Journal of
Nonlinear Science

Check for updates

Nonlinear Model Reduction for Slow–Fast Stochastic Systems Near Unknown Invariant Manifolds

Felix X.-F. Ye[1] · Sichen Yang[2] · Mauro Maggioni[2,3]

Abstract

We introduce a nonlinear stochastic model reduction technique for high-dimensional stochastic dynamical systems that have a low-dimensional invariant effective manifold with slow dynamics and high-dimensional, large fast modes. Given only access to a black-box simulator from which short bursts of simulation can be obtained, we design an algorithm that outputs an estimate of the invariant manifold, a process of the effective stochastic dynamics on it, which has averaged out the fast modes, and a simulator thereof. This simulator is efficient in that it exploits of the low dimension of the invariant manifold, and takes time-steps of size dependent on the regularity of the effective process, and therefore typically much larger than that of the original simulator, which had to resolve the fast modes. The algorithm and the estimation can be performed on the fly, leading to efficient exploration of the effective state space, without losing consistency with the underlying dynamics. This construction enables fast and efficient simulation of paths of the effective dynamics, together with estimation of crucial features and observables of such dynamics, including the stationary distribution, identification of metastable states, and residence times and transition rates between them.

Communicated by Alain Goriely.

✉ Felix X.-F. Ye
  xye2@albany.edu

1 Department of Mathematics and Statistics, University at Albany, Albany, USA

2 Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, USA

3 Department of Mathematics, Johns Hopkins University, Baltimore, USA

1 Introduction

Many mathematical models of dynamical systems, across the sciences, are based on ordinary and stochastic differential equations (ODEs and SDEs, respectively), with a large number of degrees of freedom, often with dynamics at very different timescales. These systems offer multiple significant challenges to their simulation and understanding, which often require collecting a large number of long trajectories to capture a wide variety of possible behaviors of the system. These challenges include: (1) the high dimension of the state space and the corresponding large number of equations; (2) many fast/stiff modes, corresponding to very rapid fluctuations (e.g., solvent molecules around a protein); and (3) metastability, with trajectories dwelling for large times in certain regions of state space (metastable states), with rare transitions between them.

These challenges often compound in a single system, making the large scale (in state space and in time) phenomena of the system, which are often of interest in applications, difficult to capture and study. Some of the properties that one often wishes to capture include the invariant manifold, around which trajectories lie; the stationary distribution, describing the large-time distribution of the system in state space; the residence times describing the distribution (or its expectation) of the time spent in a metastable state M before leaving it, once started in M; the transition rates and transition paths, containing information about the expected time and most likely paths followed by the system when transitioning between metastable states; the reaction coordinates, representing low-dimensional observables whose dynamics is approximately Markovian and predictive of transitions between metastable states; and the leading eigenvalues and eigenvectors of the generator of the process, related to transition rates and reaction coordinates, respectively (Coifman et al. 2005, 2008; Husic and Pande 2018; Klus et al. 2018; Rohrdanz et al. 2013; Bittracher et al. 2018; Kutz et al. 2016; Weinan and Eric 2004; Leimkuhler and Matthews 2015; Legoll and Lelièvre 2010, 2012; Givon et al. 2004; Alexander and Giannakis 2020).

These objects of interest intertwine geometry and dynamics and are our focus in this work: we aim at jointly estimating crucial geometric objects, such as a low-dimensional invariant manifold $\mathcal{M}_\epsilon$, and effective dynamics of the system at and above a given timescale, via estimated stochastic equations, that average out complex, high-dimensional aspects of the dynamics below that timescale. This reduced model can be more amenable to faster simulation, with low-dimensional equations and time-steps much larger than those needed by a simulator of the original system.

As with any type of model reduction, loss of information is in general unavoidable, and it has possibly dramatic consequences, among which loss of Markovianity, and to a loss of accuracy in predictions by the reduced model, especially at large times. Our approach aims at reducing these problems, at least on a suitable class of systems. We consider the problem of nonlinear model reduction for stochastic systems that, while presenting the above challenges, have features that are possibly redeeming, if appropriately exploited: fast and slow modes of evolution of the system, with a non-negligible *separation of timescales*; *a low-dimensional invariant manifold*, onto which the dynamics may be projected by averaging the fast modes, while preserving information about the large-scale/time phenomena; *fast modes that may be linearized*, but may be high-dimensional, and have large magnitude, with varying direction relative

to the invariant manifold. The objects we need to estimate are nonlinear: the invariant manifold, the corresponding reduction map onto it, and the effective stochastic equations on it.

The modality in which we have access to ground-truth trajectory data is important for algorithmic, statistical and computational considerations. We assume to be given access to the system via a black-box simulator $\mathcal{S}$ taking time-steps $\delta t \lesssim \epsilon$ (to resolve the fast modes, whose derivative has the order $1/\epsilon$), that we can call to obtain only short trajectories, of length-in-time of order $\tau \gg \epsilon$, where τ is typically of the order of the relaxation timescale of the fast modes. From a given initial condition, we use the simulator $\mathcal{S}$ to obtain a burst of N short paths, each of time length $O(\tau)$. This now classical setup (Frewen et al. 2009a) enables trivial parallelization, across initial conditions and paths from each initial conditions, and is well-suited to applications (Liu et al. 2015; Kim et al. 2015; Leimkuhler and Matthews 2015; Dietrich et al. 2021). A crucial problem is how initial conditions for these bursts are made available. When they cannot be chosen by the estimation procedure, they may be modeled as randomly drawn, ideally from a probability measure on state space that is reasonably well-distributed over the state space: it is then straightforward to estimate how many initial conditions needs to be sampled to guarantee, with high probability, coverage (see, e.g., Crosskey and Maggioni 2017). When the initial conditions may be chosen by the estimation procedure, a natural exploration–exploitation dilemma arises: refining the estimates in a region of space already populated by initial conditions by sampling more initial conditions and paths, or generate paths from initial conditions "outside" from the parts of state space already visited? And how to generate the latter? Especially when the state space is high-dimensional, and the dynamics of interest is along a low-dimensional invariant manifold, it is not trivial to sample new initial conditions. Even more so in applications, where many physical constraints are often extremely complex and unknown. We develop a simple approach, called "exploration mode", discussed below, that addresses both the exploration–exploitation dilemma and the problem of generating new initial conditions "outside" the already-visited state space; all needed is a small number (e.g., 1) of initial conditions; our numerical examples will be run in "exploration mode".

From the observations of N paths from an initial condition, we estimate locally the invariant manifold $\mathcal{M}_\epsilon$, the effective local directions of the fast modes and an oblique projection along them onto $\mathcal{M}_\epsilon$, and an effective drift and diffusion coefficient to be used for an effective Itô diffusion on $\mathcal{M}_\epsilon$. In these steps, it is crucial to avoid the curse of dimensionality, which would demand a number of observations exponential in the high dimension D of the state space. We achieve this by using simple parametric models for these local estimators, and prove that the sampling requirement scale favorably linearly in D. We then piece together these local estimators of the effective dynamics at timescales τ, to obtain a global estimate of $\mathcal{M}_\epsilon$, of a nonlinear projection onto $\mathcal{M}_\epsilon$, and of a process on $\mathcal{M}_\epsilon$ called ATLAS.

The ATLAS stochastic process $(\mathbf{z}_t^{\mathcal{A}})_{t\geq 0}$ takes place on the estimated $\mathcal{M}_\epsilon$, and aims at reproducing, at timescale τ and above, the dynamics of the original process on $\mathcal{M}_\epsilon$, after averaging out the fast modes. The ATLAS process may be simulated much more efficiently than the original process, as it is lower dimensional and amenable to be simulated with time-steps of size $\tau \gg \epsilon$, instead of size $\lesssim \epsilon$ as in $\mathcal{S}$. We demonstrate

this construction numerically on several systems that display several different salient features: nonlinear slow manifolds, lack of a global map to globally linear slow and fast variables, linear and curvilinear fast modes, with and without a clear separation of timescales between fast and slow modes.

As we mentioned above, in many applications (e.g., in molecular dynamics) a "large enough" set of initial conditions, at which to collect bursts of paths, is not known. A key contribution of this work is to introduce a construction for ATLAS in *"exploration mode"*, where we initially construct ATLAS from a very small number of initial conditions, and update it on the fly by collecting new bursts of simulations from $\mathcal{S}$, started at automatically well-chosen initial conditions, whenever ATLAS trajectories leave an ever-increasing "domain of competency" of the current ATLAS. This yields an increasing family of ATLASes, *each consistent with the previous ones and with the original dynamics, on ever-increasing subsets of the state space*, without over-sampling already explored regions. To efficiently and consistently explore the effective state space for the system is a crucial ability of ATLAS, achieved with techniques very different from existing ones, which typically are based on biasing the dynamics, and trading exploration with fidelity to the original dynamics (Chiavazzo et al. 2017; Frewen et al. 2009b; Tribello et al. 2014; Zheng et al. 2013; Chen et al. 2015).

All our numerical experiments are run exclusively in exploration mode and demonstrate that ATLAS accurately reproduces features of the dynamics at medium and large timescales and enables the efficient construction of Markov state models (MSMs), and of approximations of important observables, such as eigenvalues and eigenvectors of the generator of the dynamics. These in turn may be used for further reduction of the dynamics at very large timescales, estimating transition rates, and yielding low-dimensional embeddings of $\mathcal{M}_\epsilon$.

2 Fast–Slow SDEs with Slow Nonlinear Manifolds

A classical model of fast–slow SDEs is

$$\begin{cases} \mathrm{d}\mathbf{x}_t = g(\mathbf{x}_t, \mathbf{y}_t)\mathrm{d}t + G(\mathbf{x}_t, \mathbf{y}_t)\,\mathrm{d}U_t \\ \mathrm{d}\mathbf{y}_t = \dfrac{1}{\epsilon} f(\mathbf{x}_t, \mathbf{y}_t)\mathrm{d}t + \dfrac{1}{\sqrt{\epsilon}} F(\mathbf{x}_t, \mathbf{y}_t)\,\mathrm{d}V_t \end{cases}, \tag{2.1}$$

where $\epsilon > 0$ is a small parameter, determining the separation of timescales, $\mathbf{y}_t \in \mathbb{R}^{D-d}$ and $\mathbf{x}_t \in \mathbb{R}^d$ are, respectively, the fast and slow variables, and $(U_t)_{t\geq 0}$, $(V_t)_{t\geq 0}$ are independent Wiener processes in $\mathbb{R}^d$ and $\mathbb{R}^{D-d}$ respectively. The drift coefficients f, g and the diffusion coefficients F, G are assumed to be regular, e.g., twice-differentiable. Systems governed by this type of equations have been extensively studied (Pavliotis and Stuart 2008; Berglund and Gentz 2006; van Kampen 1981; Gardiner 2009). We are interested in the situation, common in applications, where the ambient dimension D is much larger than "intrinsic" dimension d of the slow variables. The time-step δt in the original simulator $\mathcal{S}$ of Eq. (2.1) is typically $\lesssim \epsilon$ to ensure accuracy and stability of the numerical scheme, making it computationally onerous. This constraint

on the time-step is generally applicable only to explicit schemes, motivating continued research in implicit schemes, which, however, in high dimensions still appear to be computationally prohibitive. While the techniques we introduce are also applicable to ODEs with minor changes, including for example substituting bursts of stochastic paths by bursts of deterministic paths started by stochastically perturbed initial conditions, we focus on SDEs to streamline the presentation. Also, recall that fast–slow ODEs may be approximated by SDEs, at least in the limit $\epsilon \to 0$ (Pavliotis and Stuart 2008).

2.1 The Slow Manifold and Averaged Equations on it

The fast variable $\mathbf{y}$ is assumed to relax, at a timescale $O(\tau)$, and stay close to the slow manifold $\mathcal{M}_0^{\mathbf{x}} := \{(\mathbf{x}, \mathbf{y}^{\star}(\mathbf{x})) : f(\mathbf{x}, \mathbf{y}^{\star}(\mathbf{x})) = 0\}$ of the corresponding deterministic system (with $F, G \equiv 0$), if the slow manifold is asymptotically stable. Geometric singular perturbation theory implies the existence of an invariant manifold $\mathcal{M}_\epsilon^{\mathbf{x}}$, close to $\mathcal{M}_0^{\mathbf{x}}$, see Berglund and Gentz (2006), Berglund and Gentz (2003), Kuehn (2015) and Appendix A.1. Under suitable further conditions, one then obtains a reduced set of equations on $\mathcal{M}_\epsilon^{\mathbf{x}}$, with the drift and diffusion coefficients on $\mathcal{M}_\epsilon^{\mathbf{x}}$ obtained by locally *averaging*, at each $\mathbf{x}_0 \in \mathcal{M}_\epsilon^{\mathbf{x}}$, those in Eq. (2.1) against the conditional invariant measure $\nu(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$ of the fast modes (Pavliotis and Stuart 2008; Berglund and Gentz 2006). The technical assumptions needed can be far from trivial, e.g., often G is assumed independent of $\mathbf{y}$ (Yu and Veretennikov 1991; Givon et al. 2006; Givon 2007). The *reduced equations* are

$$\mathrm{d}\bar{\mathbf{x}}_t = \bar{g}(\bar{\mathbf{x}}_t)\mathrm{d}t + \bar{G}(\bar{\mathbf{x}}_t)\mathrm{d}U_t, \qquad \bar{\mathbf{y}}_t = \bar{\mathbf{y}}(\bar{\mathbf{x}}_t, \epsilon), \tag{2.2}$$

which define a process on the invariant manifold $\mathcal{M}_\epsilon^{\mathbf{x}}$, for small ϵ (see Appendix A.3 for details). Having averaged out the fast variables, the reduced dynamics deliberately lose information about the details of the dynamics of the fast variables and phenomena below the timescale τ, but it yields a low-dimensional process (in the regime of interest $d \ll D$), that reproduces the effective dynamics of the original system on $\mathcal{M}_\epsilon^{\mathbf{x}}$ at timescale of order $O(\tau)$ and, ideally, beyond.

2.2 Nonlinear Observations and Unknown Slow/Fast Variables

In the model in Eq. (2.1), the slow and fast variables are the given, linear, and orthogonal coordinates $\mathbf{x}$ and $\mathbf{y}$. In applications, however, the slow and fast variables are typically not known a priori and need to be identified, and in general they are neither linear nor orthogonal (Wechselberger 2020).

This motivates the following observation model. We view the system in Eq. (2.1) as a black-box latent local model: *Black-box* because equations are not available to us. *Latent* because we do not have access to $\mathbf{x}$ and $\mathbf{y}$, but to observations $\mathbf{z}$, ranging in $\Omega \subseteq \mathbb{R}^D$, which can be mapped to latent variables $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^D$, satisfying Eq. (2.1). *Local* because such a map is not a global map, but is in fact realized by a collection of charts $\{(\mathcal{U}_\alpha, \varphi_\alpha)\}_\alpha$, consisting of open neighborhoods $\{\mathcal{U}_\alpha\}_\alpha$ covering Ω and smooth

maps $\varphi_\alpha : \mathcal{U}_\alpha \to \mathbb{R}^D$, each invertible on its range and such that $\varphi_\alpha \circ \varphi_{\alpha'}^{-1}$ is smooth where defined, so that for every point $\mathbf{z} \in \Omega$ the local latent variables $(\mathbf{x}, \mathbf{y}) = \varphi_\alpha(\mathbf{z})$ satisfy Eq. (2.1), for $\mathcal{U}_\alpha \ni \mathbf{z}$ (there exists one such $\mathcal{U}_\alpha$ since $\{\mathcal{U}_\alpha\}_\alpha$ covers Ω). Geometrically, this is of course the natural setup for expressing that the observations $\mathbf{z}$ are on a manifold parametrized by a system of charts (called an atlas, in differential geometry). This geometric perspective is here merged with the dynamics, through the condition that the local parametrizations map the dynamics of the observed variables to a dynamics where the latent variables follow the model equations (2.1). This model is inspired and generalizes that of Singer et al. (2009) and Wechselberger (2020), where the aim was to discover an embedding of the underlying slow variables, via the lowest frequency eigenfunctions of an estimated generator of the whole process, and not necessarily a parametrization of the invariant manifold, nor effective equations on it. That approach is broadly applicable to a larger class of processes than ours, for example when the fast modes are highly nonlinear; however this comes at the price of falling victim of the curse of dimension, requiring sampling paths from an exponentially large number of initial conditions (this is not discussed in Singer et al. 2009, but it could be derived). We shall estimate the slow variables and effective equations directly, without first learning the detailed behavior of the high-dimensional fast variables, and subsequently reduce the dynamics via eigen-decompositions.

In the observed variables $\mathbf{z}$, the process $(\mathbf{x}_t, \mathbf{y}_t)_{t\geq 0}$ maps to a process $(\mathbf{z}_t)_{t\geq 0}$, where slow and fast variables are in general nonlinearly mixed, instead of being linear and orthogonal as in Eq. (2.1). The slow variables $\mathbf{z}^{\text{slw}}$ will lie on a nonlinear invariant manifold $\mathcal{M}_\epsilon$ of dimension d; trajectories will lie in a domain of concentration around $\mathcal{M}_\epsilon$ that we model as a non-self-intersecting tube around $\mathcal{M}_\epsilon$. $\mathcal{M}_\epsilon$ is close to a slow manifold $\mathcal{M}_0$ (which in general is *not* the image of $\mathcal{M}_0^{\mathbf{x}}$ under the maps φ_α^{-1}). Locally around the initial point $\mathbf{z}_0 \in \mathcal{M}_\epsilon$, one may linearize the equations for $(\mathbf{z}_t)_{0 \leq t \lesssim \tau(\mathbf{z}_0)}$ to a form similar to Eq. (2.1), with $\mathbf{x}$ replaced by slow variables $\mathbf{z}^{\text{slw}}$ and $\mathbf{y}$ replaced by fast variables $\mathbf{z}^{\text{fst}}$. Under the same linearization, $\mathbf{z}^{\text{slw}}$ is approximated as lying in the tangent space $T_{\mathbf{z}_0}\mathcal{M}_\epsilon := \text{span}(\text{col}(U_d^{\text{slw}}))$ to $\mathcal{M}_\epsilon$ at $\mathbf{z}_0$, and $\mathbf{z}^{\text{fst}}$ is approximated as lying in $\mathbb{V}_{\mathbf{z}_0}^{\text{fst}} := \text{span}(\text{col}(V_{D-d}^{\text{fst}}))$. The slow and fast directions $U_d^{\text{slw}}, V_{D-d}^{\text{fst}}$ in general vary, smoothly, with $\mathbf{z}_0$. $(\mathcal{M}_\epsilon - \mathbf{z}_0)$ is locally a graph of a function $\bar{\mathbf{z}}_\epsilon^{\text{fst}}(\cdot; \mathbf{z}_0) : T_{\mathbf{z}_0}\mathcal{M}_\epsilon \to \mathbb{V}_{\mathbf{z}_0}^{\text{fst}}$ over the slow variables $\mathbf{z}^{\text{slw}}$. We can then proceed to the reduction to equations in the slow variables $\mathbf{z}^{\text{slw}}$ only, in a form similar to Eq. (2.2), by averaging the fast variables at a prescribed timescale τ, obtaining a reduced process on $\mathcal{M}_\epsilon$.

2.3 Structure of the Local Reduced Effective Equations

We assumed that the deviation of the fast variable from the invariant manifold $\mathcal{M}_\epsilon$ lies, exactly or approximately, in $\mathbb{V}_{\mathbf{z}_0}^{\text{fst}}$; in this subspace let it be given by $\boldsymbol{\xi}_t := \mathbf{z}_t^{\text{fst}} - \bar{\mathbf{z}}_\epsilon^{\text{fst}}(P(\mathbf{z}_t^{\text{slw}}); \mathbf{z}_0)$ with $t \lesssim \tau(\mathbf{z}_0)$, where P is the projection onto $T_{\mathbf{z}_0}\mathcal{M}_\epsilon$ with kernel $\mathbb{V}_{\mathbf{z}_0}^{\text{fst}}$. The idea of averaging in fast–slow systems (Pavliotis and Stuart 2008; Freidlin et al. 2012; Givon et al. 2004; Berglund and Gentz 2006; van Kampen 1981) exploits the timescale separation between slow and fast variables: at the separation timescale $\tau(\mathbf{z}_0)$, the dynamics of $\boldsymbol{\xi}_t$ conditioned at $\mathbf{z}_t^{\text{slw}} = \mathbf{z}_0$ reaches its quasi-equilibrium distribution $\nu(\boldsymbol{\xi}|\mathbf{z}_t^{\text{slw}} = \mathbf{z}_0)$, which we approximate by a $(D - d)$-

dimensional Gaussian distribution $\mathcal{N}(0, \Xi(\mathbf{z}_0))$. If the trace of $\Xi(\mathbf{z}_0)$ is large, the fast oscillations of $\boldsymbol{\xi}_t$ around $\mathcal{M}_\epsilon$ have large expected amplitude.

We discuss in Sect. 3 how $\tau(\mathbf{z}_0)$ may be estimated from observations of short trajectories. We assume throughout, but only in order to simplify the presentation, that $\tau(\mathbf{z}_0)$ can be chosen to be the same at all locations $\mathbf{z}_0$: we simply denote it as τ, and is assumed as given. By *stochastic averaging*, in these coordinates, the **reduced** stochastic dynamics on the slow variables is obtained by averaging the drift and diffusion terms by this quasi-equilibrium distribution, leading to reduced SDEs

$$d\bar{\mathbf{z}}_t^{\mathrm{slw}} = b(\bar{\mathbf{z}}_t^{\mathrm{slw}})dt + H(\bar{\mathbf{z}}_t^{\mathrm{slw}})dU_t, \tag{2.3}$$

similar to those in Eq. (2.2). These SDEs may be viewed in intrinsic coordinates, or in Cartesian coordinates in the ambient space $\mathbb{R}^D$, with $\mathbf{z}_t^{\mathrm{slw}} \in \mathbb{R}^D$ but on $\mathcal{M}_\epsilon$, $b \in \mathbb{R}^D$ a vector field on $\mathcal{M}_\epsilon$, and $H \in \mathbb{R}^{D \times d}$ acting on a Wiener process U_t in $\mathbb{R}^d$.

In classical stochastic averaging, it considers $\epsilon \to 0$ and the separation timescale τ is typically of order ϵ, see, e.g., Givon et al. (2006); Givon (2007), Liu (2010) and Zhang et al. (2018). Here, instead, motivated by applications, we consider ϵ fixed, unknown and unused, and τ fixed, known or estimated, and larger than ϵ, and independent thereof. The dynamics of $\mathbf{z}_t^{\mathrm{slw}}$ is low-dimensional, taking place on $\mathcal{M}_\epsilon$, and represents the reduced **effective** dynamics at timescales τ and beyond, having averaged out the fast transients of the high-dimensional process $(\boldsymbol{\xi}_t)_{t \geq 0}$ at timescales $\lesssim \tau$. Simulating $\mathbf{z}_t^{\mathrm{slw}}$ requires a *time-step independent of ϵ*, often much larger than ϵ, and only dependent on the regularity of $\mathcal{M}_\epsilon$ and of the regularity of the effective drift b and diffusion coefficient H on $\mathcal{M}_\epsilon$.

Our goal is to estimate a process, called ATLAS, that approximates $\bar{\mathbf{z}}_t^{\mathrm{slw}}$, on an estimated $\hat{\mathcal{M}}_\epsilon$, given observations of bursts of short trajectories; in all our examples the simulator of ATLAS will take time-steps exactly equal to τ.

2.4 ATLAS: Learning a Reduced Effective Model

Given observations of multiple bursts of short trajectories, of time length $O(\tau)$, around each of a collection of initial points $\{\mathbf{z}_0^l\}_{l=1,\ldots,L} \subset \mathbb{R}^D$, we estimate: the local slow variables by estimating a point $\mathbf{z}^l$ on $\mathcal{M}_\epsilon$ and a local tangent space $T_{\mathbf{z}^l}\mathcal{M}_\epsilon$ to $\mathcal{M}_\epsilon$ at $\mathbf{z}^l$; a subspace $\mathbb{V}_{\mathbf{z}^l}^{\mathrm{fst}}$ transversal to $\mathcal{M}_\epsilon$ at $\mathbf{z}^l$ containing the linearized fast modes $\boldsymbol{\xi}$; effective drift and diffusion coefficients for $\mathbf{z}_t^{\mathrm{slw}}$ in $T_{\mathbf{z}^l}\mathcal{M}_\epsilon$ as in Eq. (2.3) for the effective dynamics of the slow variables around $\mathbf{z}^l$ at the timescale τ. These objects are completely local, around each $\mathbf{z}^l$. Since a global reduction step bringing the equations to the standard form Eq. (2.1) may not be possible, for example because of global topological obstructions (e.g., a slow manifold consisting of a circle cannot be mapped globally to a linear coordinate), in the spirit of the very definition of manifolds and their atlases, we will "glue" together the estimated local charts and equations into a set of charts and smoothly coordinated equations, generating a process, called ATLAS, and a corresponding simulator for obtaining global paths on the estimated invariant manifold. ATLAS, given enough data and under suitable assumptions on the dynamics, estimates in a consistent and accurate fashion the local

---

Input: $\mathcal{S}$: original simulator of the system; μ_0 probability measure for initial conditions; τ: timescale for reduction; L: number of initial conditions.

Output: $\mathcal{A}$: simulator of the effective process $(\mathbf{z}_t^{\mathcal{A}})_{t \geq 0}$.

  - Estimate local parameters of effective dynamics   [Sect. 3.3]

  - Estimate geometric properties of $\mathcal{M}_\epsilon$ [Sect. 3.4]

  - Construct ATLAS using the above estimators   [Sect. 3.5]

Algorithm 1: High-level pseudo-code for ATLAS construction

dynamics and its statistics, we also demonstrate in our numerical experiments that important long-time observables, including the stationary distribution and mean residence times in regions, metastable and not, of state space, are accurately estimated by ATLAS.

3 ATLAS Construction

During construction, ATLAS is assumed to have access to a black-box simulator $\mathcal{S}$, that takes as input an initial condition $\mathbf{z}_0$ and a time t_0, typically $O(\tau)$, and returns a path $(\mathbf{z}_t)_{t \in [0, t_0]}$, driven by the latent equations as in Eq. (2.1). The construction proceeds in multiple steps, see Algorithm 1.

Before describing the details, we present an example.

Example: Fast/slow system around a pinched sphere. This system is used as a reference throughout our discussion, construction and testing of ATLAS. The process is an Itô diffusion on a "smoothly pinched" two-dimensional sphere centered at the origin (the invariant manifold $\hat{\mathcal{M}}_\epsilon \subseteq \mathbb{R}^3$, see Fig. 2), perturbed by very rapid fluctuations in the radial direction. These fast modes are (a) *large* (equal to a significant fraction of the reach of $\mathcal{M}_\epsilon$); (b) a.e. *not* orthogonal to $\mathcal{M}_\epsilon$, see Figs. 1 (step 3), and 7; and (c) may be approximated by a radial Ornstein–Uhlenbeck (O-U) process. For these reasons, a local PCA of an ensemble of short trajectories would fail to estimate the local tangent plane to $\mathcal{M}_\epsilon$. Given τ, at least as large as the timescale of relaxation of the fast modes, the correct effective equations on $\mathcal{M}_\epsilon$ should be obtained by averaging along an appropriate oblique (radial, in this case) projection onto $\mathcal{M}_\epsilon$.

We remark that neither the original system (in $\mathbb{R}^3$) nor the slow system on $\mathcal{M}_\epsilon$ is driven by overdamped Langevin equations: the drift is not the gradient of a potential, the diffusion coefficient is not constant, and the process is not reversible. Therefore, methods-based approximations by an overdamped Langevin equations, such as those in Coifman et al. (2005), Coifman et al. (2008), Singer et al. (2009) and Rohrdanz et al. (2011), would be biased, and likely inaccurate. The effective dynamics on $\mathcal{M}_\epsilon$ has two high probability regions, separated by regions of large volume where drift is small compared to diffusion ("entropic barriers"), which could make standard approximations of those inaccurate (Bicout and Szabo 2000).

To give some intuition about local geometric and dynamical quantities that play a fundamental role in this system, and more generally for systems that motivate our constructions, we show in Fig. 7 a portion of the $\hat{\mathcal{M}}_\epsilon$ about a point $\mathbf{z}_0$, a corresponding trajectory of the system started at $\mathbf{z}_0$, and several key directions in $\mathbb{R}^3$: the normal to $\hat{\mathcal{M}}_\epsilon$ at $\mathbf{z}_0$, the estimated effective direction of the fast modes at timescale τ, which is
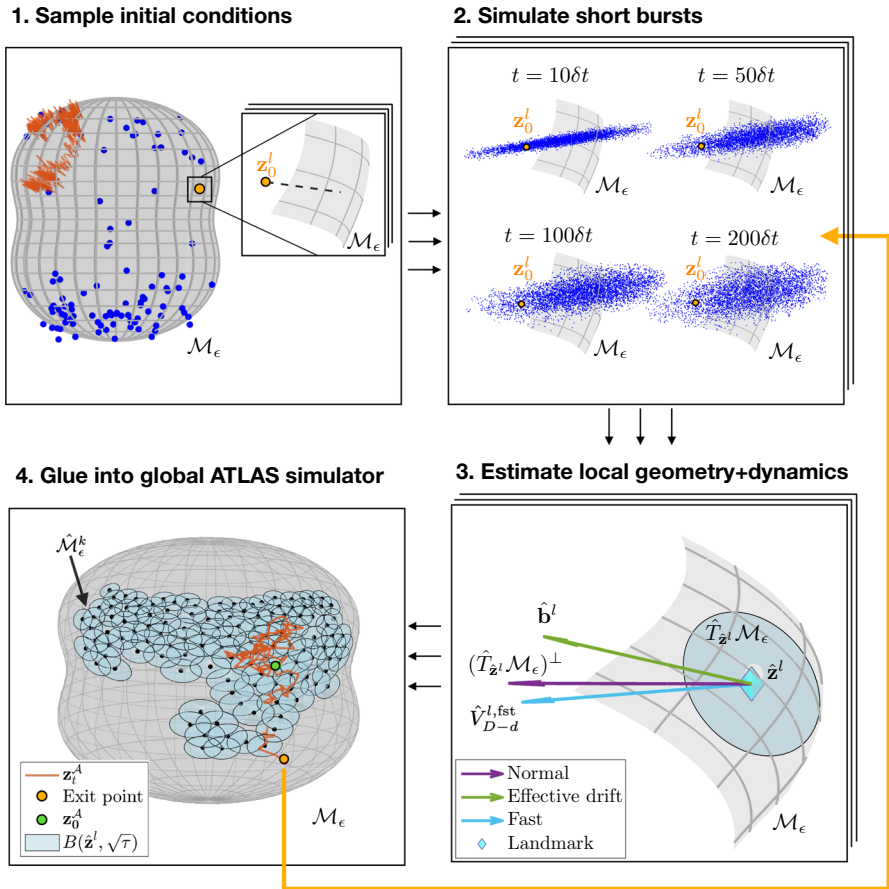
Fig. 1 High-level overview of the steps in the construction of ATLAS, for a system exhibiting fast large oscillations around an invariant manifold $\mathcal{M}_\epsilon$ shaped as a pinched sphere. **1. Sample initial conditions**, according to a given probability measure. Typically initial conditions (ICs) do not lie on $\mathcal{M}_\epsilon$ (see zoom-in portion: IC represented by orange dot is not on $\mathcal{M}_\epsilon$), nor are they well-distributed throughout state space. A sample path of the original system is shown, which oscillates around $\mathcal{M}_\epsilon$ with large amplitude. **2. Simulate short bursts**: from each IC, in parallel, a burst of short trajectories, of time length comparable to the relaxation timescale τ of the fast modes, is obtained from a black-box simulator of the original system (if τ is not given, it may be estimated from these trajectories; in this example $\tau = 200\delta t$). In this example the fast modes have large amplitude, but our technique will correctly determine fast and slow directions. **3. Estimate local geometry and dynamics**: from the trajectories in each burst, the local geometry of $\mathcal{M}_\epsilon$ (including a landmark $\hat{\mathbf{z}}^l$ on $\hat{\mathcal{M}}_\epsilon$, an affine tangent space $\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon$, and an oblique projection $\hat{P}^l(\mathbf{z})$ onto it with kernel span(col($\hat{V}_{D-d}^{l,\text{fst}}$), as well as parameters of the reduced effective dynamics (drift $\hat{\mathbf{b}}^l$ and diffusion coefficient $\hat{\Lambda}_d^l$), are estimated. **4. Glue into global ATLAS simulator**: the local geometry and dynamics estimators are glued together into a global ATLAS, with an associated simulator of a reduced effective ATLAS process $(\mathbf{z}_t^A)_{t\geq 0}$. We display estimated local tangent ellipses corresponding to balls in a diffusion-induced local metric, a sample path from the ATLAS simulator. However, at this stage ATLAS may still not cover $\mathcal{M}_\epsilon$: in "exploration mode" it will simulate paths that stop when reaching the boundary of the current "domain of expertise" (orange point), collect new bursts at those locations (as in step 2), update ATLAS with the estimators from the new region (as in step 3), and then resume the simulation(Color figure online)
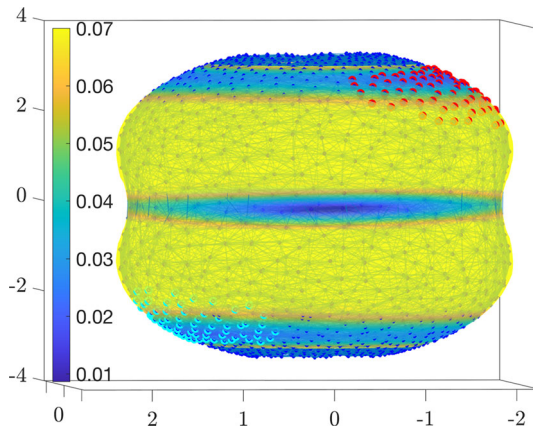
Fig. 2 Representation of the estimated pinched sphere $\hat{\mathcal{M}}_\epsilon$, together with the landmarks $\{\hat{z}^l\}_l$ (blue dots) and their local connectivity graph, all as constructed by ATLAS during long exploration. The dynamics in this case is not reversible, and the fast modes have large standard deviation (comparable to the reach of $\hat{\mathcal{M}}_\epsilon$) and are a.e. not orthogonal to $\hat{\mathcal{M}}_\epsilon$ (see Fig. 7). The surface color is the norm of the effective drift as a function on $\hat{\mathcal{M}}_\epsilon$. The landmarks are about 10^{-2}-close to $\hat{\mathcal{M}}_\epsilon$, see Table 4. The regions around the poles are very rarely visited, as the drift is unbounded there, creating a repulsion. In this visualization we have truncated those regions as landmarks become denser and denser. The landmarks marked in red and cyan represent the regions have high probability (Color figure online)

significantly different from the normal direction. We also depict the direction of the estimated effective (Itô) drift, which, as expected, is not (and, in fact, far from being) tangent to $\hat{\mathcal{M}}_\epsilon$. These depicted objects are exactly those estimated in the ATLAS construction, from local bursts of simulations.

Finally, the global geometric approximation of $\mathcal{M}_\epsilon$ and effective ATLAS process are assembled. An ATLAS path, used for exploration, is shown in Fig. 1; the accuracy of ATLAS is demonstrated in various metrics, from the geometric approximation of $\mathcal{M}_\epsilon$ to the approximation of effective drift and diffusion coefficients, to the accuracy of estimation of statistics of the process such as mean residence time in relatively small regions of state space and in metastable states (see Sect. 7.1).

3.1 Main Steps in the Construction

We are given access to a black-box simulator $\mathcal{S}$ of the process $\mathbf{z}_t$, a probability measure μ_0 on the state space of the system, a separation timescale τ, and a dimension d for the invariant manifold $\mathcal{M}_\epsilon$. We will discuss later how to proceed in the very important case when μ_0 is not, or insufficiently, provided ("exploration mode", Sect. 4), and how to estimate τ (in Sect. 3.3) and d (in Appendix C). We output ATLAS, consisting of a process $(\mathbf{z}_t^{\mathcal{A}})_{t \geq 0}$, and a corresponding simulator, approximating the effective dynamics of $\mathbf{z}_t$ on $\mathcal{M}_\epsilon$ at the timescale τ and beyond.

We sample L initial conditions $\{\mathbf{z}_0^l\}_{l=1,\dots,L} \sim_{\text{i.i.d.}} \mu_0$, and for each l we use $\mathcal{S}$ to obtain a burst $\mathcal{B}^l$ of N trajectories $\{\mathbf{z}_t^{l,n}\}_{n=1,\dots,N}$, each of time length $O(\tau)$, starting at $\mathbf{z}_0^l$. The time-step δt of $\mathcal{S}$ is typically $\lesssim \epsilon \ll \tau$ and we may think of the output of

$\mathcal{S}$ as if it was in continuous time. For each l, we focus now on the local construction around $\mathbf{z}_0^l$, given the single burst $\mathcal{B}^l$: at timescale τ, the invariant manifold is locally approximated by an estimated effective tangent space $\hat{T}_{\hat{\mathbf{z}}^l} \mathcal{M}_\epsilon$ of dimension d, at a suitably estimated point $\hat{\mathbf{z}}^l$; the deviation $\boldsymbol{\xi}_t^0$ from $\mathcal{M}_\epsilon$ reaches equilibrium before time τ, and we approximate the dynamics of the slow variable $(\mathbf{z}_t^{\mathrm{slw}})_{t \geq 0}$ on $\mathcal{M}_\epsilon$ around $\hat{\mathbf{z}}^l$ by an Itô diffusion process on $\hat{T}_{\hat{\mathbf{z}}^l} \mathcal{M}_\epsilon$ as in (2.3), which requires us to estimate an affine *oblique* projection $\hat{P}^l$ along the fast modes and onto $\hat{T}_{\hat{\mathbf{z}}^l} \mathcal{M}_\epsilon$, an effective drift $\hat{\mathbf{b}}^l$ in $\mathbb{R}^D$ (in the Itô formulation, the drift is in general not tangent to $\mathcal{M}_\epsilon$, see Figs. 1 and 7) and an effective diffusion coefficient $\hat{\Lambda}_d^l$ in $\hat{T}_{\hat{\mathbf{z}}^l} \mathcal{M}_\epsilon$.

3.2 Local Low-Order Moments of the Dynamics

The behavior of the time-dependent mean and covariance of the process started at $\mathbf{z}_0^l$ reveals crucial local properties of the geometry of the dynamics and of the slow/fast manifolds: at times t comparable to τ (but, typically, not smaller nor larger), we assume that these approximations hold:

$$
\begin{aligned}
\mathbf{m}_t^l &:= \mathbb{E}\left[\mathbf{z}_t^l \,\middle|\, \mathbf{z}_0^l\right] = \mathbf{z}_0^{l,\mathrm{slw}} + \mathbf{b}^l t + O(\epsilon), \\
C(\mathbf{z}_t^l | \mathbf{z}_0^l) &:= \mathrm{cov}(\mathbf{z}_t^l | \mathbf{z}_0^l) = \Gamma^l + \Lambda^l t + O(\epsilon),
\end{aligned}
\tag{3.1}
$$

where $\Gamma^l \succeq 0$ has rank $D - d$, and represents an averaging, at timescale τ, of the covariance of the fast modes $\Xi(\mathbf{z}_0^{l,\mathrm{slw}})$; $\Lambda^l = H^l(H^l)^T \succeq 0$ has rank d is the diffusivity of the effective reduced slow dynamics at $\mathbf{z}_0^l$. The span of Λ^l and Γ^l approximate, respectively, the tangent space $T_{\mathbf{z}_0^{l,\mathrm{slw}}} \mathcal{M}_\epsilon$ to $\mathcal{M}_\epsilon$ at $\mathbf{z}_0^{l,\mathrm{slw}}$ and, respectively, $\mathbb{V}_{\mathbf{z}_0^{l,\mathrm{slw}}}^{\mathrm{fst}}$, which will be the kernel of a projection (in general *not* orthogonal) onto $T_{\mathbf{z}_0^{l,\mathrm{slw}}} \mathcal{M}_\epsilon$. These expressions result from averaging the fast modes at timescale τ; in particular, the memory of the effective reduced slow dynamics is (approximately) forgotten.

The quantities above are unknown, and we estimate them from the observations from the burst $\mathcal{B}^l$:

$$
\hat{\mathbf{m}}_t^l := \frac{1}{N} \sum_{n=1}^N \mathbf{z}_t^{l,n}, \quad \hat{C}_t^l := \frac{1}{N-1} \sum_{n=1}^N (\mathbf{z}_t^{l,n} - \hat{\mathbf{m}}_t^l)(\mathbf{z}_t^{l,n} - \hat{\mathbf{m}}_t^l)^T.
\tag{3.2}
$$

These empirical quantities, estimated from burst data, are consistent estimators of the true local mean and covariance of the process, with an approximation of order $\sqrt{\frac{d+d_f}{N}}$, where d is the dimension of $\mathcal{M}_\epsilon$, and d_f is the number of fast modes of large amplitude - we discuss this further in Sect. 5.

We are now ready to introduce the ATLAS construction, which proceeds in three main steps in Algorithm 1, detailed in Sects. 3.3, 3.4, 3.5 respectively; see Appendix B for details.

Fig. 3 Estimating τ from a burst of trajectories: $||\hat{\mathbf{m}}_t^l||$ and $\mathrm{tr}(\hat{C}_t^l)$ should both behave as linear functions of t at the timescale of interest, as per Eq. (3.1)
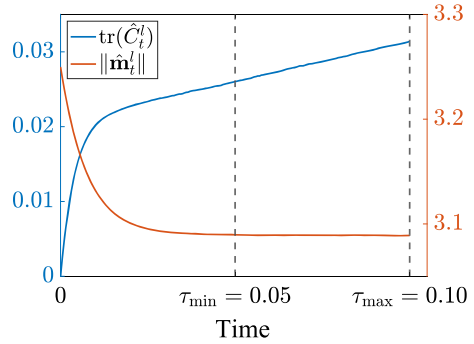
3.3 Estimation of Local Parameters of the Effective Dynamics

From each burst $\mathcal{B}_l$, $l = 1, \ldots, L$, of short simulations we compute several key quantities for constructing an approximation to the invariant manifold $\mathcal{M}_\epsilon$ and to the reduced effective stochastic dynamics on it. The relationships in Eq. (3.1) suggest that the local effective drift and diffusion coefficients for the local slow variables may be estimated from $\hat{\mathbf{m}}_t^l$ and $\hat{C}_t^l$ for $t \approx \tau$. The diffusion coefficient should also yield an estimate for local tangent plane to the $\mathcal{M}_\epsilon$, giving the local slow variables.

Separation Timescale τ

When not given, we estimate τ from the behavior of $||\hat{\mathbf{m}}_t^l||$ and $\mathrm{tr}(\hat{C}_t^l)$ as a function of t: for each burst, we obtain the time interval where these two quantities behave linearly, as per Eq. (3.1) (see fig. 3). We let $[\tau_{\min}, \tau_{\max}]$ be the intersection of such intervals over all l's, which is nonempty since we assume there exists a common relaxation time of the fast modes τ valid throughout invariant manifold (our techniques do extend to location-dependent τ^l).

Drift Coefficient of the Effective Dynamics

The estimated drift $\hat{\mathbf{b}}^l$ is obtained as the slope (in t) of $\mathbf{m}_t^l$ in Eq. (3.1), via a weighted linear regression: with $\{t_m\}_{m=1}^M$ equispaced in $[\tau_{\min}, \tau_{\max}]$,

$$\hat{\mathbf{b}}^l := \frac{\sum_{m=1}^M (\hat{\mathbf{m}}_{t_m}^l - \bar{\mathbf{m}}_M^l)(t_m - \bar{t}_M)}{\sum_{m=1}^M (t_m - \bar{t}_M)^2}, \tag{3.3}$$

where $\bar{\mathbf{m}}_M^l := \frac{1}{M} \sum_{m=1}^M \hat{\mathbf{m}}_{t_m}^l$ and $\bar{t}_M = \frac{1}{M} \sum_{m=1}^M t_m$. Figures 2 and 7 show norm and direction of $\hat{\mathbf{b}}^l$ for the pinched sphere system.

Diffusion Coefficient of the Effective Dynamics and Local Slow Variables

Similarly, the local diffusivity Λ^l of the slow effective dynamics is estimated as the slope (in t) of $C(\mathbf{z}_t^l | \mathbf{z}_0^l)$ in Eq. (3.1):

$$
\hat{\Lambda}^l := \frac{\sum_{m=1}^{M} (\hat{C}_{t_m}^l - \bar{C}_M^l)(t_m - \bar{t}_M)}{\sum_{m=1}^{M} (t_m - \bar{t}_M)^2} \,, \tag{3.4}
$$

where $\bar{C}_M^l = \frac{1}{M} \sum_{i=1}^{M} \hat{C}_{t_m}^l$. While $\hat{\Lambda}^l$ is typically not low-rank, for N large enough, with high probability (w.h.p.), its top d singular values may be well-separated from the others, yielding an estimate of the intrinsic dimension of invariant manifold (a dynamics-driven analogue of Multiscale SVD Little et al. 2017); this is case in our examples (see Figs. 7, 8, 9). We project $\hat{\Lambda}^l$ onto the space of rank d matrices by truncated SVD:

$$
\hat{\Lambda}_d^l := \mathrm{Proj}_{\mathrm{rk}(d)} \hat{\Lambda}^l = \hat{U}_d^{l,\mathrm{slw}} \hat{\Sigma}_d^l (\hat{U}_d^{l,\mathrm{slw}})^T, \tag{3.5}
$$

where $\mathrm{Proj}_{\mathrm{rk}(d)}$ denotes the projection onto rank d matrices (positive semidefinite in this case), $\hat{U}_d^{l,\mathrm{slw}} \in \mathbb{R}^{D \times d}$ orthogonal and $\Sigma_d^l \in \mathbb{R}^{d \times d}$ diagonal with the first d singular values of $\hat{\Lambda}_d^l$.[1] Let $\hat{H}_d^l := (\hat{\Lambda}_d^l)^{\frac{1}{2}}$ be the (positive) square root of $\hat{\Lambda}_d^l$.

Covariance of the Fast Dynamics

While $\hat{\Lambda}_d^l$ suffices to estimate a local tangent plane to $\mathcal{M}_\epsilon$, the affine projection P^l of the fast dynamics onto that plane, consistent with the dynamics, requires more information, as it is typically *not* an orthogonal projection. To estimate the kernel of P^l, i.e., the set of directions "along which" the dynamics near $\hat{\mathbf{z}}^l$ should be projected, we first estimate the covariance matrix $\hat{\Gamma}^l$ in Eq. (3.1) as

$$
\hat{\Gamma}^l := \bar{C}_M^l - \hat{\Lambda}^l \bar{t}_M \,, \tag{3.6}
$$

and then let the estimated fast directions to be the span of the $D - d$ eigenvectors of $\hat{\Gamma}^l$ with largest eigenvalues, which we group as columns of an orthogonal matrix $\hat{V}_{D-d}^{l,\mathrm{fst}}$. See Figs. 1 and 7 for the case of the pinched sphere system. Since $\sigma_{D-d+1}(\Gamma) = 0$ (see Eq. 3.1), $\sigma_{D-d+1}(\hat{\Gamma}^l) \ll \sigma_{D-d}(\hat{\Gamma}^l)$ w.h.p., for N large enough. In practice, not all $D - d$ dimension may be fast modes, and we may truncate at the first $d_f \leq D - d$ significant eigenvectors (e.g., in the oscillating half-moon system below).

---

[1] We note here that we tried other approaches toward estimating $\hat{\Lambda}_d^l$, for example by attempting to solve a least squares problem in the space of positive definite matrices of rank d directly, ideally with respect to a natural Riemannian metric on that space. While natural, this was significantly more computationally expensive, and it did not lead to significantly different results.

3.4 Construction of a Sketch of the invariant manifold $\mathcal{M}_\epsilon$

We now utilize the quantities estimated above to construct a sketch $\hat{\mathcal{M}}_\epsilon$ of the invariant manifold, consisting of a set of portions of well-distributed affine approximate tangent planes.

Landmarks

The initial conditions $\{\mathbf{z}_0^l\}_{l=1}^L$ of the bursts $\{\mathcal{B}_l\}_{l=1}^L$ are not assumed to be on the unknown $\mathcal{M}_\epsilon$, nor well-distributed on it. We construct a set of points, called landmarks, on our estimate of $\mathcal{M}_\epsilon$. From Eq. (3.1), replacing the quantities involved by their empirical counterparts estimated above, for each $l = 1, \ldots, L$ we define the landmark $\hat{\mathbf{z}}^l$ as

$$\hat{\mathbf{z}}^l := \bar{\mathbf{m}}_M^l - \hat{\mathbf{b}}^l \bar{t}_M \, . \tag{3.7}$$

Local Tangent Plane to $\mathcal{M}_\epsilon$

We obtain an approximate tangent space $\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon$ at $\hat{\mathbf{z}}^l$:

$$\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon := \mathrm{span}(\mathrm{cols}(\hat{U}_d^{l,\mathrm{slw}})). \tag{3.8}$$

Dynamics-Driven Oblique Projections onto a Local Tangent Plane

We obtain the oblique affine projection $\hat{P}^l$ onto $\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon$ centered at $\hat{\mathbf{z}}^l$:

$$\hat{P}^l(\mathbf{z}) := \hat{U}_d^{l,\mathrm{slw}}(\hat{U}_d^{l,\mathrm{slw}})^T(\hat{E}^l(\hat{E}^l)^T)^\dagger(\mathbf{z} - \hat{\mathbf{z}}^l) + \hat{\mathbf{z}}^l \, , \tag{3.9}$$

with kernel $\mathrm{span}(\mathrm{cols}(\hat{V}_{D-d}^{l,\mathrm{fst}}))$, where $\hat{E}^l := [\hat{U}_d^{l,\mathrm{slw}}, \hat{V}_{D-d}^{l,\mathrm{fst}}]$ and $\dagger$ is the Moore–Penrose inverse.

A Dynamics-Adapted Metric

We introduce a quasi-metric adapted to the dynamics; then we discard "redundant" landmarks that are too close to others in order to create a well-distributed set of landmarks, which, together with the approximate tangent planes estimated above, gives a parsimonious sketch of the estimated invariant manifold $\hat{\mathcal{M}}_\epsilon$.

Consider the local Mahalanobis similarity based on the quadratic form associated with the effective diffusivity $\hat{\Lambda}_d^l$

$$\hat{\bar{\rho}}^2(\mathbf{z}, \hat{\mathbf{z}}^l) := \frac{1}{\chi_d^2(p)}(\hat{P}^l(\mathbf{z}) - \hat{\mathbf{z}}^l)^T(\hat{\Lambda}_d^l)^\dagger(\hat{P}^l(\mathbf{z}) - \hat{\mathbf{z}}^l), \tag{3.10}$$

for $\mathbf{z}$ such that $\|\mathbf{z} - \hat{\mathbf{z}}^l\| \lesssim R\sqrt{\tau}$; otherwise we set $\hat{\tilde{\rho}}(\mathbf{z}, \hat{\mathbf{z}}^l) = R\sqrt{\tau}$. In practice, R is set as 10. $\chi_d^2(p)$ is the quantile function at level p of the χ^2 distribution with d degrees of freedom (we set $p = 0.95$ throughout).

Unlike Euclidean distance, $\hat{\tilde{\rho}}$ accounts for the anisotropy of the dynamics on $\hat{\mathcal{M}}_\epsilon$; a similar distance, without the oblique projection, was used, for example, with different objectives, in Coifman et al. (2005) for manifold learning, and in Singer et al. (2009) in the context of dynamical systems. We symmetrize $\hat{\tilde{\rho}}$ on $\hat{\mathcal{M}}_\epsilon$ by letting $\hat{\rho}(\hat{\mathbf{z}}^{l'}, \hat{\mathbf{z}}^l) :=$ $\max\{\hat{\tilde{\rho}}(\hat{\mathbf{z}}^{l'}, \hat{\mathbf{z}}^l), \hat{\tilde{\rho}}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^{l'})\}$. A $\sqrt{t}$-neighborhood of $\hat{\mathbf{z}}^l$ is defined as $B(\hat{\mathbf{z}}^l, \sqrt{t}) := \{\mathbf{z} \in T_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon : \hat{\tilde{\rho}}(\mathbf{z}, \hat{\mathbf{z}}^l) < \sqrt{t}\}$: it approximates the set of points reachable from $\hat{\mathbf{z}}^l$ in time $\approx t$ with probability at least p. This distance disregards the drift term; this choice reduces the asymmetry in the definition of $\hat{\tilde{\rho}}$ and in the quasi-metric property and is reasonable for diffusion-dominated dynamics see Appendix A.1). Figure 1 (4th inset) visualizes the ellipsoids corresponding to the quadratic form induced by $\hat{\Lambda}_d^l$, for the pinched sphere system.

A Well-Distributed Net of Landmarks

We now reduce the number of landmarks to a near-minimal number that still, together with their corresponding neighborhoods of radius $O(\sqrt{\tau})$, cover the estimated invariant manifold $\hat{\mathcal{M}}_\epsilon$, by discarding landmarks that are too close to each other. Before this process, we assume here that the collection of $\sqrt{\tau/2}$-neighborhoods, of the landmarks $\{\hat{\mathbf{z}}^l\}_{l=1}^L$ covers the invariant manifold $\hat{\mathcal{M}}_\epsilon$. While this typically cannot be ascertained in practice, by default ATLAS will be run in "exploration mode", which augments $\hat{\mathcal{M}}_\epsilon$ on the fly (see Sect. 4): in that case our arguments here apply to the current $\hat{\mathcal{M}}_\epsilon$ during exploration. For $l = 1, \ldots,$ we remove $\hat{\mathbf{z}}^{l'}$ if $\hat{\rho}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^{l'}) \leq (1 - 1/\sqrt{2})\hat{\kappa}\sqrt{\tau}$ and $l' > l$, where $\hat{\kappa}$ is the scaling constant. In practice, $\hat{\kappa}$ is set as 1. When this procedure terminates, we are left with $L' \leq L$ landmarks (unused landmarks are discarded). We assume, without loss of generality, that $\{\hat{\mathbf{z}}^l\}_{l=1}^{L'}$ is the reduced collection of landmarks, and to simplify the notation, we will let $L' = L$ in what follows. These landmarks, under suitable assumptions, (1) are well-separated: for $l \neq l'$, $\hat{\rho}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^{l'}) \gtrsim \sqrt{\tau}$; (2) provide a $O(\hat{\kappa}\sqrt{\tau})$-cover for $\mathcal{M}_\epsilon$, in the sense that for any $\mathbf{z} \in \hat{\mathcal{M}}_\epsilon$ there exists l s.t. $\hat{\tilde{\rho}}(\mathbf{z}, \hat{\mathbf{z}}^l) \lesssim \sqrt{\tau}$; (3) are well-distributed, in the sense that for any $\hat{\mathbf{z}}^l \in \mathcal{T}$, where $\mathcal{T}$ is any connected component of the invariant manifold $\mathcal{M}_\epsilon$ satisfying suitable constraint, there exists l' such that $\hat{\rho}(\hat{\mathbf{z}}^{l'}, \hat{\mathbf{z}}^l) \lesssim \sqrt{\tau}$. These constants $R, \hat{\kappa}$ used above could be made explicit, but they depend on quantities typically unknown in practice, such as the curvature of $\mathcal{M}_\epsilon$.

These steps define a collection of charts, each centered at one of the landmarks $\hat{\mathbf{z}}^l$, with an associated oblique projection $\hat{P}^l$ with range $\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon$, an effective drift $\hat{\mathbf{b}}^l \in \mathbb{R}^D$ and effective diffusivity $\hat{\Lambda}_d^l \in \hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon$. The reader may wish to revisit Figs. 1, 2, 4 and 7, that visualize these objects.

3.5 The ATLAS Process and Simulator for the Reduced Effective (Slow) Dynamics

The final step is to smoothly connect both the geometric and dynamic objects estimated so far at the landmarks, in order to obtain a smooth effective invariant manifold $\hat{\mathcal{M}}_\epsilon$ and an Itô process constrained on it, with the fast modes averaged at the timescale τ, together with a numerical scheme for its simulation. This smoothing is achieved by a weighted average: for $\mathbf{z} \in \mathbb{R}^D$, we let $w^l(\mathbf{z}) := \exp(-\hat{\bar{\rho}}(\mathbf{z}, \hat{\mathbf{z}}^l)/\sqrt{\tau})$ and $\mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z}) := \{\hat{\mathbf{z}}^l : \hat{\bar{\rho}}(\mathbf{z}, \hat{\mathbf{z}}^l) \leq 2\hat{\kappa}C\sqrt{\tau}\}$, $Z(\mathbf{z}) := \sum_{l \in \mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})} w^l(\mathbf{z})$, and then define:

$$\hat{P}^{\mathcal{A}}(\mathbf{z}) := \frac{1}{Z(\mathbf{z})} \sum_{l \in \mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})} \hat{P}^l(\mathbf{z}) w^l(\mathbf{z}), \tag{3.11}$$

$$\hat{\mathbf{b}}^{\mathcal{A}}(\mathbf{z}) := \frac{1}{Z(\mathbf{z})} \sum_{l \in \mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})} \hat{\mathbf{b}}^l w^l(\mathbf{z}), \tag{3.12}$$

$$\hat{\Lambda}^{\mathcal{A}}(\mathbf{z}) := \frac{1}{Z(\mathbf{z})} \sum_{l \in \mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})} \hat{\Lambda}_d^l w^l(\mathbf{z}), \tag{3.13}$$

$$\hat{H}_d^{\mathcal{A}}(\mathbf{z}) := (\text{Proj}_{\text{rk}(d)} \hat{\Lambda}^{\mathcal{A}}(\mathbf{z}))^{\frac{1}{2}}. \tag{3.14}$$

This defines the ATLAS stochastic process $\mathbf{z}_t^{\mathcal{A}}$, on the estimated invariant manifold $\hat{\mathcal{M}}_\epsilon := \hat{P}^{\mathcal{A}}(\mathbb{R}^D)$, as the Itô diffusion with drift $\hat{\mathbf{b}}^{\mathcal{A}}$ and diffusion coefficient $\hat{H}_d^{\mathcal{A}}$. To simulate this process we use the Euler–Maruyama scheme with re-projection on $\hat{\mathcal{M}}_\epsilon$, with time-step $\lambda\tau$ and $\Delta W_{\lambda\tau} \sim \mathcal{N}(0, \lambda\tau I_d)$:

$$\mathbf{z}_{t+\lambda\tau}^{\mathcal{A}} = \hat{P}^{\mathcal{A}}(\mathbf{z}_t^{\mathcal{A}} + \hat{\mathbf{b}}^{\mathcal{A}}(\mathbf{z}_t^{\mathcal{A}})\lambda\tau + \hat{H}_d^{\mathcal{A}}(\mathbf{z}_t^{\mathcal{A}})\Delta W_{\lambda\tau}), \tag{3.15}$$

In all our experiments, $\lambda = 1$, i.e., the time-step of the ATLAS simulator is equal to the timescale τ; in particular, it is independent of, and may be much larger than, the time-step δt of the original black-box simulator, which is typically $\lesssim \epsilon$.

3.6 Refinements to the Estimation Procedure

When the initial conditions for the bursts are far away from the invariant manifold $\mathcal{M}_\epsilon$ (e.g., in the oscillating half-moon example below), or the timescale of separation in the original system is not very large (e.g., in butane example below), it may take a time longer than τ to relax onto $\mathcal{M}_\epsilon$. In these cases we perform multiple rounds of the estimation phases, starting each round with initial conditions for the bursts given by the landmarks estimated in the previous round. This refinement process stops when the relative differences of estimated parameters is within 5% (in our examples, this is achieved in less than 10 rounds).

One further step of refinement, after the above ones, may be needed when the linear approximation of the geometry or of the dynamics may not be locally accurate, because $\mathcal{M}_\epsilon$ has large curvature or the effective drift term has large gradient (e.g., in the oscillating half-moons and butane examples below). In this case, since the estimated drift term is computed at the timescale $[\tau_{\min}, \tau_{\max}]$, the landmarks in the final stage are refined to be $\hat{\mathbf{z}}^l := \bar{\mathbf{m}}_M^l - \hat{\mathbf{b}}^l(\bar{t}_M - \tau_{\min})$, instead of Eq. (3.7).

4 ATLAS in Exploration Mode

In the construction of ATLAS presented above, initial conditions for the bursts of simulation were sampled from a provided measure μ_0 on the state space. Ideally such measure is well-distributed on $\mathcal{M}_\epsilon$, e.g., close to the stationary distribution of the effective reduced process, or, at least, such that the set of local means $\bar{\mathbf{m}}^l_M$ of bursts started i.i.d. from μ_0 are well-distributed on $\mathcal{M}_\epsilon$. However, this is too much to hope for in many practical situations, and it is highly desirable to be able handle more general μ_0's.

There are at least two, not-mutually-exclusive, ways of proceeding. The first one is to use any of many existing techniques aimed at efficiently sampling the effective state space, to obtain a μ_0. The literature on this subject is vast, including, among many (Chiavazzo et al. 2017; Frewen et al. 2009b; Tribello et al. 2014; Zheng et al. 2013; Chen et al. 2015). These techniques often design a bias of the dynamics to ensure rapid exploration, yielding samples with coverage of the effective state space. In some remarkable cases the samples from this biased process allow to recover statistics of the original dynamics, e.g., the stationary distribution in the case of MCMC. However, even when the stationary distribution is recovered, the biased dynamics often does not preserve other dynamic properties such as mean residence times or transition rates, important in applications. Yet other techniques in this broad family require a target statistics to be computed, and then are designed to achieve accuracy in the estimate of that statistics, but in general no other ones. There is typically a strong tension between the attempt to speed up exploration, and the ability to correct the biased sampling to obtain consistent estimates of statistics of interest. We note that, crucially, this tension does *not* arise in ATLAS: μ_0 needs to have coverage, but has no relationship with the dynamics; it is only used as a starting point for ATLAS, which will then estimates consistently the effective dynamics, and from it the stationary distribution and many other dynamical properties, such as mean residence times and transition rates, as demonstrated in the numerical experiments in Sect. 7.

The second way is to extend ATLAS to run in *exploration mode*: upon a first round of learning starting from a μ_0 with poor coverage, ATLAS runs trajectories till they exit the current, partial estimate of invariant manifold, and updates itself, accurately but efficiently, by simulating new paths of the original process at those new exit locations. In detail, suppose we have constructed ATLAS $\mathcal{A}_1$, starting from a set of bursts $\{\mathcal{B}^l\}_{l=1}^L$, therefore obtaining the process $(\mathbf{z}_t^{\mathcal{A}_1})_{t \geq 0}$ on $\hat{\mathcal{M}}_\epsilon^1$, perhaps from a small number L (even $L = 1$) of initial conditions, which are poorly distributed in the state space (e.g., supported in one or a few metastable states). While simulating $(\mathbf{z}_t^{\mathcal{A}_1})_{t \geq 0}$, ATLAS checks if the distance between $\mathbf{z}_t^{\mathcal{A}_1}$ and its closest landmark, in the $\hat{\hat{\rho}}$ "metric", is larger than some threshold $d_{\text{thr}} = O(\sqrt{\tau})$: if this is the case, then $\mathbf{z}_t^{\mathcal{A}_1}$ is beyond the current "domain of expertise" of the ATLAS $\mathcal{A}_1$. We now stop that process, and run a new burst $\mathcal{B}_*$ of simulations from this "exit point" $\mathbf{z}_t^{\mathcal{A}_1}$, estimate the local quantities of the dynamics there, and add a new landmark $\hat{\mathbf{z}}^{l+1}$, with its associated tangent plane, projection, and estimated effective drift and diffusion coefficients. This local and efficient update yields a new ATLAS process $(\mathbf{z}_t^{\mathcal{A}_2})_{t \geq 0}$ on an "enlarged" $\hat{\mathcal{M}}_\epsilon^2$. This procedure is repeated, creating ATLAS processes that capture ever-increasing approximations $\hat{\mathcal{M}}_\epsilon^1 \subset \cdots \subset \hat{\mathcal{M}}_\epsilon^k \subset \ldots$ of $\mathcal{M}_\epsilon$, discovering rarer and rarer events,

till a given computational budget (for example expressed in terms of total number of bursts used, which is the most expensive component in the construction) is exhausted, or long-enough trajectories are simulated with $(\mathbf{z}_t^{\mathcal{A}_k})_{t \geq 0}$ without leaving $\hat{\mathcal{M}}_\epsilon^k$. See details in Appendix B. *All our numerical experiments will be performed with ATLAS in exploration mode.*

We provide a pertinent visualization in step 4 of Fig. 1: at stage k of exploration, we represent $\hat{\mathcal{M}}_\epsilon^k$, and a trajectory of $(\mathbf{z}_t^{\mathcal{A}_k})_{t \geq 0}$ which at some point leaves $\hat{\mathcal{M}}_\epsilon^k$ (orange dot in the figure). The current ATLAS $\mathcal{A}_k$ stops there, and then will obtain a new burst of paths starting at that location, extract local estimates of the geometric and dynamics quantities needed, obtain $\hat{\mathcal{M}}_\epsilon^{k+1}$, and then continue. In the figure, the path is linearly interpolated between the points $(\mathbf{z}^{\mathcal{A}}(p\lambda\tau))_p$; note that the ATLAS time-step $\lambda\tau$ leads (here, and in the other examples, $\lambda = 1$); as expected, to steps of length comparable to that of the axes of the diffusion ellipsoids representing the level set τ of the quadratic form $\hat{\Lambda}_d^l$.

Note that these additions happen during ATLAS simulations (with their large $\lambda\tau$ time-steps), taking advantage of the computational speed gains achieved by the ATLAS simulator. Therefore, in real clock time, the regions of $\mathcal{M}_\epsilon$ that are rarely visited with the original simulator are more likely to be sampled quickly in ATLAS exploration mode. Note how this differs from existing techniques based on ideas of importance sampling: at no point is the dynamics of ATLAS biased, and at any stage the dynamics is consistent with the underlying effective dynamics, by construction. This procedure can be very effective, measured in real clock time, in discovering relatively rare events, such as transitions between metastable states, while updating ATLAS, seamlessly. We also note that this procedure is parallelizable across multiple paths of the current ATLAS, provided one checks that the regions being added are far enough from each other to avoid the simulation of the bursts of simulations at nearby locations: these "conflicts" are likely rare in high-dimensional state spaces, at least till ATLAS has explored the vast majority of it.

Finally, it is certainly possible to apply a multitude of techniques and heuristics (e.g., see Chiavazzo et al. 2017 and references therein) to bias the ATLAS simulator itself during exploration, i.e., combine the two approaches described in this section; once new regions are explored with a biased ATLAS, and charts created (with new local bursts initialized in those regions) and incorporated into ATLAS, then the ATLAS simulator can be run in unbiased mode and will be consistent with the effective dynamics. This decoupling of exploration and consistent estimation of the dynamics is a crucial property of ATLAS, and it is very efficient as the information from the expensive simulations of bursts of trajectories is fully reused. This strategy together with parallel learning across multiple paths are particularly helpful when the effective dynamics have a significant amount of meta-stable states.

5 Properties of ATLAS

Avoiding the Curse of Dimension

The input to ATLAS is random, so are the L initial conditions and the N paths in each burst. It is natural to ask how many short trajectories in each burst are needed to make sure the random error of all the local quantities estimated by ATLAS is small w.h.p. In particular, it is important to assess how N should scale with the dimensions D of the state space, d of $\mathcal{M}_\epsilon$, d_f of the fast modes with large magnitude. Using concentration inequalities for high-dimensional vectors and matrices that are concentrated near low-dimensional spaces (Vershynin 2018), it is possible to show that the approximation error between the empirical estimates of $\hat{\mathbf{b}}^l$, $\hat{\Lambda}^l_d$ and $\hat{\mathbf{z}}^l$ is smaller than η, with high probability, as soon as $N \gtrsim d(d + d_f)/\eta^2$, where d_f is the number of fast modes with large magnitude.

The approximation error of the direction of the fast modes appears to require larger sample size, e.g., $N \gtrsim (d + d_f)^2 \ln D/\eta^2$ samples appear needed to obtain, w.h.p., $\|\sin \Theta(\hat{V}^{l,\text{fst}}_{D-d}, V^{l,\text{fst}}_{D-d})\|_\text{F} \le \eta$. It is worthwhile to note, though, that this still depends only very weakly on D, and only quadratically on the intrinsic dimension d, and it is still quite benign as soon as $d_f \ll D$.

To summarize, there is no curse of dimensionality—i.e., a requirement of a number of samples exponential in D—when estimating the local quantities above.

Robustness to Model Error and Nonlinearities in the Fast Modes

The type of stochastic systems for which ATLAS is expected to perform well have been described in Sect. 2. In particular, locally we assumed the existence of an invariant manifold in the observed space, on which the slow dynamics $(\mathbf{z}^{\text{slw}}_t)_{t \ge 0}$ takes place, and such that the fast dynamics $(\mathbf{z}^{\text{fst}}_t)_{0 \le t \le \tau}$ conditioned on $\mathbf{z}^{\text{slw}}_t = \mathbf{z}_0$ approximately is an O-U process on the subspace $\mathbb{V}^{\text{fst}}_{\mathbf{z}_0}$. In the latent space, the SDEs are those in Eq. (2.1), and one then linearizes locally the observation map φ (or, rather, φ_α, as in Sect. 2) to obtain approximate local equations in the $\mathbf{z}$ variables. ATLAS averages the observed process $(\mathbf{z}_t)_{t \ge 0}$ at timescale τ to obtain the reduced effective dynamics on $\hat{\mathcal{M}}_\epsilon$.

ATLAS is quite robust to these assumptions. One reason is that ATLAS uses mainly information at timescale τ: details, such as nonlinearities, or lack of regularity, of the original process $\mathbf{z}_t$ below that timescale are averaged out, possibly leading to effective processes at the timescale that are amenable to approximation by ATLAS. In a forthcoming work, we prove results in this direction, under technical conditions on the coefficients f, g, F, G of the SDEs Eq. (2.1), on the regularity of the map φ, and on the stationary measure $\nu(\boldsymbol{\xi}|\mathbf{z}^{\text{slw}}_t = \mathbf{z}_0)$ of the (fast) displacement process $\boldsymbol{\xi}$ conditioned on $\mathbf{z}_0$.

This robustness is reflected in the results for both the second and third examples we consider in Sect. 7, which are both significant perturbations of the basic model. In the "oscillating half-moons" example, a high-dimensional analogue of an example considered in Singer et al. (2009) and Dsilva et al. (2016), the fast displacement process $\boldsymbol{\xi}$ is nonlinear and not constrained to an affine subspace, but on a curved

"half-moon"-shaped manifold. In the butane model, the fast mode has both large and small nonlinear components; its slow manifold is also highly curved, with effective drift having large gradient. Nevertheless, ATLAS provides accurate estimates of the behavior of these systems, both at timescale τ and at very large timescales, providing accurate estimates of the stationary distribution, mean residence time and transition rates between metastable states.

Computational Complexity and Simulation Speed-Up

The input data to ATLAS are the bursts of trajectories $\{\mathcal{B}_l\}_{l=1}^{L}$, of time length $\approx \tau$. The cost of obtaining one time-step from the black-box simulator is at least of order D^2. The time-step δt of the simulator needs to be $\lesssim \epsilon$ due to having to resolve the fast modes. The total number of short paths collected is equal to #landmarks×#paths per landmark= $L \times N$. Therefore the total computational cost of obtaining the bursts is at least $O(\frac{\tau}{\epsilon} D^2 L N / c)$, where c is the number of parallel cores. Constructing ATLAS requires $O(D^2 d N)$ calculations to estimate local means, covariances, effective drift, effective diffusion coefficients, landmarks and tangent planes; $O(C^d D L \log L)$ for constructing and organizing the landmarks using, for example, cover trees (Beygelzimer et al. 2006). A time-step of the ATLAS simulator as in Eq. (3.15), which has time length $\lambda\tau$, has cost $O(C^d D d^2)$ by using iterative SVD combining Eq. (3.13,3.14), where C^d corresponds to the number of landmarks in $\mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})$. Therefore, simulating a path of time length T would have computational cost $O(D^2 T / \epsilon)$ with the original simulator, and $O(C^d D d^2 T / \tau)$ with ATLAS. This is a dramatic speed-up when $\epsilon \ll \tau$ and $d \ll D$. This is very useful when many long paths are needed to estimate dynamical quantities of interest.

6 Applications of ATLAS

Estimation of Large-Time Dynamical Properties

ATLAS may be used to simulate long paths efficiently, and therefore estimate important properties of the system, such as its stationary distribution, residence times from certain regions of state space (e.g., metastable states), and transition rates between them. Our numerical experiments in Sect. 7 show that such large-time quantities may be estimated accurately by ATLAS, even when run in exploration mode. Note that ATLAS is constructed using only local information, at timescale τ, that may be easily collected in parallel; yet the effects of the multiple estimation and numerical simulation errors do not appear to compound in these estimates of large-time quantities (Crosskey and Maggioni 2017).

ATLAS, Approximate Generators, Eigenfunctions and Eigenvalues

The ATLAS process $(\mathbf{z}_t^{\mathcal{A}})_{t \geq 0}$ may be used to approximate the generator of the effective slow dynamics and its spectral components, including eigenvalues and eigenvectors,

especially the low-frequency ones. It may serve as a black-box for matrix–vector multiplication in iterative eigensolvers. In general, ATLAS may be used to compute approximations of $\mathbb{E}[h(\mathbf{z}_t^{\text{slw}})]$, for sufficiently regular observables h.

Markov State Models (MSMs) from ATLAS

In MSMs (Husic and Pande 2018) one constructs (1) a partition of state space $\{C_k\}_{k=1}^K$ and (2) a Markov transition matrix $P \in \mathbb{R}^{K \times K}$ with $P_{kk'}$ being the probability of transitioning from C_k to $C_{k'}$ in one MSM time-step. MSMs may be "large-timescale MSMs", where each C_k corresponds to a metastable state and the MSM models the rare transitions between them, and "small-timescale MSMs", where K is large and the C_k's are small regions of state space.

Large-timescale MSMs may be constructed if the metastable states are known and a large number of transitions between them are observed. Since these transitions are rare, by definition of metastability, this construction is very expensive in general; however, ATLAS can help identifying metastable states and estimating transition rates efficiently.

Small-timescale MSMs are very flexible tools, and as $K \to \infty$ the transition matrix P approximates in a suitable sense the generator of the process, and convergence is (under suitable assumptions) strong enough to guarantee convergence of the slow eigenfunctions of P to those of the generator of the process. These eigenfunctions, and the corresponding eigenvalues, yield important information about the process, including metastable states. However, the construction of the local clusters C_k is crucial, and many recipes exist (Pérez-Hernández et al. 2013; Kutz et al. 2016). This is a challenging task and typically cursed by the ambient dimension D. Many existing techniques require, in order to be of any practical value, the a priori knowledge of a suitable small number of slow variables on which the process is projected, and in which the construction of the C_k's is performed (Husic and Pande 2018; Klus et al. 2018). In particular, we are not aware of techniques for efficiently constructing the C_k's in the situation where there are many fast modes, possibly with large amplitude. In this context, ATLAS naturally constructs the small-timescale MSMs (at timescale τ), in a principled and well-organized fashion, with soft instead of hard partitions, which may diminish the memory effect. ATLAS also uses dynamics-adapted oblique projections and the corresponding estimated local invariant manifold to reduce the dimension, without needing slow variables as inputs. In our experiments, C_k's in the MSM correspond to the Voronoi cells, in the $\hat{\tilde{\rho}}$ "metric", of the landmarks[2], and the transition matrix is estimated by running ATLAS trajectories of length $O(\tau)$ (see appendix B). We may use the small-timescale MSMs to compute approximate slow eigenfunctions and eigenvalues of the system and estimate the number and locations of metastable states and then construct the large-timescale MSMs.

---

[2] Of course, we do not explicitly construct such cells; we only need a function mapping a state $\mathbf{z}$ to the index of cell it belongs to, and this is achieved by finding the nearest landmark in the $\hat{\tilde{\rho}}$ "metric".

7 Numerical Experiments

We construct ATLAS for three model systems: "pinched sphere", "oscillating half-moons" and "butane model". We evaluate its performance in multiple ways: first of all, against analytically derived reduced models with analytical approximations to the slow manifold $\mathcal{M}_0$, effective drift and diffusion coefficients (see Appendix C). For the first two examples, the effective dynamics are calculated in the limit $\epsilon \to 0$; for butane, the effective dynamics are chosen to be the dihedral angle dynamics. It is important to remark that these are not the true effective dynamics on the invariant manifold $\mathcal{M}_\epsilon$ at timescale τ, which is what ATLAS approximates, and are also not amenable to analytical calculation for finite ϵ. Although with this caveat, we regard them as analytical approximations sufficient as a first check on the quality of the ATLAS process for the local statistics, and report in Table 4 the estimator errors for drift, diffusion, invariant manifold and tangent spaces, between ATLAS and these analytically derived reduced models (details in Appendix D).

We also study the accuracy of ATLAS in estimating key medium- and large-time statistics of the dynamics, in particular the stationary distribution, mean residence times (MRTs) and transition rates for metastable states, and MRTs in regions of state space that are not necessarily metastable. In each example, we repeat the construction of ATLAS 10 times, to assess the variability over the random observed data.

We visualize the invariant manifold $\hat{\mathcal{M}}_\epsilon$ for each example, as well as key quantities including the stationary distribution and eigenfunctions of MSMs; in these plots we use suitable parametrizations (that of course were not used nor known to ATLAS). Further details and figures for the models are available in Appendix C.

7.1 Pinched sphere System

We start with the pinched sphere system, introduced in Sect. 3. Its governing equations, expressed in spherical (latent) coordinates, are

$$
\begin{aligned}
\mathrm{d}r &= -\frac{c_1(r - R(\theta))}{\epsilon r}\mathrm{d}t + \frac{c_2}{\sqrt{\epsilon}r}\mathrm{d}W_1 \\
\mathrm{d}\theta &= \frac{c_3\cos(3\theta)}{r\sin(\theta)}\mathrm{d}t + \frac{c_4\sin(\theta)}{r}\mathrm{d}W_2 \\
\mathrm{d}\phi &= \frac{c_5\sin(\phi + \theta)}{r}\mathrm{d}t + \frac{c_6}{r}\mathrm{d}W_3.
\end{aligned}
\tag{7.1}
$$

The fast variable is the radial coordinate r; the slow variables are the angles ϕ, θ. The slow manifold (in the limit $\epsilon \to 0$) is $R(\theta) = \sqrt{a_1 + a_2\cos^2(\theta)}$, visualized in Fig. 2. The observations **z** are in Cartesian coordinates, each of which contains a mix of nonlinearly coupled slow and fast components. Note that the drift diverges near the poles, creating a strong repulsion, and is relatively small in other wide regions of the state space, creating entropic barriers (Bicout and Szabo 2000).

The dominant local PCA mode only captures the fast direction, due to its large amplitude, and fails to identify the slow variables, which are also not orthogonal
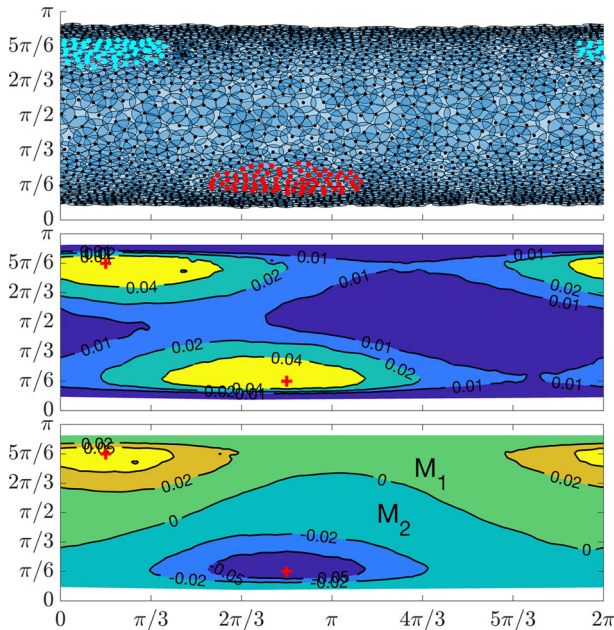
Fig. 4 Pinched sphere visualized in the plane of the parametrization (ϕ, θ) of $\mathcal{M}_\epsilon$. **Top**: landmarks and their neighborhoods $B(\hat{z}^l, \sqrt{\tau})$. We also show two sets S_{red} and S_{cyan} around the fixed points in the reduced deterministic system, marked with red crosses in the other insets. **Center**: eigenfunction φ_1 of the MSM, with eigenvalue $\lambda_1 = 1$. **Bottom**: eigenfunction φ_2, with eigenvalue $\lambda_2 = 0.9995$

to the fast ones. ATLAS successfully estimates that the invariant manifold is two-dimensional, and identifies the separation timescale τ (see Appendix C.1 and Fig. 7). ATLAS yields an accurate estimation of the effective drift and diffusion terms, as well as of $\mathcal{M}_\epsilon$ (see Table 4). We visualize in Fig. 4 the $\sqrt{\tau}$-neighborhoods over $\mathcal{M}_\epsilon$ (unwrapped in the (ϕ, θ) coordinates for clarity), reflecting the ellipsoids associated with the diffusion coefficient. At the time we terminate exploration, as expected the only regions that are not covered are those around the south and north poles, which are very rarely visited.

The TICA method (Molgedey and Schuster 1994; Pérez-Hernández et al. 2013) is global and indicates that all observed coordinates are important; in particular the common approach of constructing MSMs in the TICA coordinates would be cursed by the ambient dimension. Here we construct MSMs using ATLAS. The top two eigenfunctions of the transition matrix of an MSM constructed from ATLAS on $\hat{\mathcal{M}}_\epsilon$ are visualized in Fig. 4. The first eigenfunction φ_1 is (up to rescaling) the invariant distribution; the level set $\varphi_2 = 0$ partitions the state space into two metastable states M_1 and M_2. We also let $C_1 := \{\varphi_2 > +0.02\}$ and $C_2 := \{\varphi_2 < -0.02\}$; initial conditions for paths used in the computation of mean residence times (MRTs) will be from $S_{\text{cyan}} := \{\varphi_2 > 0.05\}$ and $S_{\text{red}} := \{\varphi_2 < -0.05\}$ (see Fig. 2 and 4), where φ_1 is large.

In Table 1 we report the accuracy of ATLAS in estimating the MRTs in M_1, C_1 (resp. M_2, C_2) starting from set S_{cyan} (resp., S_{red}). ATLAS yields $\leq 2\%$ relative error for these quantities, with runtime at least 6 times smaller than original simulator

Table 1 Mean residence times (MRTs) for Pinched sphere

| IC | First exit from C_1 or C_2 | | | First exit from M_1 or M_2 | | |
	Sim. Original $\mathcal{S}$	ATLAS	ATLAS orth. proj.	Original $\mathcal{S}$	ATLAS	ATLAS orth. proj.
S_{Cyan}	30.6 ±0.4	29.7 ±0.4	28.9 ±0.4	905.6 ±14.0	903.1 ±14.1	712.2 ±10.9
S_{Red}	30.6 ±0.5	30.2 ±0.5	29.4 ±0.4	1009.3 ±15.6	1016.5 ±15.8	791.2 ±12.2
Runtime	0.39 h	0.02 h	0.02 h	4.15 h	0.66 h	0.52 h
S_{Cyan}	N/A	0.002 ±0.010	−0.026 ±0.008	N/A	0.010 ±0.008	−0.212 ±0.008
S_{Red}	N/A	0.003 ±0.013	−0.019 ±0.010	N/A	0.011 ±0.009	−0.211 ±0.007

Top: MRTs from initial conditions in IC set till exit from specified sets; bottom: corresponding relative errors of MRTs

$\mathcal{S}$; training time is about 21hrs. Of course, the transition rates between metastable states, which are determined by the mean residence times for double-well systems, are also very accurate. Using orthogonal projections, instead of the ATLAS oblique projections, leads to a significant loss of accuracy in long-time observables (e.g., exit times from M_1, M_2). The estimated L^1-norm of the difference of the density of the invariant distribution between original and ATLAS simulators is 0.107 ± 0.009.

7.2 Oscillating Half-Moons

This is a multiscale stochastic system in $\mathbb{R}^2 \times \mathbb{R}^{18}$ that generalizes the one in Singer et al. (2009) to high dimensions. Its governing equations in latent coordinates are:

$$
\begin{aligned}
\mathrm{d}\theta &= (a_1 + a_2 \sin(2\theta) + a_3 \cos(\theta)) \, \mathrm{d}t + a_4 \mathrm{d}W_1, \\
\mathrm{d}r_1 &= \frac{b_1}{\epsilon}(1 - r_1) \, \mathrm{d}t + \frac{b_2}{\sqrt{\epsilon}} \mathrm{d}W_2, \\
\mathrm{d}u_i &= \frac{b_3}{\epsilon}(-u_i)\mathrm{d}t + \frac{b_4}{\sqrt{\epsilon}}\mathrm{d}W_i, \quad i = 3, \ldots, 20.
\end{aligned}
\tag{7.2}
$$

The observables in Cartesian $\mathbb{R}^{20}$ by

$$
z_1 = r_1 \cos(\theta + r_1 - 1), \, z_2 = r_1 \sin(\theta + r_1 - 1), \, z_i = r_1 + u_i, \tag{7.3}
$$

for $i = 3, \ldots, 20$. The dynamics of the angle θ is that of an uneven double-well system with metastable states M_{Left} and M_{Right} around $\theta = \pm\pi/2$. The radial variable r and other u_i's evolve as O–U processes. The fast variables $r, u_2, \ldots, u_{19}$ are nonlinearly coupled in the observed Cartesian coordinates.

A typical trajectory exhibits fast oscillations with a half-moon shape, far from a radial direction, while evolving slowly along the circular slow manifold driven by the double-well potential and diffusion along it (see in Appendix C.2).

Local PCA again fails to detect the slow manifold (see Fig. 8). Notwithstanding the lack of linearity of the fast modes, ATLAS accurately identifies the invariant manifold and the effective dynamics on it, see Table 4. While the relative error of estimated drift and covariance matrix seems large (32% and 11%, resp.), if the error is measured only in the first two important coordinates—since the error in the other 18 dimensions does not contribute to effective observations—then these relative errors drop to 19% and 6% (resp.). The invariant distribution estimated by ATLAS is very close to the one by original simulator $\mathcal{S}$ (see Fig. 5), with the estimated L^1-norm of the difference of their densities is 0.098 ± 0.006. The main reason for the small translational bias in the estimate of the stationary distribution is that the fast modes do not fully relax at the timescale τ, and increasing τ is not an option in this case due to the high curvature of the $\mathcal{M}_\epsilon$. As reported in Table 2, the estimated MRTs in the metastable states are quite accurate, and so are the transition rates. The training time for ATLAS is about 17hrs; runtime for estimating the large-time quantities above is less than half that of original simulator $\mathcal{S}$.
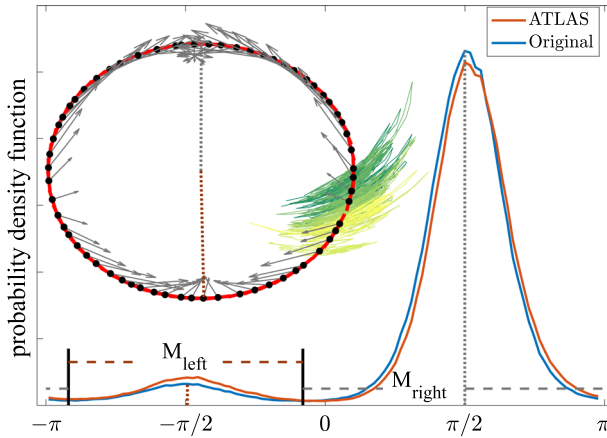
Fig. 5 Oscillating half-moons. The short illustrative trajectory of time 1×10^2 is plotted in (z_1, z_2) and colored according to the time t. The landmarks (black dots) and their neighborhood (red lines) and effective drift direction (gray arrows) in (z_1, z_2) are plotted in the left. In the right, the smoothed histograms from the trajectories of time length 8×10^6 generated by $\mathcal{S}$ and ATLAS simulator, projected with ATLAS's projection, are plotted in the coordinate of angle θ

Table 2 Mean residence time (MRT) for oscillating half-moons

	Sim.		
IC	Original $\mathcal{S}$	ATLAS	ATLAS orth. proj.
M_{Left}	918 ± 11	902 ± 11	897 ± 11
M_{Right}	12552 ± 142	11221 ± 128	10855 ± 123
Runtime	$2.49\,\text{h}$	$1.05\,\text{h}$	$1.05\,\text{h}$
M_{Left}	N/A	-0.02 ± 0.01	-0.02 ± 0.01
M_{Right}	N/A	-0.13 ± 0.01	-0.13 ± 0.01

Top: MRTs from initial conditions in IC set till exit from specified sets; bottom: corresponding relative errors of MRTs

7.3 Butane Model

This is a model for the butane molecule, inspired by molecular dynamics (Legoll and Lelièvre 2012; Schappals et al. 2017), in the form of overdamped Langevin equations in $\mathbb{R}^6$ (see Appendix C.3). The dihedral angle ϕ, which determines the distance of two outer carbons groups, is usually considered to be the slow variable. TICA however flags two coordinates, x_4 and z_4, as important coordinates; in the plane that they span three metastable states M_{trans}, $M_{\text{bot-cis}}$ and $M_{\text{top-cis}}$, concentrated around a circular $\mathcal{M}_\epsilon$, are apparent (see Fig. 6). ATLAS identifies that the slow variable is one-dimensional, accurately estimates the tangent line direction and $\mathcal{M}_\epsilon$. The relative error of the estimated drift in the (x_4, z_4) plane are on average 9%, vs. 20% in all 6 dimensions reported in Table 4. The 5 fast variables are almost orthogonal to the slow variable (as suggested in Legoll and Lelièvre 2012): we therefore expect the local orthogonal projections to work as well as the oblique ones. The top three eigenfunctions of an
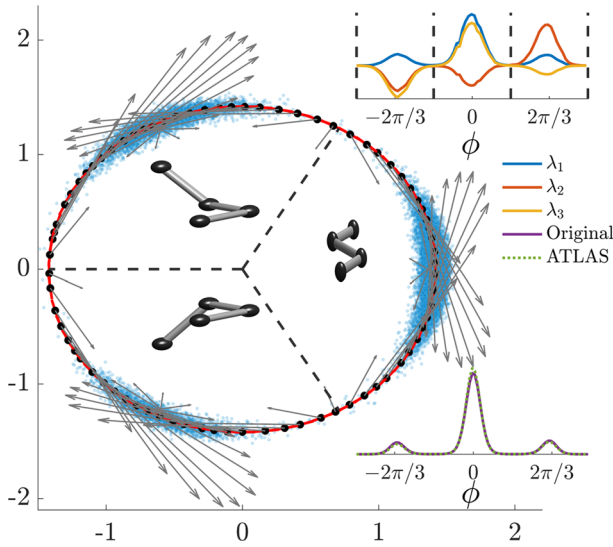
Fig. 6 Butane. The points in the sample trajectory with time length 20 simulated by the original simulator are scattered in blue dots in (x_4, z_4). Three data clusters corresponds to $M_{\text{top-cis}}$ (upper left), $M_{\text{bot-cis}}$ (lower left) and M_{trans} (right) metastable states. The landmarks (black dots) and their neighborhood (red lines) and effective drift direction (gray arrows) in (x_4, z_4) are plotted. In the upper right, the top three eigenfunctions of the transition matrices are plotted in the coordinate of ϕ with $\lambda_1 = 1, \lambda_2 = 0.9999$ and $\lambda_3 = 0.9999$. In the bottom right, the kernel-fitted invariant distribution from the trajectory of time length 500 generated by the original simulator and ATLAS simulator are plotted in the coordinate of ϕ
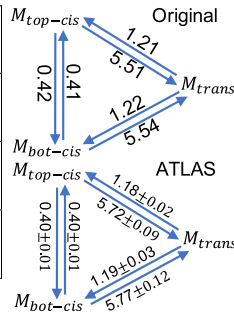
MSM estimated by ATLAS simulator identify these three metastable regions on the slow manifold, see Fig. 6 and Appendix C.3. The invariant distribution of the ATLAS process has density very close, on $\hat{\mathcal{M}}_\epsilon$, to the one generated by the original simulator, with the estimated L^1-norm of the difference of the density 0.060 ± 0.013. The results reported in Table 3 show that the mean residence times in the three metastable states, estimated with ATLAS, are within 4% relative error, with a runtime is about 68% of the original simulator. All estimated reaction rate constants are within 5% relative error. The training time of ATLAS is about 13hrs.

8 Conclusion

We have introduced a nonlinear nonparametric technique for reduction of fast-slow stochastic systems, that given a timescale τ and access to short trajectories from a black-box simulator, estimates an invariant manifold and an effective stochastic process, called ATLAS, on it, that averages the original system below the timescale τ. The simulator for ATLAS has time-step of order τ, typically much larger than the time-step of the original simulator δt (which depends on the fastest timescale), and is intrinsically low-dimensional, making it possible to compute efficiently many long paths of the effective dynamics, and compute approximations to important quantities, such as stationary distributions, mean residence times, and transition rates. We have

Table 3 Mean residence time (MRT) and reaction rates for butane

Sim. / IC	Original $\mathcal{S}$	ATLAS	Ave. rel. error
M_{trans}	0.412 ±0.005	0.414 ±0.005	0.02 ±0.01
$M_{\text{top-cis}}$	0.168 ±0.002	0.166 ±0.002	−0.03 ±0.02
$M_{\text{bot-cis}}$	0.168 ±0.002	0.158 ±0.002	−0.03 ±0.02
Runtime	5.61h	3.84h	

Left: the metastable states are given in ϕ variable: $M_{\text{trans}} = (-\frac{\pi}{3}, \frac{\pi}{3})$, $M_{\text{top-cis}} = (\frac{\pi}{3}, \pi)$ and cis $M_{\text{bot-cis}} = (-\pi, -\frac{\pi}{3})$. **Right:** mean transition rate constants between metastable states, with corresponding confidence intervals (estimated over multiple ATLASes)

Table 4 Summary of error analysis

Models	Errors AbsErr($\hat{\mathcal{M}}_\epsilon$)	AbsErr($\hat{T}_{\hat{z}^l}$)	relErr($\hat{\mathbf{b}}$)	relErr($\hat{\Lambda}$)
Pinched sphere	$7(3) \times 10^{-3}$	0.02(0.03)	0.11(0.10)	0.04(0.03)
	$3(1) \times 10^{-2}$	0.03(0.01)	0.10(0.08)	0.05(0.03)
Oscillating half-moons	$9.63(0.04) \times 10^{-3}$	0.10(0.06)	0.33(0.23)	0.12(0.07)
	$5(2) \times 10^{-3}$	0.09(0.05)	0.32(0.18)	0.11(0.05)
Butane	$6(2) \times 10^{-3}$	0.014(0.007)	0.17(0.12)	0.06(0.02)
	$5(2) \times 10^{-3}$	0.011(0.007)	0.21(0.13)	0.08(0.01)

In each example, we report the mean and standard derivation (in parentheses) of error terms for the quantities evaluated at landmarks (white rows), and at points along ATLAS trajectories (gray rows); see Appendix D

shown that, under suitable conditions, the estimation of ATLAS is not cursed by the dimension of the state space, and that ATLAS is robust to certain model errors.

This technique significantly extends the one introduced in Crosskey and Maggioni (2017) by correctly handling (1) large fast modes, instead of only very small fast oscillations around a slow manifold, which could be estimated by local PCA, (2) fast modes that are not orthogonal to the slow manifolds, (3) smoothly interpolating all estimated geometric and dynamics quantities, increasing the accuracy of the estimation. Last but not least, it is designed to efficiently run in exploration mode, without loss of accuracy.

The literature on model reduction, averaging and homogenization is vast, see, e.g., Pavliotis and Stuart (2008), Hartmann et al. (2020), Husic and Pande (2018), Maria Bruna et al. (2014) and Givon et al. (2004). Unlike existing techniques, here we do not require: previous knowledge of reaction coordinates or of the slow variables, which we estimate directly; linearity of the slow variables (as in PCA/PODs Holmes et al. 2012); that the fast modes are small (as in local PCA/PODs Holmes et al. 2012 or DMD Rowley et al. 2009; Kutz et al. 2016 or TICA Molgedey and Schuster 1994; Pérez-Hernández et al. 2013), nor that they are orthogonal to the slow manifold, nor

that they can be globally defined (as in manifold learning techniques such as Coifman et al. (2008), Rohrdanz et al. (2011), and Singer et al. (2009) and many others), which either requires the absence of even simple topological obstructions (loops) or require a possibly arbitrarily large number of additional coordinates. We also do not require to sample long trajectories, and in exploration mode we do not require a set of sufficiently well-behaved initial conditions; unlike exploration techniques such as Chiavazzo et al. (2017) (and references therein). These techniques can fail (and they do in our examples) to correctly parametrize the invariant manifold, or (not exclusive) the effective dynamics, or would be cursed by the dimension of the state space. Our ATLAS algorithm estimates consistently and accurately the effective dynamics and its invariant manifold in an exploration scheme, which by itself is useful in many cases. Our reduction onto the estimated invariant manifold is nonlinear, and the estimation of both the invariant manifold and of the Itô diffusion is locally parametric, in order to reduce the local sample size required for a given accuracy, but globally nonparametric.

The setting of our work, where a latent slow–fast system in a natural linear coordinate system is observed through a nonlinear observation map, is inspired by the works (Singer et al. 2009; Dsilva et al. 2016). These works start with a latent model significantly simpler than that in Eq. (2.1), and their objective is to learn the map back to the latent space, or at least to the slow variables in the latent space, from bursts of trajectories in observed space. That problem is tackled under significantly stronger assumptions on the latent system, and the approach is typically cursed by the ambient dimension, mainly because it seeks the reduction to slow variables *after* having constructed an approximation to the full system in the state space. In our work we first locally estimate a reduced system, and do so parsimoniously, by using a rather minimal set of parametric tools, and avoiding the curse of dimensionality.

Extensions to higher order equations, such as Langevin equations, more general local models and nonlinear open neighborhood, incorporating symmetries and conserved quantities, considering non-Gaussian noise and combination with rare sampling techniques are currently being explored.

Data Availability Data deposition: The software package implementing the proposed algorithms can be found on https://github.com/yexf308/ATLAS.

A: Assumption, Linear Approximation, and Averaging

We briefly review here the definitions of slow and invariant manifolds, some of the very basic expansions in geometric singular perturbation theory that motivate our key linearized model in Eq. (3.1), and the assumptions underlying them.

As a matter of notation, $E_{i.}$ denotes the i-th row of a matrix E, and $E_{.j}$ denotes the j-th column of a matrix E.

A.1: Assumption

The following assumptions 1-4 ensure that for the original latent stochastic system Eq. (2.1) there exists a uniformly asymptotically stable invariant manifold $\mathcal{M}_\epsilon^{\mathbf{x}}$ (Berglund and Gentz 2006, 2003; Kuehn 2015).

Assumption 1 Domain and differentiability: $f \in \mathcal{C}^2(\mathcal{D}, \mathbb{R}^{D-d})$, $g \in \mathcal{C}^2(\mathcal{D}, \mathbb{R}^d)$ and $F \in \mathcal{C}^1(\mathcal{D}, \mathbb{R}^{(D-d)\times(D-d)})$, $G \in \mathcal{C}^1(\mathcal{D}, \mathbb{R}^{d\times d})$, where $\mathcal{D}$ is an open subset of $\mathbb{R}^d \times \mathbb{R}^{D-d}$. We further assume that f, g, F, G are bounded in sup-norm by a constant M within $\mathcal{D}$.

Assumption 2 Slow manifold: there is a connected open subset $\mathcal{D}_0 \subset \mathbb{R}^d$ and a continuous function $\mathbf{y}^\star : \mathcal{D}_0 \to \mathbb{R}^{D-d}$ such that

$$\mathcal{M}_0^{\mathbf{x}} = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{D} : \mathbf{y} = \mathbf{y}^\star(\mathbf{x}), \mathbf{x} \in \mathcal{D}_0\}$$

is a slow manifold of the deterministic system, that is, $(\mathbf{x}, \mathbf{y}^\star(\mathbf{x})) \in \mathcal{D}$ and $f(\mathbf{x}, \mathbf{y}^\star(\mathbf{x})) = 0$ for all $\mathbf{x} \in \mathcal{D}_0$.

Assumption 3 Stability: the slow manifold is uniformly asymptotically stable, that is, all eigenvalues of the Jacobian matrix

$$A^\star(\mathbf{x}) = \partial_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}^\star(\mathbf{x}))$$

have negative real parts, uniformly bounded away from 0 for all $\mathbf{x} \in \mathcal{D}_0$.

Assumption 4 Non-degeneracy: the diffusivity matrix $F(\mathbf{x}, \mathbf{y})F(\mathbf{x}, \mathbf{y})^T$ is positive definite.

Under these assumptions, Fenichel's theorem guarantees the existence of an invariant manifold (also called adiabatic manifold) (Berglund and Gentz 2006, 2003; Jones 1995)

$$\mathcal{M}_\epsilon^{\mathbf{x}} = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{D} : \mathbf{y} = \bar{\mathbf{y}}(\mathbf{x}, \epsilon), \mathbf{x} \in \mathcal{D}_0\},$$

in a neighborhood of which trajectories concentrate for an extended time w.h.p. Also, $\mathcal{M}_\epsilon^{\mathbf{x}}$ is close to the slow manifold $\mathcal{M}_0^{\mathbf{x}}$ in the sense that $\bar{\mathbf{y}}(\mathbf{x}, \epsilon) = \mathbf{y}^\star(\mathbf{x}) + O(\epsilon)$.

The next assumption 5 imposes that the effect of the drift term is small relative to the effect of the diffusion term.

Assumption 5 Diffusion-dominated dynamics: for any $1 \leq l \leq L$, $\sqrt{\sigma_d(\Lambda^l)} \gg \|\mathbf{b}^l\|\sqrt{\tau}$.

This assumption allows us to simplify the construction of the diffusion-adapted, Mahalanobis-like metric. Indeed, the $\sqrt{\tau}$-neighborhood of the landmark $B(\mathbf{z}^l, \sqrt{\tau})$ that we use is not exactly the same as the estimated $p\%$ confidence region of finding

the effective reduced stochastic system at time $\sqrt{\tau}$, started at $\mathbf{z}^l$: that would be better approximated by

$$\tilde{B}(\mathbf{z}^l, \sqrt{\tau}) := \{\mathbf{v} : \frac{1}{\mathcal{X}_d^2(p)}(\mathbf{v} - \mathbf{z}_0^{l,\text{slw}} - \mathbf{b}^l\tau)^T(\Lambda^l)^\dagger(\mathbf{v} - \mathbf{z}_0^{l,\text{slw}} - \mathbf{b}^l\tau) < \tau\}.$$

However, with the assumption of diffusion-dominated dynamics, the approximation we use is satisfactory. Indeed, the boundary of $B(\mathbf{z}^l, \sqrt{\tau})$ is a hyperellipsoid of dimension d embedded in $\mathbf{R}^D$. Columns of $U_d^{l,\text{slw}}$, which are the eigenvectors of Λ^l, define the principle axes of the hyperellipsoid. $\sigma_1(\Lambda^l), \sigma_2(\Lambda^l), \cdots, \sigma_d(\Lambda^l)$ are proportional to squares of the lengths of the semi-axes. The vertices of the hyperellipsoid at time t are $\mathbf{z}_0^{l,\text{slw}} \pm \sqrt{\chi_d^2(p)\sigma_i(\Lambda^l)t}\left(U_d^{l,\text{slw}}\right)_{\cdot i}$ for $i = 1, 2, \cdots, d$, so the minimum length of semi-axes at $t = \tau$ for the hyperellipsoid is $\sqrt{\chi_d^2(p)\sigma_d(\Lambda^l)\tau}$. In the mean time, the center of the hyperellipsoid is moved by $\|\mathbf{b}^l\|\tau$ if we use $\tilde{B}(\mathbf{z}^l, \sqrt{\tau})$ instead of $B(\mathbf{z}^l, \sqrt{\tau})$. Assumption 5 guarantees that the movement of the center is negligible relative to the length of the semi-axes of the hyperellipsoid.

A.2: Linear Approximation

In Eq. (3.1) we assumed linear approximations of the time-dependent expectation $\mathbf{m}_t^l$ and covariance $C(\mathbf{z}_t^l | \mathbf{z}_0^l)$. The slopes of these quantities, as a function of time, are $\mathbf{b}^l$ and Λ^l, and the intercepts of these quantities are $\mathbf{z}_0^l$ and Γ^l. In this section, we provide some mathematical intuitions for this assumption, following the exposition of Berglund and Gentz (2006), to which we refer the reader for further details. We start by considering the system in the latent space, and its linear approximation (2.1) near the invariant manifold. First, we define the deviation of sample paths from the invariant manifold: $\zeta_t := \mathbf{y}_t - \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)$. An application of Itô's formula implies that the fast dynamics part ζ_t satisfies

$$d\zeta_t = \frac{1}{\epsilon}\Big(f(\mathbf{x}_t, \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon) + \zeta_t) - \epsilon\partial_\mathbf{x}\bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)g(\mathbf{x}_t, \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon) + \zeta_t) - \epsilon c^{\text{Itô}}(\mathbf{x}_t, \zeta_t)\Big)dt$$
$$+ \frac{1}{\sqrt{\epsilon}}\Big[-\sqrt{\epsilon}\partial_\mathbf{x}\bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)G(\mathbf{x}_t, \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon) + \zeta_t) \ F(\mathbf{x}_t, \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon) + \zeta_t)\Big]dW_t,$$

$$\text{(A.1)}$$

where $c^{\text{Itô}}(\mathbf{x}, \zeta) := \frac{1}{2}\sum_{j,k=1}^d \frac{\partial^2\bar{\mathbf{y}}(\mathbf{x},\epsilon)}{\partial x_j\partial x_k}G_{j\cdot}(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon) + \zeta)G_{k\cdot}^T(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon) + \zeta)$, and $W_t = [U_t; V_t]$, (Here ";" denotes concatenation of column vectors). Always following Berglund and Gentz (2006), *if* we ignore the Itô-term and use the fact that the drift term vanishes when $\zeta = 0$, we derive that the invariant manifold $\bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)$ should satisfy the following PDE:

$$f(\mathbf{x}_t, \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)) = \epsilon\partial_\mathbf{x}\bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)g(\mathbf{x}_t, \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)). \tag{A.2}$$

Therefore, by Taylor expansion and Eq. (A.2), we have the following linear approximation $\tilde{\zeta}_t$ of the fast dynamics ζ_t in Eq. (A.1):

$$d\tilde{\zeta}_t = \frac{1}{\epsilon} A(\mathbf{x}_t, \epsilon)\tilde{\zeta}_t dt + \frac{1}{\sqrt{\epsilon}} F_0(\mathbf{x}_t, \epsilon) dW_t, \tag{A.3}$$

where

$$A(\mathbf{x}, \epsilon) := \partial_{\mathbf{y}} f(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon)) - \epsilon \partial_{\mathbf{x}} \bar{\mathbf{y}}(\mathbf{x}, \epsilon) \partial_{\mathbf{y}} g(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon)) \quad,$$

$$F_0(\mathbf{x}, \epsilon) := \frac{1}{\sqrt{\epsilon}} \left[-\sqrt{\epsilon} \partial_{\mathbf{x}} \bar{\mathbf{y}}(\mathbf{x}, \epsilon) G(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon)) \ F(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon)) \right].$$

As for the slow dynamics $[\mathbf{x}_t; \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)]$, we have

$$d \begin{bmatrix} \mathbf{x}_t \\ \bar{\mathbf{y}}(\mathbf{x}_t, \epsilon) \end{bmatrix} = g^{\text{slw}}(\mathbf{x}_t) dt + G^{\text{slw}}(\mathbf{x}_t) dU_t, \tag{A.4}$$

where $g^{\text{slw}}(\mathbf{x}) := B(\mathbf{x}) g(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon)) + \tilde{c}^{\text{Itô}}(\mathbf{x})$, $G^{\text{slw}}(\mathbf{x}) := B(\mathbf{x}) G(\mathbf{x}, \bar{\mathbf{y}}(\mathbf{x}, \epsilon))$, and here $B(\mathbf{x}) := \begin{bmatrix} I_{d \times d} \\ \partial_{\mathbf{x}} \bar{\mathbf{y}}(\mathbf{x}, \epsilon) \end{bmatrix} \in \mathbb{R}^{D \times d}$, the Itô term $\tilde{c}^{\text{Itô}}(\mathbf{x}) = \begin{bmatrix} O_{d \times (D-d)} \\ I_{(D-d) \times (D-d)} \end{bmatrix} c^{\text{Itô}}(\mathbf{x}, 0)$.
Here of course $\bar{\mathbf{y}}(\mathbf{x}_t, \epsilon)$ is slaved to $\mathbf{x}_t$, and we call the dynamics of $\mathbf{x}_t$ only the slow reduced dynamics happening in $\mathbb{R}^d$.

The dynamics of $\tilde{\zeta}_t$ shown in Eq. (A.3) is a high dimensional Ornstein–Uhlenbeck process, and thus its expectation $\mathbb{E}\left[\tilde{\zeta}_t^l \middle| \mathbf{z}_0^l\right]$ and covariance $\text{cov}(\tilde{\zeta}_t^l | \mathbf{z}_0^l)$ at landmark l stabilize exponentially fast to $\mathbf{0}$ and some matrix $\Theta(\mathbf{x}_0^l)$. In particular, the error is of order $O(\epsilon)$ once t reaches the timescale of separation $\tau \gg \epsilon$, where here we have utilized assumption 3. At the same time, at the timescale of separation τ, the slow dynamics does not change significantly. In particular, in a neighborhood of a landmark $\mathbf{z}^l$, the drift and diffusion coefficients $g^{\text{slw}}(\mathbf{x}_t^l)$, $G^{\text{slw}}(\mathbf{x}_t^l)$ in Eq. (A.4) do not change too much, and we may treat the slow dynamics as having nearly constant drift and diffusion coefficients. These equations and reasoning justify the form of reduced equations obtain by averaging, in the limit $\epsilon \to 0$, of the form of Eq. (2.2).

Now, notice that we have the following decomposition of the latent coordinate $\mathbf{w}_t^l$ into coordinates for fast dynamics ζ_t^l and coordinates for slow dynamics $[\mathbf{x}_t^l; \bar{\mathbf{y}}(\mathbf{x}_t^l, \epsilon)]$:

$$\mathbf{w}_t^l = \begin{bmatrix} O_{d \times (D-d)} \\ I_{(D-d) \times (D-d)} \end{bmatrix} \zeta_t^l + \begin{bmatrix} \mathbf{x}_t^l \\ \bar{\mathbf{y}}(\mathbf{x}_t^l, \epsilon) \end{bmatrix}. \tag{A.5}$$

Then, according to previous analysis regarding timescale of separation as well as Eq. (A.5), we have the following linear approximations of the latent coordinate at

times t comparable to τ:

$$\mathbb{E}\left[\mathbf{w}_t^l\middle|\mathbf{w}_0^l\right] = \begin{bmatrix} \mathbf{x}_0^l \\ \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon) \end{bmatrix} + g^{\text{slw}}(\mathbf{x}_0^l)t + O(\epsilon),$$

$$\text{cov}(\mathbf{w}_t^l|\mathbf{w}_0^l) = \begin{bmatrix} O_{d\times(D-d)} \\ I_{(D-d)\times(D-d)} \end{bmatrix} \Theta(\mathbf{x}_0^l) \begin{bmatrix} O_{d\times(D-d)} \\ I_{(D-d)\times(D-d)} \end{bmatrix}^T \tag{A.6}$$

$$+ G^{\text{slw}}(\mathbf{x}_0^l)\left[G^{\text{slw}}(\mathbf{x}_0^l)\right]^T t + O(\epsilon).$$

These equations motivate Eq. (3.1), but in the latent space.

We now proceed to consider the situation in the observed variables, locally around a fixed point $\mathbf{z}_t^l$; in particular we consider the behavior of the time-dependent expectation $\mathbf{m}_t^l$ and covariance $C(\mathbf{z}_t^l|\mathbf{z}_0^l)$. Assume the local dynamics around landmark $\mathbf{z}^l$ is within the chart $(\mathcal{U}_{\alpha(l)}, \varphi_{\alpha(l)})$, and let $\varphi_{\alpha(l)}(\mathbf{z}_t^l) = \mathbf{w}_t^l = [\mathbf{x}_t^l; \mathbf{y}_t^l]$. We assume that the 1-st order Taylor approximation of $\varphi_{\alpha(l)}^{-1}$ is accurate enough, which corresponds to the map φ having sufficiently small Hessian, or restricting the size of the neighborhood under consideration to be small enough. We can then assume that for t comparable to τ, we have:

$$\varphi_{\alpha(l)}^{-1}(\mathbf{x}_t^l, \mathbf{y}_t^l) = \varphi_{\alpha(l)}^{-1}(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))$$

$$+ J^T(\varphi_{\alpha(l)}^{-1})\,|_{(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))}\left(\begin{bmatrix} \mathbf{x}_t^l \\ \mathbf{y}_t^l \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0^l \\ \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon) \end{bmatrix}\right) + O(\epsilon).$$

Then, according to Eq. (A.6), we have the linear approximations in the observation space of the form

$$\mathbf{z}_0^{l,\text{slw}} = \varphi_{\alpha(l)}^{-1}(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon)),$$

$$\mathbf{b}^l = J^T(\varphi_{\alpha(l)}^{-1})\,|_{(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))}\, g^{\text{slw}}(\mathbf{x}_0^l),$$

$$\Gamma^l = J^T(\varphi_{\alpha(l)}^{-1})\,|_{(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))} \begin{bmatrix} I_{(D-d)\times(D-d)} \\ O_{d\times(D-d)} \end{bmatrix}$$

$$\Theta(\mathbf{x}_0^l) \begin{bmatrix} I_{(D-d)\times(D-d)} \\ O_{d\times(D-d)} \end{bmatrix}^T J(\varphi_{\alpha(l)}^{-1})\,|_{(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))},$$

$$\Lambda^l = J^T(\varphi_{\alpha(l)}^{-1})\,|_{(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))}\, G^{\text{slw}}(\mathbf{x}_0^l)\left[G^{\text{slw}}(\mathbf{x}_0^l)\right]^T J(\varphi_{\alpha(l)}^{-1})\,|_{(\mathbf{x}_0^l, \bar{\mathbf{y}}(\mathbf{x}_0^l, \epsilon))},$$

which justify the crucial approximation in Eq. (3.1), which motivates all our local estimators.

A.3: Averaging

In this section, we briefly review the idea of stochastic averaging, which is a classical method to analyze fast/slow systems, see for example Freidlin et al. (2012) and Pavliotis and Stuart (2008) for a comprehensive review of this subject. See in particular,

theorem 2.1 in chapter 7 of Freidlin et al. (2012), known as the averaging principle,
Givon et al. (2004) for a survey of several approaches to the problem of extracting
effective dynamics including averaging, where similarities and differences between
these approaches are highlighted.

From the theory perspective, there exists huge body of literature on stochastic
averaging, where strong and weak convergence results are provided under different
regularity assumptions on slow–fast SDE coefficients, see, e.g., Givon et al. (2006),
Khas'minskii (1966), Bakhtin and Kifer (2004), Kifer (2005), Li (2008) and Yu and
Veretennikov (1991). Besides these, many works also study the rate of convergence
of the original process to the reduced one (e.g., as a function of ϵ); see for example
Pavliotis and Stuart (2008), Givon (2007), Khasminskii and Yin (2004), Liu (2010),
Vanden-Eijnden (2003), Zhang et al. (2018), Röckner et al. (2019), Abourashchi and
Veretennikov (2010), Has'minskii (1966) and Weinan et al. (2005). In particular, the
recent work (Röckner et al. 2019) provides a very general, robust and unified method
for establishing the averaging principle, involving both strong and weak convergence,
for slow–fast SDEs with irregular coefficients and under the fully coupled case (i.e.,
the diffusion coefficient in the slow equation can depend on the fast term) for weak con-
vergence (but not for strong convergence). This leads to simplifications and extensions
of previously established results. It also shows that the strong and weak convergence
rates depend only on the regularity of all the coefficients with respect to the slow
variable.

In addition, a very recent work (Hartmann et al. 2020) provides quantitative results
on the connection of averaging to coarse-graining and effective dynamics in multiscale
studies. It also presents a detailed comparison of the averaging and the conditional
expectation approach in the case of (non-reversible) Ornstein–Uhlenbeck (O-U) pro-
cesses and isolate sufficient conditions under which the two approaches agree.

Now, we briefly review the formulation of averaging, and we start from considering
in the latent space. From Eq. (2.1), we define the averaged approximation around the
invariant manifold, as in Eq. (2.2), with the averaged coefficients given by Röckner
et al. (2019):

$$\bar{g}(\mathbf{x}) := \int_{\mathbb{R}^{D-d}} g(\mathbf{x}, \mathbf{y}) \nu(\mathrm{d}\mathbf{y}|\mathbf{x}) \quad , \quad \bar{G}(\mathbf{x}) := \sqrt{\int_{\mathbb{R}^{D-d}} G(\mathbf{x}, \mathbf{y}) G^T(\mathbf{x}, \mathbf{y}) \nu(\mathrm{d}\mathbf{y}|\mathbf{x})} \, . \tag{A.7}$$

Here the conditional invariant measure $\nu(\mathbf{y}|\mathbf{x})$, as mentioned in Eq. (2.2), is the unique
invariant measure of the process $\mathbf{Y}_t^{\mathbf{x}}$ (Röckner et al. 2019), governed by the equations
"frozen" at $\mathbf{x}$:

$$\begin{cases} \mathrm{d}\mathbf{Y}_t^{\mathbf{x}} = f(\mathbf{x}, \mathbf{Y}_t^{\mathbf{x}})\mathrm{d}t + F(\mathbf{x}, \mathbf{Y}_t^{\mathbf{x}})\mathrm{d}V_t \\ \mathbf{Y}_0^{\mathbf{x}} = \mathbf{y} \end{cases} .$$

Now we move to the observation space. In the local coordinates discussed in Sect. 2,
split between the local tangent plane to the invariant manifold and the affine subspace
containing the fast variables, we can decouple slow and fast variables. We then have

the following local SDEs written in the variables $\mathbf{z}^{\text{slw}}$ and $\boldsymbol{\xi}$, derived from Eq. (2.1) by applying Itô's formula:

$$\begin{cases} d\mathbf{z}_t^{\text{slw}} = \tilde{g}(\mathbf{z}_t^{\text{slw}}, \boldsymbol{\xi}_t)dt + \tilde{G}(\mathbf{z}_t^{\text{slw}}, \boldsymbol{\xi}_t)dU_t \\ d\boldsymbol{\xi}_t = \dfrac{1}{\epsilon}\tilde{f}(\mathbf{z}_t^{\text{slw}}, \boldsymbol{\xi}_t)dt + \dfrac{1}{\sqrt{\epsilon}}\tilde{F}(\mathbf{z}_t^{\text{slw}}, \boldsymbol{\xi}_t)dV_t \end{cases}.$$

From these SDEs in local coordinates, we can define the reduced SDEs, as in Eq. (2.3), with the averaged coefficients given by

$$b(\mathbf{z}^{\text{slw}}) = \int_{\mathbb{R}^D} \tilde{g}(\mathbf{z}^{\text{slw}}, \boldsymbol{\xi})\nu(d\boldsymbol{\xi}|\mathbf{z}^{\text{slw}}) \ ,$$

$$H(\mathbf{z}^{\text{slw}}) = \sqrt{\int_{\mathbb{R}^D} \tilde{G}(\mathbf{z}^{\text{slw}}, \boldsymbol{\xi})\tilde{G}^T(\mathbf{z}^{\text{slw}}, \boldsymbol{\xi})\nu(d\boldsymbol{\xi}|\mathbf{z}^{\text{slw}})}. \qquad (A.8)$$

Here again, the conditioned invariant measure $\nu(\boldsymbol{\xi}|\mathbf{z}^{\text{slw}})$, as mentioned in Eq. (2.3), is the unique invariant measure of the process $\boldsymbol{\xi}_t^{\mathbf{z}^{\text{slw}}}$, governed by the equations "frozen" at $\mathbf{z}^{\text{slw}}$:

$$\begin{cases} d\boldsymbol{\xi}_t^{\mathbf{z}^{\text{slw}}} = \tilde{f}(\mathbf{z}^{\text{slw}}, \boldsymbol{\xi}_t^{\mathbf{z}^{\text{slw}}})dt + \tilde{F}(\mathbf{z}^{\text{slw}}, \boldsymbol{\xi}_t^{\mathbf{z}^{\text{slw}}})dV_t \\ \boldsymbol{\xi}_0^{\mathbf{z}^{\text{slw}}} = \boldsymbol{\xi} \end{cases}.$$

As mentioned in Sect. 2, these reduced SDEs as in Eq. (2.3) may be viewed in intrinsic coordinates, or in Cartesian coordinates in the ambient space $\mathbb{R}^D$, with $\mathbf{z}_t^{\text{slw}} \in \mathbb{R}^D$ but on $\mathcal{M}_\epsilon$, $b \in \mathbb{R}^D$ a vector field on $\mathcal{M}_\epsilon$, and $H \in \mathbb{R}^{D \times d}$ acting on a Wiener process U_t in $\mathbb{R}^d$.

B: Algorithms

We provide here detailed pseudo-code for the construction of ATLAS, see Algorithms 2, 3, 4, 5, 6, 7. The algorithm follows closely the theory, with minor caveats, having to do with constants that are assumed known, while in practice they would either need to be estimated, or set by the user using external information.

    We also discuss several minor modifications to the algorithms.

1. In algorithm 6, the neighbor landmarks of the current point $\mathbf{z}_t$, $\mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z}_t)$ are approximated by the nearest landmark and its neighbors, $\{k_t, \mathcal{N}(k_t)\}$. Then we need an efficient method to search the nearest landmark k_{t+1} for the next point $\mathbf{z}_{t+1}$ in the "metric" $\hat{\tilde{\rho}}$. we update the nearest landmark by only calculating the distance of $\mathbf{z}_{t+1}$ to the current nearest landmark and its neighbors, and repeat this procedure until the nearest landmark remains the same. This procedure avoids the global search which could be very expensive. Then when simulating ATLAS process, there is no need to check whether $\|\mathbf{z} - \hat{\mathbf{z}}^l\| \leq \hat{R}_{\max}$ when calculating $\hat{\tilde{\rho}}(\mathbf{z}_t, \hat{\mathbf{z}}^l)$, since the current point $\mathbf{z}_t$ is always close enough to the neighbor landmarks.

2. In algorithm 5 and algorithm 7 we say that when $\hat{\rho}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^k) < d_{\mathrm{con}}$, we add k to $\mathcal{N}(l)$ and add l to $\mathcal{N}(k)$: in practice, we relax this condition to $\min(\hat{\tilde{\rho}}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^k), \hat{\tilde{\rho}}(\hat{\mathbf{z}}^k, \hat{\mathbf{z}}^l)) < d_{\mathrm{con}}$, in order to include more landmarks. This slight modification is particularly useful when the condition that the process is diffusion-dominated (see assumption 5 in Appendix A) does not hold.

3. In algorithm 3 and algorithm 6, there is no need to explicitly calculate $\hat{\Lambda}^{\mathcal{A}}(\mathbf{z})$ since we can use the iterative algorithm to output top d singular values and corresponding singular vectors to estimate the diffusion coefficient $\hat{H}_d^{\mathcal{A}}(\mathbf{z})$ with the cost of $O(C^d D d^2)$ where C^d is the number of landmarks in $\mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})$. In the $d = 1$ scenarios (i.e., oscillating half-moons and butane), the average number of landmarks in $\mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})$ is approximated 4 and for the $d = 2$ example (Pinched Spheres), this average increases to about 8. However, if d is not relatively small, the number of landmarks in $\mathcal{N}_\tau^{\mathcal{A}}(\mathbf{z})$ may grow exponentially as d, so it might be challenging to simulate trajectories with ATLAS simulator. When the ambient dimension D is very large, one could use randomized SVD in Eqs. (3.5, 3.13, 3.14) to further significantly lower the computational complexity in projection to rank d (Liberty et al. 2007). In all three models, we didn't use this approach since the biggest D that we tested is 20 and the most time consuming part is not this projection step.

4. When refinement of the landmark is necessary in algorithm 4, we perform this only in the last round of refinement. The reason that we don't use this correction at each round of refinement is it will consistently move the landmark position toward the direction of the effective drift and the region like around the saddle point will be not covered by the neighborhoods of landmarks.

Input: $\mathcal{S}(\mathbf{z}_0, N, t_0)$: original simulator generates N trajectories starting from $\mathbf{z}_0$ of time length t_0; μ_0: probability measure for initial conditions; L: number of initial conditions; $[\tau_{\min}, \tau_{\max}]$: interval for regression; τ: timescale for reduction; d: intrinsic dimension; d_{con}: connection threshold.

Output: ATLAS $\mathcal{A}$: structure containing data to evaluate $\hat{P}^{\mathcal{A}}, \hat{\mathbf{b}}^{\mathcal{A}}, \hat{\Lambda}^{\mathcal{A}}, \hat{H}_d^{\mathcal{A}}$ at any $\mathbf{z}$, and simulate $\mathbf{z}_t^{\mathcal{A}}$ on $\hat{\mathcal{M}}_\epsilon$.

1: Sample L initial conditions $\{\mathbf{z}_0^l\}_{l=1,\dots,L} \sim_{\text{i.i.d.}} \mu_0$
2: **for** $l = 1, \dots, L$ **do**
3:     $\mathcal{B}^l \leftarrow \mathcal{S}(\mathbf{z}_0^l, N, \tau_{\max})$.
4:     $[\{\hat{\mathbf{b}}^l, \hat{\Lambda}_d^l, \hat{H}_d^l, \bar{\mathbf{m}}_M^l, \bar{t}_M, \hat{V}_{D-d}^{l,\mathrm{fst}}, \hat{U}_d^{l,\mathrm{slw}}\}_l] \leftarrow$ LearningDynamics$(\mathcal{B}^l, \tau_{\min}, \tau_{\max}, d)$: estimation of local parameters of effective dynamics.
5:     $[\{\hat{\mathbf{z}}^l, \hat{T}_{\hat{\mathbf{z}}^l} \mathcal{M}_\epsilon, \hat{P}^l(\mathbf{z})\}_l] \leftarrow$ LearningGeometry$(\hat{\mathbf{b}}^l, \bar{\mathbf{m}}_M^l, \bar{t}_M, \hat{V}_{D-d}^{l,\mathrm{fst}}, \hat{U}_d^{l,\mathrm{slw}})$: estimation of local parameters of $\mathcal{M}_\epsilon$.
6: **end for**
7: $\{\hat{\mathbf{z}}^l\}_{l=1}^{L'} \leftarrow$ ConstructIM$(\{\hat{\mathbf{z}}^l\}_{l=1}^L, \tau, d_{\mathrm{con}})$: construct sketch of the invariant manifold $\mathcal{M}_\epsilon$.
8: $\hat{P}^{\mathcal{A}}, \hat{\mathbf{b}}^{\mathcal{A}}, \hat{\Lambda}^{\mathcal{A}}, \hat{H}_d^{\mathcal{A}} \leftarrow$ assemble interpolated quantities using the estimated quantities above and Eqs. (3.11–3.14).

Algorithm 2: Pseudo-code for ATLAS construction

$[\{\hat{\mathbf{b}}^l, \hat{\Lambda}^l_d, \hat{H}^l_d, \bar{\mathbf{m}}^l_M, \bar{t}_M, \hat{V}^{l,\text{fst}}_{D-d}, \hat{U}^{l,\text{slw}}_d\}_l] \leftarrow \text{LearningDynamics}(\mathcal{B}^l, \tau_{\min}, \tau_{\max}, d)$

Input: $\mathcal{B}^l$, the local burst; $[\tau_{\min}, \tau_{\max}]$, interval for regression; d, intrinsic dimension.

Output: Estimated drift, diffusivity matrix and diffusion coefficient $\hat{\mathbf{b}}^l, \hat{\Lambda}^l_d, \hat{H}^l_d$ of landmarks.

1: $t_1, \ldots, t_M$: M equispace points in $[\tau_{\min}, \tau_{\max}]$.

2: $\hat{\mathbf{m}}^l_t \leftarrow \frac{1}{N} \sum_{n=1}^N \mathbf{z}^{l,n}_t, \hat{C}^l_t \leftarrow \frac{1}{N-1} \sum_{n=1}^N (\mathbf{z}^{l,n}_t - \hat{\mathbf{m}}^l_t) \otimes (\mathbf{z}^{l,n}_t - \hat{\mathbf{m}}^l_t)$: empirical means and covariances, Eq. (3.2).

3: $\bar{\mathbf{m}}^l_M \leftarrow \frac{1}{M} \sum_{m=1}^M \hat{\mathbf{m}}^{l,N}_{t_m}, \bar{C}^l_M \leftarrow \frac{1}{M} \sum_{i=1}^M \hat{C}^l_{t_m}, \bar{t}_M \leftarrow \frac{1}{M} \sum_{m=1}^M t_m$.

4: $\hat{\mathbf{b}}^l \leftarrow \sum_{m=1}^M (\hat{\mathbf{m}}^{l,N}_{t_m} - \bar{\mathbf{m}}^l_M)(t_m - \bar{t}_M)/\sum_{m=1}^M (t_m - \bar{t}_M)^2$: estimated drift term, Eq. (3.3).

5: $\hat{\Lambda}^l \leftarrow \sum_{m=1}^M (\hat{C}^l_{t_m} - \bar{C}^l_M)(t_m - \bar{t}_M)/\sum_{m=1}^M (t_m - \bar{t}_M)^2$

6: $\hat{\Lambda}^l_d \leftarrow \text{Proj}_{\text{rk}(d)}(\hat{\Lambda}^l) = \hat{U}^{l,\text{slw}}_d \hat{\Sigma}_d (\hat{U}^{l,\text{slw}}_d)^T$: estimated diffusivity matrix, Eq. (3.4).

7: $\hat{H}^l_d \leftarrow (\hat{\Lambda}^l_d)^{\frac{1}{2}} = \hat{U}^{l,\text{slw}}_d \sqrt{\hat{\Sigma}_d}$: estimated diffusion coefficient, Eq. (3.5)

8: $\hat{\Gamma}^l_{D-d} \leftarrow \text{Proj}_{\text{rk}(D-d)}(\bar{C}^l_M - \hat{\Lambda}^l \bar{t}_M) = \hat{V}^{l,\text{fst}}_{D-d} \tilde{\Sigma}_{D-d}(\hat{V}^{l,\text{fst}}_{D-d})^T$: estimated covariance matrix of the fast modes, Eq. (3.6).

Algorithm 3: Estimation of local parameters of effective dynamics [Sect. 3.3]

$[\{\hat{\mathbf{z}}^l, \hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon, \hat{P}^l(\mathbf{z})\}_l] \leftarrow \text{LearningGeometry}(\hat{\mathbf{b}}^l, \bar{\mathbf{m}}^l_M, \bar{t}_M, \hat{V}^{l,\text{fst}}_{D-d}, \hat{U}^{l,\text{slw}}_d)$

Input: Estimated quantities in algorithm 3;

Output: $\hat{\mathbf{z}}^l$: landmarks; $\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon$: estimated tangent planes; $\hat{P}^l(\mathbf{z})$: oblique projection along the fast mode.

1: $\hat{\mathbf{z}}^l \leftarrow \bar{\mathbf{m}}^l_M - \hat{\mathbf{b}}^l \bar{t}_M$: estimated landmarks, Eq. (3.7).

2: $\hat{P}^l(\mathbf{z}) \leftarrow \hat{U}^{l,\text{slw}}_d (\hat{U}^{l,\text{slw}}_d)^T (EE^T)^\dagger (\mathbf{z} - \hat{\mathbf{z}}^l) + \hat{\mathbf{z}}^l$, where $E = [\hat{U}^{l,\text{slw}}_d, \hat{V}^{l,\text{fst}}_{D-d}]$: estimated oblique affine projections along the fast modes, Eq. (3.9).

3: $\hat{T}_{\hat{\mathbf{z}}^l}\mathcal{M}_\epsilon \leftarrow \text{span}(\text{cols}(\hat{U}^{l,\text{slw}}_d))$: estimated local tangent plane, Eq. (3.8).

Algorithm 4: Estimation of local geometric parameters of invariant manifold $\mathcal{M}_\epsilon$ [Sect. 3.4]

Input: landmarks $\{\hat{\mathbf{z}}^l\}_{l=1}^L$ and their estimated quantities; τ, timescale for reduction; d_{con}: connection threshold.

Output: updated landmarks $\{\hat{\mathbf{z}}^l\}_{l=1}^{L'}$ and their neighbors $\mathcal{N}(l)$.

1: **for** $l = 1, \ldots, L$ **do**

2: 　**for** $k = l + 1, \ldots, L$ **do**

3: 　　**if** $\hat{\rho}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^k) < \left(1 - \sqrt{\frac{1}{2}}\right) \sqrt{\tau}$ **then**

4: 　　　Remove $\hat{\mathbf{z}}^k$.

5: 　　**else if** $\hat{\rho}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^k) < d_{\text{con}}$ **then**

6: 　　　Add k to $\mathcal{N}(l)$, add l to $\mathcal{N}(k)$.

7: 　　**end if**

8: 　**end for**

9: **end for**

Algorithm 5: Construct sketch of the invariant manifold $\mathcal{M}_\epsilon$, $\{\hat{\mathbf{z}}^l\}_{l=1}^{L'} \leftarrow \text{ConstructIM}(\{\hat{\mathbf{z}}^l\}_{l=1}^L, \tau, d_{\text{con}})$ [Sect. 3.4]

Input: landmarks $\{\hat{\mathbf{z}}^l\}_{l=1}^{L'}$ and their estimated quantities; λ: time-step in ATLAS simulator in the unit of τ; $\mathbf{z}_t$: current point; k_t: the nearest landmark of current point

Output: $\mathbf{z}_{t+1}$, next point; k_{t+1}, the nearest landmark of current point in the "metric" $\hat{\tilde{\rho}}$

1: $\omega^l(\mathbf{z}_t) \leftarrow \exp(-\hat{\tilde{\rho}}(\mathbf{z}_t, \hat{\mathbf{z}}^l)/\sqrt{\tau})$ for $l \in \{k_t, \mathcal{N}(k_t)\}$, $Z(\mathbf{z}_t) \leftarrow \sum_{l \in \{k_t, \mathcal{N}(k_t)\}} \omega^l(\mathbf{z}_t)$.

2: $\mathbf{z}_t^{\mathcal{A}} \leftarrow \frac{1}{Z(\mathbf{z}_t)} \sum_{l \in \{k_t, \mathcal{N}(k_t)\}} \hat{P}^l(\mathbf{z}) w^l(\mathbf{z}_t)$, $\hat{\mathbf{b}}^{\mathcal{A}}(\mathbf{z}_t) \leftarrow \frac{1}{Z(\mathbf{z}_t)} \sum_{l \in \{k_t, \mathcal{N}(k_t)\}} \hat{\mathbf{b}}^l w^l(\mathbf{z}_t)$ in Eqs. (3.11, 3.12).

3: $\hat{\Lambda}^{\mathcal{A}}(\mathbf{z}_t) \leftarrow \frac{1}{Z(\mathbf{z}_t)} \sum_{l \in \{k_t, \mathcal{N}(k_t)\}} \hat{\Lambda}^l w^l(\mathbf{z}_t)$, $\hat{H}_d^{\mathcal{A}}(\mathbf{z}_t) \leftarrow (\text{Proj}_{\text{rk}(d)} \hat{\Lambda}^{\mathcal{A}}(\mathbf{z}_t))^{\frac{1}{2}}$ in Eqs. (3.13, 3.14).

4: $\mathbf{z}_{t+1} \leftarrow \mathbf{z}^{\mathcal{A}} + \hat{\mathbf{b}}^{\mathcal{A}}(\mathbf{z}_t)\lambda\tau + \hat{H}_d^{\mathcal{A}}(\mathbf{z}_t)\Delta W_{\lambda\tau}$, where $\Delta W_{\lambda\tau} \sim \mathcal{N}(0, \lambda\tau I_d)$.

5: Update the nearest landmark k_{t+1} for the next point $\mathbf{z}_{t+1}$.

Algorithm 6: Simulate one time-step with ATLAS simulator, $[\mathbf{z}_{t+1}, k_{t+1}] \leftarrow$ ATLAStime-step$(\mathbf{z}_t, k_t, \{\hat{\mathbf{z}}^l\}_{l=1}^{L'}, \lambda)$ [Sect. 3.5]

Input: $\mathcal{S}(\mathbf{z}_0, N, t_0)$: original simulator generates N trajectories starting from $\mathbf{z}_0$ of time length t_0; μ_0, probability measure for initial conditions; $T_{\max}$: the maximum time-steps; L_{int}: number of initial conditions; $[\tau_{\min}, \tau_{\max}]$: interval for regression; τ: timescale for reduction; d: intrinsic dimension; d_{con}: connection threshold; d_{thr}: exploration threshold; λ: time-step in ATLAS simulator in the unit of τ; $\mathbf{z}$: current point; k: the nearest landmark of the current point.

Output: $\mathbf{z}$: next point; k: the nearest landmark of next point; updated landmarks $\{\hat{\mathbf{z}}^l\}$ and their estimated quantities; ATLAS $\mathcal{A}$.

1: Construct ATLAS $\mathcal{A}$ with L_{int} initial conditions by performing algorithm 2 and construct L landmarks and their estimators.

2: Start the initial condition from the first landmark: $k_0 \leftarrow 1$, $\mathbf{z}_0 \leftarrow \hat{\mathbf{z}}^1$.

3: **while** $t \leq T_{\max}$ **do**

4:    **if** $\min_{j \in \{k_t, \mathcal{N}(k_t)\}} \hat{\tilde{\rho}}(\mathbf{z}_t, \hat{\mathbf{z}}^j) > d_{\text{thr}}$ **then**

5:       $l \leftarrow L + 1$, $\mathbf{z}_0^l \leftarrow \mathbf{z}_t$, $\mathcal{B}^l \leftarrow \mathcal{S}(\mathbf{z}_0^l, N, \tau_{\max})$,

6:       $[\{\hat{\mathbf{b}}^l, \hat{\Lambda}_d^l, \hat{H}_d^l, \bar{\mathbf{m}}_M^l, \bar{t}_M, \hat{V}_{D-d}^{l,\text{fst}}, \hat{U}_d^{l,\text{slw}}\}_l] \leftarrow$ LearningDynamics$(\mathcal{B}^l, \tau_{\min}, \tau_{\max}, d)$,

7:       $[\{\hat{\mathbf{z}}^l, \hat{T}_{\hat{\mathbf{z}}^l} \mathcal{M}_\epsilon, \hat{P}^l(\mathbf{z})\}_l] \leftarrow$ LearningGeometry$(\hat{\mathbf{b}}^l, \bar{\mathbf{m}}_M^l, \bar{t}_M, \hat{V}_{D-d}^{l,\text{fst}}, \hat{U}_d^{l,\text{slw}})$.

8:       **for** $k = 1, \ldots, L$ **do**

9:          **if** $\hat{\rho}(\hat{\mathbf{z}}^l, \hat{\mathbf{z}}^k) < d_{\text{con}}$ **then**

10:             Add k to $\mathcal{N}(l)$, add l to $\mathcal{N}(k)$.

11:          **end if**

12:       **end for**

13:       Update ATLAS $\mathcal{A}$ and $L \leftarrow L + 1$.

14:       $\mathbf{z}_t \leftarrow \hat{\mathbf{z}}^l$, $k_t \leftarrow l$.

15:    **end if**

16:    Simulate one time-step with ATLAS simulator: $[\mathbf{z}_{t+1}, k_{t+1}] =$ ATLAStime-step$(\mathbf{z}_t, k_t, \{\hat{\mathbf{z}}^l\}_{l=1}^L, \lambda)$.

17:    $t \leftarrow t + 1$.

18: **end while**

Algorithm 7: Construct ATLAS simulator in exploration mode [Sect. 4]

Input: landmarks $\{\hat{\mathbf{z}}^l\}_{l=1}^L$ and their estimated quantities; N_{msm}: number of initial points to construct MSM, dt_{msm}: time of the short paths to construct MSM, τ: timescale for reduction.

Output: M: transition matrix for the Markov State Model constructed from ATLAS

1: **for** $i = 1, \ldots, L$ **do**

2:    **for** $j = 1, \ldots, N_{\text{msm}}$ **do**

3:       $[\mathbf{z}, k(j)] \leftarrow$ ATLAStime-step$(\hat{\mathbf{z}}^i, i, \{\hat{\mathbf{z}}^l\}_{l=1}^L, dt_{\text{msm}}/\tau)$

4:    **end for**

5:    $M_{il} \leftarrow \frac{\#(k==l)}{N_{\text{msm}}}$

6: **end for**

Algorithm 8: Construct Markov state models from ATLAS [Sect. 6]

C: Examples

In this section we discuss in depth details and results of ATLAS in the numerical examples in Sect. 7. In Table 5 we report the parameters of the models and for the construction of ATLAS.

In our numerical tests on the accuracy of ATLAS for the approximation of invariant distribution, we proceed as follows. We will generate two sufficiently long trajectories with both the original and ATLAS simulator, and we take samples from each of the two trajectories. Then, we apply the projection $\hat{P}^l$ to the samples obtained from the original simulator. Here we only use the oblique projection which is consistent with the ATLAS simulation method, and does not require any knowledge of the latent space. For plotting purposes only, we visualize the smoothed histograms by binning the projected samples according to latent slow variables, or other specified coordinates, for both the original and ATLAS simulator. The L^1- and L^2-norm of the difference of their approximate probability densities are calculated directly from the histograms. The bin widths should be of order $\sqrt{\tau}$, consistently with the spirit of ATLAS simulator, which averages information below timescale τ, with a constant coefficient depending on the scaling of the diffusion coefficient for the slow variable. If the latent slow variable is unknown, we could automatically cluster the samples by its nearest landmark.

In our numerical tests on the accuracy of ATLAS for the local statistics, we proceed as follows. We will generate a long trajectory with the ATLAS simulator and compare the estimated invariant manifold, estimated tangent space, estimated effective drift and diffusion terms at each point with the analytically derived reduced dynamics on the slow manifold.

In our numerical tests on the accuracy of ATLAS for the approximation of medium- and large-time observables, we proceed as follows. We will sample initial conditions with respect to the invariant measure restricted to the specified initial regions of state space, and obtain an estimate of the residence time in the target regions, with both simulators. To be specific, first, we generate a single sufficiently long trajectory with either the original or ATLAS simulator and uniformly sample N_{IC} initial conditions that are restricted to the corresponding initial regions (e.g., specified metastable states). Second, we run in parallel the dynamics with both simulators giving to each the same initial conditions, sampled above. We check whether a trajectory reaches the boundary of region at each ATLAS time-step for both simulators and record the residence time once they leave. This is to ensure the consistency of time-steps for both simulators in recording residence times. At last, we compute the mean residence time and its confidence interval for both simulators. To ensure the robustness of the ATLAS algorithm, we repeat the construction of ATLAS (which is random with its input, the observed bursts) and sampling of residence time ten times, and calculate the confidence interval of the relative error of the mean resident time.

It is nontrivial to identify the metastable regions in fast-slow stochastic systems when D is very large. With ATLAS one can easily build up Markov state models(MSMs) and construct the transition matrix for the effective reduced process. First, the number of landmarks naturally become associated to the states of MSMs; more precisely state l of the MSMs is the set of points on the invariant manifold whose

nearest landmark, in the $\hat{\tilde{\rho}}$, is $\hat{\mathbf{z}}^l$. Starting from the node of the same i-th landmark we use ATLAS simulator to simulate in parallel N_{msm} short trajectories of time τ, which is one ATLAS time-step. Let N_{ij} be the number of trajectories whose end positions land in the same state of the j-th landmark. The probability transition from i-th landmark to j-th landmark is estimated as $M_{ij} := N_{ij}/N$, see algorithm 8. The eigenvalues and eigenvectors are approximations to the spectrum and eigenfunctions of the transfer operator of the reduced stochastic system, $\exp(\tau\mathcal{L})$, where $\mathcal{L}$ is the generator of the reduced effective process. Since we don't assume the reversibility, the spectrum and eigenvectors could be imaginary. The top left eigenvector of the transition matrix is proportional to the invariant distribution of reduced stochastic process and is always real. The spectral gap, which is the distance between the next dominant eigenvalue and the eigenvalue 1, indicates the decay rate of correlations and the reciprocal of the spectrum gap shows the order of ATLAS time-steps to reach the equilibrium. The number of the dominant eigenvalues are the number of metastable states. Performing some clustering or connectivity detection algorithms on landmarks with high equilibrium density can yield metastable regions. At least in the setting of real eigenvectors, the positive or negative regions of successive dominant eigenvectors can be used to identify metastable regions.

C. 1: Pinched sphere model

The governing slow–fast SDE in Cartesian coordinates becomes

$$
d\mathbf{z} = \left(J(\mathbf{z}) \cdot \mathbf{b}_1(\mathbf{z}) + \frac{1}{2} A(\mathbf{z}) \cdot \mathbf{b}_2(\mathbf{z}) \right) dt + J(\mathbf{z}) \cdot \sigma(\mathbf{z}) \begin{bmatrix} dW_1 \\ dW_2 \\ dW_3 \end{bmatrix} \tag{C.1}
$$

where

$$
J = \begin{bmatrix} \frac{z_1}{\|\mathbf{z}\|} & \frac{z_3 z_1}{\sqrt{z_1^2+z_2^2}} & -z_2 \\ \frac{z_2}{\|\mathbf{z}\|} & \frac{z_3 z_2}{\sqrt{z_1^2+z_2^2}} & z_1 \\ \frac{z_3}{\|\mathbf{z}\|} & -\sqrt{z_1^2+z_2^2} & 0 \end{bmatrix},
$$

$$
\mathbf{b}_1 = \begin{bmatrix} -\frac{c_1}{\epsilon} \frac{1-\sqrt{a_1+a_2 z_3^2/\|\mathbf{z}\|^2}}{\|\mathbf{z}\|} \\ c_3 \frac{4 z_3^3/\|\mathbf{z}\|^3 - 3 z_3/\|\mathbf{z}\|}{\sqrt{z_1^2+z_2^2}} \\ c_5 \left(\frac{z_2 z_3}{\sqrt{z_1^2+z_2^2}\|\mathbf{z}\|^2} + \frac{z_1}{\|\mathbf{z}\|^2} \right) \end{bmatrix}, \quad A = \begin{bmatrix} -z_1 & -z_1 \\ -z_2 & -z_2 \\ -z_3 & 0 \end{bmatrix}, \tag{C.2}
$$

$$
\mathbf{b}_2 = \begin{bmatrix} \frac{c_4^2(z_1^2+z_2^2)}{\|\mathbf{z}\|^4} \\ c_6^2/\|\mathbf{z}\|^2 \end{bmatrix}, \quad \sigma = \mathrm{diag} \begin{bmatrix} c_2/(\sqrt{\epsilon}\|\mathbf{z}\|) \\ c_4 \frac{\sqrt{z_1^2+z_2^2}}{\|\mathbf{z}\|^2} \\ c_6/\|\mathbf{z}\| \end{bmatrix}
$$

Table 5 Parameters of three examples

	Model parameters	δt	τ	$[\tau_{\min}, \tau_{\max}]$	N	Refine	N_{refine}
Pinched sphere	$\epsilon = 5 \times 10^{-3}$, $a_1 = 4$, $a_2 = 8$, $c_1 = 2$, $c_2 = 5 \times 10^{-1}$, $c_3 = 5 \times 10^{-2}$, $c_4 = 4 \times 10^{-1}$, $c_5 = 5 \times 10^{-2}$, $c_6 = 4 \times 10^{-1}$	5×10^{-4}	$200\delta t$	$[100\delta t, 200\delta t]$	4×10^5	No	NA
Oscillating half-moons	$\epsilon = 1 \times 10^{-2}$, $a_1 = 0$, $a_2 = 5 \times 10^{-3}$, $a_3 = 2.5 \times 10^{-3}$, $a_4 = 6 \times 10^{-2}$, $b_1 = 4 \times 10^{-2}$, $b_2 = 3.5 \times 10^{-2}$, $b_3 = 5 \times 10^{-2}$, $b_4 = 2 \times 10^{-2}$	5×10^{-2}	$20\delta t$	$[20\delta t, 25\delta t]$	8×10^6	Yes	1.6×10^7
butane	$l = 1.53$, $k_2 = 3.19225 \times 10^5$, $k_3 = 6.25 \times 10^4$, $\theta_{\text{eq}} = 1.9548$, $c_1 = 2.03782 \times 10^3$, $\beta = 4 \times 10^{-3}$, $c_2 = 1.5852 \times 10^2$, $c_3 = -3.22770 \times 10^3$	1×10^{-6}	$10\delta t$	$[10\delta t, 15\delta t]$	1.6×10^6	Yes	1.6×10^7

	L_{int}	$\chi_d^2(0.95)$	$\hat{\kappa}$	$\hat{R}_{\max}$	λ	d_{con}	d_{thr}	N_{msm}	dt_{msm}	N_{IC}
Pinched sphere	100	5.991	1	1	1	$3\sqrt{\tau}$	$1.05\sqrt{\tau}$	10^6	τ	15000
Oscillating half-moons	10	1.96	1	0.4	1	$4\sqrt{\tau}$	$0.95\sqrt{\tau}$	10^6	τ	30,000
butane	2	1.96	1	1.5	1	$4\sqrt{\tau}$	$0.95\sqrt{\tau}$	10^6	τ	30,000

Without the noise term, the deterministic counterpart of this system has two stable fixed points, $(\theta^*, \phi^*) = (\pi/6, 5\pi/6), (5\pi/6, \pi/6)$, which are marked in Fig. 4. Note under the current parameter settings, $c_6 \gg c_5$ and $c_4 \gg c_3$, the assumption of diffusion-dominated dynamics is satisfied. We claimed in the main text that this system is not reversible, and we demonstrate it here. The corresponding Fokker Planck equation of the governing equation in Cartesian coordinates is

$$\partial_t p(t, \mathbf{z}) = -\nabla \cdot \left[(J(\mathbf{z})\mathbf{b}_1(\mathbf{z}) + \frac{1}{2} A(\mathbf{z})\mathbf{b}_2(\mathbf{z})) p(t, \mathbf{z}) \right] + \nabla^2 [D(\mathbf{z})p(t, \mathbf{z})] . \quad \text{(C.3)}$$

where $D(\mathbf{z}) = \frac{1}{2}(J(\mathbf{z})\sigma^2(\mathbf{z})J^T(\mathbf{z}))$. We transform the equation into the symmetric form

$$\partial_t p(t, \mathbf{z}) = -\nabla \cdot \left[(J(\mathbf{z})\mathbf{b}_1(\mathbf{z}) + \frac{1}{2} A(\mathbf{z})\mathbf{b}_2(\mathbf{z}) - \nabla D(\mathbf{z})) p(t, \mathbf{z}) \right]$$
$$+ \nabla \cdot [D(\mathbf{z})\nabla p(t, \mathbf{z})] . \quad \text{(C.4)}$$

where $\nabla D(\mathbf{z}) = \left[\sum_{j=1}^{3} \frac{\partial}{\partial z_j} D_{1j}(\mathbf{z}), \sum_{j=1}^{3} \frac{\partial}{\partial z_j} D_{2j}(\mathbf{z}), \sum_{j=1}^{3} \frac{\partial}{\partial z_j} D_{3j}(\mathbf{z}) \right]^T$.

The theorem 3.3.7 in Jiang et al. (2004) indicates that the system is reversible if and only if the force $\mathbf{b}_3(\mathbf{z}) = D^{-1}(\mathbf{z})(J(\mathbf{z})b_1(\mathbf{z}) + \frac{1}{2} A(\mathbf{z})b_2(\mathbf{z}) - \nabla D(\mathbf{z}))$ is a conservative vector field. We calculate the curl of $\mathbf{b}_3(\mathbf{z})$ is

$$\nabla \times \mathbf{b}_3(\mathbf{z})$$
$$= \left[z_1 f_1(z_1, z_2) + z_2 f_2(z_1, z_2, z_3), z_2 f_1(z_1, z_2) - z_1 f_2(z_1, z_2, z_3), \frac{2c_5 z_1}{c_6^2(z_1^2 + z_2^2)} \right]^T$$
$$\text{(C.5)}$$

where

$$f_1(z_1, z_2) = -\frac{2c_5 z_2}{c_6^2(z_1^2 + z_2^2)^{3/2}}, \qquad f_2(z_1, z_2, z_3) = \frac{2c_3 z_3(3(z_1^2 + z_2^2) - z_3^2)}{c_4^2(z_1^2 + z_2^2)^2 \|\mathbf{z}\|}$$
$$- \frac{2a_2 c_1 z_3}{c_2^2 \|\mathbf{z}\|^2 \sqrt{a_1 + \frac{a_2 z_3^2}{\|\mathbf{z}\|^2}}} \quad \text{(C.6)}$$

The curl is not zero if all coefficients are positive. So this system is not reversible.

Identifying the slow manifold and the effective stochastic dynamics. The slow manifold $\mathcal{M}_0^{\mathbf{x}}$ in spherical coordinate is $r^*(\theta) = R(\theta)$ in the limit of $\epsilon \to 0$. Since Cartesian coordinates z_1, z_2, z_3 are linear with the radius r, the slow manifold $\mathcal{M}_0$ in Cartesian coordinates is $[R(\theta)\sin(\theta)\cos(\phi), R(\theta)\sin(\theta)\sin(\phi), R(\theta)\cos(\theta)]^T$ in the limit of $\epsilon \to 0$. In this example, $\mathcal{M}_0$ is the image of $\mathcal{M}_0^{\mathbf{x}}$ under the coordinate transformation. It is uniformly asymptotically stable and it has negative and positive curvature at different points (with the chosen set of parameter a_1, a_2). The reduced dynamics (in the limit $\epsilon \to 0$) on the slow manifold $\mathcal{M}_0$, in Cartesian coordinates, is

given as

$$d\mathbf{z} = \mathbf{b}^{\text{eff}}dt + H^{\text{eff}}\begin{bmatrix} dW_2 \\ dW_3 \end{bmatrix} \tag{C.7}$$

where

$$\mathbf{b}^{\text{eff}} = \begin{bmatrix} \frac{d}{d\theta}(R(\theta)\sin(\theta))\cos(\phi) & -R(\theta)\sin(\theta)\sin(\phi) \\ \frac{d}{d\theta}(R(\theta)\sin(\theta))\sin(\phi) & R(\theta)\sin(\theta)\cos(\phi) \\ \frac{d}{d\theta}(R(\theta)\cos(\theta)) & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{c_3\cos(3\theta)}{R(\theta)\sin(\theta)} \\ \frac{c_5\sin(\phi+\theta)}{R(\theta)} \end{bmatrix}$$

$$+ \frac{1}{2}\begin{bmatrix} \frac{d^2}{d\theta^2}(R(\theta)\sin(\theta))\frac{c_4^2\sin^2(\theta)\cos(\phi)}{R^2(\theta)} - \frac{c_6^2\sin(\theta)\cos(\phi)}{R(\theta)} \\ \frac{d^2}{d\theta^2}(R(\theta)\sin(\theta))\frac{c_4^2\sin^2(\theta)\sin(\phi)}{R^2(\theta)} - \frac{c_6^2\sin(\theta)\sin(\phi)}{R(\theta)} \\ \frac{d^2}{d\theta^2}(R(\theta)\cos(\theta))\frac{c_4^2\sin^2(\theta)}{R^2(\theta)} \end{bmatrix},$$

$$H^{\text{eff}} = \begin{bmatrix} \frac{d}{d\theta}(R(\theta)\sin(\theta))\cos(\phi) & -R(\theta)\sin(\theta)\sin(\phi) \\ \frac{d}{d\theta}(R(\theta)\sin(\theta))\sin(\phi) & R(\theta)\sin(\theta)\cos(\phi) \\ \frac{d}{d\theta}(R(\theta)\cos(\theta)) & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{c_4\sin(\theta)}{R(\theta)} & 0 \\ 0 & \frac{c_6}{R(\theta)} \end{bmatrix}$$

$$\tag{C.8}$$

and $\theta = \arctan2(\sqrt{z_1^2 + z_2^2}, z_3)$, $\phi = \text{mod}(\arctan2(z_2, z_1), 2\pi)$. This result is not exactly the effective dynamics (averaged at the finite timescale τ) on the invariant manifold $\mathcal{M}_\epsilon$, but it provides us some good reference. Here, and in the other numerical experiments, we will in fact consider this as ground truth, for measuring the accuracy of the ATLAS estimators of geometric objects (e.g., $\mathcal{M}_\epsilon$) and dynamics quantities (e.g., $\mathbf{b}$ and Λ). Note that when we measure the accuracy of other quantities such as mean residence times and accuracy of the stationary distributions, these approximations are not used, and the ATLAS estimates are compared directly with those obtained from the original simulator.

Estimating the relevant timescale τ. In Fig. 7 we visualize the behavior of $\text{tr}(\hat{C}_t^l)$ and $\|\hat{\mathbf{m}}_t^l\|$ as a function of time: we observe that they initially behave nonlinearly, but then transit to a linear regime, at time about 0.025, consistent with the approximations in Eq. (3.1). In this example, we choose $(\tau_{\min}, \tau_{\max}) = (0.05, 0.10)$ and the timescale of separation $\tau = 0.10$.

Estimating dimension and tangent spaces of $\mathcal{M}_0$, and direction of the fast modes. We all report in Fig. 7 the behavior of the singular values of $\hat{C}_\tau^l$, $\hat{\Gamma}_{D-d}^l$ and $\hat{\Lambda}_d^l$ at one landmark in descending order. The dominant singular values are visualized with cross markers. Local Principal Component Analysis (PCA) corresponds to the analysis of $\hat{C}_N(\tau)$: it exhibits 1 dominant singular value, with the corresponding singular vector close to the direction of fast variables, because the large fluctuations of the fast modes dominates. The other two singular vectors of $\hat{C}_N(\tau)$, which are necessarily, in PCA, orthogonal to the leading mode, are not the directions of slow variables at most regions. The number of the dominant singular values of $\hat{\Gamma}_{D-d}^l$ is 1 and its corresponding singular vector correctly estimates the direction of the fast variable. Finally, the number of the dominant singular values of $\hat{\Lambda}_d^l$ is 2, equal to the correct dimension of the slow
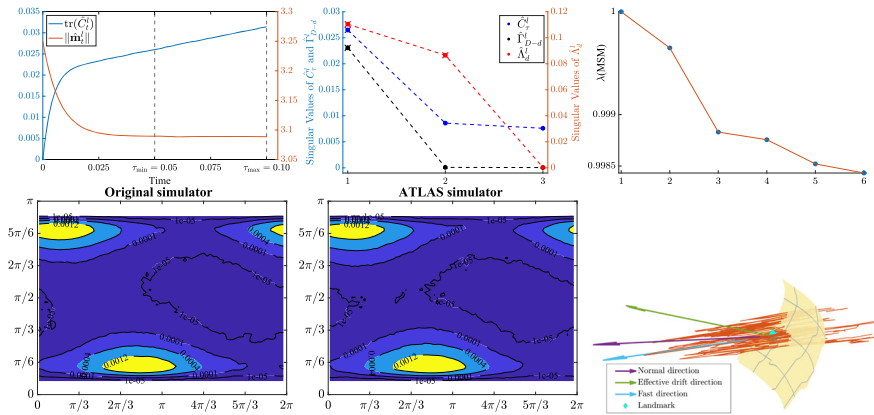
Fig. 7 Pinched sphere: **Top right:** plot of trace of empirical covariance and norm of empirical mean, $\text{tr}(\hat{C}_t^l)$ and $\|\hat{\mathbf{m}}_t^l\|$ verse time at one landmark. **Top center:** singular values of $\hat{C}_\tau^l$, $\hat{\Gamma}_{D-d}^l$ and $\hat{\Lambda}_d^l$ at one landmark, the dominant singular values are marked as cross labels. **Top left:** top six eigenvalues of the Markov transition matrix. **Bottom left:** the contour plot of kernel-fitted invariant distribution in the coordinate (ϕ, θ) from the trajectory of time length 10^6 generated by the original simulator. **Bottom center:** the contour plot of kernel-fitted invariant distribution in the coordinate (ϕ, θ) from the trajectory of time length 10^6 generated by the ATLAS simulator. **Bottom right:** A trajectory of time length 2.5τ starting from the marked landmark is simulated with the original simulator is shown. We depict a portion of the key estimated objects: the invariant manifold $\mathcal{M}_\epsilon$, drift at $\mathbf{z}_0$, fast direction, and normal direction. Note that the fast direction is far from the normal one, and the (Itô) drift is far from being tangent to $\mathcal{M}_\epsilon$, and also that $\mathcal{M}_\epsilon$ is locally a graph over the tangent plane at $\mathbf{z}_0$

manifold, and the span of their corresponding singular vectors correctly estimates the tangent space of the slow manifold. On average, it requires 1585 charts to fully describe the invariant manifold.

Identifying metastable states, and estimation of large-time properties of the process. We assume we do not know the metastable states, nor the number of such states. The top six eigenvalues of the Markov transition matrix (see Fig. 7) are real and they exhibit a clear spectral gap after the top two eigenvalues, which indicates, correctly, that this system has two metastable states. We simulate trajectories of time length 10^6 with the original simulator and with the ATLAS simulator, and from those we estimate the invariant distribution by using smoothed histograms with bins constructed in the latent coordinates (ϕ, θ), which parametrize the invariant manifold. We visualize them in Fig. 7: from the contour plot of both distributions they appear very close, and indeed in the L^1-norm of the difference of two densities (i.e., the total variation distance between the distributions) is 0.107 ± 0.009, and the L^2-norm of the difference of densities (a more robust but less stringent distance) is 0.0034 ± 0.0004.

Estimation of residence times. In the stage of estimating residence times, we generate a single long trajectory of time length 5×10^5 with the ATLAS simulator (this time length is much larger than the residence times to be estimated), from which we uniformly sample N_{IC} initial conditions that are in $S_{cyan} = \{\varphi_2 > 0.05\}$ and $S_{red} = \{\varphi_2 < -0.05\}$. Here one can use either original simulator or ATLAS simulator, however, the ATLAS simulator is much faster. In this example, we test the medium and large residence time by defining the boundary of the residence set as $\{\varphi_2 > 0.02$

and $\{\varphi_2 < -0.02\}$ in the first experiments, and, the metastable states M_1 and M_2 in the second experiment. We say that a point on invariant manifold is in a set if the closest landmark of the point is in the corresponding set (with this definition, the set is slightly different from the corresponding sublevel set or superlevel set of an eigenfunction).

C.2: Oscillating Half-Moons

Identifying the slow manifold and the effective stochastic dynamics. With the setting of parameter, our model is reversible since $a_1 = 0$ (Ge and Qian 2012). In the latent variable space, the fast variables are r_1 and u_i and the slow variable is θ. In the limit $\epsilon \to 0$, the fast variables relax to the equilibrium at $r_1 = 1$ and $u_i = 0$ so the slow manifold $\mathcal{M}_0^x$ in the latent variable is $r_1 = 1$, which is the unit circle. The "local" invariant distribution of the fast variable r_1 and u_i are

$$v(r_1|\theta) = \sqrt{\frac{b_1}{b_2^2 \pi}} \exp\left[-\frac{b_1}{b_2^2} (r_1 - 1)^2 \right], \qquad v(u_i|\theta) = \sqrt{\frac{b_3}{b_4^2 \pi}} \exp\left[-\frac{b_3}{b_4^2} (u_i)^2 \right]. \tag{C.9}$$

With the current parameter setup, there is very small probability that r_1 can go to the negative side but we will ignore the possibility in our analytical calculations.

In the example, the fast variables are nonlinearly coupled in the observed Cartesian coordinates, so the slow manifold $\mathcal{M}_0$ in Cartesian coordinates is not the image of $\mathcal{M}_0^x$ under the coordinate transformation. In the limit $\epsilon \to 0$, the landmark position in Cartesian coordinate for given θ is

$$\bar{z}_1(\theta) := \mathbb{E}(z_1|\theta) = \int_{-\infty}^{+\infty} r_1 \cos(r_1 + \theta - 1) v(r_1|\theta) dr_1$$

$$= \exp\left(-\frac{b_2^2}{4b_1} \right) \sqrt{1 + \left(\frac{b_2^2}{2b_1} \right)^2} \cos(\theta + \theta_s), \tag{C.10}$$

$$\bar{z}_2(\theta) := \mathbb{E}(z_2|\theta) = \int_{-\infty}^{+\infty} r_1 \sin(r_1 + \theta - 1) v(r_1|\theta) dr_1$$

$$= \exp\left(-\frac{b_2^2}{4b_1} \right) \sqrt{1 + \left(\frac{b_2^2}{2b_1} \right)^2} \sin(\theta + \theta_s), \tag{C.11}$$

$$\bar{z}_i(\theta) := \mathbb{E}(z_i|\theta) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} (r_1 + u_i) v(r_1|\theta) dr_1 \right) v(u_i|\theta) du_i = 1. \tag{C.12}$$

where the angle shift θ_s is $\theta_s = \arctan\left(\frac{b_2^2}{2b_1} \right)$. Then the slow manifold $\mathcal{M}_0$ is the circle

embedded in $(\bar{z}_1, \bar{z}_2, 1, \ldots, 1)$ with the radius $\bar{r} = \exp\left(-\frac{b_2^2}{4b_1} \right) \sqrt{1 + \left(\frac{b_2^2}{2b_1} \right)^2}$ and the

angle is shifted by θ_s compared to the standard angle in Cartesian coordinates. This shift

is due to the nonlinearity of the fast modes, in particularly their curvature. With current parameter setup, the radius is approximately $\bar{r} = 0.9925$ and θ_s is approximately 0.0153.

The distance of the point from the slow manifold $\mathcal{M}_0$, dist$(\mathbf{z}, \mathcal{M}_0)$
$= \sqrt{(\sqrt{z_1^2 + z_2^2} - \bar{r})^2 + \sum_{i=3}^{20}(z_i - 1)^2}$. In Cartesian coordinate, the effective dynamics of the first and second coordinate, $\bar{z}_1, \bar{z}_2$ are

$$
\mathrm{d}\begin{bmatrix} \bar{z}_1(\theta) \\ \bar{z}_2(\theta) \end{bmatrix} = \left((a_1 + a_2 \sin(2\theta) + a_3 \cos(\theta)) \begin{bmatrix} -\bar{z}_2(\theta) \\ \bar{z}_1(\theta) \end{bmatrix} - \frac{a_4^2}{2} \begin{bmatrix} \bar{z}_1(\theta) \\ \bar{z}_2(\theta) \end{bmatrix} \right) \mathrm{d}t
$$
$$
+ a_4 \begin{bmatrix} -\bar{z}_2(\theta) \\ \bar{z}_1(\theta) \end{bmatrix} \mathrm{d}W_t \tag{C.13}
$$

The effective dynamics on the other 18 coordinates has zero drift term and zero diffusion term.

Estimating the relevant timescale τ. In Fig. 8, $\mathrm{tr}(\hat{C}_t^l)$ reaches the linear regime at $t = 0.5$, however, the norm of the empirical mean $\|\hat{\mathbf{m}}_t^l\|$ behaves linearly only after $t = 1.0$ (this is true also of the first and the second coordinates which are also TICA coordinates, $\hat{m}_1^l, \hat{m}_2^l$). This is consistent with the large fast modes, and high curvature of the slow manifold; the training interval that we choose is $[\tau_{\min}, \tau_{\max}] = [1.0, 1.25]$, and $\tau = 1.0$.

Estimating dimension and tangent spaces of $\mathcal{M}_0$, and direction of the fast modes. We report in Fig. 8 the behavior of the singular values of $\hat{C}_t^l$, $\hat{\Gamma}_{D-d}^l$ and $\hat{\Lambda}_d^l$ at one landmark in descending order. The dominant singular values are visualized with cross markers. Although the number of the dominant singular value of covariance matrix $\hat{C}_\tau^l$ and the diffusivity matrix $\hat{\Lambda}^l$ are 1, the corresponding singular vector of $\hat{C}_\tau^l$ is not the slow direction, but close to fast direction, dooming a naïve approach based on local PCA. On the other hand, the dominant singular vector of $\hat{\Lambda}_d^l$ correctly estimates the tangent direction of the slow manifold. The number of the dominant singular values of covariance matrix $\hat{\Gamma}_{D-d}^l$ is 1, which is the correct number of fast variables. The scatter plot of the samples from paths of a burst, at $t = 0.5, 1.0, 1.25$ in the global TICA coordinates (z_1, z_2), visualizes how the data cloud is stretched out nonlinearly away from the slow manifold. Due to the relatively big curvature, the part of slow manifold that these data clouds covers at different time clearly show the nonlinear effects. A local linear approximation on dynamics and geometry might not be accurate at these scales, making the refinement of the landmark positions necessary, for one round of refinement as discussed in Sect. 3. This procedure significantly reduces the bias introduced by the local linear approximation. To further reduce the effect of the nonlinearity, we choose the effective timescale τ as lower bound of the training interval. Here the scale of separation is large enough, multiple rounds of refinement are not necessary. On average, it requires 65 charts to fully describe the invariant manifold.

Identifying metastable states, and estimation of large-time properties of the process. From the top six eigenvalues of the transition matrix of a Markov state model constructed from ATLAS, we note the significant gap between the second and third eigenvalues, which correctly indicates the system has two metastable
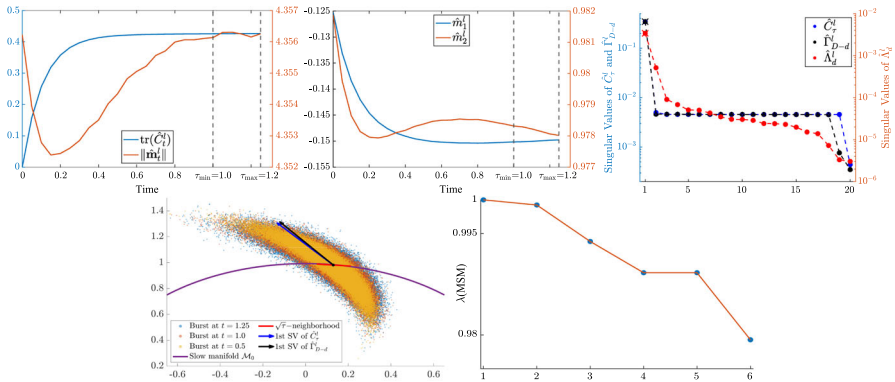
Fig. 8 Oscillating half-moons: **Top left:** plot of trace of empirical covariance and norm of empirical mean, $\mathrm{tr}(\hat{C}_t^l)$, $\|\hat{\mathbf{m}}_t^l\|$ verse time at one landmark. **Top center:** plot the first and second coordinates of the empirical mean $\hat{\mathbf{m}}_t^l$ verse time at one landmark. **Top right:** singular values in log scale of $\hat{C}_\tau^l$, $\hat{\Gamma}_{D-d}^l$ and $\hat{\Lambda}_d^l$ at one landmark, the dominant singular values are marked as cross labels. **Bottom left:** scatter of the burst in the (z_1, z_2) coordinate at $t = 0.5, 1.0, 1.25$, together with the slow manifold, $\sqrt{\tau}$-neighborhood and dominant singular vectors of $\hat{C}_\tau^l$, $\hat{\Gamma}_{D-d}^l$. **Bottom right:** top six eigenvalues of the Markov transition matrix

states. In this example, we explicitly provide the regions of the two metastable states, $M_{\mathrm{Left}} = \{\theta \in \left(-2\tan^{-1}(4+\sqrt{15}), -2\tan^{-1}(4-\sqrt{15})\right)\}$ and $M_{\mathrm{Right}} = \{\theta \in \left(-2\tan^{-1}(4-\sqrt{15}), -2\tan^{-1}(4+\sqrt{15})+2\pi\right)\}$ in the latent variable θ, that parametrizes the slow manifold.

We simulate trajectories of time 8×10^6 with the original simulator and with the ATLAS simulator, and from those we plot the invariant distributions in the latent variable θ. The standard bin width is $a_4\sqrt{\tau} = 0.06$ and we use it for the plot and calculation of the difference of the two distributions, which are very close (see Fig. 5). The L^1-norm of the difference of two approximated densities is 0.098 ± 0.006 and L^2-norm of the difference is 0.015 ± 0.001. Another observation is that albeit ATLAS is only expected to be accurate for the standard bin width, we also attempted other smaller bin widths, all the way down to 0.0019, which is much smaller than $a_4\sqrt{\tau}$, and observed no increase in the estimation error, thanks to the regularity of such distributions.

Estimation of residence times. In the stage of simulating residence time, we generate a single long trajectory of time 2.4×10^6 by the original simulator and uniformly sample N_{IC} initial conditions in each metastable state. The boundary of the residence set is the same as the region of the metastable state. In this example, we say the point is in the residence set if its latent variable θ is in the corresponding metastable state.

C.3: Butane

Governing equations for butane dynamics. We consider the overdamped Langevin dynamics of the butane molecule. The positions are denoted as $q^i \in \mathbb{R}^3$ for $1 \le i \le 4$.

To remove the rigid body motion invariant, we set

$$q^1 = \begin{bmatrix} x_1 \\ y_1 \\ 0 \end{bmatrix}, q^2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, q^3 = \begin{bmatrix} 0 \\ y_3 \\ 0 \end{bmatrix}, q^4 = \begin{bmatrix} x_4 \\ y_4 \\ z_4 \end{bmatrix}.$$

The potential energy is given as

$$V = \sum_{i=1}^{3} V_{\text{bond}}\left(\|q^{i+1} - q^i\|\right) + \sum_{i=1}^{2} V_{\text{angle}}\left(\theta_i\right) + V_{\text{torsion}}(\phi).$$

where θ_1, θ_2 are the angles formed by the three first atoms and the three last atoms respectively. ϕ is the dihedral angle, i.e, the angle between the plane on which the first three atoms lay and the plane on which the three last atoms lay. These potential functions are $V_{\text{bond}}(l) = \frac{k_2}{2}\left(l - l_{eq}\right)^2$, $V_{\text{angle}}(\theta) = \frac{k_3}{2}\left(\theta - \theta_{eq}\right)^2$ and $V_{\text{torsion}}(\phi) = c_1 \cos\phi + c_2 \cos^2\phi + c_3 \cos^3\phi$. The numerical values for the constants are those in Schappals et al. (2017). The overdamped Langevin dynamics on the state space $\mathbb{R}^6$ is

$$d\mathbf{z} = -\nabla V(\mathbf{z})dt + \sigma dW_t \tag{C.14}$$

where $\mathbf{z} = \begin{bmatrix} x_1 & y_1 & y_3 & x_4 & y_4 & z_4 \end{bmatrix}^T$ and the diffusion coefficient is $\sigma = \sqrt{2\beta^{-1}}$. The dihedral angle ϕ has the explicit form when $x_1 < 0$,

$$\cos(\phi) = \frac{(\vec{v}_{43} \times \vec{v}_{32}) \cdot (\vec{v}_{12} \times \vec{v}_{23})}{|\vec{v}_{43} \times \vec{v}_{32}| \cdot |\vec{v}_{12} \times \vec{v}_{23}|} = \frac{x_4}{\sqrt{x_4^2 + z_4^2}}, \tag{C.15}$$

where $\vec{v}_{ij} = q^j - q^i$. If we define the counterclockwise rotation as positive, then $x_4 = l\sin(\theta_{eq})\cos(\phi)$ and $z_4 = l\sin(\theta_{eq})\sin(\phi)$. If $x_1 < 0$, $y_3 > 0$, the explicit form of the potential is

$$\begin{aligned}
V(x_1, y_1, y_3, x_4, y_4, z_4) &= c_1\frac{x_4}{(x_4^2 + z_4^2)^{1/2}} + c_2\frac{x_4^2}{x_4^2 + z_4^2} + c_3\frac{x_4^3}{(x_4^2 + z_4^2)^{3/2}} \\
&+ \frac{1}{2}k_2\left(\left(\sqrt{x_1^2 + y_1^2} - l\right)^2 + \left(\sqrt{y_3^2} - l\right)^2 + \left(\sqrt{x_4^2 + z_4^2 + (y_3 - y_4)^2} - l\right)^2\right) \\
&+ \frac{1}{2}k_3\left(\left(\theta_{eq} - \arccos\left(\frac{y_1}{\sqrt{x_1^2 + y_1^2}}\right)\right)^2 \right. \\
&+ \left.\left(\theta_{eq} - \arccos\left(\frac{(y_3 - y_4)}{\sqrt{x_4^2 + (y_3 - y_4)^2 + z_4^2}}\right)\right)^2\right).
\end{aligned} \tag{C.16}$$

Identifying the slow manifold and the effective stochastic dynamics. The dihedral angle ϕ is usually chosen as the slow variable of the butane dynamics and we will have the slow manifold $\mathcal{M}_0$, which is a circle embedded in $\mathbb{R}^6$, if $x_1 > 0$, $y_3 < 0$,

$$\mathcal{M}_0 = \left\{ \left[-l\sin(\theta_{eq}),\ l\cos(\theta_{eq}),\ l,\ x_4,\ l - l\cos(\theta_{eq}),\ z_4 \right]^T \text{ with } x_4^2 + z_4^2 = l^2 \sin^2(\theta_{eq}) \right\}$$

The potential energy for the dihedral angle has three local minima at $\phi = -2\pi/3$ (bot-cis), $\phi = 0$(trans) and $\phi = 2\pi/3$ (top-cis). Butane dynamics is well known for its conformation isomerism and can be treated as the unimolecular reaction of three states. The trans conformer has lower energy than the cis conformer, so the trans state is more stable than the cis state. The distance of the point from the slow manifold $\mathcal{M}_0$ is calculated as follows,

$$\text{dist}(\mathbf{z}, \mathcal{M}_0)$$
$$= \sqrt{(x_1 + l\sin(\theta_{eq}))^2 + (y_1 - l\cos(\theta_{eq}))^2 + (y_3 - l)^2 + (y_4 + l\cos(\theta_{eq}) - l)^2 + \left(\sqrt{x_4^2 + z_4^2} - l\sin(\theta_{eq})\right)^2} \quad \text{(C.17)}$$

The proposed effective stochastic dynamics of the dihedral angle ϕ is

$$d\phi_t = -\nabla V_{\text{torsion}}(\phi_t)dt + \sqrt{2\beta^{-1}}\sigma(\phi_t)dW_t \quad \text{(C.18)}$$

As suggested in Legoll and Lelièvre (2012), the $\sigma(\phi_t)$ is $\sigma^2(\phi_t) = \mathbb{E}\left(|\nabla\phi|^2(\mathbf{y})|\phi(\mathbf{y}) = \phi_t\right)$. In this case, $\sigma(\phi_t)$ can be explicitly calculated here, $\sigma(\phi_t) = \frac{1}{l\sin(\theta_{eq})}$. Then in $\mathbb{R}^6$, the explicit form of the effective stochastic dynamics of the fourth and sixth coordinates, x_4, z_4 are

$$\partial_t x_4 = \left(-z_4^2 \frac{c_1\left(x_4^2 + z_4^2\right) + 2c_2 x_4\sqrt{x_4^2 + z_4^2} + 3c_3 x_4^2}{(x_4^2 + z_4^2)^{5/2}} - \frac{x_4}{\beta l^2 \sin^2(\theta_{eq})} \right) dt - \frac{z_4\sqrt{2\beta^{-1}}}{l\sin(\theta_{eq})}dW_t,$$

$$\partial_t z_4 = \left(x_4 z_4 \frac{c_1\left(x_4^2 + z_4^2\right) + 2c_2 x_4\sqrt{x_4^2 + z_4^2} + 3c_3 x_4^2}{(x_4^2 + z_4^2)^{5/2}} - \frac{z_4}{\beta l^2 \sin^2(\theta_{eq})} \right) dt + \frac{x_4\sqrt{2\beta^{-1}}}{l\sin(\theta_{eq})}dW_t.$$

$$\text{(C.19)}$$

Other variables, x_1, y_1, y_3, y_4 have no drift and no diffusion in the effective stochastic dynamics.

Estimating the relevant timescale τ. Based on the strength of the bond angle and the largest parameter in the torsion potential, the contribution from the bond and bond angle part will be quickly relaxed and the scale of separation is approximately a factor of 20, which is not very large, both in absolute terms and when compared to the other examples we considered. It is therefore necessary to perform multiple rounds of refinement to ensure the initial condition is close enough to invariant manifold. Initially, we use the time window $[\tau_{\min}, \tau_{\max}] = [4 \times 10^{-5}, 5 \times 10^{-5}]$ to learn the parameters in each landmark and proceeds with rounds of refinement until the relative differences of estimated parameters within 5% (this resulted in no more than 10 rounds), as discussed in Sect. 3. In Fig. 9, both $\text{tr}(\hat{C}_t^l)$ and $\|\hat{\mathbf{m}}_t^l\|$ reach the linear

regime at $t = 5 \times 10^{-6}$, as well as the fourth and sixth coordinates of $\hat{\mathbf{m}}_t^l$, which are TICA coordinates. We thus ran a last round of refinement with the true training interval $[1 \times 10^{-5}, 1.5 \times 10^{-5}]$ and $\tau = 1 \times 10^{-5}$.

Estimating dimension and tangent spaces of $\mathcal{M}_0$, and direction of the fast modes. We report in Fig. 9 the behavior of the singular values of $\hat{C}_\tau^l$, $\hat{\Gamma}_{D-d}^l$ and $\hat{\Lambda}_d^l$ at one landmark in descending order. In this example, the dominant singular vector of covariance matrix $\hat{C}_t^l$ matches with the one of the diffusivity matrix $\hat{\Lambda}_d^l$ and the first dominant singular vector of covariance matrix $\hat{\Gamma}_{D-d}^l$ is almost orthogonal with the slow direction. There are 5 dominant singular values for the covariance matrix $\hat{\Gamma}_{D-d}^l$, then it shows the fast variable has 5 dimensions. Similar to the oscillating half-moon example, the part of slow manifold that these data cloud covers clearly shows the nonlinear effect. Therefore, τ is chosen from the lower bound of the training interval and the special procedure to modify the landmark during the last round of refinement is necessary. On average, it requires 77 charts to fully describe the invariant manifold.

Identifying metastable states, and estimation of large-time properties of the process. From the top six eigenvalues of the transition matrix of a Markov state model constructed from ATLAS, we observe a significant gap between the third and fourth eigenvalues, correctly indicating that the system has three metastable states. As in the half-moon example, we express the metastable states in the latent (dihedral, in this case) angle ϕ: trans:=$\{\phi \in (-\frac{\pi}{3}, \frac{\pi}{3})\}$, top-cis:=$\{\phi \in (\frac{\pi}{3}, \pi)\}$ and bot-cis:=$\{\phi \in (-\pi, -\frac{\pi}{3})\}$. We simulate trajectories of time 5×10^2 with the original simulator and with the ATLAS simulator, and from those we plot both invariant distributions in the dihedral angle ϕ. The standard bin width is $\sigma\sqrt{\tau} = 0.07$ and we use it for the plot and calculation of the difference of two distributions. The L^1-norm of the difference of two approximated densities is 0.060 ± 0.013 and L^2-norm of the difference is 0.013 ± 0.003.

Estimation of residence times. In the stage of simulating residence time, we generate a single long trajectory of time 5×10^2 by the original simulator and uniformly sample N_{IC} initial conditions in each metastable state. The boundary of the residence set is the same as the region of the metastable states. In this example, we say the point is in the residence set if its dihedral angle ϕ is in the corresponding metastable state.

D: Error Analysis

The equations of the relative error of Euclidean norm of the estimated drift term $\hat{\mathbf{b}}^{\mathcal{A}}(\mathbf{z})$, estimated diffusivity matrix $\hat{\Lambda}^{\mathcal{A}}(\mathbf{z})$, and between the estimated and theoretical tangent space, are defined, respectively, as

$$\text{relErr}(\hat{\mathbf{b}}) = \frac{\|\hat{\mathbf{b}}^{\mathcal{A}}(\mathbf{z}) - \mathbf{b}(\mathbf{z})\|}{\|\mathbf{b}(\mathbf{z})\|}, \ \text{relErr}(\hat{\Lambda}) = \frac{\|\hat{\Lambda}_d^{\mathcal{A}}(\mathbf{z}) - \Lambda(\mathbf{z})\|}{\|\Lambda(\mathbf{z})\|}, \ \text{AbsErr}(\hat{T}_{\hat{\mathbf{z}}^l})$$
$$= \arccos\left(\left\|(\hat{U}_d^{l,\text{slw}})^T U_d^{l,\text{slw}}\right\|\right). \tag{D.1}$$

For the estimation error of the invariant manifold, as discussed above we compared points generated by long ATLAS trajectories, which lie on $\hat{\mathcal{M}}_\epsilon$ by definition (of $\hat{\mathcal{M}}_\epsilon$)
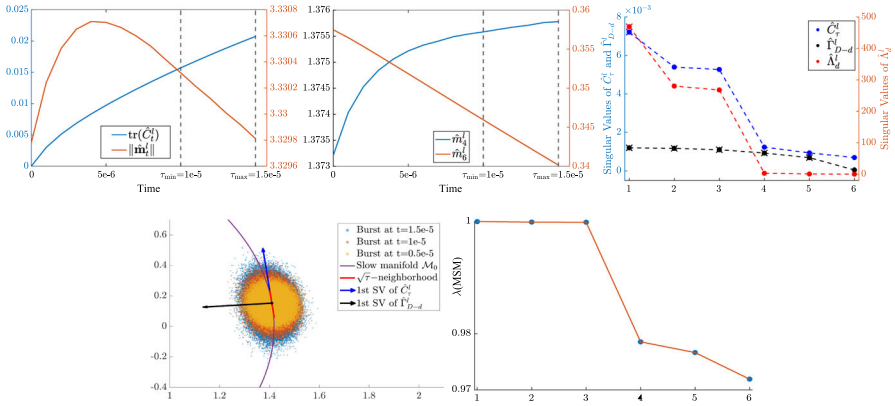
Fig. 9 butane: **Top left:** plot of trace of empirical covariance and norm of empirical mean, $\text{tr}(\hat{C}^l_t)$ and $\|\hat{\mathbf{m}}^l_t\|$ verse time at one landmark. **Top center:** plot the fourth and sixth coordinates of the empirical mean $\hat{\mathbf{m}}^l_t$ verse time at one landmark. **Top right:** singular values of $\hat{C}^l_\tau$, $\hat{\Gamma}^l_{D-d}$ and $\hat{\Lambda}^l_d$ at one landmark, the dominant singular values are marked as cross labels. **Bottom left:** scatter of the burst in the (x_4, z_4) coordinate at $t = 0.5 \times 10^{-5}, 1 \times 10^{-5}, 1.5 \times 10^{-5}$, together with the slow manifold, $\sqrt{\tau}$-neighborhood and dominant singular vectors of $\hat{C}^l_\tau$, $\hat{\Gamma}^l_{D-d}$. **Bottom right:** top six eigenvalues of the Markov transition matrix

with points on the slow manifold obtained by averaging the equations in the limit as $\epsilon \to 0$; this corresponds to the following:

- in the Pinched sphere example, from the calculations above we let $\text{AbsErr}(\hat{\mathcal{M}}_\epsilon) = |r - r^\star(\theta)|$.
- In the oscillating half-moon example, we use the distance of the point to the slow manifold,

$$\text{AbsErr}(\hat{\mathcal{M}}_\epsilon) = \sqrt{(\sqrt{z_1^2 + z_2^2} - \bar{r})^2 + \sum_{i=3}^{20}(z_i - 1)^2}$$

- In the butane example, we use the distance of the point to the slow manifold,

$$\text{AbsErr}(\hat{\mathcal{M}}_\epsilon)$$
$$= \sqrt{(x_1 + l\sin(\theta_{eq}))^2 + (y_1 - l\cos(\theta_{eq}))^2 + (y_3 - l)^2 + (y_4 + l\cos(\theta_{eq}) - l)^2 + \left(\sqrt{x_4^2 + z_4^2} - l\sin(\theta_{eq})\right)^2}.$$

References

Abourashchi, N., Veretennikov, A.Y.: On stochastic averaging and mixing. Theory Stoch. Process. **16**(1), 111–129 (2010)

Alexander, R., Giannakis, D.: Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. Physica D **409**, 132520 (2020)

Bakhtin, V., Kifer, Y.: Diffusion approximation for slow motion in fully coupled averaging. Probab. Theory Relat. Fields **129**(2), 157–181 (2004)

Berglund, N., Gentz, B.: Geometric singular perturbation theory for stochastic differential equations. J. Differ. Equ. **191**(1), 1–54 (2003)

Berglund, N., Gentz, B.: Noise-Induced Phenomena in Slow-Fast Dynamical Systems. Probability and its Applications. Springer, Berlin (2006)

Beygelzimer A., Kakade S., Langford J.: Cover trees for nearest neighbor. In: ICML, pp. 97–104 (2006)

Bicout, D.J., Szabo, A.: Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. Protein Sci. **9**(3), 452–465 (2000)

Bittracher, A., Banisch, R., Schütte, C.: Data-driven computation of molecular reaction coordinates. J. Chem. Phys. **149**(15), 154103 (2018)

Chen, M., Tang-Qing, Yu., Tuckerman, M.E.: Locating landmarks on high-dimensional free energy surfaces. Proc. Natl. Acad. Sci. USA **112**(11), 3235–3240 (2015)

Chiavazzo, E., Covino, R., Coifman, R.R., William Gear, C., Georgiou, A.S., Hummer, G., Kevrekidis, I.G.: Intrinsic map dynamics exploration for uncharted effective free-energy landscapes. Proc. Natl. Acad. Sci. USA **114**(28), E5494–E5503 (2017)

Coifman, R.R., Lafon, S., Lee, A.B., Mauro Maggioni, B., Nadler, F.W., Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proc. Natl. Acad. Sci. USA **102**(21), 7426–7431 (2005)

Coifman, R.R., Kevrekidis, I.G., Lafon, S., Maggioni, M., Nadler, B.: Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. Multiscale Model. Simul. **7**(2), 842–864 (2008)

Crosskey, M.C., Maggioni, M.: Atlas: a geometric approach to learning high-dimensional stochastic systems near manifolds. Multiscale Model. Simul. **15**(1), 110–156 (2017)

Dietrich, F., Makeev, A., Kevrekidis, G., Evangelou, N., Bertalan, T., Reich, S., Kevrekidis, I.G.: Learning effective stochastic differential equations from microscopic simulations: combining stochastic numerics and deep learning (2021)

Dsilva, C.J., Ronen Talmon, C., Gear, W., Coifman, R.R., Kevrekidis, I.G.: Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems. SIAM J. Appl. Dyn. Syst. **15**(3), 1327–1351 (2016)

Freidlin, M.I., Szucs, J., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Grundlehren der mathematischen Wissenschaften. Springer, Berlin (2012)

Frewen, T.A., Hummer, G., Kevrekidis, I.G.: Exploration of effective potential landscapes using coarse reverse integration. J. Chem. Phys. **131**(13), 134104 (2009)

Frewen, T.A., Hummer, G., Kevrekidis, I.G.: Exploration of effective potential landscapes using coarse reverse integration. J. Chem. Phys. **131**(13), 134104 (2009)

Gardiner, C.: Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer Series in Synergetics, Springer, Berlin (2009)

Ge, H., Qian, H.: Landscapes of non-gradient dynamics without detailed balance: stable limit cycles and multiple attractors. Chaos An Interdiscip. J. Nonlinear Sci. **22**(2), 023140 (2012)

Givon, D.: Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems. Multiscale Model. Simul. **6**(2), 577–594 (2007)

Givon, D., Kupferman, R., Stuart, A.: Extracting macroscopic dynamics: model problems and algorithms. Nonlinearity **17**(6), R55–R127 (2004)

Givon, D., Kevrekidis, I.G., Kupferman, R.: Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. Commun. Math. Sci. **4**(4), 707–729 (2006)

Hartmann, C., Neureither, L., Sharma, U.: Coarse graining of nonreversible stochastic differential equations: quantitative results and connections to averaging. SIAM J. Math. Anal. **52**(3), 2689–2733 (2020)

Has'minskii, R.Z.: On stochastic processes defined by differential equations with a small parameter. Theory Probab. Appl. **11**(2), 211–228 (1966)

Holmes, P., Lumley, J.L., Berkooz, G., Rowley, C.W.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge Monographs on Mechanics, 2nd edn. Cambridge University Press, Cambridge (2012)

Husic, B.E., Pande, V.S.: Markov state models: from an art to a science. J. Am. Chem. Soc. **140**(7), 2386–2396 (2018). (**PMID: 29323881**)

Jiang, D.-Q., Qian, M., Qian, M.-P.: Mathematical Theory of Nonequilibrium Steady States, Volume 1833 of Lecture Notes in Mathematics. Springer, Berlin, (2004). On the frontier of probability and dynamical systems

Jones, C.K.R.T.: Geometric singular perturbation theory. In: Dynamical systems (Montecatini Terme, 1994), Volume 1609 of Lecture Notes in Mathematics, pp. 44–118. Springer, Berlin (1995)

Khas'minskii, R.Z.: A limit theorem for the solutions of differential equations with random right-hand sides. Theory Probab. Appl. **11**(3), 390–406 (1966)

Khasminskii, R.Z., Yin, G.: On averaging principles: an asymptotic expansion approach. SIAM J. Math. Anal. **35**(6), 1534–1560 (2004)

Kifer, Y.: Another proof of the averaging principle for fully coupled dynamical systems with hyperbolic fast motions. Discrete Contin. Dyn. Syst. A **13**(5), 1187–1201 (2005)

Kim, S.B., Dsilva, C.J., Kevrekidis, I.G., Debenedetti, P.G.: Systematic characterization of protein folding pathways using diffusion maps: application to trp-cage miniprotein. J. Chem. Phys. **142**(8), 085101 (2015)

Klus, S., Nüske, F., Koltai, P., Hao, W., Kevrekidis, I., Schütte, C., Noé, F.: Data-driven model reduction and transfer operator approximation. J. Nonlinear Sci. **28**(3), 985–1010 (2018)

Kuehn, C.: Multiple time Scale Dynamics, Volume 191 of Applied Mathematical Sciences. Springer, Berlin (2015)

Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: Dynamic Mode Decomposition. SIAM (2016)

Legoll, F., Lelièvre, T.: Effective dynamics using conditional expectations. Nonlinearity **23**(9), 2131–2163 (2010)

Legoll, F., Lelièvre, T.: Some remarks on free energy and coarse-graining. In Numerical analysis of multiscale computations, Volume 82 of Lecture Notes Computing Science Engineering, pp. 279–329. Springer (2012)

Leimkuhler, B., Matthews, C.: Molecular Dynamics: With Deterministic and Stochastic Numerical Methods. Interdisciplinary Applied Mathematics. Springer, Berlin (2015)

Li, X.-M.: An averaging principle for a completely integrable stochastic Hamiltonian system. Nonlinearity **21**(4), 803–822 (2008)

Liberty, E., Woolfe, F., Martinsson, P.-G., Rokhlin, V., Tygert, M.: Randomized algorithms for the low-rank approximation of matrices. Proc. Natl. Acad. Sci. USA **104**(51), 20167–20172 (2007)

Little, A.V., Maggioni, M., Rosasco, L.: Multiscale geometric methods for data sets i: multiscale svd, noise and curvature. Appl. Comput. Harmon. Anal. **43**(3), 504–567 (2017)

Liu, D.: Strong convergence of principle of averaging for multiscale stochastic dynamical systems. Commun. Math. Sci. **8**(4), 999–1020 (2010)

Liu, P., Siettos, C.I., Gear, C.W., Kevrekidis, I.G.: Equation-free model reduction in agent-based computations: coarse-grained bifurcation and variable-free rare event analysis. Math. Model. Nat. Phenom. **10**(3), 71–90 (2015)

Maria Bruna, S., Chapman, J., Smith, M.J.: Model reduction for slow-fast stochastic systems with metastable behaviour. J. Chem. Phys. **140**(17), 174107 (2014)

Molgedey, L., Schuster, H.G.: Separation of a mixture of independent signals using time delayed correlations. Phys. Rev. Lett. **72**, 3634–3637 (1994)

Pavliotis, G.A., Stuart, A.: Multiscale Methods: Averaging and Homogenization. Texts in Applied Mathematics, Springer, Berlin (2008)

Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., Noé, F.: Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. **139**(1), 015102 (2013)

Röckner, M., Sun, X., Xie, L.: Strong and weak convergence in the averaging principle for sdes with hölder coefficients (2019)

Rohrdanz, M.A., Zheng, W., Maggioni, M., Clementi, C.: Determination of reaction coordinates via locally scaled diffusion map. J. Chem. Phys. **134**(12), 124116 (2011)

Rohrdanz, M.A., Zheng, W., Clementi, C.: Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. Annu. Rev. Phys. Chem. **64**(1), 295–316 (2013)

Rowley, C.W., Mezič, I., Bagheri, S., Schlatter, P., Henningson, D.S.: Spectral analysis of nonlinear flows. J. Fluid Mech. **641**, 115–127 (2009)

Schappals, M., Mecklenfeld, A., Kröger, L., Botan, V., Köster, A., Stephan, S., García, E.J., Rutkai, G., Raabe, G., Klein, P., Leonhard, K., Glass, C.W., Lenhard, J., Vrabec, J., Hasse, H.: Round robin study: molecular simulation of thermodynamic properties from models with internal degrees of freedom. J. Chem. Theory Comput. **13**(9), 4270–4280 (2017)

Singer, A., Erban, R., Kevrekidis, I.G., Coifman, R.R.: Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. Proc. Natl. Acad. Sci. USA **106**(38), 16090–16095 (2009)

Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G.: Plumed 2: new feathers for an old bird. Comput. Phys. Commun. **185**(2), 604–613 (2014)

van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, Volume 888 of Lecture Notes in Mathematics. North-Holland Publishing Co. (1981)

Vanden-Eijnden, E.: Numerical techniques for multi-scale dynamical systems with stochastic effects. Commun. Math. Sci. **1** (2003)

Vershynin, R.: Deviations of Random Matrices and Geometric Consequences, pp. 216–231. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2018)

Wechselberger, M.: Geometric Singular Perturbation Theory Beyond the Standard form, Volume 6 of Frontiers in Applied Dynamical Systems: Reviews and Tutorials. Springer, Cham (2020)

Weinan, E., Vanden-Eijnden, E.: Metastability, conformation dynamics, and transition pathways in complex systems. In: Multiscale Modelling and Simulation, vol. 39 of Lecture Notes Computing Science Engineering, pp. 35–68. Springer (2004)

Weinan, E., Liu, D., Vanden-Eijnden, E.: Analysis of multiscale methods for stochastic differential equations. Commun. Pure Appl. Math. **58**(11), 1544–1585 (2005)

Yu, A., Veretennikov: On the averaging principle for systems of stochastic differential equations. Math. USSR Sb. **69**(1), 271–284 (1991)

Zhang, B., Hongbo, F., Wan, L., Liu, J.: Weak order in averaging principle for stochastic differential equations with jumps. Adv. Differ. Equ. **1**, 2018 (2018)

Zheng, W., Rohrdanz, M.A., Clementi, C.: Rapid exploration of configuration space with diffusion-map-directed molecular dynamics. J. Phys. Chem. B **117**(42), 12769–12776 (2013)