# DIFFUSION MAPS, REDUCTION COORDINATES AND LOW DIMENSIONAL REPRESENTATION OF STOCHASTIC SYSTEMS

R.R. COIFMAN[*], I.G. KEVREKIDIS[†], S. LAFON[*‡], M. MAGGIONI[*§], AND B. NADLER[¶]

**Abstract.**
The concise representation of complex high dimensional stochastic systems via a few reduced coordinates is an important problem in computational physics, chemistry and biology. In this paper we use the first few eigenfunctions of the backward Fokker-Planck diffusion operator as a coarse grained low dimensional representation for the long term evolution of a stochastic system, and show that they are optimal under a certain mean squared error criterion. We denote the mapping from physical space to these eigenfunctions as the *diffusion map*. While in high dimensional systems these eigenfunctions are difficult to compute numerically by conventional methods such as finite differences or finite elements, we describe a simple computational data-driven method to approximate them from a large set of simulated data. Our method is based on defining an appropriately weighted graph on the set of simulated data, and computing the first few eigenvectors and eigenvalues of the corresponding random walk matrix on this graph. Thus, our algorithm incorporates the local geometry and density at each point into a global picture that merges in a natural way data from different simulation runs. Furthermore, we describe lifting and restriction operators between the diffusion map space and the original space. These operators facilitate the description of the coarse-grained dynamics, possibly in the form of a low-dimensional effective free energy surface parameterized by the diffusion map reduction coordinates. They also enable a systematic exploration of such effective free energy surfaces through the design of additional "intelligently biased" computational experiments. We conclude by demonstrating our method on a few examples.

**Key words.** Diffusion maps, dimensional reduction, stochastic dynamical systems, Fokker Planck operator, metastable states, normalized graph Laplacian.

**AMS subject classifications.** 60H10, 60J60, 62M05

**1. Introduction.** Systems of stochastic differential equations (SDE's) are commonly used as models for the time evolution of many chemical, physical and biological systems of interacting particles [22, 45, 52]. There are two main approaches to the study of such systems. The first is by detailed Brownian Dynamics (BD) or other stochastic simulations, which follow the motion of each particle (or more generally variable) in the system and generate one or more long trajectories. The second is via analysis of the time evolution of the probability densities of these trajectories using the numerical solution of the corresponding time dependent Fokker-Planck (FP) partial differential equation.

For typical high dimensional systems, both approaches suffer from severe limitations, when applied directly. The main limitation of standard BD simulations is the *scale gap* between the atomistic time scale of single particle motions, at which the SDE's are formulated, and the macroscopic time scales that characterize the long term evolution and equilibration of these systems. This scale gap puts severe constraints on detailed simulations due to the requirement to accurately integrate the fastest motions and degrees of freedom in the system, such as fast chemical reactions and particle-particle collisions. Therefore, the time step in detailed simulations is typically constrained to be orders of magnitude smaller than the characteristic times of

[*]Department of Mathematics, Yale University, New Haven, CT, 06520.
[†]Chemical Engineering and PACM, Princeton University, Princeton, NJ 08544.
[‡]Currently at Google, Inc.
[§]Currently at department of mathematics, Duke University, Durham, NC 27708.
[¶]Corresponding author, Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, 76100, Israel. `boaz.nadler@weizmann.ac.il`.

the phenomena we wish to study. Moreover, for systems with well defined metastable states and relatively rare transitions between them, direct simulations spend the majority of computer resources exploring the motion "within" the metastable states and only an exponentially small part exploring the transitions "between them", which are often the quantity of interest.

The main limitation of standard computational methods that solve the FP equation is the curse of dimensionality. While for dimension $n \leq 3$ the FP equation can typically be solved numerically, in higher dimensional systems the solution of the relevant partial differential equation is practically impossible by standard methods such as finite differences or finite elements, since the number of grid points grows like $(1/h)^n$ where $h$ is the grid spacing. We note, however, that for some high dimensional systems this direct computation is still possible with the construction of sparse grids [9].

While in both approaches the detailed time evolution of a stochastic system requires a high dimensional description with many degrees of freedom, often its long term or coarse grained evolution is of a *low dimensional* nature. The main challenges in this case are the identification of dynamically meaningful slow variables, or *reduction coordinates*[1], and the description of the effective dynamics of the system in this low dimensional representation. The main requirements for good reduction coordinates is that they are dynamically meaningful, in the sense that we can write an effective SDE for the long-term dynamics of the system based on these coordinates. Thus, on a coarse enough time scale, the dynamics of the reduction coordinates is approximately *Markovian*, without further dependence on the fine details of the high dimensional system.

In some systems, either the form of the equations, prior experience, or some underlying physical intuition help determine good reduction coordinates. Then appropriate equations can be formulated in these variables, and in some special cases their exact form can even be found by rigorous mathematics based on the Mori-Zwanzig projection approach [24, 55]. In more complex cases where a rigorous derivation of the dynamics is mathematically intractable, many numerical approaches to solve these tasks have been suggested in the literature, such as transition path sampling, the nudged elastic band, the string method, the transfer operator approach, Perron cluster analysis and many others, see [16, 17, 18, 19, 20, 26, 46], and references therein. In addition, given further knowledge about the system, such as a good dividing surface between reactant and product regions, several algorithms for the efficient computation of the transition rates have been developed [44, 37, 53]. Despite these results, still in many high dimensional systems useful low dimensional representations are far from obvious.

In this paper we show that there is an intimate connection between the eigenfunctions of the backward FP operator and useful global low dimensional representations, and hence propose to use the first few of these eigenfunctions as reduction coordinates. We show that the first few eigenfunctions are *optimal* under a mean square error criterion for the approximation of probability densities in a suitable Hilbert space, and denote the mapping from the physical space to the first few eigenfunctions as a *diffusion map*. As in the case of the time-dependent FP equation, the compu-

---

[1]We make a distinction between reduction coordinates of a general dynamical system, and the *reaction coordinate* of chemical physics, which is typically a single variable that quantifies the progress of a reaction. As we describe in the paper, reduction coordinates may also be introduced in the absence of a chemical reaction and without well defined reactant and product regions.

tation of the eigenfunctions of the FP operator is practically impossible by standard space discretization methods. In this paper we present a different approach, which approximates these first few eigenfunctions from a large set of simulated data points. Our algorithm is based on the definition of a weighted graph on the simulated points and the subsequent computation of the first few eigenvalues and eigenvectors of a random walk on this graph. The proposed method does not take into account the time ordering of the simulated points, and can therefore easily merge data from different simulation runs (with different initial conditions, different initial seeds of the random number generator, etc.). As proven theoretically and shown in a few illustrative examples, in the presence of a spectral gap, the description of the system by the first few eigenfunctions gives a dynamically meaningful low dimensional representation.

Furthermore, taking a step beyond data analysis and a low dimensional representation, we describe restriction and lifting operators between the original space and the diffusion map space. These operators enable efficient extraction of the macroscopic dynamics in this lower dimensional representation. Specifically, following the equation-free coarse molecular dynamics approach [31, 32, 33], we propose to explore the effective free energy and diffusion coefficients as a function of the diffusion map coordinates by a series of multiple short simulations appropriately initialized at given values of these reduction coordinates. This methodology thus outlines a systematic manner to bridge the scale gap and estimate macroscopic dynamics and quantities of interest, such as mean exit times, transition probabilities, etc.

The paper is organized as follows. In section 2 we describe our problem and present a concise review of known results in the theory of stochastic differential equations, making the paper reasonably self-contained. In section 3 we define the diffusion distance between different configurations of a stochastic system and its relation to the eigenfunctions of the FP operator and to a low dimensional representation of the system. Section 4 describes an algorithm to approximate the diffusion map from discrete data, as well as restriction and lifting operators that allow communication between the two spaces. In section 5 we present applications of our method to a few illustrative examples. We conclude in section 6 with a summary and discussion.

**2. Problem Setup.**

**2.1. The Langevin Equation.** Consider a stochastic system with $n$ variables, confined for simplicity to a finite compact connected region $\Omega \subseteq \mathbb{R}^n$ with smooth reflecting boundaries. We assume that the time evolution of the system, described by its state $\boldsymbol{x}(t)$ at time $t$ $(\boldsymbol{x}(t) \in \Omega)$, follows a first order stochastic differential equation (SDE) written in non-dimensional form as

$$(2.1) \qquad \dot{\boldsymbol{x}} = -\nabla U(\boldsymbol{x}) + \sqrt{2/\beta}\, \dot{\boldsymbol{w}}$$

where $U(\boldsymbol{x})$ is the potential energy of a configuration $\boldsymbol{x}$, $\beta = 1/k_B T$ is a thermal factor, and $\boldsymbol{w}(t)$ is standard Brownian motion in $n$ dimensions. We assume the potential $U(\boldsymbol{x})$ to be smooth and in particular bounded from above and below. However, much of what follows, with suitable technical modifications, could be derived under more general conditions, for example for a non-compact region $\Omega$, or a potential $U$ not necessarily smooth or bounded, as long as the condition

$$(2.2) \qquad \int_{\Omega} e^{-\beta U(\boldsymbol{x})} d\boldsymbol{x} < \infty$$

is satisfied and under the assumption that the process is ergodic.

In this paper we focus on systems whose long time evolution is of a low dimensional nature. This is the case, for example, in systems governed by rare events where the potential $U$ has a few deep wells separated by high barriers, or in systems with well defined low dimensional manifolds where the potential $U$ contains steep gradients in all directions normal to the manifold, thus effectively constraining the system to approximately lie on it. The task at hand is to find good low dimensional representations of such systems and the characteristics of their coarse grained dynamics in this representation. In the context of systems governed by rare events, typical system level tasks include the identification of the metastable configurations and the transition pathways and rates between them.

**2.2. Forward and Backward Fokker-Planck Equations.** Integration of the SDE (2.1) produces random paths whose ensemble defines time dependent probability distributions on $\Omega$. To study the dynamics of the system, it is convenient to consider the time evolution of these probability distributions. Specifically, from the theory of stochastic processes [22, 45], the transition probability density $p(\boldsymbol{x}, t | \boldsymbol{x}_0, 0)$ of finding the system at location $\boldsymbol{x}$ at time $t$, given an initial location $\boldsymbol{x}_0$ at time $t = 0$ satisfies the forward Fokker-Planck (also known as Smoluchowski) equation

$$(2.3) \qquad \frac{\partial p}{\partial t} = \mathcal{L}p = \frac{1}{\beta}\Delta p + \nabla \cdot (p\nabla U)$$

defined in $(\boldsymbol{x}, t) \in \Omega \times \mathbb{R}^+$, with reflecting (no flux) boundary conditions on $\partial\Omega$.

Under the smoothness assumption on the potential $U$ and the compactness assumption on the domain $\Omega$ in which the Fokker-Planck equation (FPE) is defined, the operator $\mathcal{L}$ has a discrete spectrum of non-positive eigenvalues $\{-\lambda_j\}_{j=0}^\infty$, with $\lambda_0 = 0 > -\lambda_1 \geq -\lambda_2 \geq \ldots$, with a single accumulation point at $-\infty$ and with associated eigenfunctions $\{\varphi_j\}_{j=0}^\infty$ [13]. The solution of (2.3) can be written as

$$(2.4) \qquad p(\boldsymbol{x}, t | \boldsymbol{x}_0, 0) = \sum_{j=0}^\infty a_j e^{-\lambda_j t}\varphi_j(\boldsymbol{x})$$

where the coefficients $a_j$ depend on the initial conditions at time $t = 0$. Under fairly general conditions on the potential $U$ and the region $\Omega$, the eigenfunctions $\varphi_j$ are smooth bounded functions, and the sum in (2.4) converges uniformly in $\boldsymbol{x}$ for all times $t > t_0$ with $t_0 > 0$, see for example [15]. The eigenfunction $\varphi_0(\boldsymbol{x})$ corresponding to the eigenvalue $\lambda_0 = 0$ is given by the Boltzmann equilibrium distribution

$$(2.5) \qquad \varphi_0(\boldsymbol{x}) = C_\beta e^{-\beta U(\boldsymbol{x})}$$

where $C_\beta$ is a temperature dependent normalization factor.

Since the stochastic process $\boldsymbol{x}(t)$ is ergodic, then regardless of the initial configuration $\boldsymbol{x}_0 \in \Omega$,

$$(2.6) \qquad \lim_{t\to\infty} p(\boldsymbol{x}, t | \boldsymbol{x}_0, 0) = \varphi_0(\boldsymbol{x})$$

which means that $a_0 = 1$. Thus, according to (2.4) the *approach* to the equilibrium density $\varphi_0(\boldsymbol{x})$ is governed by the next eigenfunctions $\{\varphi_j\}_{j\geq 1}$, and their corresponding eigenvalues $\lambda_j$ and coefficients $a_j$.

A different way to study the approach to equilibrium is to consider the time evolution of averages of functions defined on $\Omega$. Let $f : \Omega \to \mathbb{R}$ be a smooth function in $L_2(\Omega)$, and define

$$(2.7) \qquad g(\boldsymbol{x}, t) = \mathbb{E}\{f(\boldsymbol{x}(t)) \,|\, \boldsymbol{x}(0) = \boldsymbol{x}\}.$$

4

Then, $g$ satisfies the backward Fokker-Planck equation, also known as the Chapman-Kolmogorov equation,

$$(2.8) \qquad \frac{\partial g}{\partial t} = \mathcal{L}^* g = \frac{1}{\beta} \Delta g - \nabla g \cdot \nabla U$$

in the domain $(\boldsymbol{x}, t) \in \Omega \times \mathbb{R}^+$, with initial conditions

$$(2.9) \qquad g(\boldsymbol{x}, 0) = f(\boldsymbol{x}).$$

The operator $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$ under the standard inner product in $L_2(\Omega)$,

$$(2.10) \qquad \langle u, v \rangle = \int_\Omega u(\boldsymbol{x}) v(\boldsymbol{x}) d\boldsymbol{x}$$

that is $\langle \mathcal{L}u, v \rangle = \langle u, \mathcal{L}^* v \rangle$. Therefore, $\mathcal{L}^*$ has the same eigenvalues $\{-\lambda_j\}_{j \geq 0}$ as $\mathcal{L}$ with corresponding eigenfunctions $\psi_j(\boldsymbol{x})$, and the solution to (2.8) can be written as

$$(2.11) \qquad g(\boldsymbol{x}, t) = \sum_j b_j e^{-\lambda_j t} \psi_j(\boldsymbol{x}).$$

The eigenfunction corresponding to $\lambda_0 = 0$ is the constant function $\psi_0(\boldsymbol{x}) = 1$. Thus

$$(2.12) \qquad \lim_{t \to \infty} g(\boldsymbol{x}, t) = b_0$$

with the approach to this equilibrium constant governed by the next eigenfunctions and eigenvalues $\{\psi_j, \lambda_j\}$, for $j \geq 1$.

The operators $\mathcal{L}$ and $\mathcal{L}^*$ are adjoint and thus the two sets of eigenfunctions $\varphi_j$ and $\psi_j$ can, and from now on will be normalized to be bi-orthonormal

$$(2.13) \qquad \langle \varphi_i, \psi_j \rangle = \delta_{i,j}.$$

Under this normalization, the coefficients $a_j, b_j$ are given by

$$(2.14) \qquad b_j = \int_\Omega f(\boldsymbol{x}) \varphi_j(\boldsymbol{x}) d\boldsymbol{x}$$

and

$$(2.15) \qquad a_j = \int_\Omega p(\boldsymbol{x}, 0) \psi_j(\boldsymbol{x}) d\boldsymbol{x} = \psi_j(\boldsymbol{x}_0).$$

One last theoretical result of interest is the connection between the eigenfunctions $\varphi_j$ and $\psi_j$. The transformation $p(\boldsymbol{x}) = e^{-U(\boldsymbol{x})} g(\boldsymbol{x})$ gives

$$(2.16) \qquad \mathcal{L}p = e^{-U} \mathcal{L}^* g.$$

Therefore, up to a normalization constant

$$(2.17) \qquad \psi_j(\boldsymbol{x}) = \varphi_j(\boldsymbol{x}) e^{U(\boldsymbol{x})} = \varphi_j(\boldsymbol{x}) / \varphi_0(\boldsymbol{x}).$$

Furthermore, under the normalization (2.13), the eigenfunctions $\varphi_j$ of the operator $\mathcal{L}$ are orthonormal in $L_2(\Omega, w(\boldsymbol{x}))$, where the inner product is with respect to the weight function $w(\boldsymbol{x}) = 1/\varphi_0(\boldsymbol{x})$,

$$(2.18) \qquad \langle u, v \rangle_w = \int_\Omega u(\boldsymbol{x}) v(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x}.$$

**3. Diffusion Distances and Diffusion Maps.** Eigenfunction expansions of the forward and backward Fokker-Planck operators as described by eqs. (2.4) and (2.11) have been used extensively in the theoretical and numerical study of stochastic differential equations (see for example the books [22, 43] and references therein). In this section we show how they can be used in a natural way to perform a dynamically meaningful *dimensional reduction* for the long term time evolution of high dimensional stochastic systems.

Reducing the dimension of a stochastic system implies lumping entire sets of different configurations into a single representative. Therefore, a crucial step in performing dimensional reduction of a stochastic system is to define a meaningful measure of dynamical proximity between different configurations. We propose a measure for this proximity using the evolution of probability densities that start from different configurations. Specifically, for any point $\boldsymbol{y} \in \Omega$ let $p(\boldsymbol{x}, t | \boldsymbol{y})$ denote the solution of (2.3) with initial condition $p(\boldsymbol{x}, 0 | \boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y})$. We define the *diffusion distance* at time $t$ between any two points $\boldsymbol{x}_0, \boldsymbol{x}_1 \in \Omega$ as the distance between the corresponding probability densities at time $t$, initialized at $\boldsymbol{x}_0$ or at $\boldsymbol{x}_1$. Naturally, we measure this distance in the Hilbert space $L_2(\Omega, w)$, with the weight function $w(\boldsymbol{x}) = 1/\varphi_0(\boldsymbol{x})$

$$(3.1) \qquad D_t^2(\boldsymbol{x}_0, \boldsymbol{x}_1) = \|p(\boldsymbol{x}, t | \boldsymbol{x}_0) - p(\boldsymbol{x}, t | \boldsymbol{x}_1)\|_{L_2(w)}^2.$$

Combining (2.4), (2.15) and (2.17) we obtain that

$$(3.2) \qquad D_t^2(\boldsymbol{x}_0, \boldsymbol{x}_1) = \sum_{j \geq 1} e^{-2\lambda_j t} |\psi_j(\boldsymbol{x}_0) - \psi_j(\boldsymbol{x}_1)|^2.$$

Note that summation starts from $j = 1$ since the term with $j = 0$ that corresponds to $\lambda_0 = 0$ cancels out.

Furthermore, we define the $k$-dimensional *diffusion map* at time $t$ as the non-linear mapping from the original space of configurations to the Euclidean space with coordinates defined by the values of the first $k$ eigenfunctions,

$$(3.3) \qquad \Psi_{k,t}(\boldsymbol{x}) := \left( e^{-\lambda_1 t} \psi_1(\boldsymbol{x}), e^{-\lambda_1 t} \psi_2(\boldsymbol{x}), \dots, e^{-\lambda_1 t} \psi_k(\boldsymbol{x}) \right).$$

Eq. (3.2) shows that the diffusion distance, which is a natural measure to assess the dynamical proximity of two configurations of the system, is equal to the Euclidean distance between the corresponding diffusion map coordinates (with $k = \infty$).

In principle, equality in (3.2) holds only for $k = \infty$. However, in practice, many stochastic systems exhibit a *spectral gap* with only a few eigenvalues close to zero, and the rest converging to infinity. In these cases, to estimate the expected evolution of a stochastic dynamical system currently located at $\boldsymbol{x}$, it suffices to know only the first few eigenvalues and eigenfunctions $\{(\lambda_1, \psi_1(\boldsymbol{x})), (\lambda_2, \psi_2(\boldsymbol{x})), \dots, (\lambda_k, \psi_k(\boldsymbol{x}))\}$. Then, the diffusion map (with a small number $k$) is a lower dimensional representation of the system that captures the essential features for its expected long term dynamical evolution. Note that the number of required coordinates, $k$, depends on the time scale relevant to the observer and on the required accuracy of the approximation. The larger the observer's time scale, the smaller the number of coordinates needed to describe the evolution on this time scale.

Moreover, it is possible to show that this mapping is *optimal* among all possible $k$-dimensional mappings, in the following mean squared error sense. Consider a $k$-dimensional approximation of the transition probability densities $p(\boldsymbol{x}, t | \boldsymbol{y})$, which

measure the approach to equilibrium from the starting point $\boldsymbol{y}$. A $k$-dimensional approximation is a linear expansion of the form

$$(3.4) \qquad p_k(\boldsymbol{x}, t|\boldsymbol{y}) = \sum_{j=0}^{k-1} \alpha_j(\boldsymbol{y})v_j(\boldsymbol{x})$$

where the functions $v_j$ are orthonormal in $L_2(\Omega, w)$. Then the following theorem holds.

**Theorem:** *The optimal $k$-dimensional approximation of $p(\boldsymbol{x}, t|\boldsymbol{y})$ that minimizes the mean squared norm of the approximation error*

$$(3.5) \qquad \min \mathbb{E}_{\boldsymbol{y}}\{\|p(\boldsymbol{x}, t|\boldsymbol{y}) - p_k(\boldsymbol{x}, t|\boldsymbol{y})\|^2_{L_2(\Omega, w)}\}$$

*where averaging is over all initial points $\boldsymbol{y}$ sampled according to the equilibrium density $\varphi_0(\boldsymbol{y})$, is given by $v_j(\boldsymbol{x}) = \varphi_j(\boldsymbol{x})$ and $\alpha_j(\boldsymbol{y}) = e^{-\lambda_j t}\psi_j(\boldsymbol{y})$. Therefore, the optimal $k$-dimensional approximation is given by the truncated sum*

$$(3.6) \qquad p_k(\boldsymbol{x}, t|\boldsymbol{y}) = \sum_{j=0}^{k-1} e^{-\lambda_j t}\psi_j(\boldsymbol{y})\varphi_j(\boldsymbol{x}).$$

The proof, given in the Appendix, is essentially a continuum version of the well known PCA/SVD decomposition.

**3.1. Eigenfunctions and Reduction Coordinates.** In typical applications in physics, chemistry and biology, while the original system under study is high dimensional with many fast and coupled degrees of freedom, its long term evolution, or the evolution of some of its statistical properties are low dimensional. In such cases, and specifically in systems whose dynamics are governed by rare events, one seeks to describe the coarse grained evolution of the system by only a few slow variables or reduction coordinates, which we shall denote by $\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), \ldots, r_k(\boldsymbol{x}))$.

One criterion for $\boldsymbol{r}(\boldsymbol{x})$ to be good reduction coordinates is that regardless of their initial values, the remaining coordinates or their statistics are *slaved* to $\boldsymbol{r}(\boldsymbol{x})$ and their statistics quickly equilibrate to some probability distribution (invariant measure), which may depend on the values of $\boldsymbol{r}(\boldsymbol{x})$. In this case, for time scales longer than the relaxation time of the (remaining) fast variables, the time evolution of these reduction coordinates is approximately *Markovian*,

$$(3.7) \qquad \frac{d\boldsymbol{r}}{dt} = F(\boldsymbol{r}, \boldsymbol{w})$$

where $F$ depends only on the reduction coordinates $\boldsymbol{r}$ and on a stochastic noise vector $\boldsymbol{w}(t)$. The above relation is sometimes denoted in the literature as a *closure*, in the sense that the dynamics of $\boldsymbol{r}$ are closed within themselves: no additional information about the original configuration $\boldsymbol{x}$ is required to describe the time evolution of $\boldsymbol{r}$. Eq. (3.7) is obviously an approximation whose interpretation is short memory, or weak dependence on the past.

In the context of chemical or biological reactions, these reduction coordinates should in addition provide a meaningful measure of the progress of a reaction [6]. We also note that if (3.7) is one dimensional with a known $F$ (e.g., only a single reduction coordinate), it can easily be solved semi-analytically or numerically.

Given a few reduction coordinates $\boldsymbol{r}(\boldsymbol{x})$, the standard theoretical method to compute their temporal evolution is via the Mori-Zwanzig projection operator approach

[55, 24]. In some situations, the decomposition into slow reduction coordinates and remaining fast variables is *assumed* known a-priori and/or given explicitly by the forms of the equations, see for example [24, 51, 28]. In some of these cases, rigorous mathematical results regarding the closure properties of the slow variables can be proven and the specific form of the function $F$ in (3.7) can be derived. In more complicated problems, where the decomposition into slow and fast variables is not clear and rigorous mathematical results are intractable, one typically uses physical or chemical intuition to suggest good reduction coordinates and simulations to empirically study their dynamics, see for example [31, 34, 20]. The main practical problem, however, is that in the majority of cases concerning complex high dimensional systems such an a-priori intuition is unavailable. Moreover, poor choices of reduction coordinates fail to satisfy the closure property (3.7), so that their dynamics are non-Markovian with long memory effects.

The following lemma shows that in the presence of a spectral gap, the first few eigenfunctions are natural reduction coordinates.

**Lemma:** *Consider the SDE (2.1) with a spectral gap, e.g., with $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{k-1} \ll \lambda_k$; then for time scales $t \gg \max(1/\lambda_k, t_0)$, where $t_0$ is defined below, the first few diffusion map coordinates $(\psi_0(\boldsymbol{x}), \psi_1(\boldsymbol{x}), \psi_2(\boldsymbol{x}), \ldots, \psi_{k-1}(\boldsymbol{x}))$ are approximately Markovian.*

**Proof:** Let $p_k(\boldsymbol{x}, t \,|\, \boldsymbol{y})$ denote the approximation of the transition density using the first $k$ diffusion map coordinates,

$$(3.8) \qquad p_k(\boldsymbol{x}, t \,|\, \boldsymbol{y}) = \varphi_0(\boldsymbol{x}) + \sum_{j=1}^{k-1} e^{-\lambda_j t} \psi_j(\boldsymbol{y}) \varphi_j(\boldsymbol{x}).$$

These coordinates are approximately Markovian, if regardless of the initial location $\boldsymbol{y}$, $p_k(\boldsymbol{x}, t | \boldsymbol{y}) = f(\psi_0(\boldsymbol{y}), \ldots \psi_{k-1}(\boldsymbol{y}))$ is close to $p(\boldsymbol{x}, t | \boldsymbol{y})$. We now show that this indeed holds, in the sense of the $L_1$ norm. Rather that proving directly that $\|p - p_k\|_1$ is small we instead consider the function $\varphi_0 + p - p_k$. First we show that for sufficiently large times this is a probability density. Then, we bound its relative entropy with respect to $\varphi_0$, which bounds the required $L_1$ norm.

The proof proceeds as follows: First, note that

$$\int p_k(\boldsymbol{x}, t | \boldsymbol{y}) d\boldsymbol{x} = 1.$$

Further, since both $\varphi_j$ and $\psi_j$ are bounded, then for large enough times the finite sum in (3.8) is strictly smaller than $\varphi_0(\boldsymbol{x})$ and so $p_k \geq 0$ is indeed a probability density. We define $t_0$ as the minimal time for which $|p - p_k| \leq \varphi_0(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$.

For times $t \geq t_0$, $\varphi_0 + p - p_k$ is a probability density, and we consider its free energy, also known as its relative entropy with respect to $\varphi_0 = e^{-\beta U}$,

$$(3.9) \qquad H(\varphi_0 + p - p_k | \varphi_0) = \int_\Omega (\varphi_0 + p - p_k) \log \left( \frac{\varphi_0 + p - p_k}{\varphi_0} \right) d\boldsymbol{x}.$$

The relative entropy (3.9) bounds the $L_1$ distance between the two densities $p - p_k$ in light of the well known Csiszár-Kullback-Pinsker inequality

$$(3.10) \qquad \|\rho_1 - \rho_2\|_1^2 \leq 2H(\rho_1 | \rho_2)$$

Inserting (2.4) into (3.9) and using the inequality $\log(1 + x) \leq x$ gives

$$H(\varphi_0 + p - p_k|\varphi_0) \leq \int \frac{(p - p_k)^2}{\varphi_0} d\boldsymbol{x}.$$

Opening the brackets and applying the orthogonality conditions on the eigenfunctions $\varphi_j(\boldsymbol{x})$ gives

(3.11) $$H(\varphi_0 + p - p_k|\varphi_0) \leq \sum_{j \geq k} e^{-2\lambda_j t} \psi_j^2(\boldsymbol{y}).$$

The sum $\sum_j \psi_j^2(\boldsymbol{y})$ is infinite. However, using the equality

$$p(\boldsymbol{y}, t|\boldsymbol{y}) = \varphi_0(\boldsymbol{y}) \left[ 1 + \sum_{j \geq 1} e^{-\lambda_j t} \psi_j^2(\boldsymbol{y}) e \right]$$

and (3.8), for times $t > t_0$ we can bound the right hand side of (3.11) by

(3.12) $$H \leq \frac{p(\boldsymbol{y}, t|\boldsymbol{y}) - p_k(\boldsymbol{y}, t|\boldsymbol{y})}{\varphi_0(\boldsymbol{y})} e^{-\lambda_k t} \leq e^{-\lambda_k t}.$$

The last equation shows that for $t \gg \max(1/\lambda_k, t_0)$ the relative entropy is indeed small. However, note that the decay of $H$ depends also on the initial point $\boldsymbol{y}$, and so the quality of the approximation by the first $k$ diffusion map coordinates depends also on the density at $\boldsymbol{y}$. For trajectories that start at extremely rare (low density) points $\boldsymbol{y}$ with $\varphi_0(\boldsymbol{y}) \ll 1$, the approximation with only $k$ coordinates may not be appropriate. On the other hand, for computations of various macroscopic quantities, such as mean exit times, we will mostly be interested in low dimensional approximations at the "bottom" of the metastable states and near the saddle points connecting them, and not at other points with much lower densities. Finally, we note that relative entropy and more general logarithmic Sobolev inequalities have been extensively used both to bound the heat kernel (Green's function) of elliptic operators as well as to study the decay to equilibrium of more general Fokker-Planck type equations, see [15, 35].

**3.2. Spectral Gaps in Reversible Diffusions.** The eigenfunction expansion and the lemma above show that for dimensionality reduction using the first $k$ diffusion map coordinates one of the following two conditions is required. Either there is an explicit spectral gap $\lambda_{k+1} \gg \lambda_k$, or the next few diffusion map coordinates $\psi_{k+1}, \ldots, \psi_{k+m}$ are uniquely determined by the first $k$ diffusion map coordinates and in addition $\lambda_{k+m+1} \gg \lambda_k$, so again there is an effective spectral gap.

Both cases represent a separation of time scales. The most common setting for the first case ($\lambda_{k+1} \gg \lambda_k$) is overdamped diffusion in an energy landscape having $k$ potential wells. In this case, the $k-1$ smallest non-zero eigenvalues of $\mathcal{L}$ are inversely proportional to mean exit times from the different wells, and thus very close to zero. The remaining eigenvalues capture the relaxation times in each of the individual wells, and are thus significantly larger [36].

The second case can occur for example when there is a slow variable $x_1$ without any deep well in its effective potential, but the dynamics of the remaining variables are much faster, so they effectively equilibrate or are slaved to the slow variable. In this case the first few eigenfunctions are all functions of the slow variable $x_1$, and

so there is an effective spectral gap till the first eigenfunction that depends on the remaining fast variables $x_2, \ldots, x_n$.

There are at least two cases where dimensionality reduction using the first $k$ diffusion map coordinates may fail. The first is when there is a hierarchy of spectral gaps and the other is when there is no clear separation of time scales. For example, for diffusion in a multi-well potential, if there is one wide potential well whose internal relaxation time is significantly larger than the exit times between two other smaller wells, the first few eigenfunctions may fail to capture the dynamics between the small wells [40]. In such cases, the reduced dynamics in the first few coordinates may be strongly misleading.

**Example:** Consider the simple case of a high dimensional system with two deep wells centered at $\boldsymbol{x}_0$ and at $\boldsymbol{x}_1$, separated by a high potential barrier with a single saddle point. This system is governed by rare events which are the transitions between the two wells. Therefore, there is a single very small positive eigenvalue $\lambda_1 > 0$ with a spectral gap to the next eigenvalue $\lambda_2 \gg \lambda_1$. The eigenvalue $\lambda_1$ is intimately connected to the characteristic equilibration time of the system, and in the asymptotic limit of small noise or low temperature (as $\beta \to \infty$) it is given by $1/\bar{\tau}$ where $\bar{\tau}$ is the mean first passage time for the system to surmount the highest potential barrier on its way towards the deepest well [36]. Moreover, it follows from large deviation theory that in the asymptotic limit of small noise ($\beta \to \infty$), almost all of the transition paths are concentrated along a smooth curve $\gamma$ connecting the two wells and passing through the saddle point [45]. This curve is defined by the characteristics of the eikonal equation, which in the case of a potential system is the solution of the ordinary differential equation

$$(3.13) \qquad \frac{d\gamma(t)}{dt} = \nabla U(\gamma(t))$$

with initial condition $\gamma(0) = \boldsymbol{x}_s + \varepsilon \boldsymbol{n}$ where $\boldsymbol{x}_s$ is the location of the saddle, $\varepsilon$ is small and $\boldsymbol{n}$ is the direction corresponding to the negative eigenvalue of the Hessian of $U$ at $\boldsymbol{x}_s$. These results imply that it is possible to write an effective one dimensional SDE in any variable $s$ that parameterizes the curve $\gamma$ (for example, its arclength)

$$(3.14) \qquad \frac{ds(t)}{dt} = f(s) + \sqrt{2D(s)}\dot{w}(t).$$

A "natural" reduction coordinate for this system is thus the arclength $s$.

In our approach for this case we would need only the first non-trivial backward eigenfunction $\psi_1$. This eigenfunction is a monotonic function of the arclength of $\gamma$, which is approximately equal to a constant $c_1$ in one well, a different constant $c_2$ in the other well, with a rapid transition between these two values within an internal boundary layer near the saddle point of the potential energy $U(\boldsymbol{x})$ [36]. Therefore, while with the diffusion map approach our reduction coordinate is not the curve arclength itself, it is a monotonic function of it (see also section 5 for a specific numerical example).

**4. Diffusion Maps from Discrete Data.** For simple low dimensional stochastic systems, the eigenfunctions of the FP operator can be computed explicitly by standard numerical methods. While in higher dimensions this is difficult if not practically impossible, simulations of sample paths of the corresponding SDE (2.1) are relatively easily performed, and millions of configurations can be sampled. From this

set of data, the minima of the potential $U(\boldsymbol{x})$ as well as the statistics of the transition times between the metastable states can be approximated. Obviously, since the system spends most of the time in the deep wells and only an exponentially small fraction in transitions between them, direct simulations require very long runs to obtain good statistical estimates of the mean exit times and of the transition probabilities between metastable states. Alternatively, if prior knowledge regarding locations of metastable states and dividing surfaces are known, sophisticated biasing techniques can be employed to sample the interesting paths.

In this section, following our earlier work [39, 10] we present a method to calculate a discrete approximation of the first few of these eigenfunctions given enough simulated data. Given a set $\mathcal{S} = \{\boldsymbol{x}_i\}_{i=1}^N$ of simulated data points from the SDE (2.1), possibly from many independent simulation runs, the algorithm works as follows:

1. Choose an $\varepsilon > 0$, which is a measure of the local neighborhood of any point in the dataset; this is an adjustable parameter of the algorithm. Consider the kernel $K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2 / 2\varepsilon^2)$;

2. For each $\boldsymbol{x} \in \mathcal{S}$, compute the quantity $p_\varepsilon(\boldsymbol{x}) = \sum_j K(\boldsymbol{x}, \boldsymbol{x}_j)$, and construct the following matrix

$$(4.1) \qquad \tilde{K}_{i,j} = \frac{K(\boldsymbol{x}_i, \boldsymbol{x}_j)}{\sqrt{p_\varepsilon(\boldsymbol{x}_i) p_\varepsilon(\boldsymbol{x}_j)}}.$$

3. Define $D_i = \sum_j \tilde{K}_{i,j}$ and construct a Markov matrix $M = D^{-1}\tilde{K}$, with

$$M_{i,j} = \frac{\tilde{K}_{i,j}}{D_i}.$$

4. Compute the first few eigenvalues and right eigenvectors of $M$, $(Mv = \lambda v)$.

In [39, 12, 25, 48, 4] (and references therein) it is shown that for points $\boldsymbol{x}_i$ randomly sampled from a probability density $\varphi_0(\boldsymbol{x}) = C_\beta e^{-\beta U(\boldsymbol{x})}$, as the number of points $N \to \infty$, and as $\varepsilon \to 0$, the discrete operator $(M - I)/\varepsilon$ converges (in probability) to the backward FP operator (2.8). Thus, the eigenfunctions of the FP operator can be approximated from large simulated datasets also for high dimensional systems where standard discretization methods are not feasible. We present some examples in section 5.

**Remarks:**

1) The quantity $p_\varepsilon(\boldsymbol{x})$ is, up to normalization, a density estimation at $\boldsymbol{x}$. For systems where the energy of a configuration $U(\boldsymbol{x})$ is known explicitly, the points $\{\boldsymbol{x}_i\}$ do not have to be sampled with density proportional to their equilibrium density. For example, if configurations are sampled by a biased procedure, then formula (4.1) is replaced by

$$(4.2) \qquad \tilde{K}_{i,j} = \frac{K(\boldsymbol{x}_i, \boldsymbol{x}_j)}{\sqrt{e^{-\beta U(\boldsymbol{x}_i)} e^{-\beta U(\boldsymbol{x}_j)}}}.$$

2) The matrix $M$ is adjoint to a symmetric matrix $M_s = D^{1/2} M D^{-1/2} = D^{-1/2} \tilde{K} D^{-1/2}$. Therefore, the numerical computation of the first few eigenvalues and eigenvectors can be made on $M_s$, using (typically faster and more robust) algorithms for symmetric matrices.

3) The first few eigenfunctions of the FP operator are, under general conditions, smooth functions of $\boldsymbol{x}$. Therefore, in practice, there is no need to compute the eigenvectors of the huge $N \times N$ matrix $M$ from all the data. Instead, it is possible to

*sub-sample* the dataset, for example retaining only points which are at least a distance of $\delta$ apart (where $\delta$ is a meta-parameter). This leads to the computation of eigenvectors of a much smaller (and sparse) matrix. Moreover, there exist fast multiscale algorithms to compute the eigenvectors of such matrices [8, 11].

4) As discussed in [38], there is an interesting connection between the study of dynamical systems and data clustering in the machine learning community. The basic observation is that metastable states in dynamical systems correspond to well defined clusters. Therefore, machine learning methods to identify clusters from large datasets whose dynamical equations either are unknown or do not even exist (for example, as in document clustering), can equally be used to identify metastable states from large datasets arising from simulations of dynamical systems. Indeed, various variants of the algorithm described above have been suggested for spectral clustering, see for example [54, 3, 47]. However, none of these works considered the underlying relation to the asymptotic computation of the eigenfunctions of a Fokker-Planck operator. As a side issue, our analysis provides a mathematical explanation for the success and limitations of these methods in their tasks of identifying clusters [38, 40]. Finally, we note that similar approaches have also been used for clustering of conformational states of dynamical systems [18, 19, 29].

5) In our formulation, eq. (2.1) we assumed equal noise strengths in all variables. If however, the noise strength is different for different coordinates, as in

$$(4.3) \qquad dx_i = -\gamma_i \frac{dU}{dx_i} dt + \sqrt{2\gamma_i}\dot{w}_i$$

then in order to consistently approximate the corresponding Fokker-Planck operator, we simply need to change our local Euclidean metric to a *weighted* anisotropic one,

$$(4.4) \qquad \|\boldsymbol{x}\| = \sum_{i=1}^{n} x_i^2/\gamma_i \ .$$

Even in the more general case where the diffusion tensor may depend on the configuration $\boldsymbol{x}$, detection of slow variables is possible by using anisotropic diffusion maps [49].

6) The variable $\varepsilon$ is a meta-parameter of the algorithm. The theoretical analysis shows that in the limit of infinite data, as $\varepsilon \to 0$, the finite Markov matrix approximates the Fokker-Planck operator. However, for a finite simulation dataset $\varepsilon$ must remain finite. It should be small, so that indeed we only consider local nearby points in the construction of the graph, but not too small due to the finite size of the dataset. The work [48] presents a possible criterion for the choice of $\varepsilon$.

**4.1. Dynamics on the Diffusion Map Coordinates.** Beyond the benefits of dimensional reduction, the projection of the system onto the diffusion map coordinates also allows systematic design of computational experiments, where biased simulations are initialized at chosen values of the diffusion map coordinates, thus allowing efficient exploration of the dynamics of the system in these coordinates. Specifically, we propose to study the dynamics of the system projected onto these few coordinates by the "lift-run-restrict" scheme, as shown in figure 4.1.

This scheme requires three operators. The first is a *lifting* operator which, given values for the diffusion map coordinates, produces random configurations consistent with these coordinates. The second is a black-box simulator which runs the dynamics of the system from such an initial configuration, and the third is a *restriction* operator,
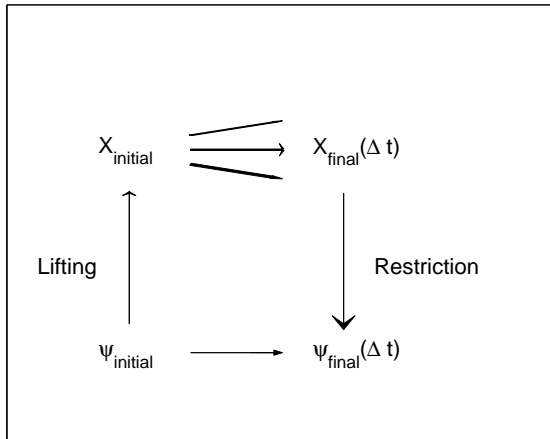
FIG. 4.1. *The lift-run-restrict scheme on the diffusion map coordinates.*

which computes values of the diffusion map coordinates for new points not contained in the original dataset.

Given these operators, the lift-run-restrict approach works as follows: First, we create a large dataset of simulated configurations and compute the diffusion map. Suppose, for simplicity, that the system can be projected into the first diffusion map coordinate $\psi_1$. That is, we assume there exists an effective drift $f(\psi_1)$ and an effective diffusion coefficient $D(\psi_1)$, such that the time evolution of $\psi_1(\boldsymbol{x}(t))$ approximately satisfies the Îto SDE

$$(4.5) \qquad \dot{\psi}_1 = f(\psi_1) + \sqrt{2D(\psi_1)}\dot{w}.$$

To estimate these drift and diffusion coefficients, we discretize the possible values of $\psi_1$ into $m$ bins centered at $a_1 < a_2 < \ldots < a_m$. For each value $a_j$, using the lifting operator we create $k$ random initial configurations $\boldsymbol{x}$ such that $\psi_1(\boldsymbol{x}) = a_j$. For each such configuration we simulate the full system dynamics for a time $\tau$, which is long in comparison to the fast modes of the system, but still much shorter than the global equilibration time (the MFPT, for example). At the end of each such run, using the restriction operator we compute the value of $\psi_1(\boldsymbol{x}(\tau))$. The drift and diffusion coefficients are now estimated from the statistics of $\psi_1(\boldsymbol{x}(\tau))$, for example by maximum likelihood methods [1].

Given a reduction coordinate, this scheme was suggested in [31] and applied to many problems, where identification of a reduced coordinate was relatively easy. Similar approaches are also described by Brandt [7] and by Vanden-Eijnden [51]. The novelty of our approach is that we apply this method on the (non-linear) diffusion map coordinates which are themselves computed from the data, whereas other approaches require a-priori knowledge of the reduction coordinates which are typically chosen to be a single phase space variable, or, possibly, a linear combination of several of them.

According to our approach, instead of the original high dimensional process $\boldsymbol{x}(t)$, we study the low dimensional process $\{\psi_1(\boldsymbol{x}(t)), \ldots, \psi_k(\boldsymbol{x}(t))\}$. According to Ito's formula, each component of this diffusion map satisfies the following SDE

$$(4.6) \qquad d\psi_j(\boldsymbol{x}(t)) = -\lambda_j \psi_j(\boldsymbol{x}(t))dt + \sqrt{\frac{2}{\beta}}\|\nabla_{\boldsymbol{x}}\psi_j(\boldsymbol{x}(t))\|dw_j.$$

13

Therefore, the condition for the first diffusion map coordinate to be approximately Markovian, is that $\|\nabla_{\boldsymbol{x}}\psi_1(\boldsymbol{x})\|$ is approximately some function of only $\psi_1$.

**4.2. Lifting and Restriction Operators.** Our approach can be viewed as a non-linear principal component analysis (PCA). For linear PCA, the translation between the original high dimensional space and the low dimensional space spanned by the first few eigenvectors of the covariance matrix is simple, since it amounts to the computation of linear dot products. In our case, since our mapping is non-linear, more complicated operators that translate between the two spaces are needed (e.g., see also [42]).

**Restriction Operator:** Let $\boldsymbol{x}$ be a new point, not necessarily in the original set $\mathcal{S}$. Our aim is to compute $\psi_1(\boldsymbol{x}), \psi_2(\boldsymbol{x}), \ldots, \psi_k(\boldsymbol{x})$, without re-computation of the eigenvalues and eigenvectors of the matrix $M$ with the point $\boldsymbol{x}$ added to it. To this end we note that by definition, for any point $\boldsymbol{x}_k \in \mathcal{S}$,

$$(4.7) \qquad \lambda_j \psi_j(\boldsymbol{x}_k) = \sum_{\boldsymbol{y} \in \mathcal{S}} M(\boldsymbol{x}_k, \boldsymbol{y})\psi_j(\boldsymbol{y}).$$

For a new point $\boldsymbol{x}$ the same formula should hold. However, while the left hand side is unknown, the right hand side can be approximated, which gives the following formula for the extension of the diffusion map values to new points,

$$(4.8) \qquad \psi_j(\boldsymbol{x}) = \frac{1}{\lambda_j} \sum_{\boldsymbol{y} \in \mathcal{S}} M(\boldsymbol{x}, \boldsymbol{y})\psi_j(\boldsymbol{y}).$$

This formula is known as the Nyström extension and has been widely used in many contexts, see for example [5, 2]. In machine learning this formula plays an important role, as it enables learning the classification of new test points, not present in the training set.

**Lifting Operator:** For simplicity we describe a lifting operator based only on the first diffusion map coordinate. Similar operators can be constructed for higher dimensional versions. Let $\theta$ be a value for the first diffusion map coordinate, and let $p_\theta(\boldsymbol{x})$ denote the marginal probability density of the equilibrium density $p(\boldsymbol{x})$ under the constraint $\psi(\boldsymbol{x}) = \theta$. Our aim is to construct a lifting operator that outputs random configurations $\boldsymbol{x}$, according to $p_\theta$.

Given the restriction operator above, there is a simple method to approximately sample from this density, by running constrained stochastic dynamics as in (2.1), but with a modified potential

$$(4.9) \qquad U_\kappa(\boldsymbol{x}) = U(\boldsymbol{x}) + \kappa(\psi_1(\boldsymbol{x}) - \theta)^2.$$

The equilibrium density corresponding to this potential is given by $C_{\beta,\kappa}e^{-\beta U_\kappa(\boldsymbol{x})}$. where $C_{\beta,\kappa}$ is a normalization factor. For large enough values of $\kappa$ this density is sharply peaked at points $\boldsymbol{x}$ satisfying the required constraint $\psi_1(\boldsymbol{x}) = \theta$. Starting from any point $\boldsymbol{x}$, as time $t \to \infty$ the simulated points approach this equilibrium density. By restricting to those points with $\psi_1(\boldsymbol{x}) = \theta$ up to a specified precision, we effectively sample from the density $p_\theta$ as required. Moreover, since $\psi_1$ is a slow coordinate, the equilibration time of the modified system is still of the same order of magnitude as the fast relaxation time $\tau_R$. Therefore, to sample from $p_\theta$ it is not
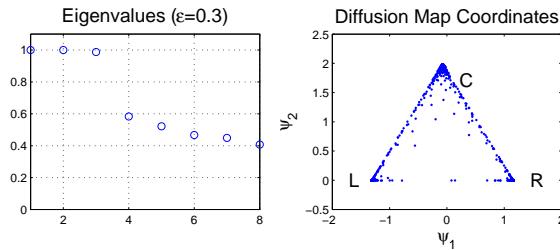
FIG. 5.1. *(Left) Eigenvalues of the diffusion map and (right) projection of simulated data onto the first two diffusion map coordinates for the three well system.*

required to run long simulations with the above constrained dynamics. We remark that simulating the SDE (2.1) with the modified potential (4.9) can be viewed as a specific realization of biased umbrella sampling.

**4.3. Active Phase Space Exploration.** As described above, the diffusion map coordinates are approximated using currently available simulation data. If the interesting parts of the phase space have been appropriately sampled, then the computed diffusion maps will provide a reasonable reduction coordinate. There are, however, two possible scenarios that deserve special treatment. The first is a case where the state space has been only partly explored, for example only some of the metastable states have been sampled. The second is when the transition regions have been explored only sparsely. These two cases can be handled with the aid of lifting and restriction operators.

In the first case, where only part of the phase space has been explored, the current computation of the diffusion map coordinates can be used to bias the simulation and initialize new simulations near the boundaries of the currently explored phase space. This can lead to faster discovery of new metastable states. In the second case, once a rough estimate of the diffusion map and thus of the transition region is known, it can be used to bias simulations and sample many more points near the transition region, leading to an improved recomputation of the diffusion map coordinates. We illustrate both scenarios in example 5.1 below. We would like to remark, however, that this purely data-driven approach has its limitations. For example, due to the curse of dimensionality, it may fail if there are too many potential directions to explore that are orthogonal to the directions spanned by the first few diffusion map coordinates. These issues require further study.

**5. Examples.**

**5.1. A Three Well Potential.** Consider the following three well potential in two dimensions

$$(5.1) \quad U(x,y) = 3e^{-x^2-(y-1/3)^2} - 3e^{-x^2-(y-5/3)^2} - 5e^{-(x-1)^2-y^2} - 5e^{-(x+1)^2-y^2}$$

constrained to the domain $(x,y) \in \Omega = [-4,4] \times [-4,4]$. This potential has two deep wells at $x_L = (-1.113, -0.03685)$, $x_R = (1.113, -0.03685)$, and another shallow well at $x_C = (0, 1.7566)$. Therefore, there are two possible paths to move from $x_L$ to $x_R$: A direct path, and one that passes through the central well $x_C$. Properties of (2.1) with this potential were studied by various authors (see [41, 17] and references therein).

In figure 5.1 we present simulation results for this potential with temperature parameter $\beta = 4.0$. On the left we plot the eigenvalues of the Markov matrix $M$
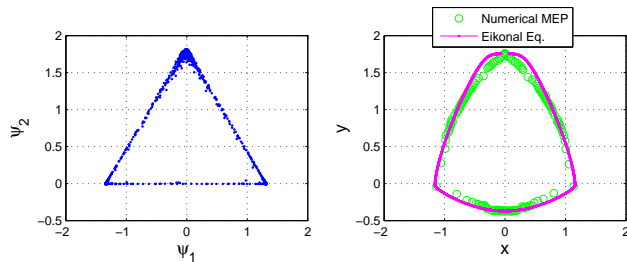
15

FIG. 5.2. *(Left) Diffusion map coordinates after removal of highly improbable points and insertion of probable points along the path. (Right) Theoretical (purple line) vs. numerical minimum energy path computed using the diffusion map coordinates (green dots).*

with parameter $\varepsilon = 0.30$. As expected, besides the trivial eigenvalue $\lambda = 1$ there are exactly two additional eigenvalues very close to 1 with a spectral gap between $\lambda_2$ and $\lambda_3$. On the right we plot the projection of the simulated data onto the first two eigenfunctions $\psi_1$ and $\psi_2$. As seen from this figure, other than at the bottom of the wells where the dynamics are two dimensional, this projection reduces the dimensionality of the system from 2-D to 1-D.

As seen from this figure, there are not many points along the direct path between $x_L$ and $x_R$. With the aid of the lifting and restriction operators we produce many more points there. As a result we obtain a more uniform covering of the triangle of pathways on the diffusion map space, as shown in figure 5.2 (left). In addition, we can also numerically find the location of the saddle point and the most probable path as follows: For each point $(\psi_1, \psi_2)$ on the 1-D curve connecting two metastable wells in the diffusion map space, we look for the point $(x, y)$ in physical space with the minimum potential energy. This is a numerical scheme to approximate the minimum energy path (MEP), and its comparison to the theoretical Eikonal equation is shown in figure 5.2 (right).

Finally, another task we can perform with the diffusion map coordinates is a "clever" exploration of phase space, or "climbing out from a potential well". To illustrate, consider the same system but at a very low temperature ($\beta = 10$), and with an initial knowledge of only a single metastable state, the one at $x_L$. In this case, starting from $x_L$ a direct simulation would require an extremely long time to exit this well and find the other metastable states. However, using the diffusion map approach, we simulate many short paths inside the well, subsample them, compute the diffusion map, and then restart paths at locations whose *diffusion distance* is the largest from the initial point at the center of the well. This procedure allows a systematic exploration of points further and further away from the left well, leading to a discovery of the central well orders of magnitude faster than by a standard simulation. A snapshot of this process for the first 6 iterations is shown in figure 5.3. As seen from the figure, after only 6 iterations, in which we ran our stochastic simulator for a short time of only $\tau = 0.15$, we discover the central well. For comparison purposes, for $\beta = 10$ the mean exit time from the left well is approximately $\bar{\tau} = O(10^{10})$. Finally, we remark that this approach can also be easily used in conjunction with other acceleration methods, such as the parallel replica method and temperature-accelerated dynamics [53]. In principle one can even accelerate this type of computation by smoothly projecting backward in time (reverse coarse projective integration) [23, 34].
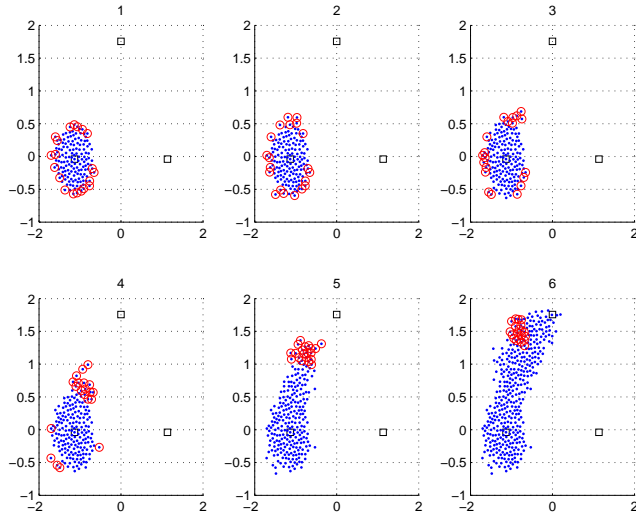
FIG. 5.3. *Iterative climbing up from the well at $x_L$. At each new iteration we start from the 20 most distant points in diffusion distance from $x_L$ (red circles). The subsampled set of simulated points at each iteration is denoted by blue dots, whereas the black squares mark the potential minima corresponding to the three wells.*

**5.2. An Attracting Manifold Example.** Consider the SDE (2.1) with the following potential in 2-D,

$$(5.2) \qquad U(x,y) = \left(\frac{x-3}{2}\right)^4 + \mu\left(y - \left(1 - \frac{1}{1+x}\right)\right)^2$$

with $\beta = 1$, defined in $\Omega = \{(x,y)|x > 0\}$ with reflecting boundary conditions at $x = 0$.

When $\mu$ is large, for example $\mu = 40$, most points lie close to the manifold (1-D curve in our example) defined by

$$(5.3) \qquad y = 1 - \frac{1}{1+x}.$$

Since $\mu$ is large, a trajectory initialized at a large distance away from this manifold will first quickly approach the manifold and then perform random dynamics on it. Therefore, the effective dynamics are essentially one-dimensional, constrained to the interval $x \in [0,6]$ by the first term in (5.2). In this specific example the first variable $x$ which parameterizes the manifold, is also a "slow" variable, in the sense that an approximate equation can be written for the evolution of $x(t)$, which does not involve knowledge of $y(t)$. However, not every single variable that parameterizes the manifold is a slow variable. Suppose for example that instead of observing points $(x, y)$, we observe points $(w, z)$ in a rotated coordinate system with $\theta = 45^o$,

$$(5.4) \qquad \begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Even though $w$ also parameterizes the manifold, it is not a slow variable. The evolution of $w(t)$ depends not only on $w(0)$ but also on the initial value $z(0)$. In figure 5.4 we
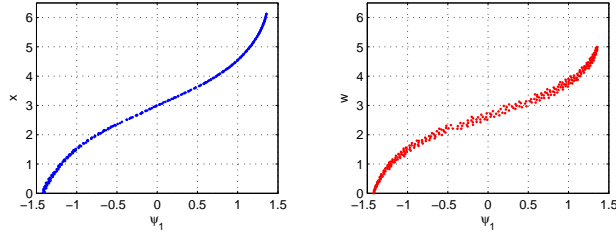
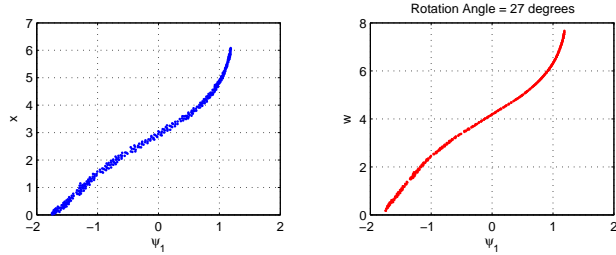FIG. 5.4. *The observables $x$ and $w$ vs. the reduction coordinate $\psi_1$ for the potential (5.2).*



FIG. 5.5. *The observables $x$ and $w$ vs. the reduction coordinate $\psi_1$ for the potential (5.5).*

plot the two different observables $x$ and $w$ vs. the reduction coordinate $\psi_1$ computed using points sampled from simulated paths of the SDE. As shown in the figure, $x$ is almost 1-to-1 with $\psi_1$, whereas $w$ is not (notice the fatness of the plot), indicating that indeed $w$ is not a good reduction coordinate.

Suppose we observed points in the $(w, z)$ coordinate system. Can we still find the slow variable $x$ ? By computing the diffusion map, this simply amounts to searching for a rotation angle $\theta$ which makes the variable $w\cos(\theta) + z\sin(\theta)$ as closest to one-to-one with $\psi_1$ as possible.

As an example, consider the potential

$$(5.5) \qquad U(x,y) = \left(\frac{x-3}{2}\right)^4 + \mu\left(y - 2\sqrt{x}\right)^2 .$$

In this case, for large $\mu$ the attracting manifold is $y = 2\sqrt{x}$. However, even though both $x$ as well as $y$ parameterize the manifold, neither of them are slow variables. By assuming that some linear combination of $x$ and $y$ is a slow variable, we can optimize a 1-1 measure between $\psi_1$ and a rotated version of $(x, y)$ and find that at an angle $\theta = 27^o$, the variable $w = x\cos\theta + y\sin\theta$ is approximately 1-to-1 with $\psi_1$, and is thus approximately a slow variable. In figure 5.5 we plot both $x$ and $w$ vs. $\psi_1$. Notice the fatness in the $x - \psi_1$ plot vs. the almost perfect one-to-one correspondence between $w$ and $\psi_1$.

**5.3. The Seven Lennard-Jones Sphere System.** Consider the system of seven Lennard-Jones spheres in the plane, with a potential given by all pair-wise interactions

$$(5.6) \qquad U(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_7) = \sum_{i<j} U_{LJ}(\boldsymbol{x}_i - \boldsymbol{x}_j)$$
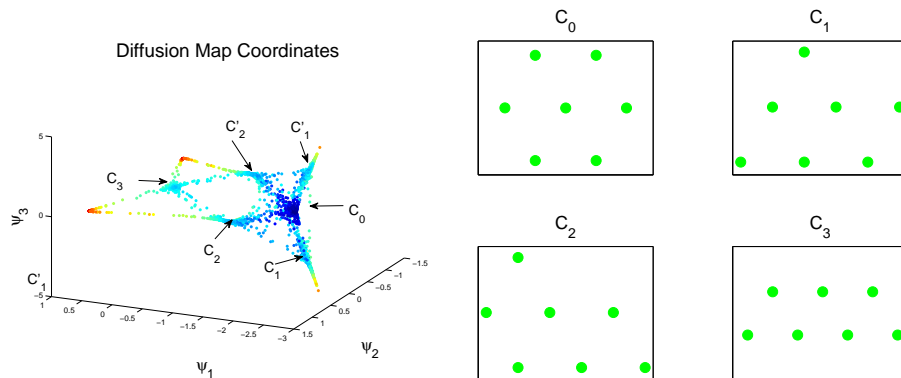
18

Fig. 5.6. *Non-linear dimensional reduction in terms of the first three diffusion map coordinates for the 7 Lennard-Jones spheres system (Left), and 4 metastable states (Right). The diffusion map has a symmetric shape since to each of the states $C_1, C_2$ there are also the reflected states $C'_1, C'_2$.*

where $U_{LJ}$ is the standard 6-12 Lennard-Jones potential. This system is 14 dimensional (or 12 dimensional after mean centering each configuration), and has a few metastable states with rare transitions between them [16, 20]. In figure 5.6 we present the first three diffusion map coordinates computed using data from a deterministic molecular dynamics simulation of this system at constant energy. The computations were based on more than 1 million configurations which were subsampled and reweighted (according to their empirical density) to about 3600 representative configurations. The similarity between configurations is based on a Gaussian kernel with width $\varepsilon = 0.2$. The distance between configurations was computed as follows, to be invariant under translations and rotations, but not reflections (hence the symmetry in the figure). Let $C_1$ and $C_2$ denote two configurations, mean centered and aligned along their principal axes of inertia. We denote by $C_1(i)$ the location of the $i$-th particle in that configuration. For a distance between configurations we considered the Hausdorff distance between the two sets,

$$d(C_1, C_2) = \max\{\max_{1 \le i \le 7} \min_{1 \le j \le 7} \|C_1(i) - C_2(j)\|, \max_{1 \le i \le 7} \min_{1 \le j \le 7} \|C_2(i) - C_1(j)\|\}$$

Given a large set of configurations, the complexity of the computation is of the order of $|S_{sub}| \times |S| \times W_{dist}$ where $S_{sub}$ is the size of the subsampled set, $|S|$ is the size of the original set of configurations and $W_{dist}$ is the complexity of distance calculation between a pair of configurations, plus the complexity of eigenvalue and eigenvector computations for the $|S_{sub}| \times |S_{sub}|$ matrix.

The projection of the system onto the first three diffusion map coordinates clearly shows the different metastable states and the approximately one-dimensional pathways between some of them. For example, it clearly shows that there are essentially no direct transitions between the state $C_0$ and the state $C_3$. A detailed diffusion map study of the dynamics of this system will be described in a separate publication.

**6. Summary and Discussion.** The problem of finding good low dimensional descriptions for high dimensional dynamical systems is an important and active area of research. In this paper we presented a rational and principled approach to this problem for the case of reversible diffusions, suggesting the eigenfunctions of the backward FP operator as good reduction coordinates. We then presented algorithms

to approximate the first few of these eigenfunctions from the exploration of the long-term dynamics using a given large set of simulated data. We also presented lifting and restriction operators that enable to study the effective dynamics on these coordinates by a multiple series of appropriately initialized short simulation runs. An essential requirement for this approach is the presence of a spectral gap between the eigenvalues so that the time scales of the coarse variables are well separated from the ones on the molecular/atomic motion. This is also the standard assumption for many other reduction methods. Further research is required to study more complex multi-scale systems, characterized by either a hierarchy of spectral gaps or by many different mixed time and length scales.

We note that while these eigenfunctions are optimal under a mean squared error criterion, other criteria for good reduction coordinates are obviously possible. For example, within the context of chemical reactions between a well defined reactant and product region, the optimal reaction coordinate is the set of isocommittor surfaces, with each surface defined as all points with the same probability $q$ to reach the product region before reaching the reactant region. This reaction coordinate is the solution of the backward FP equation with boundary conditions of zero in the reactant region and one in the product region. As with the eigenfunctions of the FP operator, it is difficult if not impossible to compute in high dimensions. Yet, many works propose optimizing surfaces and finding a good reaction coordinate based on this principle, see for example [6, 21] and references therein.

Our computational approach is closely related to the transfer operator approach [46], which also computes an approximation to the eigenfunctions of the FP operator [27], and to Perron cluster analysis [18, 19]. We remark that these methods are applicable to other dynamical systems, not necessarily described microscopically by diffusion in a potential field. Results analogous to those of Section 3 may obviously be more difficult to prove in such cases. We also note that [27] in fact considered the more general case of non-reversible diffusions and proved that the backward eigenfunctions can be used to partition the space into metastable states in this case as well. In comparison, there are two distinct features of our approach. The first is that our approximation scheme merges all simulated data points into account without using their time indexing. The second difference is that we propose to use the eigenfunctions as new low dimensional coordinates, on which it is possible to study the long term evolution of the system via the lift-run-restrict scheme. Our work is also related to the application of other dimensionality reduction techniques to data from molecular dynamics simulations, such as [14]. The main differences in our work is that the proposed dimensionality reduction is intrinsically related to the dynamics, and has provably good properties in approximating long-term behavior of the system.

In many chemical systems the transitions between metastable states reflect some underlying mechanism. In our approach we find a parametrization of the transition pathways and of the saddle points on them in the diffusion map space. Further work is needed to understand the corresponding mechanism in the physical configuration space (for example, if the transition is due to an infrequent flip in a given bond). This may be achieved via the restriction and lifting operators between the two spaces, for example, by computing the minimum energy path (MEP) between two metastable states.

Finally, in this paper we did not discuss issues of complexity and convergence rates, nor did we analyze the approximation errors and the number of points and iterations needed to achieve a given accuracy. These important mathematical issues

will be considered in future work.

**Appendix.** Proof of Theorem 1: By induction. First consider the zeroth order approximation $p_0 = \alpha_0(\boldsymbol{y})v_0(\boldsymbol{x})$. Since $p_0$ must be a probability density function, it follows that $\alpha_0(\boldsymbol{y}) = 1$. Using the orthonormality conditions of $\varphi_j$ and $\psi_j$, we obtain that $v_0(\boldsymbol{x}) = \varphi_0(\boldsymbol{x})$.

Now consider $p_1 = \varphi_0(\boldsymbol{x}) + \alpha_1(\boldsymbol{y})v_1(\boldsymbol{x})$. We decompose $v_1$ in the basis $\varphi_j$ and $\alpha_1$ in the basis $\psi_k$,

$$v_1(\boldsymbol{x}) = \sum_{j \geq 1} e^{-\lambda_j t} a_j \varphi_j(\boldsymbol{x})$$

$$\alpha_1(\boldsymbol{y}) = \sum_{k \geq 1} b_k \psi_k(\boldsymbol{y})$$

where $a_j$ and $b_k$ are yet undetermined. Inserting these expressions into (3.2) gives that

$$(6.1) \qquad \mathbb{E}_{\boldsymbol{y}}\{\|p - \hat{p}\|_{L_2(\Omega,w)}\} = \sum_j e^{-2\lambda_j t}(1 - 2a_j b_j) + \sum_j e^{-2\lambda_j t} a_j^2 \sum_k b_k^2.$$

Differentiation with respect to $a_j$ gives that

$$(6.2) \qquad\qquad\qquad\qquad a_j = b_j/B$$

where $B = \sum b_j^2$. Plugging this back into the equation gives minimization of the following quantity

$$(6.3) \qquad\qquad \min_{b_j} \sum_{j \geq 1} e^{-2\lambda_j t}\left(1 - \frac{b_j^2}{B}\right).$$

Since $\lambda_j$ is an increasing sequence, minimization is obtained with $b_1 = 1$ and $b_j = 0$ for all $j > 1$.

## REFERENCES

[1] Y. AÏT-SAHALIA AND P.A. MYKLAND, *Estimating Diffusions with Discretely and Possibly Randomly Spaced Data: A General Theory,* Annals of Statistics, (2004) 32: 2186-2222.

[2] C. H. BAKER, *The numerical treatment of integral equations*, Oxford, Clarendon Press, 1977.

[3] M. BELKIN AND P. NIYOGI, *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering*, In Advances in Neural Information Processing Systems, Vol. 14, 2002.

[4] M. BELKIN AND P. NIYOGI, *Convergence of Laplacian Eigenmaps*, In Advances in Neural Information Processing Systems, Vol 18, 2006.

[5] Y. BENGIO, O. DELALLEAU, N. LE ROUX, J.-F. PAIEMENT, P. VINCENT AND M. OUIMET, *Learning Eigenfunctions Links Spectral Embedding and Kernel PCA*, Neur. Comp., 16(10):2197-2219, 2004.

[6] R. Best and G. Hummer, *Reaction coordinates and rates from transition paths*, Proc. Nat. Acad. Sci., (2005) 102(19):6732-6737.

[7] A. Brandt, *Multiscale solvers and systematic upscaling in computational physics*, Proc. Conf. on Comp. Phys. 2004, September 2004, Genova, Italy.

[8] A. Brandt, S. McCormick, J. Ruge, *Multigrid methods for differential eigenproblems,* SIAM J. Sci. Comp. 4(2): 244-260, 1983.

[9] H.-J. Bungartz and M. Griebel, Sparse grids. *Acta Numerica,* (2004) 13:1-123.

[10] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni,B. Nadler, F. Warner and S. Zucker, *Geometric Diffusion as a tool for harmonic analysis and structure definition of data: diffusion maps*, Proc. Nat. Acad. Sci., (2005) Vol. 102(21):7426-31.

[11] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni,B. Nadler, F. Warner and S. Zucker, *Geometric Diffusion as a tool for harmonic analysis and structure definition of data: multiscale methods*, Proc. Nat. Acad. Sci., (2005) Vol. 102(21):7426-31.

[12] R.R. Coifman and S. Lafon, *Diffusion maps*, App. Comp. Harm. Anal., (2006) 21:5-30.

[13] R. Courant and D. Hilbert, *Methods of Mathematical Physics, II*, Wiley-Interscience, NY, 1964.

[14] Das, P., Moll, M., Stamati, H., Kavraki, L.E. and Clementi, C., *Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction*, Proc. Nat. Acad. Sci., (2006), Vol. 103(26):9885-9890.

[15] E.B. Davies, *Heat Kernels and Spectral Theory*, Cambridge University Press, 1989.

[16] C. Dellago, P. Bolhuis, and D. Chandler, *Efficient Transition Path Sampling: Applications to Lennard-Jones cluster rearrangements*, J. Chem. Phys., (1998) 108:9236.

[17] M. Dellnitz, M. Hessel-von Molo, Ph. Metzner, R. Preis and C. Schütte, *Graph algorithms for dynamical systems*, In Mielke, A. (ed.): Analysis, Modeling and Simulation of Multiscale Problems, Springer Verlag, Heidelberg, pp. 619-646, 2006.

[18] P. Deuflhard, W. Huisinga, A. Fischer, Ch. Schütte, *Identification of almost invariant aggregates in reversibel nearly uncoupled Markov chains*, Lin. Alg. and Appl. (2000), 315:39-59.

[19] P. Deuflhard and M. Weber, *Robust Perron cluster analysis in conformation dynamics*, Lin. Alg. and Appl. (2005) 398:161-184.

[20] W. E., W. Ren and E. Vanden-Eijnden, *String Method for the study of rare events*, Physical Review B, 66:052301 (2002).

[21] W. E. W. Ren and E. Vanden-Eijnden, *Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes*, Chem. Phys. Lett. (43):242-247, 2005.

[22] C. W. Gardiner, *Handbook of stochastic methods*, third edition, Springer NY, 2004.

[23] C. W. Gear and I. G. Kevrekidis, *Computing in the past with forward integration*, Physics Letters A, 321:335-343, 2004.

[24] D. Givon, R. Kupferman and A. Stuart, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, (2004) 17:R55-R127 .

[25] M. Hein, J. Audibert and U. von Luxburg, *From graphs to manifolds - weak and strong pointwise consistency of graph laplacians*, Conference on Learning Theory (2005).

[26] G. Henkelman and H. Jónsson, *Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points*,J. Chem. Phys., (2000) 113:9978-9985.

[27] W. Huisinga, S.P. Meyn and C. Schütte, *Phase transitions and metastability in Markovian and molecular systems*, Annals. of App. Prob. 14(1): 419-458 (2004).

[28] W. Huisinga, C. Schutte and A.M. Stuart, *Extracting Macroscopic Stochastic Dynamics: Model Problems*, Comm. Pure App. Math. 56:234-269 (2003).

[29] W. Huisinga, C. Best, R. Roitzsch, C. Schutte and F. Cordes, *From simulation data to conformational ensembles: structure and dynamics based methods*, J. Comp. Chem., 20:1760-1774, 1999.

[30] G. Hummer, *Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations*, New J. of Phys. 7:34 (2005).

[31] G. Hummer and I.G. Kevrekidis, *Coarse molecular dynamics of a peptide fragment: free energy, kinetics and long-time dynamics computations*, J. Chem. Phys. (2003) 118:10762.

22

[32] I.G. Kevrekidis, C.W. Gear and G. Hummer, *Equation-free: The computer-aided analysis of comptex multiscale systems*, AICHE JOURNAL (2004) 50(7):1346-1355.

[33] I.G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and K. Theodoropoulos, *Equation-Free Coarse-Grained Multiscale Computation: Enabling Microscopic Simulators to Perform System-Level Tasks*, Comm. Math. Sciences (2003) 1(4):715.

[34] D.I. Kopelevich, A.Z. Panagiotopoulos and I.G. Kevrekidis, *Coarse-grained kinetic computations for rare events: Application to micelle formation*, J. Chem. Phys.,122(4):044908 (2005).

[35] P. A. Markowich, C. Villani, *On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis*, Matematica Contemporanea (SBM), vol. 19, pp.1-31, 2000.

[36] B.J. Matkowsky and Z. Schuss, *Eigenvalues of the Fokker-Planck operator and the approach to equilibrium for diffusions in potential fields*, SIAM J. App. Math. (1981) 40(2):242-254.

[37] D. Moroni, P.G. Bolhuis and T.S. van Erp, *Rate constants for diffusive processes by partial path sampling*, J. Chem. Phys., 120(9):4055-4065 (2004).

[38] B. Nadler, S. Lafon, I.G. Kevrekidis and R.R. Coifman, *Diffusion Maps, Spectral Clustering and Eigenfunctions of Fokker-Planck Operators*, in Advances in Neural Information Processing Systems 18, Y. Weiss, B. Schölkopf and J. Platt (Editors), (2006) 955–962.

[39] B. Nadler, S. Lafon, R.R. Coifman and I. G. Kevrekidis, *Diffusion Maps, Spectral Clustering, and the Reaction Coordinates of Dynamical Systems*, Journal of Applied and Computational Harmonic Analysis (2006) 21:113-127.

[40] B. Nadler, M. Galun, *Fundamental Limitations of Spectral Clustering*, in Advanced in Neural Information Processing Systems 19, B. Schölkopf and J. Platt and T. Hoffman (Editors), (2007) 1017–1024.

[41] S. Park, M.K. Sener, D. Lu and K. Schulten, *Reaction paths based on mean first passage times*, J. Chem. Phys. 119(3): 1313-9 (2003).

[42] R. Erban, T. Frewen, X. Wang, T. Elston, R.R. Coifman, B. Nadler and I.G. Kevrekidis, *Variable-free exploration of stochastic models: a gene regulatory network example*, J. Chem. PHys., 126, 155103 (2007).

[43] H. Risken, The Fokker-Planck Equation: Methods of Solution and Applications (Springer Series in Synergetics, V. 18), 2nd edition, 1996.

[44] M.J. Ruiz-Montero, D. Frenkel and J.J. Brey, *Efficient schemes to compute diffusive barrier crossing rates*, Mol. Phys. 90:925-942 (1997).

[45] Z. Schuss, *Theory and application of stochastic differential equations*, Wiley, New-York, 1980.

[46] C. Schutte, A. Fischer, W. Huisinga and P. Deuflhard, *A direct approach to conformational dynamics based on hybrid monte carlo*, J. Comp. Phys. 151:146-168 (1999).

[47] J. Shi and J. Malik, *Normalized cuts and image segmentation*, PAMI, (2000), Vol. 22.

[48] A. Singer, *From graph to manifold Laplacian: the convergence rate*, Applied and Computational Harmonic Analysis, (2006) 21(1):135-144.

[49] A. Singer, R. Erban, I.G. Kevrekidis and R.R. Coifman, *Detecting the slow manifold by anisotropic diffusion maps*, submitted.

[50] S. Sriraman, I.G. Kevrekidis and G. Hummer, *Coarse master equation from bayesian analysis of replica molecular dynamics*,  J. Phys. Chem. B, 2005.

[51] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Comm. Math. Sci., (2003) 1(2), 385-391.

[52] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, 2001.

[53] A.F. Voter, F. Montalenti and T.C. Germann, *Extending the time scale in atomistic simulation of materials*, Annu. Rev. Mat. Res., 32:321-46, 2002.

[54] Y. Weiss, *Segmentation using eigenvectors: a unifying view*, ICCV (1999).

[55] R. Zwanzig, *Non-equilibrium statistical mechanics*, Oxford University Press, 2001.