

and/or sign of the eigenvectors are important. For the work presented in Chapter 4, we must normalize the eigenvectors from different data sets so that the resulting embeddings are consistent. We first scale the eigenvectors so that  $\|\phi_i\| = m$  (where  $m$  is the number of data points) to make the embedding coordinates invariant to the size of the data set. Still, the computed embedding eigenvectors, even for two identical data sets, may differ by a sign. Reconciling the signs for the embeddings of different data sets can be rationally done in several ways and is somewhat problem-specific. For example, if the mean of the embedding is sufficiently far from 0, we can require  $\langle \phi_i \rangle > 0$ ; alternatively, if there is a common region sampled by two data sets obtained from the same system, the sign of each eigenvector can be chosen to optimize the consistency of the embeddings of the common region data. Similarly in Chapter 5, the sign of the eigenvector is important and is determined after analysis using *a priori* knowledge of the system dynamics.

## 2.2 The Mahalanobis Distance

An essential component of manifold learning algorithms (such as diffusion maps) is having an informative distance metric between data points, as one of the key assumptions is that points which have a small distance are close on the manifold. However, often the Euclidean distance is not meaningful or informative, and we require a more sophisticated metric. Here, we will discuss the Mahalanobis distance [92], a metric which has some nice properties that will allow us to analyze data from multiscale stochastic systems in Chapter 3 and data from chemical simulations in Chapter 4.

Let  $\widehat{\mathbf{C}}(t)$  be the estimated covariance matrix associated with the measured sample  $\mathbf{y}(t)$  (the specifics will be discussed in subsequent chapters). We define a Riemannian metric between a pair of samples using the associated covariance matrices as

$$\|\mathbf{y}(t_i) - \mathbf{y}(t_j)\|_M^2 = \frac{1}{2}(\mathbf{y}(t_i) - \mathbf{y}(t_j))^T (\widehat{\mathbf{C}}(t_i) + \widehat{\mathbf{C}}(t_j))^\dagger (\mathbf{y}(t_i) - \mathbf{y}(t_j)); \quad (2.9)$$

Seeded Content – [Weisstein, Eric W. "Angle." From MathWorld--A Wolfram Web Resource.](http://mathworld.wolfram.com/Angle.html) <http://mathworld.wolfram.com/Angle.html>

Given two [intersecting lines](#) or [line segments](#), the amount of [rotation](#) about the point of intersection (the [vertex](#)) required to bring one into correspondence with the other is called the angle  $\theta$  between them. The term "plane angle" is sometimes used to distinguish angles in a plane from [solid angles](#) measured in space (International Standards Organization 1982, p. 5).

The term "angle" can also be applied to the rotational offset between intersecting planes about their common line of intersection, in which case the angle is called the [dihedral angle](#) of the planes.

Angles are usually measured in [degrees](#) (denoted  $^\circ$ ), [radians](#) (denoted rad, or without a unit), or sometimes [gradians](#) (denoted grad).

The concept of an angle can be generalized from the [circle](#) to the [sphere](#), in which case it is known as [solid angle](#). The fraction of a [sphere](#) subtended by an object (its solid angle) is measured in [steradians](#), with the entire [sphere](#) corresponding to  $4\pi$  [steradians](#).

One full rotation in these three measures corresponds to  $360^\circ$ ,  $2\pi$  rad, or 400 grad. Half a full [rotation](#) is called a [straight angle](#), and a [quarter](#) of a full rotation is called a [right angle](#). An angle less than a [right angle](#) is called an [acute angle](#), an angle greater than a [right angle](#) (but less than a [straight angle](#)) is called an [obtuse angle](#), and an angle greater than a [straight angle](#) (but less than a [full angle](#)) is called a [reflex angle](#).

The use of [degrees](#) to measure angles harks back to the Babylonians, whose [sexagesimal](#) number system was based on the number 60.  $360^\circ$  likely arises from the Babylonian year, which was composed of 360 days (12 months of 30 days each). The [degree](#) is further divided into 60 [arc minutes](#), and an [arc minute](#) into 60 [arc seconds](#). A more natural measure of an angle is the [radian](#). It has the property that the [arc length](#) around a [circle](#) is simply given by the radian angle measure times the [circle radius](#). The [radian](#) is also the most useful angle measure in [calculus](#) because the [derivative](#) of [trigonometric](#) functions such as

$$\frac{d}{dx} \sin x = \cos x$$

does not require the insertion of multiplicative constants like  $\pi/180$ . [Gradians](#) are sometimes used in surveying (they have the nice property that a [right angle](#) is exactly 100 [gradians](#)), but are encountered infrequently, if at all, in mathematics.

A ruled [semicircle](#) used for measuring and drawing angles is called a [protractor](#). A [compass](#) can also be used to draw circular [arcs](#) of some angular extent.

### **Wolfram Web Resources**

Mathematica » The #1 tool for creating Demonstrations and anything technical.

Wolfram|Alpha » Explore anything with the first computational knowledge engine.

Wolfram Demonstrations Project » Explore thousands of free applications across science, mathematics, engineering, technology, business, art, finance, social sciences, and more.

Computerbasedmath.org » Join the initiative for modernizing math education.

Online Integral Calculator » Solve integrals with Wolfram|Alpha.

Step-by-step Solutions » Walk through homework problems step-by-step from beginning to end. Hints help you try the next step on your own.

Wolfram Problem Generator » Unlimited random practice problems and answers with built-in Step-by-step solutions. Practice online or make a printable study sheet.

Wolfram Education Portal » Collection of teaching and learning tools built by Wolfram education experts: dynamic textbook, lesson plans, widgets, interactive Demonstrations, and more.

Wolfram Language » Knowledge-based programming for everyone.

rotational symmetries. Angular synchronization [126] is an algorithm for registering a data set given pairwise alignment information. Vector diffusion maps [129] combines angular synchronization and diffusion maps to both register and uncover structure in a data set in a single computation. We demonstrate these two algorithms using a synthetic data set whose relatively simple dynamics allow us to easily visualize and illustrate the main features of the different algorithms. Motivated by the geometry of the *Drosophila* embryo images in Chapter 5, we construct a sequence of concentration profiles defined on a ring, and rotate each ring randomly around its center; an example is shown in Figure 2.2A. Rotation of the ring corresponds to shifting (with periodic boundary conditions) the one-dimensional concentration profile shown at the bottom of Figure 2.2A (the symmetry group is  $SO(2)$ , the group of all two-dimensional proper rotations). Each concentration profile is a noisy Gaussian (shown in Figure 2.2B), and the Gaussians increase in intensity as a function of “time”. We discretize the profiles into 100 points, so our numerical data will be 100-dimensional vectors (the corresponding symmetry group for the discretized profiles is  $\mathbb{Z}_{100}$ , the group of integers modulo 100). Figure 2.2C shows the entire data set; the concentration profiles have been stacked in an array, so that each row corresponds to a single profile. Because the profiles are unregistered and unordered, the underlying dynamics (a Gaussian whose amplitude grows in time) are not readily apparent.

The angular synchronization algorithm aligns pairs of data points with respect to a given symmetry group from pairs of alignment measurements [126]. Let  $\mathbf{y}(1), \dots, \mathbf{y}(m)$  denote the signals that we wish to align with respect to rotations; each signal is a function defined on the unit circle (on the plane). First assume that each signal  $\mathbf{y}(i)$  is a *noisy* rotated copy of the underlying signal  $\mathbf{y}_{true}$  (which we are *not* given), such that

$$\mathbf{y}(i) = f(\mathbf{y}_{true}, \theta_i) + \xi(i) \tag{2.12}$$

eigenvectors, our problem formulation accounts for not only pairwise alignment information, but also these higher-order considerations.

## 2.4 Vector Diffusion Maps

Vector diffusion maps combines the algorithms of angular synchronization and diffusion maps into a single computation that both removes rotational symmetries and parameterizes the data *modulo* these symmetries [129]. Given data points  $\mathbf{y}(1), \dots, \mathbf{y}(m)$ , one first constructs the matrix  $\mathbf{S} \in \mathbb{R}^{md_{rot} \times md_{rot}}$ , with the  $i, j^{th}$  block of  $\mathbf{S}$ ,  $\mathbf{S}_{ij}$ , defined as

$$\mathbf{S}_{ij} = \mathbf{A}_{ij} \mathbf{H}_{ij} \quad (2.22)$$

where  $\mathbf{A}_{ij} \in \mathbb{R}$  (defined in Section 2.1.2) pertains to the diffusion kernel between data points, and  $\mathbf{H}_{ij} \in \mathbb{R}^{d_{rot} \times d_{rot}}$  (defined in (2.16)) pertains to the pairwise alignment between data points. It is important to note that distance  $\|\mathbf{y}(i) - \mathbf{y}(j)\|$  used in the diffusion kernel in (2.3) is the distance between data points *after* pairwise alignment, i.e., the minimum distance between all possible shifts of the two data points. In the language of symmetry groups, this distance is a metric between the orbits induced by the relevant symmetry group.

One then computes the eigenvalues  $\lambda_0, \lambda_1, \dots, \lambda_{md_{rot}-1}$  and eigenvectors  $\phi_0, \phi_1, \dots, \phi_{md_{rot}-1}$  of  $\mathbf{S}$ , ordered such that  $|\lambda_0| \geq |\lambda_1| \geq \dots \geq |\lambda_{md_{rot}-1}|$ . These eigenvectors contain information about *both* the optimal rotations (the “synchronization” component) and the variation of the data *after* the spatial symmetries have been removed. Assuming that the data (after symmetries have been factored out) are relatively closely clustered, it is reasonable to expect, as in angular synchronization, that the top (block) eigenvector of  $\mathbf{S}$  contains approximations of the optimal rotations, which can be computed in the same way from (2.19). We then expect subsequent eigenvectors to contain information about the main direction(s) of data variability modulo the geometric symmetries.

In general, the vector diffusion maps embedding coordinates are given by

$$\psi_{k,l}(i) = \langle \phi_k(i), \phi_l(i) \rangle, \quad (2.23)$$

where  $\phi_k(i) \in \mathbb{R}^{d_{rot}}$  denotes the  $i^{th}$  block of  $\phi_k$ . If we assume that the rotations and the dynamics are uncoupled and therefore separable, then the eigenvectors of  $\mathbf{S}$  have the following structure: each block eigenvector contains estimates of the optimal rotations (up to a constant rotation) multiplied by the corresponding embedding coordinate (a scalar). As the first diffusion maps coordinate is constant over the data, the first block eigenvector contains only the optimal rotations. The second block eigenvector (eigenvectors  $d_{rot}$  through  $2d_{rot} - 1$ ) contains the optimal rotations, each multiplied by their second diffusion maps coordinate. We can therefore recover this diffusion maps coordinate by taking inner products of the columns of the second block eigenvector with columns of the first block eigenvector. The  $j^{th}$  embedding coordinate will be given by  $\psi_{k,l}$ , where  $jd_{rot} < k \leq (j+1)d_{rot} - 1$  and  $0 \leq l \leq d_{rot} - 1$ , and we select  $k, l$  such that the coordinate  $\psi_{k,l}$  has the largest variability, i.e., the  $j^{th}$  coordinate is  $\psi_{k,l}$ , where  $k, l$  is the solution to

$$\begin{aligned} & \max_{\substack{jd_{rot} \leq k \leq (j+1)d_{rot} - 1 \\ 0 \leq l \leq d_{rot} - 1}} \sum_i \psi_{k,l}(i)^2. \end{aligned} \quad (2.24)$$

We can use the eigenvalues from vector diffusion maps to help deduce the dimensionality of the data. In diffusion maps, the largest eigenvalue will always be 1 and correspond to the trivial (constant) eigenvector, and  $|\lambda_k|$  gives a measure of the importance of coordinate  $\phi_k$ . We therefore expect to see a “spectral gap” in the eigenvalues which separates the meaningful coordinates from those corresponding to noise (modulo higher harmonics; see Chapter 6 and [48]). In vector diffusion maps, the importance of each coordinate is measured by the product of the corresponding eigenvalues (i.e., the importance of  $\psi_{k,l}$  is given by  $|\lambda_k \lambda_l|$ ). We again

inherent time scale separation; in this case, it is possible, in principle, to obtain a reduced system of differential equations in only the slow variables *without memory terms*. Such an analysis crucially hinges on knowing in which variables (or, more generally, functions of variables) one can write such a reduced system of slow evolution equations.

Moving averages and subsampling have often been used in simple cases as appropriate functions of variables in which to formulate slow lower-dimensional models [104]. However, if the underlying dynamics are sufficiently nonlinear, such statistics may fail to capture the relevant structures and time scales within the data (see Figure 3.1 for a schematic illustration). For well-studied systems, one often has some *a priori* knowledge of the appropriate observables (such as phase field variables) with which to formulate the reduced dynamics [27, 145]. However, such observables may not be immediately obvious upon inspection for new complex systems, and so we require an automated approach to construct such slow variables.

Given an explicit system of ordinary differential equations, one can make numerical approximations, such as the quasi-steady state approximation [121] or the partial equilibrium approximation [55], to reduce the system dimensionality without introducing memory terms. There has been some recent analytical work on extending and generalizing such ideas to more complex systems of equations [3, 21, 34, 38, 60, 104, 131]. However, in many instances, closed form, analytical models are not given explicitly, but can only be inferred from simulation and/or experimental data. We therefore turn to data-driven techniques to analyze such systems and uncover the relevant dynamical modes. In particular, we will use a manifold-learning based approach, as such methods can accommodate nonlinear structures in high-dimensional data.

The core of most manifold learning methods is having a notion of similarity between data points, usually through a distance metric [12, 30, 32, 111, 138]. The distances are then integrated into a global parameterization of the data, typically through the solution of an eigenproblem. In this chapter, we will analyze multiple time scale stochastic dynamical

systems using data-driven methods. Standard “off-the-shelf” manifold learning techniques which utilize the Euclidean distance are not appropriate for analyzing data from such multiscale systems, since this metric does not account for the disparate time scales. Research efforts have addressed the construction of more informative distance metrics, which are less sensitive to noise and can better recover the true underlying structure in the data by suppressing unimportant sources of variability [14, 56, 117, 125, 146]. The Mahalanobis distance is one such metric. It was shown that the Mahalanobis distance can remove the effect of *observing* the underlying system variables through a complex, nonlinear function [43, 127, 135]. Here, we will show the analogy between removing the effects of such nonlinear observation functions (in the context of data analysis), and reducing a dynamical system to remove the effects of the fast variables. Our approach will build a parameterization of the data which is consistent with the underlying slow variables. Because our approach is data-driven, we require no explicit description of the model, and can extract the underlying slow variables from either simulation or experimental data. Furthermore, the approach implicitly identifies the slow variables within the data and does not require any *a priori* knowledge of the fast or slow variability sources. Even when the underlying dynamical system is complex with nonlinear coupling between the fast and slow variables, we will show that our approach has the potential to isolate the underlying slow modes.

We will present detailed analysis for our method, and provide conditions under which it will successfully recover the slow variables. Furthermore, based on this analysis, we will present data-driven protocols to tune the parameters of the method appropriately. Our presentation and discussion will address two-time-scale stochastic systems; however, we claim that our framework and analysis readily extends to systems with multiple time scale separations.



conditions on  $a_i(\mathbf{x})$ , the same can be said for more general drift functions, where  $\mu_i$  are the eigenvalues of the Jacobian of  $\mathbf{a}(\mathbf{x}) = \begin{bmatrix} a_1(\mathbf{x}) & \cdots & a_{d_1}(\mathbf{x}) & a_{d_1+1}(\mathbf{x})/\epsilon & \cdots & a_d(\mathbf{x})/\epsilon \end{bmatrix}^T$  [142]. Therefore, (3.1) defines an  $d$ -dimensional stochastic system with  $d_1$  slow variables and  $d - d_1$  fast variables, and  $\epsilon$  defines the time scale separation. The ratio of the powers of  $\epsilon$  in the drift and diffusion terms in (3.1) is essential, as we require the square of the diffusivity to be of the same order as the drift as  $\epsilon \rightarrow 0$  [13]. If the diffusivity is larger, then, as  $\epsilon \rightarrow 0$ , the equilibrium measure will be unbounded. Conversely, if the diffusivity is smaller, the equilibrium measure will go to 0 as  $\epsilon \rightarrow 0$ .

Assuming the sample average of  $a_i(\mathbf{x})$  converges to a distribution which is only a function of the slow variables, then by the averaging principle [53], we can write a reduced SDE in *only* the slow variables  $x_1, \dots, x_{d_1}$ . The aim of our work is to show how we can detect such slow variables *automatically* from data, in order to help inform modeling efforts and aid in the writing of such reduced stochastic models. In general, we are not given the variables  $\mathbf{x}(t)$  from the original SDE system, but instead, we are given some *observations* in the form  $\mathbf{y}(t) = \mathbf{f}(\mathbf{x}(t))$ . We assume that  $\mathbf{f} : \mathbb{R}^d \mapsto \mathbb{R}^n$ ,  $d \leq n$ , is a deterministic (possibly nonlinear) function whose image is a  $d$ -dimensional manifold  $\mathcal{M}_d$  in  $\mathbb{R}^n$ . For our analysis, we require  $\mathbf{g} = \mathbf{f}^{-1}$  to be well-defined on  $\mathcal{M}_d$ , and both  $\mathbf{f}$  and  $\mathbf{g}$  to be continuously differentiable to fourth order. Given data  $\mathbf{y}(t_1), \dots, \mathbf{y}(t_m)$  on  $\mathcal{M}_d$  we would like to recover a parameterization of the data that is one-to-one with the slow variables  $x_1, \dots, x_{d_1}$ .

In order to recover the slow variables from data, we will utilize a local metric that collapses the fast directions. Typically, such a metric averages out the fast variables. However, simple averages are inadequate to describe data which is observed through a complicated nonlinear function. Instead, we propose to use the Mahalanobis distance, which measures distances normalized by the respective variances in each local principal direction. Using this metric, we still retain information about both the fast and slow directions and can more clearly observe complex dynamic behavior within the data set.

- [85] Stéphane Lafon, Yosi Keller, and Ronald R Coifman. Data fusion and multicue data matching by diffusion maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1784–1797, 2006.
- [86] Themis Lazaridis and Martin Karplus. “New view” of protein folding reconciled with the old through multiple unfolding simulations. *Science*, 278(5345):1928–1931, 1997.
- [87] Tony Lelièvre, Francis Nier, and Grigorios A Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.
- [88] Maria Leptin. Gastrulation movements: The logic and the nuts and bolts. *Developmental Cell*, 8(3):305–320, 2005.
- [89] Bomyi Lim, Núria Samper, Hang Lu, Christine Rushlow, Gerardo Jiménez, and Stanislav Y Shvartsman. Kinetics of gene derepression by ERK signaling. *Proceedings of the National Academy of Sciences*, 110(25):10330–10335, 2013.
- [90] R. J. Loncharich, B. R. Brooks, and R. W. Pastor. Langevin dynamics of peptides - the frictional dependence of isomerization rates of N-Acetylalanine-N'-Methylamide. *Biopolymers*, 32(5):523–535, 1992.
- [91] Ao Ma and Aaron R Dinner. Automatic method for identifying reaction coordinates in complex systems. *The Journal of Physical Chemistry B*, 109(14):6769–6779, 2005.
- [92] Prasanta Chandra Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936.
- [93] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.
- [94] Georges Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.
- [95] Georges Matheron. The intrinsic random functions and their applications. *Advances in Applied Probability*, pages 439–468, 1973.
- [96] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert-Jan Marrink. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.
- [97] Hazime Mori. Transport, collective motion, and Brownian motion. *Progress of Theoretical Physics*, 33(3):423–455, 1965.
- [98] Boaz Nadler, Stéphane Lafon, Ronald R Coifman, and Ioannis G Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, 2006.
- [99] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.