Detecting the slow manifold by anisotropic diffusion maps

Amit Singer[*†] **Radek Erban**[‡] **Ioannis G. Kevrekidis**[§] **Ronald R. Coifman**[†]

[†]Yale University, Department of Mathematics, New Haven, CT 06520,[‡]University of Oxford, Mathematical Institute, 24-29 St. Giles', Oxford, OX1 3LB, United Kingdom, and [§]Chemical Engineering and PACM, Princeton University, Princeton, NJ 08544, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Non-linear independent component analysis is combined with diffusion map data analysis techniques to detect good observables in high-dimensional dynamic data. The widely applicable procedure, a crucial step in model reduction approaches, is illustrated on stochastic chemical reaction network simulations.

diffusion maps | slow manifold | dimensionality reduction | chemical reactions

Evolution of dynamical systems often occurs on two or more time scales. A simple deterministic example is given by the coupled system of ODEs

$$\mathrm{d}u/\mathrm{d}t = \alpha(u,v), \quad [1]$$

$$\mathrm{d}v/\mathrm{d}t = \tau^{-1}\beta(u,v), \quad [2]$$

with the small parameter $0 < \tau \ll 1$ and $\alpha(u,v)$ and $\beta(u,v)$ are $O(1)$. For any given initial condition (u_0, v_0), already at $t = O(\tau)$ the system approaches a new value (u_0, v), where v satisfies the asymptotic relation $\beta(u_0, v) = 0$. Although the system is fully described by two coordinates, the relation $\beta(u,v) = 0$ defines a slow one-dimensional manifold which approximates the slow dynamics for $t \gg \tau$. In this example it is clear that v is the fast variable while u is the slow one. Projecting onto the slow manifold here is rather easy: the fast foliation is simply "vertical", i.e. $u = \text{const}$. However, when we observe the system in terms of the variables $x = x(u,v)$ and $y = y(u,v)$ which are unknown non-linear functions of u and v, then the "observables" x and y have both fast and slow dynamics. Projecting onto the slow manifold becomes nontrivial, because the transformation from (x,y) to (u,v) is unknown. Detecting the existence of a slow manifold under these conditions and projecting onto it are important in any model reduction technique. Knowledge of a good parametrization of such a slow manifold is a crucial component of the equation-free framework for modeling and computation of complex/multiscale systems [1, 2, 3].

In recent years, diffusion maps [4, 5, 6, 7, 8, 9] have been used to detect low-dimensional, nonlinear manifolds underlying high-dimensional data sets. In this paper we combine diffusion maps with recent tools from non-linear independent component analysis [10] to detect slow variables in high-dimensional data arising from dynamic model simulations. The proposed algorithm takes into account the time dependence of the data, whereas in the diffusion map approach the time labeling of the data points is not included. We demonstrate our algorithm for stochastic simulators arising in the context of chemical/biochemical reaction modeling.

Multiscaled chemical reactions: a toy example

Consider the reversible chemical reaction (a dimerization, which is a part of several biochemical mechanisms [11, 12]) involving two molecular species X and Y

$$X + X \underset{k_2}{\overset{k_1}{\rightleftarrows}} Y \quad [3]$$

where k_1 and k_2 are the forward and backward rate constants. The probability that an additional molecule of type Y is produced from two X molecules (resp. two molecules of X produced from one molecule of Y) in an infinitesimally small time interval $[t, t+dt]$ is $k_1 X(t)(X(t)-1)dt$ (resp. $k_2 Y(t)dt$), where $X(t)$ and $Y(t)$ are the number of molecules of type X and Y at time t [13]. The chemical reaction (3) satisfies the stoichiometric conservation law

$$X(t) + 2Y(t) = \text{const}, \quad [4]$$

so that the state vector $[X(t), Y(t)]$ is restricted to a line in the phase plane. We now couple the chemical reaction (3) with a slow production of X molecules from an external source

$$\emptyset \xrightarrow{k_3} X \quad [5]$$

where equation (5) means that the probability of the production of an additional molecule of type X in an infinitesimally small time interval $[t, t+dt]$ is $k_3 dt$; the rate constants and the initial state are chosen in such a way that the production process (5) is much slower than the dimerization reactions (3). This is the case, for example, for the following choice of parameters:

$$X(0) = 100, \;\; Y(0) = 100, \;\; k_1 = 1, \;\; k_2 = 100, \;\; k_3 = 50. \quad [6]$$

The average time to produce an additional X molecule is $k_3^{-1} = 0.02$, whereas the average times for the forward and backward dimerization are $(k_1 X(0)(X(0)-1))^{-1} \approx 10^{-4}$ and $(k_2 Y(0))^{-1} = 10^{-4}$. This implies that both X and Y are fast variables; yet their linear combination $X + 2Y$ is a slow variable. The conservation law (4) no longer holds since production was added. Instead, $X + 2Y$ is slowly growing. To confirm this fact, we simulate the time evolution of the pair $[X(t), Y(t)]$ using the Gillespie stochastic simulation algorithm (SSA) [13]. In Figure 1, we plot the time evolution of X, Y and $X + 2Y$.

This naturally leads to the following question: How to detect the slow variable $X + 2Y$ from data? A priori knowledge that we seek a linear combination of the original variables lends itself to fitting the coefficients of such a combination. Such fitting is however not possible for the general nonlinear case.

Conflict of interest footnote placeholder

Insert 'This paper was submitted directly to the PNAS office.' when applicable.

*To whom correspondence should be addressed. E-mail: amit.singer@yale.edu

Short simulation bursts

It is convenient to analyze our approach in the diffusion limit, for which the simulation is well-approximated by a stochastic differential equation (SDE). The chemical Langevin equation for the time evolution of X and Y, which is formally derived from the corresponding chemical master equation, is given in the Itô form by [14, 15, 16]

$$\begin{aligned} \mathrm{d}x &= (2k_2y - 2k_1x(x-1) + k_3)\,\mathrm{d}t \\ &\quad -2\sqrt{k_1x(x-1)}\,\mathrm{d}w_1 + 2\sqrt{k_2y}\,\mathrm{d}w_2 + \sqrt{k_3}\,\mathrm{d}w_3, \end{aligned} \quad [7]$$

$$\begin{aligned} \mathrm{d}y &= (k_1x(x-1) - k_2y)\,\mathrm{d}t \\ &\quad +\sqrt{k_1x(x-1)}\,\mathrm{d}w_1 - \sqrt{k_2y}\,\mathrm{d}w_2, \end{aligned} \quad [8]$$

where w_i $(i = 1, 2, 3)$ are standard independent Brownian motions. The approximation (7)-(8) is also characterized by a time scale separation and possesses the slow variable $x+2y$; multiplying (8) by 2 and adding to (7) gives

$$\mathrm{d}(x + 2y) = k_3\,\mathrm{d}t + \sqrt{k_3}\,\mathrm{d}w_3. \quad [9]$$

Equation (9) shows that the approximated stochastic dynamics of $x + 2y$ is decoupled from the individual dynamics of x and y, as expected from (3)-(4).

The Euler-Maruyama method for (7)-(8) suggests that in a time step Δt, the state vector $[x(t), y(t)]$ propagates to the random state vector $[x(t + \Delta t), y(t + \Delta t)]$

$$\begin{aligned} x(t+\Delta t) &\approx x(t) + (2k_2y(t) - 2k_1x(t)(x(t)-1) + k_3)\,\Delta t \\ &\quad -2\sqrt{(k_1x(t)(x(t)-1) + k_2y(t))}\,Z_1 + \sqrt{k_3}\,Z_2, \\ y(t+\Delta t) &\approx y(t) + (k_1x(t)(x(t)-1) - k_2y(t))\,\Delta t \\ &\quad + \sqrt{(k_1x(t)(x(t)-1) + k_2y(t))}\,Z_1, \end{aligned}$$

where $Z_1, Z_2 \sim \mathcal{N}(0, \Delta t)$ are independent normally distributed random variables with zero mean and variance Δt (Z_1 corresponds to the $\mathrm{d}w_1$ and $\mathrm{d}w_2$ terms in (7)-(8)). This means that if we were to run many simulations for a short time step Δt, all starting at $[x(t), y(t)]$, the trajectories would end up at random locations forming a point cloud in the phase plane. The point cloud has a bivariate normal distribution, whose center is located at $\boldsymbol{\mu} = [\mu_x, \mu_y]^T$, given by

$$\begin{aligned} \mu_x &= x(t) + (2k_2y(t) - 2k_1x(t)(x(t)-1) + k_3)\,\Delta t, \\ \mu_y &= y(t) + (k_1x(t)(x(t)-1) - k_2y(t))\,\Delta t, \end{aligned}$$

and whose 2-by-2 covariance matrix Σ is

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T,$$

where

$$\mathbf{B} = \sqrt{\Delta t}\begin{pmatrix} -2\sqrt{k_1x(t)(x(t)-1) + k_2y(t)} & \sqrt{k_3} \\ \sqrt{k_1x(t)(x(t)-1) + k_2y(t)} & 0 \end{pmatrix}.$$

The shape of the point cloud is an ellipse, because the level lines of the probability density function

$$p(x, y) = \frac{1}{2\pi\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}$$

are ellipses ($\boldsymbol{x} = [x, y]^T$). When there is a separation of time scales, the ellipses are thin and elongated. For example, for the set of parameters given in (6), the eigenvalues of $\boldsymbol{\Sigma}$ for $[x, y] = [100, 100]$ are $\sigma_1^2 \approx 10^5\Delta t$ and $\sigma_2^2 \approx 10\Delta t$. This means that the long axis of the ellipse is two orders of magnitude longer than the short axis ($\sigma_1/\sigma_2 \approx 10^2$). The eigenvector corresponding to σ_1 is approximately $[-2, 1]^T$, pointing in the direction of the fast dynamics on the line $x + 2y = \text{const}$. The second eigenvector is approximately $[1, 2]^T$, pointing in the direction of the slow dynamics.

The eigen-decomposition of the covariance matrix is simply the principal component analysis (PCA) of the local point cloud generated by the short simulation burst. We produce many short simulation bursts starting at different initialization points $[x, y]$. For each burst we perform the PCA and estimate its covariance matrix $\boldsymbol{\Sigma}_{(x,y)}$. The principal components of $\boldsymbol{\Sigma}_{(x,y)}$ are the local directions of the rapidly changing variables at $[x, y]$, whereas components with small eigenvalues correspond to the slow variables.

We wish to piece together the locally defined components into globally consistent coordinates. The toy model (3)-(5) presents no special difficulty, because the principal components of $\boldsymbol{\Sigma}_{(x,y)}$ are approximately $[-2, 1]$ and $[1, 2]$ everywhere (independent of $[x, y]$). In general, however, the slow variable may be some complicated non-linear function of the state variables. In such cases, it is not trivial to find a globally consistent slow coordinate.

Anisotropic diffusion maps

To integrate the local information into global coordinates we use anisotropic diffusion maps (ADM), introduced in [10]. Suppose $u = u(x, y) = x + 2y$ (resp. $v = v(x, y) = -2x + y$) are the slowly changing (resp. the rapidly changing variables). Together they define a map $g : (x, y) \mapsto (u, v)$ from the observable state variables x and y to the "dynamically meaningful" coordinates u and v. Alternatively, the inverse map $f \equiv g^{-1} : (u, v) \mapsto (x, y)$ is given by $x = x(u, v)$ and $y = y(u, v)$. The point cloud in the observable (x, y)-plane, generated by the short bursts, is the image under f of a similar point cloud in the inaccessible (u, v)-plane. The slow manifold (curve) in the (x, y)-plane can be thought of as the image of u-axis, $f(u, 0) = [x(u, 0), y(u, 0)]$. The ellipses in the (u, v)-plane are also thin and elongated, and they share an important property: they all have the v-axis as their long axis and the u-axis as their short axis, due to the separation of time scales. The ratio between the eigenvalues of $\boldsymbol{\Sigma}$ defines a small parameter $0 < \tau^2 \ll 1$ that measures the time scale separation. In other words, the change in u in a small time step Δt is typically τ times smaller the amount of change in v. The parameter $\tau = \tau(u)$ can also be a function of u, allowing the possibility of different variability of the rapid dynamics for different values of u. This suggests to define the scaled variable $v_\tau = \tau v$. This scaling contracts the elongated ellipse in the (u, v)-plane into a circle in the (u, v_τ)-plane.

Now that we showed how to identify ellipses in the observable space that are images of circular disks in the inaccessible space, we are in position to use the result of [10], that relates the anisotropic graph Laplacian in the observable space with the (isotropic) graph Laplacian in the inaccessible space. We formulate our method in a general setting. Then we apply it to the toy example.

The construction of the ADM is performed as follows. Suppose $\boldsymbol{x}^{(i)} \in \mathbb{R}^M$, $i = 1, \ldots, N$, are N data points in an M-dimensional data space. For every data point $\boldsymbol{x}^{(i)} =$

$[x_1^{(i)}, x_2^{(i)}, \ldots, x_M^{(i)}]$, $i = 1, \ldots, N$, we generate an ensemble of short simulation bursts initialized at the data point, i.e. $\boldsymbol{x}(0) \equiv [x_1(0), x_2(0), \ldots, x_M(0)] = \boldsymbol{x}^{(i)}$. We collect the statistics of the simulated trajectories after a short time period Δt. In particular, we compute the averaged position $\mu^{(i)} = [\mu_1^{(i)}, \ldots, \mu_M^{(i)}]$

$$\mu_j^{(i)} = \left\langle x_j(\Delta t) \,|\, \boldsymbol{x}(0) = \boldsymbol{x}^{(i)} \right\rangle, \qquad j = 1, \ldots, M, \qquad [10]$$

and the elements of the covariance matrix

$$\boldsymbol{\Sigma}^{(i)} = \{\sigma_{jk}^{(i)}\}_{j,k=1}^{M}$$

by

$$\sigma_{jk}^{(i)} = \frac{1}{\Delta t}\left[\left\langle x_j(\Delta t)\, x_k(\Delta t) \,|\, \boldsymbol{x}(0) = \boldsymbol{x}^{(i)} \right\rangle - \mu_j^{(i)}\mu_k^{(i)}\right] \qquad [11]$$

where the notation $\langle\cdot\rangle$ stands for statistical averaging over many simulated trajectories. For each data point $\boldsymbol{x}^{(i)}$, we calculate ${\boldsymbol{\Sigma}^{(i)}}^{-1}$, the inverse of the covariance matrix. We define a symmetric $\boldsymbol{\Sigma}$-dependent squared distance between pairs of data points in the observable (x, y)-space

$$d^2_{\boldsymbol{\Sigma}}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \qquad [12]$$

$$= \frac{1}{2}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)})^T \left(\left(\boldsymbol{\Sigma}^{(i)}\right)^{-1} + \left(\boldsymbol{\Sigma}^{(j)}\right)^{-1}\right)(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}).$$

Note that for the toy model (3)-(5) the distance $d_{\boldsymbol{\Sigma}}$ is a second order approximation of the Euclidean distance in the inaccessible (u, v_τ)-space

$$d^2_{\boldsymbol{\Sigma}}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) \approx (u^{(i)} - u^{(j)})^2 + \tau^2(v^{(i)} - v^{(j)})^2. \qquad [13]$$

Because τ is a small parameter, $d_{\boldsymbol{\Sigma}}$ is controlled by the difference in the slow coordinate. The approximation (13) is also valid in higher dimensions, where they may be more than one slow coordinate ($\boldsymbol{u}$) and several fast coordinates ($\boldsymbol{v}$) and the ellipse is replaced by an ellipsoid. In such cases,

$$d^2_{\boldsymbol{\Sigma}}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) \approx \|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(j)}\|^2 + \tau^2\|\boldsymbol{v}^{(i)} - \boldsymbol{v}^{(j)}\|^2. \qquad [14]$$

Therefore, the ADM based on the "dynamic proximity" $d_{\boldsymbol{\Sigma}}$ approximates the Laplacian on the slow manifold. We construct an $N \times N$ weight matrix $\mathbf{W}$

$$W_{ij} = \exp\left\{-\frac{d^2_{\Sigma}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})}{\varepsilon^2}\right\}, \qquad [15]$$

where $\varepsilon > 0$ is the single parameter of the method. The elements of the matrix $\mathbf{W}$ are all less than or equal to one. Nearby points (i.e., their projection on the slow manifold is close) have W_{ij} close to one, whereas distant points have W_{ij} close to zero. Next, we define a diagonal $N \times N$ normalization matrix $\mathbf{D}$ whose values are given by the row sums of $\mathbf{W}$

$$D_{ii} = \sum_{k=1}^{N} W_{ik}.$$

We then compute the eigenvalues and right eigenvectors of the row stochastic matrix

$$\mathbf{A} = \mathbf{D}^{-1}\mathbf{W}. \qquad [16]$$

As discussed in [17, 5, 18, 19], the leading eigenvectors may be used as a basis for a low dimensional representation of the data. To compute those eigenvectors, we use the fact that $\mathbf{A} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{1/2}$ where $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is a symmetric matrix. Hence, $\mathbf{A}$ and $\mathbf{S}$ are similar and thus have the same spectrum. Since $\mathbf{S}$ is symmetric, it has a complete set of eigenvectors $\mathbf{q}_j$, $j = 0, \ldots, N-1$, with corresponding eigenvalues

$$\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1}. \qquad [17]$$

The right eigenvectors of $\mathbf{A}$ are given by

$$\mathbf{u}_j = \mathbf{D}^{-1/2}\mathbf{q}_j. \qquad [18]$$

Since $\mathbf{A}$ is a Markov matrix, all its eigenvalues are smaller than or equal to one, with largest eigenvalue $\lambda_0 = 1$ and a corresponding trivial eigenvector $\mathbf{u}_0 = [1, 1, \ldots, 1]$. We define the low ($n$-)dimensional representation of the state vectors by the following ADM

$$\Psi_n : \boldsymbol{x}^{(i)} \to \left[u_1^{(i)}, u_2^{(i)}, \ldots, u_n^{(i)}\right]; \qquad [19]$$

that is, the point $\boldsymbol{x}^{(i)}$ is mapped to a vector containing the i-th coordinate of each of the first n leading eigenvectors of the matrix $\mathbf{A}$. The variables $u_1^{(i)}$, $u_2^{(i)}$, ..., $u_n^{(i)}$ (which are defined on the data points) are the candidate slow variables that we were looking for.

Application of ADM to the toy example

We use $N = 2000$ data points $\boldsymbol{x}^{(i)} \equiv [x_1^i, x_2^i] = [X^{(i)}, Y^{(i)}]$, $i = 1, \ldots, 2000$, uniformly sampled from the illustrative trajectory of Figure 1 (in fact, the trajectory in Figure 1 is visualized using these 2000 data points). For every data point $\boldsymbol{x}^{(i)} = [X^{(i)}, Y^{(i)}]$, $i = 1, \ldots, 2000$, we run 10^7 replicas of stochastic simulations initialized at the data point for time $\Delta t = 10^{-4}$. We estimate $\mu_j^{(i)}$ and $\sigma_{jk}^{(i)}$, $i = 1, \ldots, 2000$, $j = 1, 2$, $k = 1, 2$ by (10)–(11) as an average over 10^7 realizations. For each data point $\boldsymbol{x}^{(i)} = [X^{(i)}, Y^{(i)}]$, we also calculate the inverse covariance matrix and the symmetric $\boldsymbol{\Sigma}$-dependent squared distance $d^2_{\boldsymbol{\Sigma}}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ by (12). We construct a 2000×2000 weight matrix $\mathbf{W}$ by (15) for $\varepsilon = 0.1$ and a matrix $\mathbf{A}$ by (16). We compute the leading eigenvectors $\mathbf{u}_j$ of $\mathbf{A}$ by (18). In Figure 2, we plot our data set where the points are colored according to the first nontrivial eigenvector $\mathbf{u}_1$. We see that the eigenvector $\mathbf{u_1}$ gives a good description of slow dynamics of this system. The slow dynamics is given by function $X + 2Y$ as can be seen in the right panel of Figure 1. The plot of $\mathbf{u_1}$ vs. $X + 2Y$ is shown in middle panel of Figure 2. We again confirm that we obtained a good slow description of the system. Finally, plotting the eigenvector $\mathbf{u_1}$ vs. X confirms that X is not a good slow variable (right panel of Figure 2).

Oscillating half-moons

Next, consider the system of stochastic differential equations

$$\mathrm{d}u = a_1\,\mathrm{d}t + a_2\,\mathrm{d}w_1, \qquad [20]$$

$$\mathrm{d}v = a_3(1 - v)\,\mathrm{d}t + a_4\,\mathrm{d}w_2, \qquad [21]$$

where a_i, $i = 1, 2, 3, 4$, are constants and $\dot{w}_i$, $i = 1, 2$ are independent δ-correlated white noises (Wiener processes). We consider (20)–(21) together with the following nonlinear transformation of variables

$$x = v\cos(u + v - 1), \qquad y = v\sin(u + v - 1). \qquad [22]$$

We will assume that the observables x and y are the actual observables, while u and v are unknown. We choose the values of parameters as: $a_1 = a_2 = 10^{-3}$, $a_3 = a_4 = 10^{-1}$. The illustrative trajectory which starts at $[x(0), y(0)] = [1, 0]$ is plotted in the left panel of Figure 3. The trajectory is colored according to time. We run simulations for a longer time 8×10^4, which accounts for about 12-13 periods, and record 2000 data points at equidistant time intervals of length $8 \times 10^4/2000 = 40$. This data set is plotted in the middle panel of Figure 3. Again, points are colored according to time. We clearly see that there is no correlation between time and the slow variable (which is u MOD 2π) because of oscillations.

To apply the ADM, we run 10^6 replicas of stochastic simulations initialized at each data point $\boldsymbol{x}^{(i)} = [x^{(i)}, y^{(i)}]$ for a time step $\Delta t = 0.1$ and estimate $\mu_j^{(i)}$ and $\sigma_{jk}^{(i)}$, $i = 1, \ldots, 2000$, $j = 1, 2$, $k = 1, 2$ by (10)–(11) as an average over 10^6 realizations. For each data point $\boldsymbol{x}^{(i)} = [x^{(i)}, y^{(i)}]$, we also calculate the inverse covariance matrix and the symmetric $\boldsymbol{\Sigma}$-dependent squared distance $d^2_{\boldsymbol{\Sigma}}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ by (12). Next, we have to choose the value of parameter ε. To do that, we construct the ε-dependent 2000×2000 weight matrix $\mathbf{W} \equiv \mathbf{W}(\varepsilon)$ by (15) for several values of ε. Then we compute

$$L(\varepsilon) = \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}(\varepsilon). \qquad [23]$$

The function $L(\varepsilon)$ is plotted in the right panel of Figure 3 (it is a log-log plot) [20]. It clearly has two constant asymptotes when $\varepsilon \to 0$ and $\varepsilon \to \infty$; as we expect, these asymptotes are smoothly connected, by an approximately straight line of slope d in a log-log plot, where d is the dimension of the slow manifold. Thus, the log-log plot of $L(\varepsilon)$ suggests to choose ε where the log-log graph of $L(\varepsilon)$ appears linear. We choose $\varepsilon = 6$. We form $\mathbf{A}$ (by (16)) and compute its few leading eigenvectors $\mathbf{u}_j$ by (18). The first nontrivial eigenvector $\mathbf{u}_1$ then describes the slow dynamics of the system. The data set (colored by the values of $\mathbf{u}_1$) is plotted in Figure 4 (left panel). We see that the ADM provides a good description of the slow dynamics. Plotting $\mathbf{u}_1$ against the observable x confirms that the latter is not a good observable (middle panel of Figure 4). The slow variable is given as a nonlinear transformation of x and y which can be computed by inverting (22) locally. It is basically a function of u MOD 2π. The eigenvector $\mathbf{u}_1$ is plotted against the slow variable u MOD 2π in the right panel of Figure 4. We again confirm that we recovered the slow dynamics correctly.

Inherently non-linear chemical reactions

We consider the following set of chemical reactions

$$X \xrightarrow{k_1} X + Z, \qquad Y + Z \xrightarrow{k_2} Y, \qquad [24]$$

$$\emptyset \xrightarrow{k_3} Y, \qquad Y \xrightarrow{k_4} \emptyset, \qquad [25]$$

$$\emptyset \xrightarrow{k_5} X. \qquad [26]$$

The first two reactions (24) are production and degradation of Z (catalyzed by X and Y, respectively). The production and degradation of Z is assumed to be happening on a fast time scale. Reactions (25) are production and degradation of Y. They are assumed to occur on an intermediate time scale (i.e. slower than the fast time scale, but faster than the slow time scale). The reaction (26) is production of X which is assumed to be slow. We choose the values of the rate constants as

$$k_1 = 1000, \quad k_2 = 1, \quad k_3 = 40, \quad k_4 = 1, \quad k_5 = 1. \qquad [27]$$

This choice of rate constants guarantees that the reactions (24) are the fastest, the reactions (25) happen on a slower time scale, and the reaction (26) is the slowest. The model (24)-(26) is approximated by the ODE system for the $O(1)$ variables $x = X/100$, $y = Y/40$ and $z = Z/2500$ as follows: $\mathrm{d}x/\mathrm{d}t = k_5/100$, $\mathrm{d}y/\mathrm{d}t = k_3/40 - k_4 y$, $\mathrm{d}z/\mathrm{d}t = 100k_1 x/2500 - 40k_2 yz$. Using the parameter values (27), we obtain $\mathrm{d}x/\mathrm{d}t = x/100$, $\mathrm{d}y/\mathrm{d}t = 1 - y$, $\mathrm{d}z/\mathrm{d}t = 40(x - yz)$. The quasi-equilibrium approximation in the z-equation (fastest) is $z = x/y$, which implies the "half-moon shaped" profile (hyperbola + noise) dynamics in the Y-Z plane. The variable y changes on a faster time scale than x. Roughly speaking, the fluctuations in y lead to the dynamics in z according to the formula $z = x/y$ where x changes very slowly, as illustrated in Figure 5. We initialize the system at $[X(0), Y(0), Z(0)] = [100, 40, 2500]$ and simulate the time evolution using the Gillespie stochastic simulation algorithm. Figure 5 shows the time evolution of X (top left panel), Y (top middle panel) and Z (top right panel). The same trajectory plotted in the Y-Z plane is given in the bottom panels of Figure 5. We plot 2000 data points lying on this trajectory colored by time (bottom left panel). In the bottom middle panel of Figure 5, we provide the similar Y-Z plot where the data points are colored according to the value of X.

The set of 2000 data points (plotted in the bottom right panel of the Figure 5) is the input of the diffusion map approach. To emphasize the strength of our approach, the data points are ordered randomly in the inputting data set. In our model, the slow variable X is a non-decreasing function of time t – see Figure 5 (top left panel). Consequently, the data set recorded from the stochastic simulation is ordered according to the slow variable. In more complicated chemical examples (e.g. problems with oscillations [22]), or the oscillating half-moons from the previous example, there is no obvious relation between the "dynamic proximity" of data points and the order in which they are recorded. Our approach works in more complicated situations, because the ADM is independent of the order of the inputting data points.

We use short bursts of time $\Delta t = 5 \times 10^{-4}$ (which corresponds to approximately 100 Gillespie SSA time steps) of stochastic simulations initialized at the $N = 2000$ data points from Figure 5 (bottom right panel). For every data point $\boldsymbol{X}^{(i)} = [X^{(i)}, Y^{(i)}, Z^{(i)}]$, $i = 1, \ldots, N$, we run 10^6 replicas of stochastic simulations initialized at the data point to estimate the covariance matrix $\boldsymbol{\Sigma}^{(i)}$. We use $\varepsilon = 1$. In the right panel of Figures 6, we plot our data set (given in Figure 5 (bottom right panel)) and we color the data points according to the first nontrivial eigenvector $\mathbf{u}_1$. We see that the eigenvector $\mathbf{u_1}$ gives a good description of slow dynamics of the system (24)-(26). The slow dynamics can be described by the variable X as can be seen in the top left panel of Figure 5. The plot of $\mathbf{u_1}$ vs. X is shown in the middle panel of Figure 6. We again confirm that we obtained a good description of the slow dynamics of the system. Finally, plotting the eigenvector $\mathbf{u_1}$ vs. Y confirms that Y is not a good slow variable (right panel of Figure 6).

Summary

Finding a reduced model for dynamical systems with a large number of degrees of freedom is of great importance in many fields. Dimensional reduction methods often use similarity measures between different states of the dynamical system to reveal its low dimensional structure. Those methods are limited when the similarity measure does not take into account the time labeling of the states. We encode the time dependence into an anisotropic similarity measure using short bursts of local simulations. The resulting leading eigenvectors of the anisotropic diffusion map approximate the eigenfunctions of the Laplacian over the manifold corresponding to the dynamically meaningful slowly varying coordinates. We demonstrated the usefulness of the ADM in analyzing dynamical systems by its successful recovery of meaningful coordinates in the particular case of multiscale chemical reactions.

Acknowledgements. This research was partially supported by Centre for Mathematical Biology, St. John's College, Oxford and Linacre College, Oxford (RE). The work of AS, IGK and RRC was partially supported by DARPA.

1. Kevrekidis, I, Gear, C, & Hummer, G. (2004) *AIChE Journal* **50**, 1346–1355.
2. Kevrekidis, I, Gear, C, Hyman, J, Kevrekidis, P, Runborg, O, & Theodoropoulos, K. (2003) *Communications in Mathematical Sciences* **1**, 715–762.
3. Gear, C, Kaper, T, Kevrekidis, I, & Zagaris, A. (2005) *SIAM Journal on Applied Dynamical Systems* **4**, 711–732.
4. Belkin, M & Niyogi, P. (2002) in *Advances in Neural Information Processing Systems (NIPS)*, eds. Dietterich, T, Becker, S, & Ghahramani, Z. (MIT Press, Cambridge, MA) Vol. 14.
5. Belkin, M & Niyogi, P. (2003) *Neural Computation* **15**, 1373–1396.
6. Coifman, R, Lafon, S, Lee, A, Maggioni, M, Nadler, B, Warner, F, & Zucker, S. (2005) *Proceedings of the National Academy of Sciences USA* **102**, 7426–7431.
7. Coifman, R, Lafon, S, Lee, A, Maggioni, M, Nadler, B, Warner, F, & Zucker, S. (2005) *Proceedings of the National Academy of Sciences USA* **102**, 7432–7437.
8. Coifman, R & Lafon, S. (2006) *Applied and Computational Harmonic Analysis* **21**, 5–30.
9. Lafon, S. (2004) Ph.D. thesis (Yale University).
10. Singer, A & Coifman, R. (2007) Non linear independent component analysis with diffusion maps. Applied and Computational Harmonic Analysis.
11. Alberts, B, Johnson, A, Lewis, J, Raff, M, Roberts, K, & Walter, P. (2002) *Molecular Biology of the Cell*. (Garland Science, New York).
12. Erban, R, Kevrekidis, I, Adalsteinsson, D, & Elston, T. (2006) *Journal of Chemical Physics* **124**, 084106.
13. Gillespie, D. (1977) *Journal of Physical Chemistry* **81**, 2340–2361.
14. Gardiner, G. (1985) *Handbook of Stochastic Processes for Physics, Chemistry and Natural Sciences*. (Springer Verlag), 2 edition.
15. Gillespie, D. (2000) *Journal of Chemical Physics* **113**, 297–306.
16. van Kampen, N. (2007) *Stochastic Processes in Physics and Chemistry*. (North-Holland, Amsterdam), 3rd edition.
17. Nadler, B, Lafon, S, Coifman, R, & Kevrekidis, I. (2006) *Applied and Computational Harmonic Analysis* **21**, 113–127.
18. Nadler, B, Lafon, S, Coifman, R, & Kevrekidis, I. (2006) in *Advances in Neural Information Processing Systems 18*, eds. Weiss, Y, Schölkopf, B, & Platt, J. (MIT Press, Cambridge, MA), pp. 955–962.
19. Erban, R, Frewen, T, Wang, X, Elston, T, Coifman, R, Nadler, B, & Kevrekidis, I. (2007) *Journal of Chemical Physics* **126**, 155103.
20. Hein, M & Audibert, Y. (2005) in *Proceedings of the 22nd International Conference on Machine Learning*, eds. L. De Raedt & S. Wrobel. pp. 289–296.
21. Gillespie, D. (1992) *Markov Processes, an introduction for physical scientists*. (Academic Press, Inc., Harcourt Brace Jovanowich).
22. Tyson, J, Csikasz-Nagy, A, & Novak, B. (2002) *BioEssays* **24**, 1095–1109.

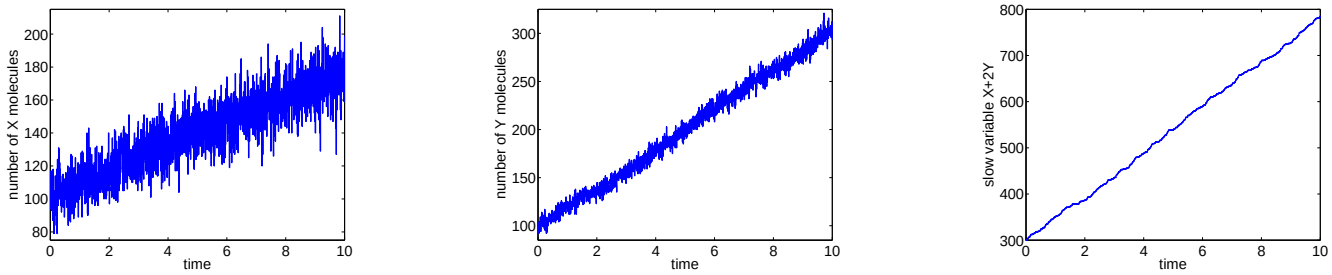

Fig. 1: *The time evolution of X, Y and $X + 2Y$ given by the stochastic simulation of the chemical system* (3) *and* (5).

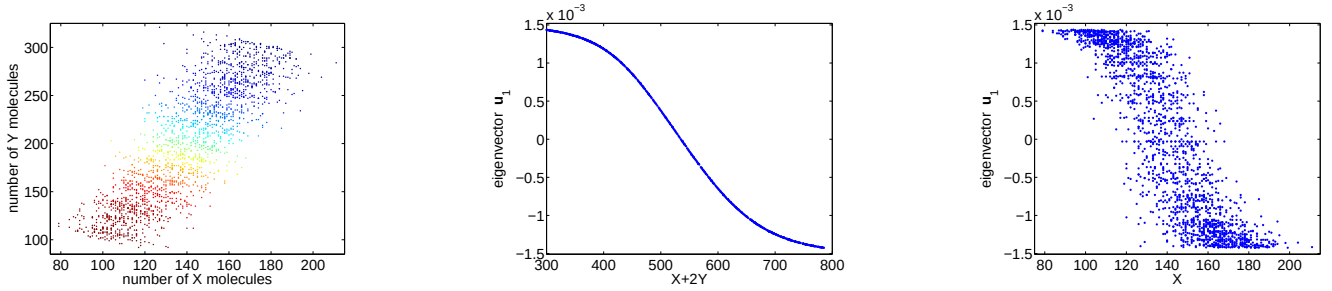

Fig. 2: *The data set with each point colored according to* $\mathbf{u}_1$ *(left panel). Vector* $\mathbf{u}_1$ *as a function of* $X + 2Y$ *(middle panel). Vector* $\mathbf{u}_1$ *as a function of* X *(right panel).*

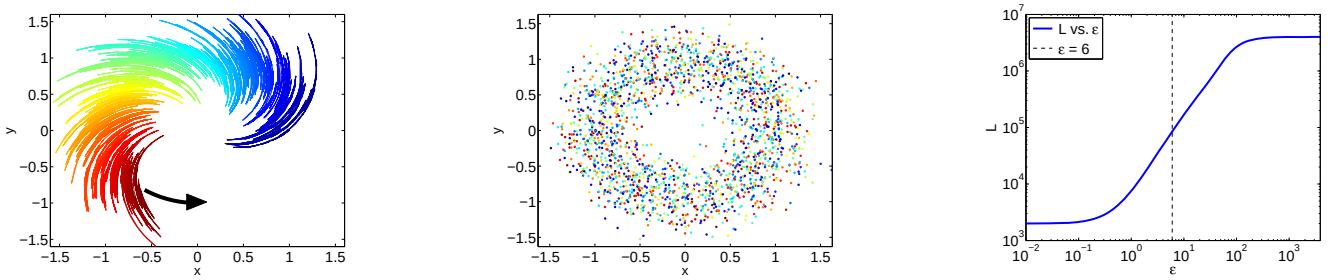

Fig. 3: *The short illustrative trajectory of* (20)–(22) *which starts at* $[x(0), y(0)] = [1, 0]$. *The trajectory is colored according to time (left panel). The representative data set sampled at equal time steps from a longer stochastic simulation. The points are colored according to time (middle panel). Plot of* $L(\varepsilon)$ *given by* (23) *(right panel).*

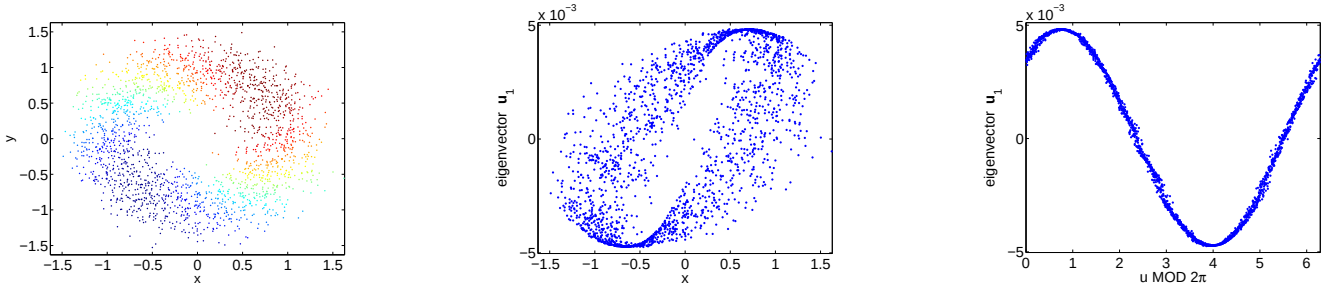

Fig. 4: *The data set with each point colored according to* $\mathbf{u}_1$ *(left panel). Vector* $\mathbf{u}_1$ *as a function of* x *(middle panel). Vector* $\mathbf{u}_1$ *as a function of* u *MOD* 2π *(right panel).*

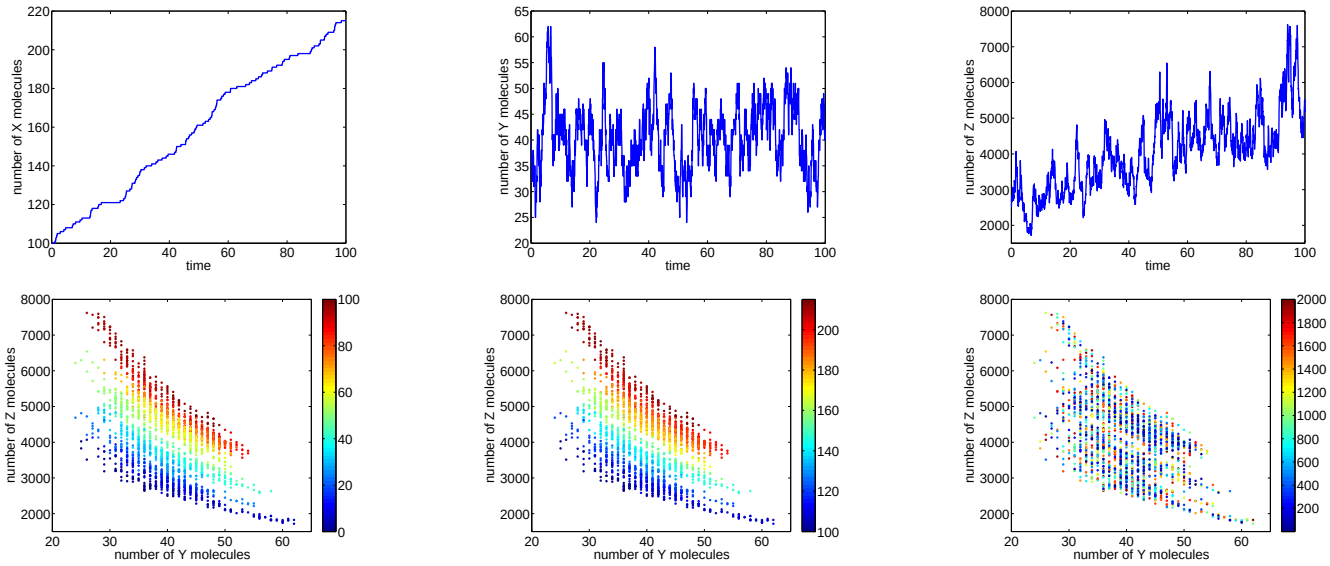

Fig. 5: *The time evolution of X (top left panel), Y (top middle panel) and Z (top right panel) given by the stochastic simulation of the chemical system* $(24)-(26)$. *The same trajectory (2000 data points, saved at equal time intervals* $\Delta t = 0.05$ *apart) plotted in the Y-Z plane is shown in the bottom panels. We color the points according to time (bottom left panel) and according to the number of X molecules (bottom middle panel). To emphasize the strength of our approach, we randomize the order of the data points – we color the resulting data set according to the order in the new list in the right panel (bottom right panel).*

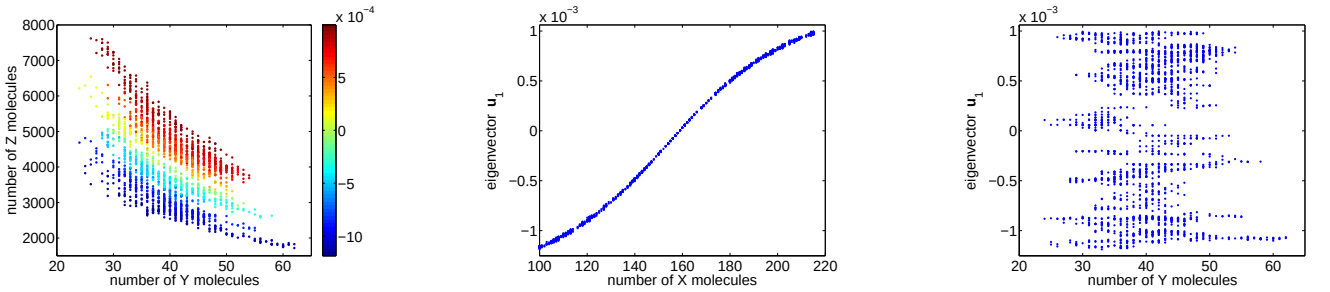

Fig. 6: *The data set in the Y-Z plane with each point colored according to $\mathbf{u}_1$ (left panel). Vector $\mathbf{u}_1$ as a function of X (middle panel). Vector $\mathbf{u}_1$ as a function of Y (right panel).*