for any $t, \epsilon > 0$. The *diffusion map* $\Psi_{\alpha,t} : \mathcal{M} \rightarrow \mathbb{R}^L$ at the data point x_i is given by

$$(3.4) \quad \Psi_{\alpha,t}(x_i) = [\lambda_1^t\psi_1(x_i), \dots, \lambda_L^t\psi_L(x_i)]^T,$$

where L is chosen large enough that λ_{L+1}^t is sufficiently negligible. Note that $(\psi_l)_i = \psi_l(x_i) = \langle \psi_l, \delta_{x_i} \rangle$ is the i th coordinate of the l th eigenvector of the matrix T .

The eigenvalues and eigenvectors of the $N \times N$ nonsymmetric matrix T need to be calculated. Fortunately, T is closely related to a symmetric matrix \hat{T} , where calculations can be done more stably and which furthermore can be readily described as a matrix of inner products, or a *Gramian matrix*. Define

$$\hat{T} = Q^{-1/2}KQ^{-1/2} = Q^{1/2}TQ^{-1/2}.$$

It is preferable to do the eigenvalue calculation with the symmetric matrix \hat{T} , since it has the same eigenvalues as T . If ξ_l are the eigenvectors of \hat{T} , then $\psi_l = Q^{-1/2}\xi_l$ are the eigenvectors and $\lambda_l^2 = a_l^{1/\epsilon}$ are the eigenvalues of $T^{1/\epsilon}$, collectively satisfying $T^{1/\epsilon}\psi_l = \lambda_l^2\psi_l$.

If we let U be the matrix whose columns are ξ_l , then $\hat{T} = U\Lambda U^T$, where $\Lambda_{ll} = \lambda_l^{2\epsilon}$. Then we let $V = Q^{-1/2}U$ be the matrix whose columns are ψ_l , and we let $R = Q^{1/2}U = QV$ be the matrix whose columns are $R_l = Q\psi_l$. This gives a biorthogonal decomposition of

$$T = Q^{-1/2}\hat{T}Q^{1/2} = Q^{-1/2}U\Lambda U^TQ^{1/2} = V\Lambda R^T = V\Lambda V^TQ.$$

Moreover, since \hat{T} is symmetric and positive definite it is a Gramian matrix, and similarly $V\Lambda V^T$ is a Gramian matrix.

We now connect $T^{t/\epsilon}$ to the Gramian matrix

$$(T^{t/\epsilon}Q^{-1})_{ij} = \sum_{l=0}^{N-1} \lambda_l^{2t}\psi_l(x_i)\psi_l(x_j) \approx \sum_{l=0}^L \lambda_l^{2t}\psi_l(x_i)\psi_l(x_j) = \langle \Psi_{\alpha,t}(x_i), \Psi_{\alpha,t}(x_j) \rangle.$$

The distance which corresponds to the inner product above is the diffusion distance in the space of functions $L^2(\mathcal{M}, d\mu/p(x))$ such that

$$(3.5) \quad D_t(x_i, x_j)^2 = \|e^{-t\mathcal{L}}\delta_{x_i} - e^{-t\mathcal{L}}\delta_{x_j}\|_{L^2(\mathcal{M}, d\mu/p(x))}^2 \approx \|\Psi_{\alpha,t}(x_i) - \Psi_{\alpha,t}(x_j)\|_2^2,$$

which is shown to satisfy the requirements of a distance in [7]. This shows that the Euclidean distance between diffusion coordinates approximates the diffusion distance on the underlying manifold \mathcal{M} . In fact, the theory of multidimensional scaling shows that the diffusion coordinates give the best preservation of the diffusion distance of any L -dimensional representation of the data [7, 21].

As mentioned above, $\mathcal{L} = \Delta$ in the special case $\alpha = 1$ in (3.3), and it is shown in [7] that $F_{\epsilon,1}^{t/\epsilon}$ approximates the heat kernel in the sense that

$$\lim_{\epsilon \rightarrow 0} F_{\epsilon,1}^{t/\epsilon} = e^{-t\Delta}.$$

The approximation $T_{ij}^{t/\epsilon} \approx \langle \delta_{x_i}, e^{-t\Delta} \delta_{x_j} \rangle$ becomes equality in the limit as $\epsilon \rightarrow 0$, where the limit assumes that the number of samples N increases to infinity simultaneously. So for $\alpha = 1$, the matrix $T^{t/\epsilon}$ is a discrete approximation to the heat kernel.

Finally, we give a new interpretation of the parameter α . Intuitively, α will determine how much influence the sampling density will have on the operator \mathcal{L} . More formally, we now show that α corresponds to a conformal change of metric and that the operator \mathcal{L} constructed by diffusion maps is related to the Laplacian with respect to the new metric. Note that the manifold \mathcal{M} inherits a Riemannian metric g from the ambient space \mathbb{R}^m and a volume form $d\text{vol} = \sqrt{|\det(g)|}d\omega$ which is given by the sampling density $\sqrt{|\det(g)|} = p = e^{-U}$. Consider the conformal change of metric $\tilde{g} = e^{4(1-\alpha)U/(n-2)}g$, and note that the Laplacian with respect to this metric is given by

$$\Delta_{\tilde{g}}\varphi = e^{-4(1-\alpha)U/(n-2)}(\Delta_g\varphi - 2(1-\alpha)\nabla\varphi \cdot \nabla U) = e^{-4(1-\alpha)U/(n-2)}\mathcal{L}\varphi.$$

This shows that the operator \mathcal{L} constructed by the diffusion map is always a scalar function multiple of a Laplacian $\Delta_{\tilde{g}}$ given by the conformal change of metric. Therefore the α parameter determines the degree to which the sampling density influences the geometry of the diffusion-mapped data. By taking $\alpha = 1$, we can remove this influence entirely and recover the Laplace–Beltrami operator $\mathcal{L} = \Delta_g$.

The results of section 3.2 will apply for arbitrary α , and we interpret the operator $F_{\epsilon,\alpha}^{t/\epsilon}$ as a generalization of the heat kernel with respect to the conformal change in metric described above. We will see in sections 3.3 and 4 that for certain dynamical systems $\alpha = 1/2$ will allow us to adapt the diffusion map to the invariant measure. After section 3.2, the rest of the article will assume $\alpha = 1/2$, and this is the value used in all examples.

3.2. Geometric interpretation of diffusion maps. In section 2 we saw how κ -weighted time-delay coordinates give an embedding of the invariant manifold into Euclidean space, where the Riemannian metric inherited from the ambient space is intrinsic to the dynamical system. Of course, while the Riemannian metric is intrinsic, the embedding itself is not and will depend on the details of the observation. Moreover, the dimension of the ambient space

will be very large, preventing efficient analysis. For the delay coordinates to be useful we need to map them to a lower-dimensional ambient space while maintaining the intrinsic geometry.

An SVD of the delay coordinates would apply a linear projection of the data set onto a lower-dimensional hyperplane of the ambient space, and thus each tangent space to the invariant manifold would be transformed by the same linear projection. Finding a single linear projection that works for each tangent space will generally require a high-dimensional target hyperplane. Roughly speaking, the SVD is attempting to maintain all of the Euclidean distances in the projection space, while *only the local Euclidean distances are relevant* to the geometry.

In this subsection, we first show that the diffusion map gives intrinsic coordinates which are independent of the observation space. Next we show that the diffusion coordinates give a mapping (an embedding for L sufficiently large) of our manifold into \mathbb{R}^L that gives the minimal distortion of the geometry of the manifold. Then we interpret the parameter t as the “scale” of our approximation to the manifold, which allows us to reduce the dimensionality while maintaining the geometry as well as possible. Finally, we will show that the parameter α corresponds to a choice of measure which will allow us to match the invariant measure of a dynamical system.

In section 3.1 we saw that, given a finite data set sampled from a manifold embedded in \mathbb{R}^N , a diffusion map constructs a discrete approximation to a heat kernel on the manifold. (When $\alpha \neq 1$, the heat kernel is with respect to a conformal change of metric.) The heat kernel on a manifold is equivalent to the Riemannian metric, so we can preserve the geometry of an embedding by preserving the heat kernel. Since the diffusion coordinates of (3.4) approximate the dominant eigenfunctions of the heat kernel, they give the best low-dimensional approximation to the heat kernel. Moreover, since the heat kernel is invariant under isometries, the diffusion coordinates will be the same for any isometric copies of the manifold. Thus, the diffusion coordinates are intrinsic and will not depend on the observation space.

Note that since the matrix $T^{t/\epsilon}Q^{-1}$ is a Gramian matrix, the diffusion coordinates (given by the eigenfunctions ψ_l) are the principal components of the associated inner product space [21]. This naturally gives two interpretations to the diffusion maps. First, in terms of kernel principal component analysis, we will see that the diffusion map maintains local distances. Second, in terms of multidimensional scaling, we will see that the diffusion map gives the minimum distortion of the nonlocal diffusion distances, which are a generalization of geodesic distances.

First, to see that the diffusion map is a kernel principal component analysis, note that the diffusion coordinates are given by the solutions of the generalized eigenvalue problem $A\psi = \lambda B\psi$, where $A_{ij} = K(x_i, x_j)$ and $B = Q$. As shown in [10], this implies that the diffusion coordinates give an embedding into \mathbb{R}^L that minimizes the functional

$$E[\Phi] = \sum_{l=1}^L \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} (\phi_l(x) - \phi_l(y))^2 \frac{K(x, y)}{Q(x)} p(x)p(y) dx dy$$

over maps $x \mapsto \Phi(x) = (\phi_1(x), \dots, \phi_L(x))$ subject to the constraint $\langle \phi_i, \phi_j \rangle = \delta_{ij}$. Since $K_{\epsilon, \alpha}$ is a localizing kernel with exponential decay, this functional penalizes moving points apart that were close in the original space. In this sense, the diffusion coordinates seek to minimizes the

local distortion of distance. This shows that the diffusion map preserves the local geometry of the embedded manifold.

Whereas the above interpretation shows that the diffusion map minimizes a stress functional, we can also interpret the diffusion map in terms of multidimensional scaling. In the previous section we saw that the matrix $T^{t/\epsilon}Q^{-1}$ contains inner products, and that there is a corresponding metric which is given by the diffusion distance from (3.5). The theory of multidimensional scaling shows that the diffusion coordinates minimize the distortion given by

$$\mathcal{P}[\Phi] = \sum_{i,j=1}^N (D_t(x_i, x_j) - \|\Phi(x_i) - \Phi(x_j)\|)^2$$

with Φ constrained as above [21]. Thus the Euclidean distance in the diffusion coordinates is as close as possible to the diffusion distances on the manifold. Optimality in terms of the functionals E and \mathcal{P} shows that the diffusion map will maintain local distances while changing long distances to approximate the intrinsic diffusion distance.

Next we provide a geometric interpretation of the parameter t . The diffusion distances were introduced in [7] and shown to be distances on the manifold for each value of the parameter t . Intuitively, the parameter t controls the scale at which we approximate the manifold. For example, as $t \rightarrow \infty$, the diffusion distance between all points approaches zero, and for t large our data set is approximated as a single point. Conversely, as $t \rightarrow 0$, the diffusion distance is related to the geodesic distance d on the manifold by the equation

$$\begin{aligned} d(x, y) &= \lim_{t \rightarrow 0} -4t \log \langle e^{-t\Delta} \delta_x, \delta_y \rangle \\ &= \lim_{t \rightarrow 0} -4t \log \left(\frac{\|e^{-t\Delta/2} \delta_x\|^2 + \|e^{-t\Delta/2} \delta_y\|^2 - D_{t/2}(x, y)^2}{2} \right). \end{aligned}$$

The geodesic distance is equivalent to the Riemannian metric, so as $t \rightarrow 0$, the diffusion distance captures all the details of the geometry. Thus, the parameter t allows us to control the scale at which we try to represent the manifold: t large gives a coarse scale representation, and t small gives the fine scale.

Of course, to represent the fine scale will require a high-dimensional diffusion map, and often the fine scale will be dominated by noise. Thus by choosing t large, we can approximate our manifold at a coarse scale that will allow a low-dimensional representation while still maintaining the geometry in the sense of the functionals E and \mathcal{P} . In the next section we will see that the t parameter also has a natural time-scale interpretation for dynamical data which further motivates the diffusion map for our application.

3.3. Time series interpretation of diffusion maps. We now extend the interpretation of diffusion maps to the case when our data has a time ordering and show that setting the sampling bias parameter $\alpha = 1/2$ will match the invariant measure. Note that a diffusion map estimates the values of the eigenfunctions ψ_l of a heat kernel at the discrete points x_i which are the input to the diffusion maps algorithm. If the input points have a time ordering, we can give the eigenfunctions a time ordering by setting $\hat{\psi}_l(t_i) = \psi_l(x_i) = \psi_l(x(t_i))$. In section 4 we will further investigate these time series. In this section we will show that if $y_i = y(t_i)$ is a

trajectory of a dynamical system with an invariant measure, then the diffusion map interprets $y(t_i)$ as the trajectory of a simplified system which has the same invariant measure.

As a simple example, assume that the data $\{x_i\} \in \mathcal{M} \subset \mathbb{R}^m$ are generated by a path of the stochastic differential equation

$$(3.6) \quad dx = -\nabla U(x) dt + \sqrt{2} dB_t,$$

where $U(x)$ is a smooth potential function on the manifold \mathcal{M} , B_t is Brownian motion, and $x_i = x(i\tau)$, where τ is a sampling interval. The backward Fokker–Planck equation for the system given by (3.6) is

$$(3.7) \quad -\frac{\partial \varphi}{\partial t} = \Delta \varphi - \nabla \varphi \cdot \nabla U,$$

and it was shown in [7] that

$$\mathcal{L}\varphi = \Delta \varphi - 2(1-\alpha)\nabla \varphi \cdot \nabla U,$$

so clearly when $\alpha = 1/2$, the diffusion map approximates the backward Fokker–Planck operator. The invariant measure of this system is given by the eigenfunction with eigenvalue zero of the corresponding forward Fokker–Planck operator,

$$\mathcal{L}^*\varphi = \Delta \varphi + \nabla(\varphi \nabla U).$$

Thus, e^{-U} is the invariant measure for (3.6) since $\mathcal{L}^*(e^{-U}) = 0$. Moreover, the invariant measure is also given by the eigenfunction $q_0(x_i) = (Q\psi_0)(x_i) \approx p(x_i) = e^{-U(x_i)}$.

Next we generalize to the case of dynamics that are more complex than those of (3.6). Assume that the data $\{y_i = y(t_i)\}$ are sampled from a trajectory of a dynamical system that has an invariant measure $\mu(y) > 0$ on a smooth manifold \mathcal{M} . Defining the potential function $U(x) = -\log \mu(x)$, equation (3.6) is a stochastic system for $x(t)$ that has the same invariant measure as $y(t)$. Note that $\mu = e^{-U}$ is the invariant measure for both $y(t)$ and $x(t)$, so in particular it is the sampling density for $y(t)$. A diffusion map with $\alpha = 1/2$ applied to the data set $\{y_i\}$ will construct the Fokker–Planck operator for (3.6) with $U = -\log \mu$. The eigenfunctions of this operator give a basis for square integrable functions on \mathcal{M} , and this basis is adapted to the invariant measure of our system. Thus, even when the trajectory is not governed by a stochastic differential equation of form (3.6), diffusion maps with $\alpha = 1/2$ will treat the data as if it were generated by (3.6) but with $U = -\log \mu$. The resulting eigenfunctions will be a generalization of the Fourier basis which is adapted to the correct invariant measure of $y(t)$ on \mathcal{M} .

4. Time-scale separation. The goal of time-scale separation is to decompose a dynamical system by a coordinate transformation such that the transformed variables are ordered by time scale and are approximately independent. In this paper we are interested in working directly with multivariate time series where no equations are known. In this section we show how a basis of eigenvectors for an operator on the state space can be used to separate time scales. In particular, we show that diffusion maps, combined with delay embeddings to reconstruct and simplify the state space, can separate time scales for a large class of interesting systems.

A diffusion map approximates eigenfunctions $\psi_l : \mathcal{M} \rightarrow \mathbb{R}$ of a heat kernel on a manifold; these are generalizations of sine and cosine. Note that the eigenfunctions of the Laplacian and the heat kernel are the same but the eigenvalues are different. We previously denoted the eigenvalues of the heat kernel by $1 = \lambda_0 \geq \lambda_1 \geq \dots > 0$, so that $\gamma_l = -\log(\lambda_l)$ are the eigenvalues of the Laplacian operator. Moreover, the eigenvalues $0 = \gamma_0 \leq \gamma_1 \leq \dots$ order the eigenfunctions according to how oscillatory they are. Intuitively, given a continuous trajectory $y(t) : \mathbb{R} \rightarrow \mathcal{M}$, the time series formed by composing $\hat{\psi}_l(t) = \psi_l(y(t))$ will be more oscillatory for large l and less oscillatory for small l . We will see that this time scale separation is most effective when the Laplacian operator is adapted to the dynamics of $y(t)$.

Consider a trajectory $y : \mathbb{R} \rightarrow \mathcal{M}$ which is reconstructed at discrete times $\{t_i\}_{i=1}^N$ by a delay embedding as $\{y(t_i)\}_{i=1}^N \subset \mathbb{R}^{r(s+1)}$. A diffusion map applied to the samples $\{y(t_i)\}$ will ignore the time ordering and treat these data points merely as a collection of samples from the manifold \mathcal{M} embedded in $\mathbb{R}^{r(s+1)}$. If μ is the invariant measure for the evolution of $y(t)$, then we have seen in section 3.3 that a diffusion map with $\alpha = 1/2$ will approximate the operator

$$\mathcal{L}(\varphi) = \Delta\varphi - \nabla\varphi \cdot \nabla(\log(\mu(x))).$$

Moreover, the diffusion map will approximate the eigenfunctions ψ_l evaluated at the sample points $\{y(t_i)\}$. Thus we can consider the diffusion mapped modes to be time series given by $\hat{\psi}_l(t_i) = \psi_l(y(t_i)) = \langle \psi_l, \delta_{y(t_i)} \rangle$. The delta function $\varphi(x, t) = \delta_{y(t)}(x)$ along the trajectory of the dynamics allows us to apply harmonic analysis on the manifold to the trajectory. Moreover, the harmonic analysis is adapted to the dynamics because it matches the invariant distribution in the limit as $t \rightarrow \infty$.

Since diffusion maps can only approximate the operator \mathcal{L} , we now assume that we can write the full evolution as a nonautonomous perturbation of \mathcal{L} so that

$$(4.1) \quad \frac{\partial \varphi}{\partial t} = -\mathcal{L}(\varphi) + \mathcal{F}(x, t).$$

We have seen that a diffusion map will produce a basis $\{\psi_l\} \subset L^2(\mathcal{M})$ consisting of eigenfunctions of \mathcal{L} such that the eigenvalues are sorted as $0 = \gamma_0 \leq \gamma_1 \leq \gamma_2 \leq \dots$. Therefore, the solution $\varphi(x, t)$ and the unknown function \mathcal{F} have a decomposition in this basis so that (4.1) becomes

$$\frac{d}{dt} \langle \varphi(x, t), \psi_l(x) \rangle = -\gamma_l \langle \varphi(x, t), \psi_l(x) \rangle + \langle \mathcal{F}(x, t), \psi_l(x) \rangle,$$

which is an ODE. Moreover, the time series $\hat{\psi}_l(t) = \langle \varphi(x, t), \psi_l(x) \rangle$ is precisely the output of our diffusion map. Setting $\hat{\mathcal{F}}(t) = \langle \mathcal{F}(x, t), \psi_l(x) \rangle$, we find that the l th diffusion map mode satisfies

$$\frac{d}{dt} \hat{\psi}_l(t) = -\gamma_l \hat{\psi}_l(t) + \hat{\mathcal{F}}(t),$$

which has the solution

$$\hat{\psi}_l(t) = ae^{-\gamma_l t} + b \int_0^t e^{-\gamma_l(t-s)} \hat{\mathcal{F}}(s) ds.$$

Thus, the eigenvalue γ_l will determine how much history from the nonautonomous term $\hat{\mathcal{F}}$ is integrated into the mode $\hat{\psi}_l$. For $\hat{\mathcal{F}}$ sufficiently regular the time scale of $\hat{\psi}_l$ will be determined

by γ_l . Of course, if $\hat{\mathcal{F}}$ is large, then we may not be able to separate time scales with this method. However, since (3.6) and (4.1) have the same invariant measure, we expect at least the slowest time scales to be well approximated. Note that this method could be improved in the future by finding better methods of approximating $\hat{\mathcal{L}}$, which would make $\hat{\mathcal{F}}$ smaller and improve the time scale separation.

Previous work on time scale separation was motivated by singular perturbation theory and focused on approximating the local dynamics. In particular, the local evolution of the slow variables was approximated by repeated simulations of the fast variables using legacy code [17] or stochastic simulators [28]. A closely related technique in [35] requires knowledge of the local fast and slow directions to decompose the tangent bundle to the state space. Since we have no knowledge of the system, the local factorizations in the observation space may not consistently identify the correct global fast and slow directions. Thus we have taken two methods of incorporating global information. First, the time-delay embedding projects onto the most stable space and incorporates the long term behavior into the local geometry. Second, by projecting onto eigenfunctions of the Laplacian, we are guaranteed a globally consistent coordinate transformation. While it is not yet clear whether this global technique will be able to achieve the higher order perturbation expansions of [35, 17], we have seen that for many interesting systems the global technique can at least project onto the slow manifold (see Example 4.1 also).

The ideas of this section are a generalization of the more formal approach taken in [9, 8]. It was shown in [8] that if a dynamical system had the simple form (3.6), then each spectral gap $\lambda_k \gg \lambda_{k+1}$ produced a coarse time-scale version of the dynamics. This means that in the coordinates $r(t) = \Psi_{1/2, s_k}(x(t))$ given by the diffusion map at scale $s_k = -1/\log(\lambda_{k+1})$ we have

$$\dot{r} = F(r, \omega)$$

for some F which depends only on the reduced coordinates r and on a stochastic noise vector ω . While this is a strong result, exactly the type we are interested in, it applies in only a very narrow context. In fact, solutions of (3.6) simply follow the gradient of a potential and are driven purely by Brownian motion. This is a reasonable context for studying systems with large numbers of particles, although for a single trajectory (3.6) precludes even periodic dynamics. In contrast, many simple physical systems follow the gradient of a potential under the influence of a nonautonomous energy input as in (4.1).

Example 4.1. As an illustration, we analyze a simple dynamical system with two time scales. The dynamics lie on a one-dimensional subset of the torus embedded in \mathbb{R}^3 and are given by

$$\begin{aligned}\dot{x} &= -y + \frac{bzx}{\sqrt{x^2 + y^2}}, \\ \dot{y} &= x + \frac{bzy}{\sqrt{x^2 + y^2}}, \\ \dot{z} &= -b(\sqrt{x^2 + y^2} - a),\end{aligned}$$

where we set $a = 6$, $b = 20$. We can write this system as a slow/fast system by changing to cylindrical coordinates, where $r^2 = x^2 + y^2$ and $\tan \theta = y/x$. Setting $\epsilon = 1/b$, the system

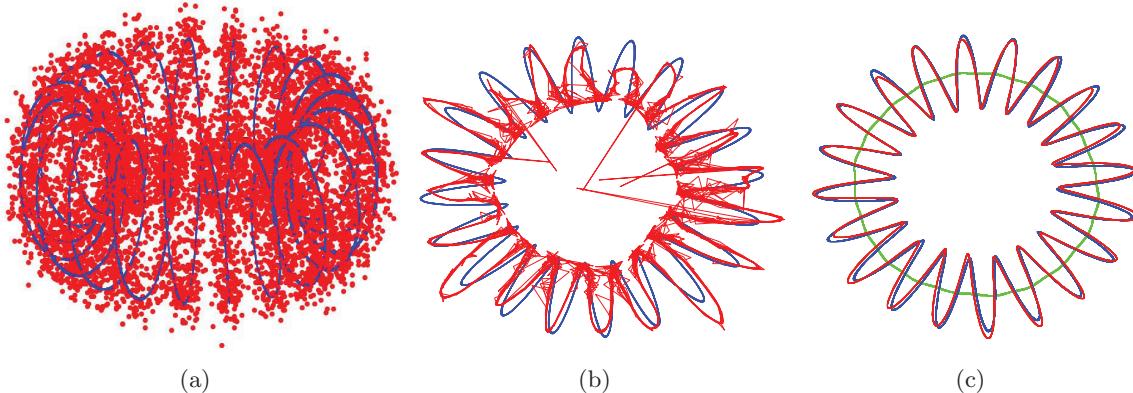


Figure 5. Reconstruction of dynamics from noisy observations using various values of κ . All plots show the (x, y) -plane dynamics in blue. (a) The noisy observations are shown in red. (b) The reconstructed dynamics for $\kappa = 1$ are shown in red. (c) Same as (b), but $\kappa = 0.02$. The projection onto the slow manifold is shown in green.

becomes

$$\begin{aligned}\dot{\theta} &= 1, \\ \epsilon\dot{r} &= z, \\ \epsilon\dot{z} &= -(r - a),\end{aligned}$$

so setting $\epsilon = 0$, we find a slow system given by $z = 0$, $r = a$, and $\dot{\theta} = 1$. The slow system simply oscillates in the (x, y) -plane, and the fast direction is primarily in the (z, r) -plane.

To test our algorithm in an equation-free setting, we use only a time series of simulated data points and do not incorporate any information from the above equations. The data will be given in the x, y, z -coordinates, and to further illustrate the advantage of the delay geometry we introduce Gaussian observational noise (see Figure 5). The ODE with initial conditions $(x(0), y(0), z(0)) = (6, 0, 2)$ was simulated for $N = 8000$ equally spaced time steps from $t = 0$ to $t = 12\pi$. We form the delay coordinates $w_i = [x_i, y_i, z_i, \dots, e^{-s\kappa}x_{i-s}, e^{-s\kappa}y_{i-s}, e^{-s\kappa}z_{i-s}]$ with $s = 500$ time delays with various values of κ to study the time-delay geometry.

Letting ψ_l be the eigenvectors (as described in section 3.1), we note that $(\psi_l)_{i-s} = \psi_l(w_i) = \langle \psi_l, \delta_{w_i} \rangle$ for $i = s+1, \dots, N$. Thus each eigenvector can be interpreted spatially (as a function of the location w_i in state space) and temporally (as a function of the time step i). In this way, we can interpret the eigenvectors as component time series, each of which represents the dynamics at a unique time scale. We can project the observed time series onto the span of the first $L + 1$ eigenfunctions by setting

$$\hat{w}_i = \sum_{l=0}^L W_l \psi_l(w_i), \quad \text{where } W_l = \langle w, q_l \rangle = \sum_{i=s+1}^N w_i q_l(w_i) = \sum_{i=s+1}^N w_i (q_l)_{i-s}.$$

For $L = 2$ we achieve the projection onto the slow manifold, as shown by the green curve in Figure 5(c). For $L = 60$ we achieve a low-pass filter that removes high frequency oscillations, as shown by the red plots in Figure 5(b) and (c). Note that to determine the slow manifold

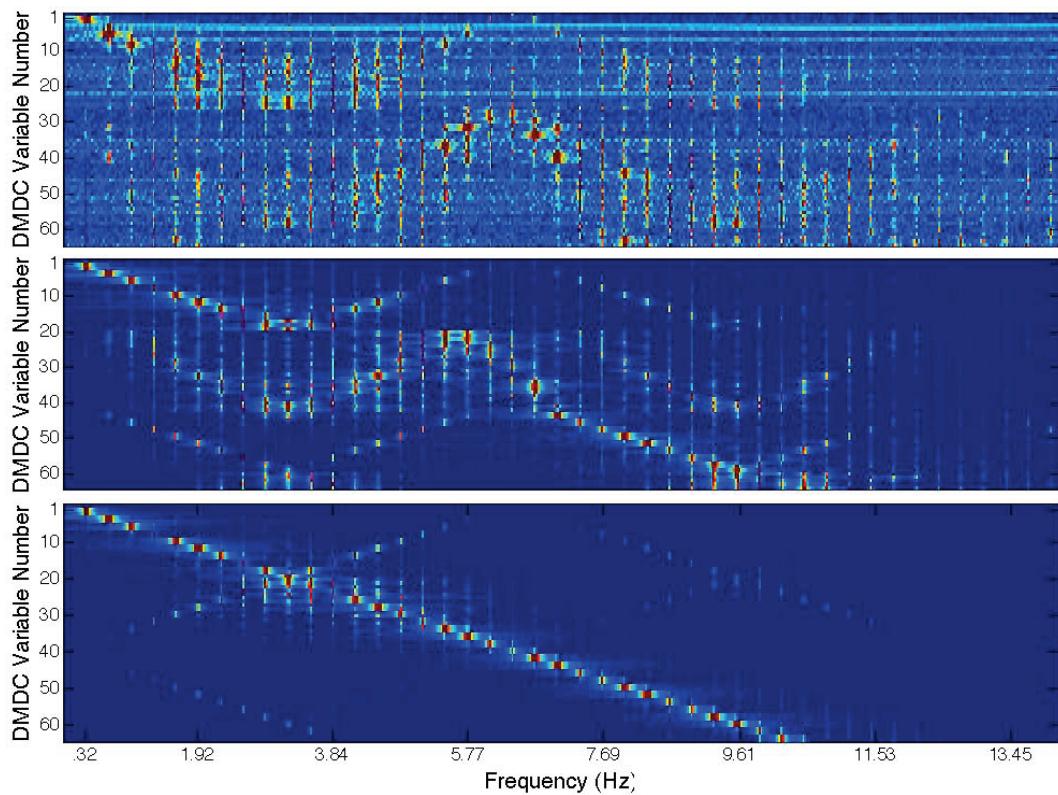


Figure 6. The Fourier spectrum of the first 64 eigenvectors of the Laplacian in the embedding geometry of Example 4.1 viewed as time series. The x -axis indicates the frequency of the oscillation, the y -axis indicates the eigenvectors, and color represents the norm of the Fourier coefficient for the relevant eigenvector at the indicated frequency. The top plot is for $\kappa = 10$, effectively no delays; the middle is for $\kappa = 0.1$; and the bottom is for $\kappa = 0.02$.

and to successfully remove noise, having the correct geometry is crucial, as shown by the failure of the reconstruction in Figure 5(b).

Furthermore, by taking the discrete Fourier spectrum of each of the time series $\psi_l(w_i)$, the time-scale separation is illustrated by the localized peaks in the Fourier spectrum of each eigenfunction and the ordering of the peaks, as shown in Figure 6. Moreover, we can see that introducing the delays improves the time-scale separation. Finally, by reconstructing the dynamics using only the first few eigenvectors, we can form the projection onto a slow time scale. These projections can reveal slow manifolds, as shown in Figure 5(c).

5. Time-scale separation for spatiotemporal dynamics. In this section we apply the proposed algorithm to video data. The first goal is to apply the algorithm to a meandering spiral wave, a reasonably complex model exhibiting spatiotemporal dynamics which are intrinsically low-dimensional. We will demonstrate how DMDC extracts meaningful low-dimensional dynamics and separates time scales. Following that, we apply the method to video of the spatiotemporal experiment referred to in Figure 1.

We assume that observations take the form of still images which are discretized versions

of $h(x)$, where x is a dynamical state of an n -dimensional attractor \mathcal{M} and $h : \mathcal{M} \rightarrow \mathbb{R}^r$, where r represents the number of pixels of the still image. The delay coordinate map is then $H : \mathcal{M} \rightarrow \mathbb{R}^{r(s+1)}$ from (2.1), where s is the number of delays, which has the effect of concatenating $s + 1$ still images.

Note that both DMDC and SVD (on the delay embedding space) produce time series. For SVD the time series are linear projections of the time-delay embedding of the images (which are observations of the state), whereas for DMDC the time series are eigenfunctions of the Laplacian on the embedding. Let x_i be the i th image in the observed video. Note that, as in Example 4.1, we can project the observed time series onto the span of the first $L + 1$ DMDC modes by setting

$$\hat{x}_i = \sum_{l=0}^L X_l \psi_l(x_i), \quad \text{where } X_l = \langle x, q_l \rangle = \sum_{i=s+1}^N x_i q_l(x_i) = \sum_{i=s+1}^N x_i (q_l)_{i-s}.$$

For video data the inner products $X_l = \langle x, q_l \rangle$ are averages of the images x_i weighted by the l th projection q_l . Thus, each X_l is an image which shows the parts of the image which are varying synchronously with the l th DMDC mode. Thus we call the image X_l the l th *component* of the decomposition. Note that SVD gives an orthonormal basis (instead of the biorthogonal system of DMDC), so for SVD the modes (analogous to $\{\psi_l\}$) and projections (analogous to $\{q_l\}$) are both given by the singular vectors of the centered delay coordinates.

5.1. Meandering spiral waves. We begin by applying DMDC to a model of meandering spiral waves. We find that resulting modes have a natural ordering which is determined by the time scale on which the mode is active. Then we compare our results to the more conventional SVD in the delay-embedding space in order to illustrate the significant advantages of the DMDC modes. Moreover, we show that time-delay coordinates are crucial to finding the best decomposition, in that an application of a diffusion map without delays is not sufficient even for ideal observations from a computational model.

Consider the reaction-diffusion system

$$\begin{aligned} u_t &= \Delta u + \frac{1}{\rho} u(1-u) \left(u - \frac{v+b}{a} \right), \\ v_t &= u - v, \end{aligned}$$

where the parameters $\rho = 0.01$, $a = 0.48$, $b = 0.01$ determine the behavior of the meandering spiral waves. In [4, 3] bifurcations of the parameter space are explored, and the system is shown to be related to a two-dimensional ODE, indicating the presence of a low-dimensional state space for the dynamics. Instead of working with the equations, our method will work directly with a sequence of still images separated by $\tau = 0.05$ (see Figure 7). The variables u and v are held to zero at the boundary, and the initial condition is $u(x, y) = v(x, y) = \sin 7x$, where the domain is $(x, y) \in [0, 1]^2$. While the behavior is robust to adding small amounts of noise to the initial condition, totally random initial conditions often die out quickly and do not lead to self-sustaining spirals.

In parallel with Example 1.2, we find that while SVD modes are organized according to variance, the DMDC modes are organized according to time scale. For example, Figure 8

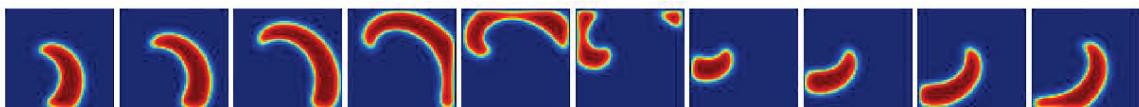


Figure 7. Images from spiral wave dynamics with $\rho = 0.01$, $a = 0.48$, $b = 0.01$ shown at time steps 0.3 apart (every sixth observed frame is shown).

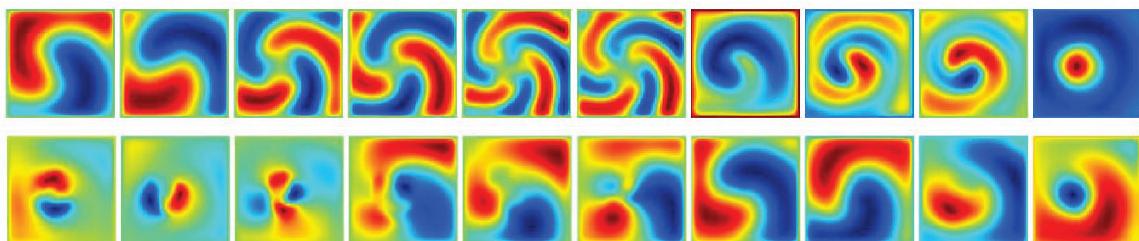


Figure 8. Top: First 10 modes of variation found from delay coordinates projected by SVD. Bottom: First 10 modes of variation found by DMDC. Images show the DMDC components (produced by averaging the spiral images weighted by an SVD or DMDC mode, respectively). Note that the first two DMDC modes are localized near the tip of the spiral and pick out the vertical and horizontal components of the slow-moving tip location. The associated video (88183_06.mov [local/web 5.44MB]) shows reconstructed dynamics using selected modes from the DMDC time-scale decomposition.

shows that the 7th and 8th DMDC modes correspond to the position of the main body of the spiral, which oscillates once per 60 frames. These modes are the first two modes found by SVD because they are the modes of largest variance in the video. In contrast, the first two DMDC modes represent the location of the tip of the spiral wave. These modes represent a slow precession in the spiral tip which oscillates every 240 frames. Thus the DMDC modes have identified an important slow mode of oscillation which has a low variance and thus would be very difficult to extract with SVD or an ad hoc technique.

To underscore this point, Figure 9 shows that the time series corresponding to the DMDC modes are ordered according to the peaks in the Fourier spectra. The figure compares the Fourier spectra of the first 64 SVD modes with the first 64 DMDC modes. Note that the SVD spectra are spread out horizontally, indicating that SVD modes contain information about many different time scales, whereas the DMDC spectra are horizontally localized, indicating that the modes have a definite time scale. Moreover, the DMDC spectra are ordered in the sense that higher order DMDC modes contain information at faster time scales. Thus, in analogy with the results of [8], projecting on the first k DMDC modes would contain all the information about oscillations on a sufficiently slow time scale. For example, the first two DMDC modes contain all the information relevant to the slow precession of the spiral tip, which has a period of approximately 240 frames. In contrast, the first two SVD modes capture the primary spiral oscillation, which occurs approximately every 60 frames.

Finally, we note that DMDC can be used for noise reduction. In Figure 10 we show an example of additive Brownian noise. In order to make a fair comparison, notice in Figure 8 that the first two SVD modes represent the main spiral oscillation and will be the best SVD modes in terms of noise reduction. Figure 8 shows that these modes also appear in

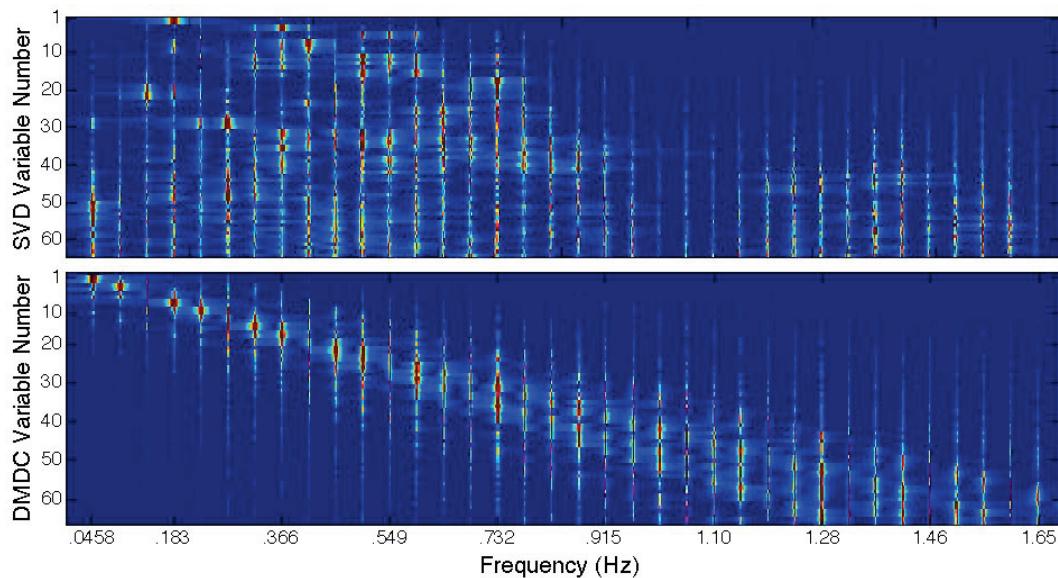


Figure 9. Top row: Fourier spectra of SVD modes reveal no time-scale ordering. Bottom row: Fourier spectra of DMDC modes show ordering according to Fourier spectral peak.

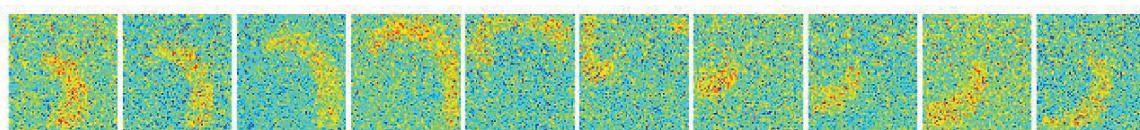


Figure 10. Same images from spiral wave dynamics as in Figure 7, but with white noise added with standard deviation of 850% of the mean.

the DMDC analysis, but occur later because of the time ordering. Thus these modes can be fairly compared for noise content, and in Figure 11 we show the dramatic noise reduction that DMDC achieves.

Intuitively, this noise reduction is achieved because DMDC is attempting to maximize the variance as measured on the manifold which the dynamics are near, whereas SVD tried to maximize the variance as measured in the ambient space in which the data is embedded, as shown in section 3.2.

A further interpretation of the noise reduction results is that DMDC provides a data-adapted Fourier transform. If the invariant measure is uniform, or if we take $\alpha = 1$ in the diffusion map analysis to eliminate sampling effects, then the DMDC modes are given by the inner product with the eigenfunctions of the Laplace–Beltrami operator on the manifold Σ . This is a well-known generalization of Fourier analysis to manifolds, and in this natural basis of generalized Fourier modes the observation noise will be concentrated in the high frequency modes. In the case where the invariant measure is nontrivial this analysis gives a possible further generalization of Fourier analysis adapted to measures on manifolds.

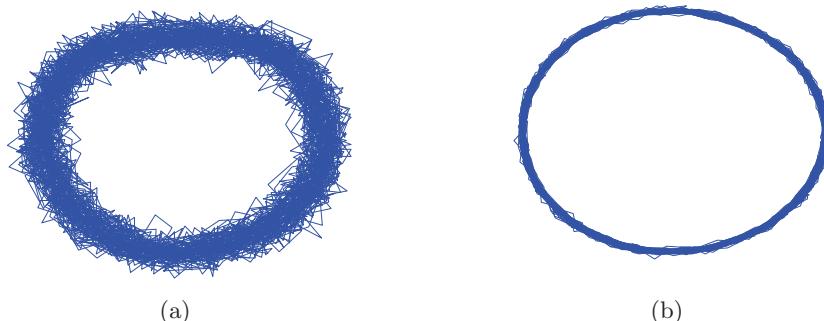


Figure 11. Embedding of noisy spiral images from Figure 10, using (a) the first two SVD modes and (b) the equivalent two DMDC modes. See the associated video (88183_07.mov [[local](#)/[web](#) 11.1MB]) for reconstructions comparing SVD and DMDC modes in an even more extreme case where the standard deviation of the white noise is 1,250% of the mean pixel value.

5.2. Time-scale separation of nematic liquid crystal videos. Electroconvection nematic liquid crystals (NLCs) display a wide range of dynamic behaviors while possessing a high degree of spatiotemporal structure. NLCs provide an excellent testbed for studying the role of coherent structures in nonequilibrium systems, which are of great interest in problems ranging from fluctuations in the power grid to weather patterns. A major motivation for studying this system is to understand the role spatiotemporal dynamics play in determining the power injected into this system. For our present purposes, we demonstrate the ability of DMDC to identify the dynamic structures supported by the liquid crystal.

Nematic liquid crystals are fluids composed of rod-like molecules whose dielectric constant and conductivity differ along the major and minor axes of the molecule. Without any externally applied electric field, the molecules are disordered; i.e., they are aligned randomly. However, in the presence of an external electric field the anisotropic dielectric constant forces the molecules to become aligned with the field, making the sample become an optical polarizer [18]. Samples made with the addition of a conductive dopant will undergo additional dynamic transitions as the applied field is increased.

At low driving voltages the charge diffuses homogeneously through the sample. As the voltage increases, perturbations will be reinforced by electrostatic and hydrostatic torques, leading to an instability. This instability ultimately leads to a periodic roll-like flow field. The flow field distorts the molecular orientation, producing a lensing effect that can be observed optically as light and dark bands, as seen in Figure 1(a). This dynamic state is similar to Rayleigh–Bénard convection, where light fluid heated from below and heavy-cooled fluid from above rises and falls, respectively, under the action of gravity. Stronger electrical drive will cause these rolls to undergo further instabilities that lead to more complex dynamic patterns, as in Figure 1(b).

NLCs support a diverse set of static and evolving patterns that depend on the structure and composition of the the liquid crystal mixture as well as driving parameters including frequency and amplitude. These patterns have been discussed at length, and there exists a large body of literature on this subject [18, 34, 12]. More complex patterns are associated

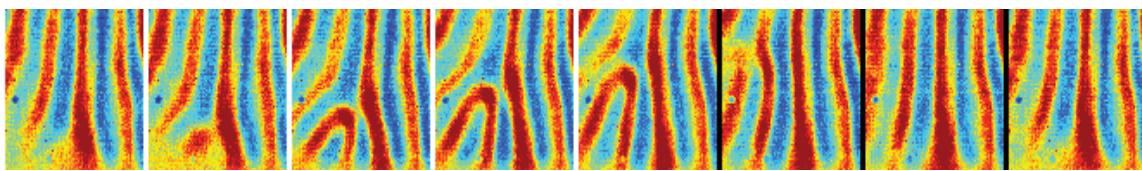


Figure 12. Eight snapshots of electroconvection in a liquid crystal sample. Successive frames are separated in time by 1.49 seconds.

with the creation and annihilation of dislocations (defects) in the roll-like pattern. The size and number of defects depend on the surface area, sample thickness, and applied electric field, respectively [34]. With a further increase in electric field the number of defects increases and the roll-like patterns are almost completely lost, regaining a nearly homogeneous state.

The samples used in the experiment are constructed from two glass microscope slides coated with the optically transparent conductor indium tin oxide (ITO). The ITO on both plates is etched so that only thin strips of the conductor remain. The intersection of these strips defines the active area of the sample, with $A = 50\mu\text{m}^2$ for results presented here. Spacers between these glass plates determine the sample thickness $d = 25\mu\text{m}$. The NLC used in this experiment is methoxybenzylidene-butyl aniline (MBBA). The sample is connected to an AC signal generator and is driven at a frequency of $f = 100\text{Hz}$. Images are acquired with a high speed camera.

Figure 12 displays snapshots of a liquid crystal sample driven just above the threshold for electroconvection (voltage = 26V). These eight images show the propagation of a defect through the sample. The sample displays some degree of spatial complexity; however, it is approximately periodic with a frequency of 0.085Hz as measured from the power spectra of the intensity fluctuations at an arbitrary point in the sample.

Figure 13 displays the spectra for the first 20 modes obtained using the SVD and DMDC algorithms (top and bottom, respectively). The spectra of most of the SVD modes show multiple frequencies, whereas the spectra for the DMDC modes each contain a single well-defined frequency reflecting the separation of time scales. Both algorithms display peak frequency of 0.085Hz for their first two dynamic modes, which is in good agreement with the frequency of oscillations as determined from the raw intensity fluctuations.

The modes are visualized by summing the entire set of images weighted by the amplitude for each mode at the corresponding time step. In Figure 14 we show the first nine modes from the SVD and DMDC decompositions. These static images show qualitatively similar patterns, though the ordering is different. However, this apparent similarity is misleading, as they display very different temporal evolution, as seen in the associated video (88183_08.mov [local/web 1.88MB]).

The dynamics of the DMDC modes reflect the fact that they are ordered by their dominant frequencies. This discrimination produces modes that contain little spatiotemporal complexity as compared to SVD modes. The lowest DMDC mode describes the slowest dynamics, which corresponds to a nearly static image seen in the first panel. In contrast, the first SVD mode displays significant spatiotemporal variations. The second and third DMDC modes capture the relatively slow horizontal translation of the vertical rolls. The fourth and fifth modes

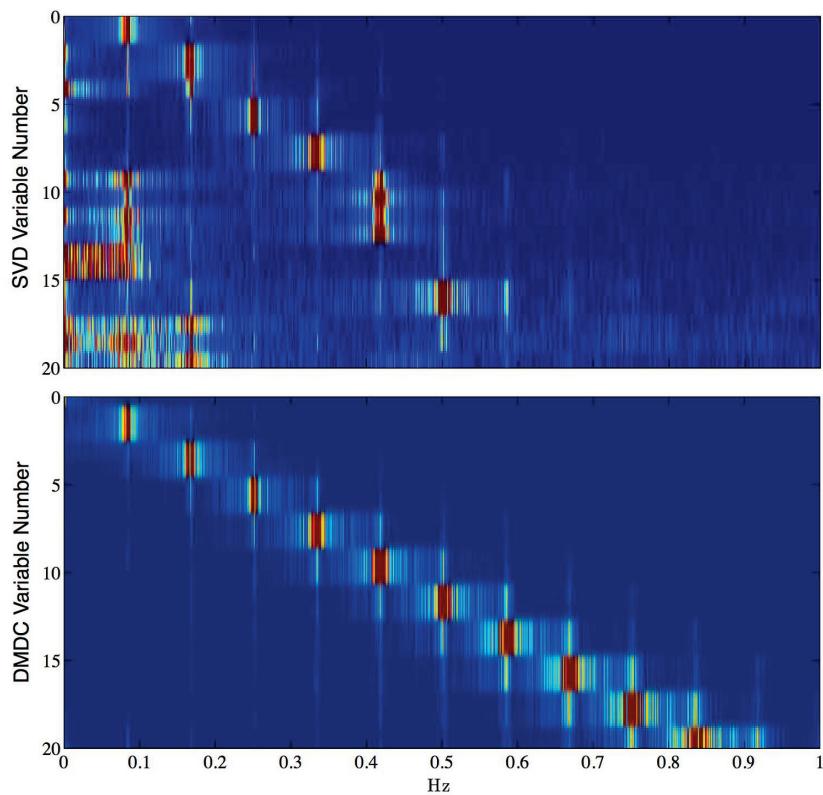


Figure 13. Top: Fourier spectra of SVD modes reveal poor time-scale localization. Bottom: Fourier spectra of DMDC modes show ordering according to largest Fourier mode.

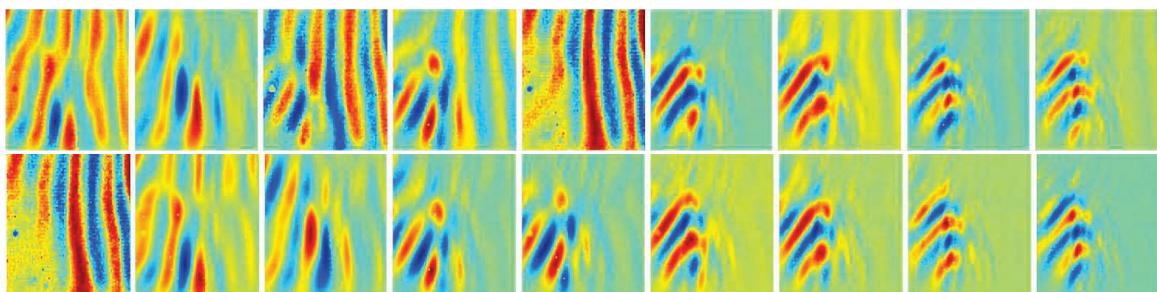


Figure 14. Top: First nine modes of variation found by SVD. Bottom: First nine modes of variation found by DMDC. Images produced by averaging the liquid crystal image weighted by n SVD or DMDC amplitudes, respectively. The associated video (88183_08.mov [local/web 1.88MB]) shows the decomposition of dynamics into DMDC modes.

capture a component of the motion up and to the left, which corresponds to part of the defect being injected into the sample from the bottom. The next highest pair of modes correspond to the nearly vertical motion of the defect traveling upward into the sample. As expected, the DMDC algorithm effectively separates modes by their time scales, giving insight into the system dynamics.

6. Discussion. The main theme of this work is that the methodology used to analyze complex data should depend on the end goal of the analysis. In the case under discussion, we are concerned with reconstructing the topology and geometry of dynamical data, without knowledge of the governing equations. Depending on the purpose, different parts of the reconstructed geometry can be considered intrinsic or extrinsic to the data, the intrinsic geometry being simply that part which relates to the features of the data which one wishes to study. Previous applications of diffusion maps [7] tend to consider the sampling density of a data set an “extrinsic” detail of the geometry, whereas the angles defined by the geometry were considered “intrinsic.”

In this paper we have applied the extrinsic/intrinsic dichotomy more broadly. Dynamical data, unlike generic data, has the additional structure of a time-ordering. There is a more fundamental notion of intrinsic geometry for dynamical systems which goes beyond the geometry of the observations. This is the geometry of the Lyapunov metric [24], which is independent of the embedding space of the dynamical system. Thus, for studying dynamical data, the natural extension of the ideas of diffusion maps is to try to isolate the intrinsic geometry of the Lyapunov metric and remove the extrinsic factors of the observed embedding. This is the focus of sections 2 and 3 of the article. Our use of diffusion maps in the algorithm focuses on its preservation of the local geometry on the reconstructed manifold. The choice of the sampling bias parameter α allows a flexible choice of measure, which is crucial to isolating the slow dynamical modes. The end goal of the DMDC algorithm is to construct a version of harmonic analysis for the observed system, which is specially adapted to the system by using the observed data to approximate its heat kernel.

While the idea of combining nonlinear dimensionality reduction with a time-delay reconstruction has been explored previously, most notably in [19], DMDC provides the theoretical connection between these techniques. Moreover, we have shown that careful adjustment of the time-delay embedding (via our weighting scheme) and the diffusion map algorithm (via careful selection of parameters) is necessary for these techniques to work together optimally.

The theory of nonlinear dimensionality reduction is quickly advancing with an emphasis on understanding the geometry of data, and the theory developed here provides the basis for future techniques to leverage the intrinsic geometry of dynamical systems. For example, it is interesting to compare this technique to the approach of local SVD in the embedded state space, mentioned in the introduction. While a local SVD can preserve, or even finely adjust, the local geometry, it currently cannot generate a global low-dimensional representation. On the other hand, a diffusion map gives a global low-dimensional representation, but it inherits the local geometry given by the embedding and cannot currently leverage the geometric information of a local SVD. Combining the global approach of diffusion maps with the extra information available from a local SVD is a question that has only begun to be addressed (for example, in [29]). This is a fundamental shortcoming of diffusion maps, and future developments should allow promising new extensions of DMDC.

Although the theory in section 2 was written for deterministic dynamics, assuming the existence of a standard Oseledots splitting, the arguments can be adapted to hold in the case of stochastic dynamics, as in the more general splitting of [2]. Here Takens’ theorem would be replaced by a stochastic version as in [30, 31]. Moreover, the basic idea of the theorem, that stable directions are magnified in backward time by the delay coordinates, may apply in

more generality, such as within multidimensional Oseledets subspaces, leading to finer notions of time scale separation.

In section 5, the DMDC algorithm is shown to be effective in identifying low-dimensional dynamics in electroconvection liquid crystal samples. In addition to the examples shown in this article, we have been able to identify multistable dynamics in this system, elicited by different stimulations. We plan to perform more measurements with different stimulation protocols to identify the full dynamics of the system.

We expect the method described in this article to be helpful in a variety of examples, from systems involving several time scales, to systems with mixed-mode oscillations [13], and spatiotemporal dynamics, where high-dimensional observations such as videos are available.

Appendix. The DMDC Algorithm. When the spatial extent of the dynamics is restricted, the intrinsic dimensionality will be small, and thus a diffusion map may be applied directly to the full images simply by taking the L^2 distance between delay embedded images in $\mathbb{R}^{r(s+1)}$ constructed in section 5. However, there are several computational considerations that can dramatically affect this analysis. By careful application of sparse data structures and extensive parallelization of the algorithm we have found that the diffusion map can be effectively applied to video data using consumer-grade hardware.

As in section 5, for each image we form a high-dimensional state vector by concatenating as many of the previous images as possible in analogy to the classical time-delay embedding. Since this creates an extremely high-dimensional vector, we need a preliminary reduction of the dimension to make the k -nearest neighbor algorithm feasible. However, we need to preserve the geometry of the embedding, so the initial reduction will be to an intermediate dimension on the order of 10^3 , much larger than the latent dimension (order of 10) but much smaller than the delay embedding dimension (order of 10^6). The preliminary dimensionality reduction is simply a linear projection from the delay embedding space to the intermediate space. This projection is chosen by generating a matrix of Gaussian random entries, which is then made orthonormal using a QR-decomposition (the R matrix is discarded). At this point we are ready to apply the diffusion map to the data represented in the intermediate space.

The full diffusion map algorithm is impractical because it requires the full matrix of Euclidean distances between all pairs of state vectors, and this requires too much memory. However, in [10] it was shown that for sufficiently high k the sparse matrix containing the k -nearest neighbors of each state vector is sufficient. This is still a computationally intensive algorithm since the state vectors are too high-dimensional to apply kd-tree or ball-tree data structures. However, the k -nearest neighbors algorithm is highly parallelizable, and it was recently shown in [16] that inexpensive graphics processing units (GPUs) could achieve a significant acceleration of the k -nearest neighbor algorithm. In our test we saw up to 200-times speed-up using the GPU over single-processor CPU applications. The remainder of the diffusion map algorithm is straightforward and is detailed in the DMDC algorithm summary below.

The parameter ϵ is the diameter of the neighborhood that is used to approximate the tangent plane to the manifold. The results of section 3 involve the limit as $\epsilon \rightarrow 0$; however, for the discrete approximations to converge we require the amount of data to go to infinity. In [7], the number of sample points is required to grow faster than $\epsilon^{n/4-1/2}$, where n is the

Diffusion Mapped Delay Coordinates (DMDC) Algorithm.

1. Let $x_i \in \mathbb{R}^r$ be the i th frame of a video with N frames of r pixels each.
2. Choose bias parameter $\alpha \in \{1/2, 1\}$, weight $\kappa > 0$, and number of delays s .
3. For $i = s+1, \dots, N$ form the state vector,

$$y_i = [x_i, e^{-\kappa} x_{i-1}, \dots, e^{-s\kappa} x_{i-s}]^T \in \mathbb{R}^{r(s+1)}.$$

4. Generate an $m \times r(s+1)$ -dimensional matrix π of Gaussian random numbers.
5. Use a QR-decomposition to orthonormalize the rows of π to form $\hat{\pi}$.
6. Form the compressed states $\hat{y}_i = \hat{\pi} y_i$.
7. For each i find the k -nearest neighbors of \hat{y}_i in \mathbb{R}^m , and let their indices be $I(i, j)$ for $j = 1, \dots, k$, ordered by increasing distance.
8. Form a sparse $(N-s) \times (N-s)$ matrix with $(N-s)k$ nonzero entries given by

$$d(i, I(i, j)) = \|\hat{y}_i - \hat{y}_{I(i, j)}\|.$$

9. Pick k_{\min} as described above, and set $\epsilon = \text{mean}_i \{d(i, I(i, k_{\min}))^2\}$.
10. Form the sparse matrix $\hat{d}(i, I(i, j)) = \exp\{-d(i, I(i, j))^2/\epsilon\}$.
11. Form the symmetric matrix $J = (\hat{d} + \hat{d}^T)/2$.
12. Form the diagonal normalization matrix $P_{ii} = \sum_j J_{ij}$.
13. Normalize to form kernel matrix $K = P^{-\alpha} J P^{-\alpha}$.
14. Form the diagonal normalization matrix $Q_{ii} = \sum_j K_{ij}$.
15. Form the symmetric matrix $\hat{T} = Q^{-1/2} K Q^{-1/2}$.
16. Find the $L+1$ largest eigenvalues a_l and associated eigenvectors ξ_l of \hat{T} .
17. Compute the eigenvalues of $\hat{T}^{1/\epsilon}$ by $\lambda_l^2 = a_l^{1/\epsilon}$.
18. Compute the eigenvectors of the matrix $T = Q^{-1} K$ by $\psi_l = Q^{-1/2} \xi_l$.
19. Compute the diffusion map variables at time scale t by

$$\Psi_{\alpha, t}(x_i) = [\lambda_1^t \psi_1(x_i), \dots, \lambda_L^t \psi_L(x_i)]^T.$$

20. The l th DMDC mode $(\psi_l)_i = \psi_l(x_{i+s})$ is a time series at time scale $\frac{-1}{\log(\lambda_l)}$.
21. Form the l th DMDC projection by $q_l = Q\psi_l$.
22. Form the l th DMDC component by $X_l = \sum_{i=s+1}^N x_i (q_l)_{i-s}$; this is an image which shows which parts of the video are relevant to the l th DMDC mode.
23. Choose any collection L_0 of indices of interesting DMDC components.
24. Construct a video by taking the i th frame to be $\hat{x}_i = \sum_{l \in L_0} \lambda_l (\psi_l)_{i-s} X_l$ for $i = s+1, \dots, N$. This projects the original video onto the modes of L_0 .

dimension of the underlying manifold. However, in most cases we have a fixed data set that we wish to analyze, so a limiting requirement on ϵ is not very useful.

There are many practical ways of choosing ϵ . The theoretical approach in [27] may become useful if this algorithm can be adapted into an iterative form which continually updates as new

data becomes available. Currently we assume that the data set is fixed, and therefore we take a pragmatic approach to choosing ϵ . Since we are assuming that the local structure can be well represented by the k -nearest neighbors, this implies that a transition probability to any state beyond the k -nearest neighbors should be near zero. On the other hand, if the transition probabilities decay too quickly, the states will quickly become isolated and the matrix will not have numerically stable eigenvalues. To balance these effects we have had very good results with $\epsilon = \text{mean}_i\{d(i, I(i, k_{\min}))\}$, where $I(i, k_{\min})$ is the index of the k_{\min} th nearest neighbor to the i th vector and $d(\cdot, \cdot)$ is the distance between the referenced data points. In general k_{\min} will be small, and we took $k_{\min} = 6$ in this paper. The choice of k_{\min} too small will lead to ϵ being too small, and the eigenvector problem will be ill posed. Choosing k_{\min} too large leads to ϵ being too large, and the approximation of the heat kernel will be poor.

As described in section 3, the parameter α chooses the influence of the sampling density on the heat kernel constructed by diffusion maps. Intuitively α is a bias parameter, and for $\alpha = 1$ the heat kernel is unbiased by the sampling density. If the video represents a dynamical system that has an invariant measure, and if the sampling density can be assumed to be the same as the invariant measure, then setting $\alpha = 1/2$ will adapt the analysis to the invariant measure. In this case, the first DMDC mode, with eigenvalue $\lambda_0 = 1$ (and hence time scale ∞), will be an approximation of the invariant measure. In this paper we used $\alpha = 1/2$ for all the examples.

We now summarize all the steps of DMDC. Note that the algorithm works unchanged when $x_i \in \mathbb{R}^r$ is a time series in any data format, where the index i represents time. In this summary we assume that the x_i are images in a video, because many elements of DMDC have a natural interpretation in this case. Steps 1–3 build the weighted delay embedding explored in section 2. Steps 4–6 are an optional precompression for extremely high-dimensional data (such as videos); without the precompression the k -nearest neighbor algorithm may be infeasible depending on the embedding dimension $r(s + 1)$. Steps 7–19 are a computationally efficient version of the diffusion maps algorithm which was introduced in [7]. Finally, in steps 20–24 we use the DMDC to separate time scales in the time series.

REFERENCES

- [1] D. AYEYELS, *Generic observability of differentiable systems*, SIAM J. Control Optim., 19 (1981), pp. 595–603.
- [2] L. ARNOLD, *Random Dynamical Systems*, Springer-Verlag, New York, 1998.
- [3] D. BARKLEY, *Spiral meandering*, in *Chemical Waves and Patterns*, R. Kapral and K. Showalter, eds., Springer-Verlag, New York, 1995, pp. 163–190.
- [4] D. BARKLEY AND I. G. KEVREKIDIS, *A dynamical systems approach to spiral-wave dynamics*, Chaos, 4 (1994), pp. 453–460.
- [5] D. S. BROOMHEAD AND G. P. KING, *Extracting qualitative dynamics from experimental data*, Phys. D, 20 (1986), pp. 217–236.
- [6] R. COIFMAN, R. ERBAN, A. SINGER, AND I. KEVREKIDIS, *Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 16090–16095.
- [7] R. COIFMAN AND S. LAFON, *Diffusion maps*, Appl. Comput. Harmonic Anal., 21 (2006), pp. 5–30.
- [8] R. COIFMAN, I. G. KEVREKIDIS, S. LAFON, M. MAGGIONI, AND B. NADLER, *Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems*, Multiscale Model. Simul., 7 (2008), pp. 842–864.

- [9] R. COIFMAN, S. LAFON, B. NADLER, AND I. KEVREKIDIS, *Diffusion maps, spectral clustering and reaction coordinates of dynamical systems*, Appl. Comput. Harmonic Anal., 21 (2006), pp. 113–127.
- [10] R. COIFMAN AND M. MAGGIONI, *Diffusion wavelets*, Appl. Comput. Harmonic Anal., 21 (2006), pp. 53–94.
- [11] G. DANGELMAYR, G. ACHARYA, J. GLEESON, I. OPREA, AND J. LADD, *Diagnosis of spatiotemporal chaos in wave-envelopes of a nematic electroconvection pattern*, Phys. Rev. E, 79 (2009), pp. 46215–46235.
- [12] M. DENNIN, D. S. CANNELL, AND G. AHLERS, *Patterns of electroconvection in a nematic liquid crystal*, Phys. Rev. E, 57 (1998), pp. 638–649.
- [13] M. DESROCHES, J. GUCKENHEIMER, B. KRAUSKOPF, C. KUEHN, H. M. OSINGA, AND M. WECHSELBERGER, *Mixed-mode oscillations with multiple time scales*, SIAM Rev., 54 (2012), pp. 211–288.
- [14] J.-P. ECKMANN AND D. RUELLE, *Ergodic theory of chaos and strange attractors*, Rev. Modern Phys., 57 (1985), pp. 617–656.
- [15] D. EGOLF, I. MELNIKOV, W. PESCHE, AND R. ECKE, *Mechanisms of extensive spatiotemporal chaos in Raleigh-Bénard convection*, Nature, 404 (2000), pp. 733–736.
- [16] V. GARCIA, E. DEBREUVE, AND M. BARLAUD, *Fast k-nearest neighbor search using GPU*, in Proceedings of the CVPR Workshop on Computer Vision on GPU, Anchorage, AK, 2008.
- [17] C. W. GEAR, T. J. KAPER, I. G. KEVREKIDIS, AND A. ZAGARIS, *Projecting to a slow manifold: Singularly perturbed systems and legacy codes*, SIAM J. Appl. Dyn. Syst., 4 (2005), pp. 711–732.
- [18] P. G. DE GENNES AND J. PROST, *The Physics of Liquid Crystals*, Clarendon Press, Oxford, UK, 1993.
- [19] D. GIANNAKIS AND A. J. MAJDA, *Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability*, Proc. Natl. Acad. Sci. USA, 109 (2012), pp. 2222–2227.
- [20] J. JOST, *Riemannian Geometry and Geometric Analysis*, Springer-Verlag, Berlin, 2002.
- [21] J. A. LEE AND M. VERLEYSEN, *Nonlinear Dimensionality Reduction*, Springer-Verlag, New York, 2007.
- [22] V. I. OSELEDETS, *A multiplicative ergodic theorem*, Trans. Moscow Math. Soc., 19 (1968), pp. 197–231.
- [23] N. H. PACKARD, J. P. CRUTCHFIELD, J. D. FARMER, AND R. S. SHAW, *Geometry from a time series*, Phys. Rev. Lett., 45 (1980), pp. 712–716.
- [24] Y. B. PESIN, *Families of invariant manifolds corresponding to nonzero characteristic exponents*, Math. USSR Izvestia, 10 (1976), pp. 1261–1305.
- [25] T. SAUER, *Time series prediction by using delay coordinate embedding*, in Time Series Prediction: Forecasting the Future and Understanding the Past, A. S. Weigend and N. A. Gershenfeld, eds., Addison-Wesley, Harlow, UK, 1994, pp. 175–193.
- [26] T. SAUER, J. A. YORKE, AND M. CASDAGLI, *Embedology*, J. Statist. Phys., 65 (1991), pp. 579–616.
- [27] A. SINGER, *From graph to manifold Laplacian: The convergence rate*, Appl. Comput. Harmonic Anal., 21 (2006), pp. 128–134.
- [28] A. SINGER, R. ERBAN, I. G. KEVREKIDIS, AND R. COIFMAN, *Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 16090–16095.
- [29] A. SINGER AND H.-T. WU, *Vector diffusion maps and the connection Laplacian*, Comm. Pure Appl. Math., 65 (2012), pp. 1067–1144.
- [30] J. STARK, *Delay embeddings for forced systems. I. Deterministic forcing*, J. Nonlinear Sci., 9 (1999), pp. 255–332.
- [31] J. STARK, D. S. BROOMHEAD, M. E. DAVIES, AND J. HUKE, *Delay embeddings for forced systems. II. Stochastic forcing*, J. Nonlinear Sci., 13 (2003), pp. 519–577.
- [32] F. TAKENS, *Detecting strange attractors in turbulence*, in Dynamical Systems and Turbulence, D. Rand and L.-S. Young, eds., Lecture Notes in Math. 898, Springer, Berlin, Heidelberg, 1981, pp. 366–381.
- [33] J. TENENBAUM, V. SILVA, AND J. LANGFORD, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2324.
- [34] T. TOH-KATONA, J. R. CRESSMAN, W. GOLBURG, AND J. GLEESON, *Persistent global power fluctuations near a dynamic transition in electroconvection*, Phys. Rev. E, 68 (2003), 103010.
- [35] A. ZAGARIS, H. G. KAPER, AND T. J. KAPER, *Two perspectives on reduction of ordinary differential equations*, Math. Nachr., 278 (2005), pp. 1629–1642.