Variational Characterization and Identification of Reaction Coordinates in Stochastic Systems

Andreas Bittracher[1], Mattes Mollenhauer[1], Péter Koltai[1], and Christof Schütte[1,2]

[1]Department of Mathematics and Computer Science, Freie Universität Berlin, Germany
[2]Zuse Institute Berlin, Germany

Abstract

Reaction coordinates are indicators of hidden, low-dimensional mechanisms that govern the long-term behavior of high-dimensional stochastic systems. We present a novel, very general characterization of these coordinates and provide conditions for their existence. We show that these conditions are fulfilled for slow-fast systems, metastable systems, and other systems with known good reaction coordinates. Further, we formulate these conditions as a variational principle, i.e., define a loss function whose minimizers are optimal reaction coordinates. Remarkably, the numerical effort required to evaluate the loss function scales only with the complexity of the underlying, low-dimensional mechanism, and not with that of the full system. In summary, we provide the theoretical foundation for an efficient computation of reaction coordinates via modern machine learning techniques.

1. Introduction

We consider high-dimensional, time- and space-continuous Markov processes. Such processes are used to model molecular dynamical systems, general interacting particle systems, and other systems that consist of many coupled but memory-free degrees of freedom. The reason researchers are interested in such systems is however often not so much the microscopic dynamics itself, but rather the phenomenon that, over long time scales, the system often exhibits much more regularity and much less complexity than the sheer number of degrees of freedom would actually allow for. An example for such a phenomenon is the famous Levianthal paradoxon for protein folding [50]: reaching the correct native state of an average-sized protein by randomly jumping between all combinatorically possible conformations (as determined by the number of backbone bond angles) would take until the heat death of the universe. In reality, however, the native state is typically reached within milliseconds. The explanation is that naturally-occurring proteins possess local (side chain) interactions that greatly restrict the backbone movement and guide the folding along low-dimensional "trenches" or "funnels". The long-term behavior is therefore determined by a latent, low-dimensional mechanism that is the result from certain restrictions of the microscopic dynamics.

This article is dedicated to the analysis of these restrictions, and to their systematic exploitation in order to identify the latent mechanism. More specifically, we will identify so-called *reaction coordinates*, that is, a low-dimensional observable of the full system that acts as a proxy for the latent mechanism. In a first step, we characterize the type of system that possesses such reaction coordinates, i.e., we identify sufficient restrictions on the microscopic dynamics. These restrictions

will turn out to be a generalization of the well-known *lumpability* condition that is required for the successful aggregation of discrete Markov chains [22]. Moreover, we will see that several types of systems for which a well-known slow mechanism exists, such as slow-fast systems or systems with a generator spectral gap, exhibit our condition.

We then present a systematic approach to identify the reaction coordinates in a bottom-up manner, assuming only knowledge about the microscopic dynamics that can be obtained by short numerical simulations. We formulate our search for optimal reaction coordinates in a variational manner, by defining a loss functional whose global minima are reaction coordinates that allow the reproduction of the system's rate-limiting sub-processes with minimal error. This loss function also can be seen as a "quality measure" for arbitrary reaction coordinates, which can be useful when comparing heuristically-defined reaction coordinates.

Finally, we show how mere knowledge of the existence of a good reaction coordinate reduces the data requirement for its numerical computation. The intuitive reason is that systems with a strong latent mechanism (of which we require no further regularity, such as linearity) express less "dynamical variety" than an arbirtrarily complex system, so less sampling data is required to "capture" this variety. To take advantage of that idea, we will derive a re-formulation of the aforementioned loss function that can be approximated by randomly choosing starting points and computing short "bursts" of numerical trajectories. For one, this yields a dimension-independent convergence rate, due to the Monte Carlo nature of this scheme. For the other, we will see that the more dominant the underlying mechanism, the smaller the prefactor of the rate.

In order to apply our proposed approach to truly high-dimensional, realistic problems, certain numerical challenges have yet to be overcome, which will be be discussed in the appropriate sections. To demonstrate our variational principle, we therefore limit ourselves to the examination of a two and three-dimensional academical slow-fast-system with one-dimensional fast component, as well as a two-dimensional metastable system. We demonstrate that the effort required to approximate the loss function indeed scales with the degree of lumpability, and that the effort is not directly connected to the dimensionality of the system.

This article is structured as follows: Section 2 compares our approach to related work. Section 3 introduces our characterization of systems with good reaction coordinates, and confirms the compatibility of that characterization with established concepts. Section 4 derives a variational principle for reaction coordinates in form of a loss function whose minimizer yields (almost) optimal reaction coordinates. Section 5 shows that approximating this loss function via a Monte Carlo method requires less dynamical samples if the system indeed possesses good reaction coordinates. Section 6 demonstrates the variational principle and approximation of the loss function by two synthetic examples. Section 7 contains the conclusions, and an outlook on future work.

2. Related work

Characterization of reaction coordinates

Existing definitions of "good" reaction coordinates were mostly derived with applications in computational physics and chemistry in mind. In that domain, one of the most popular reaction coordinate with a general dynamical interpretation is the *committor function*, which for some reactant and some product state indicates the probability to hit the latter before hitting the former [13, 29]. However, the committor can only be expected to describe the system's rate-limiting processes well for sensible choices of the reactant and product (which obviously requires a priori macro-scale knowledge of the system).

The dominant eigenfunctions of a Markovian system's transfer operator (or equivalently its Fokker Planck operator) decompose the system into linearly independent sub-processes, which equilibrate with a rate determined by the associated eigenvalue [44]. Hence, the dominant eigenfunctions have been used as reaction coordinates [31]. It has however been demonstrated that the dominant eigenfunctions themselves can be reduced further, if the associated sub-processes are in some way "nonlinearly dependent" [4]. An example of this situation will be given later in the text.

A reaction coordinate composed of dominant eigenfunctions is therefore in general not optimal in terms of its dimensionality.

The TICA (time-lagged independent component analysis) method constructs reaction coordinates as those linear combinations of the original degrees of freedom with the highest autocorrelation [33, 37]. Assuming the system is reversible, the TICA coordinates are the aforementioned eigenfunctions of the transfer operator projected onto the linear basis functions $\psi_i(x) = x_i$ [25]. They therefore suffer from the same non-optimality as the eigenfunction reaction coordinates, and in addition from non-optimality due to the overhead of linear approximation.

The most frequently analyzed special case of timescale-separated systems are *slow-fast systems* (see [36] for a text book introduction). They are characterized by the existence of a coordinate transformation such that the new coordinates can be subdivided into one quickly and one slowly moving part, and the two parts are approximately decoupled. The slow coordinates, which form a parametrization of the system's slow manifold [48], can be considered, and has been used as a good reaction coordinate of the system [46, 16]. We will see later that our characterization encompasses that of slow variables.

The reaction coordinates conceptually most alike to the definition presented in this article are the *transition manifold reaction coordinates*, proposed by some of the authors in [4] and further refined in [5]. These works characterize good reaction coordinates as a parametrization of a low-dimensional manifold in a certain function space, around which the system's transition densities cluster with progressing equilibration.

Computational strategies

In their respective original publications, most of the aforementioned reaction coordinates come with a proposed numerical scheme for their computation. The committor function, which satisfies a backward Kolmogorov equation [13], can be computed using numerical PDE solvers (although this was never proposed as a practical scheme and a vastly more efficient scheme was proposed soon-after [29]). Reaction coordinates based on transfer operator eigenfunctions (including TICA) can be computed by an eigendecomposition of a suitable discretization of that operator [9, 43, 11, 37]. Approaches that characterize reaction coordinates as parmetrization of some manifold use unsupervised manifold learning methods such as diffusion maps to learn the variables in an equation-free manner [46, 16, 4, 3].

Over time, deep-rooted relationships between the different reaction coordinates, as well as extensions and generalizations were discovered (see [24] for a partial overview), which led to alternative and more efficient schemes for their computation. While a comprehensive listing would go beyond the scope of this article, we would like to point out an emerging trend in this development, namely the formulation of a variational principle for the respective reaction coordinate. There now exist variational approaches for the committor function [23], the TICA coordinate [49], and the transfer operator eigenfunctions [30]. The driving force behind this trend is of course the desire to profit from the spectacular performance that modern deep learning and neural network-based methods have demonstrated with regard to their generalization power, robustness to overfitting and seeming immunity to the curse of dimensionality [1].

Notably missing from the above list is however a variational principle for manifold-based reaction coordinates. As mentioned before, the characterization presented in this article generalizes the transition manifold, which in turn generalizes the slow manifold, so it can be seen as a completion in that regard. The variational approach then offers the additional advantage of yielding a closed form of the reaction coordinate (in some finite-dimensional ansatz space), unlike the aforementioned geometric manifold learning algorithms, which output only discrete point-evaluations of the reaction coordinate.

Dynamical sampling

Our specific strategy to approximate the loss function from random "bursts" [15] of the dynamics shows parallels to several other computational techniques that implicitly exploit some form of

hidden regularity of the problem.

In a recent publication [6], some of the authors applied a discrete version of that strategy to time- and space-*discrete* Markov chains. There it was postulated that the long-term behavior of a large (i.e., many-state) Markov chain is essentially determined by transitions between certain *aggregates* of these states. We were able to show that the aggregates and the transition probabilities between them could be discovered from a vastly undersampled version of the transition matrix of the original chain (again obtained through random simulation bursts).

The fundamental idea behind both our discrete and continuous strategies is heavily inspired by the field of compressive sensing, see [14] for an introduction. The impressive feat of compressive sensing is its ability re-construct a signal (a high-dimensional vector) from far less samples than the Nyquist–Shannon sampling theorem would actually demand, i.e., to solve vastly underdetermined linear equations. The necessary assumption here is that the system is *sparse* in some domain (for example the frequency domain), though the location or precise number of the sparse entries does not need to be known. In a way, the dominance of a Markov process by a single low-dimensional mechanism can be interpreted as a sort of "non-linear dynamical sparsity", as there are infinitely many other mechanisms that could, but do not, influence the process.

Finally we would like to point out that the apparent similarity of our sampling strategy to randomized matrix low-rank approximation techniques [19] like the Nyström method [12] or randomized feature approximation [38] is rather superficial. While for these techniques a (nearly) low-rank structure of the target matrix is necessary to achieve low approximation error, they do not interpret this low rank as an underlying structure "generating" the matrix. Indeed, in [47] it has been argued that an (approximate) low rank is a generic property of large data matrices, and that the attribution of that rank to some "physical reason" is in general not possible. Consequently, while low-rank matrix approximation techniques offer substantial data reduction for generic large-scale data matrices, they cannot match the performance of subsampling techniques that exploit specific generating structures.

3. Characterization of good reaction coordinates

3.1. Definition of the dynamics and fundamental assumptions

Let $\mathbb{X} \subset \mathbb{R}^n$ be a Lebesgue-measurable set (the *state space*) and $(X_t)_{t\in\mathbb{R}^+}$, or short (X_t), be a time- and space-continuous Markov process on $\mathbb{X}$. Let $P^t : \mathbb{X} \times \mathscr{B} \to [0,1]$ denote the *transition probability function* of (X_t), where $\mathscr{B}$ is the Borel σ-algebra on $\mathbb{X}$, i.e.,

$$P^t[x, B] = \mathrm{Prob}[X_{t_0+t} \in B \mid X_{t_0} = x] \quad \text{for all } t_0 \geq 0.$$

For any $t > 0$ and $x \in \mathbb{X}$, $P^t(x, \cdot)$ is a probability measure on $\mathscr{B}$, and $P^t(\cdot, B)$ is a $\mathscr{B}$-measurable function for any $B \in \mathscr{B}$ [40]. Moreover, let the process be ergodic, such that a unique stationary measure $\mu : \mathscr{B} \to \mathbb{R}_0^+$ exists. We require μ to be absolutely continuous with respect to the Lebesgue measure, i.e., there exists a density $\pi : \mathbb{X} \to \mathbb{R}^+$ such that

$$\mu(B) = \int_B \pi(x)\, dx.$$

Moreover, we require that π is continuous and strictly positive.

We also require the $P^t(x, \cdot)$ to be absolutely continuous with respect to the Lebesgue measure. Thus, we may assume that there exists a family of functions $p^t : \mathbb{X} \times \mathbb{X} \to \mathbb{R}^+$ such that

$$P^t(x, B) = \int_B p^t(x, y)\, \mathrm{d}y \quad \text{for all } \tau > 0,\ x \in \mathbb{X},\ B \in \mathscr{B}. \tag{1}$$

Many classes of Markov processes are absolutely continuous, including Itô diffusions with smooth coefficients [26]. Also, we assume that the system is reversible with respect to π, i.e., the *detailed balance equation* holds:

$$p^t(x, y)\pi(x) = p^t(y, x)\pi(y) \quad \text{for all } x, y \in \mathbb{X},\ t \in \mathbb{R}^+. \tag{2}$$

The existence of good reaction coordinates will be determined by specific properties of the function p^t, hence we now examine its nature more closely. As a function of the second argument, $p^t(x,\cdot) \in L^1$ is the time-t *transition density function* of (X_t), i.e.

$$p^t(x,\cdot) = \mathrm{Law}\left(X_{t_0+t} \;\middle|\; X_{t_0} = x\right) \quad \text{for all } t_0 \geq 0.$$

On the other hand, $p^t(\cdot, y)$ as a function of the first argument is harder to interpret, and discussed less in the literature of stochastic processes. Let $L^1_\mu(\mathbb{X})$ be the space equipped with the norm

$$\|f\|_{L^1_\mu} := \int_{\mathbb{X}} f(x)\, d\mu(x).$$

We then have $p^t(\cdot, y) \in L^1_\mu$, since, by reversibility of X_t,

$$\int_{\mathbb{X}} p^t(x,y)\pi(x)\, dx = \int_{\mathbb{X}} p^t(y,x)\pi(y)\, dx \leq \|\pi\|_\infty \|p^t(y,\cdot)\|_{L^1} < \infty.$$

To distinguish it from the transition density, we call $p^t(\cdot, y)$ the time-t *transition observable* of y.

As a function of *two* arguments, we call $p^t : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ the time-t *transition kernel* of (X_t). It can be interpreted as an element of the space of functions $L^1_{\mu\times\lambda}(\mathbb{X}^2)$, where $\mu \times \lambda$ is the product measure on the space $\mathbb{X}^2$ given by the invariant measure μ and the Lebesgue measure λ on $\mathbb{X}$. For simplicity, we use the shorthand notation

$$\mathbb{K} := L^1_{\mu\times\lambda}(\mathbb{X}^2).$$

Note that by Fubini–Tonelli, we have

$$\|p(*,\cdot)\|_{\mathbb{K}} := \left\| \|p(*,\cdot)\|_{L^1(\mathbb{X})} \right\|_{L^1_\mu(\mathbb{X})} \tag{3}$$

as the norm on $\mathbb{K}$, where in (3) the inner norm applies to the argument "$\cdot$", and the outer norm applies to the argument "$*$" (this will be a convention from now on).

3.2. Lumpability and deflatability

We will now introduce two seemingly different conditions for a system/reaction coordinate pair. Each condition may individually be taken as a definition for what a good reaction coordinate is. It will turn out, however, that the two conditions are equivalent to each other for reversible systems, so a good reaction coordinate with respect to one condition is a good reaction coordinate with respect to the other.

Let $r < n$, and $\mathbb{Z} \subset \mathbb{R}^r$ a domain. Any function $\xi \in C(\mathbb{X}, \mathbb{Z})$ is called a *reaction coordinate*. We denote the z-level set of ξ by

$$\Sigma_\xi(z) := \{x \in \mathbb{X} \mid \xi(x) = z\}.$$

Further, for later use, we denote the μ_z the marginal stationary measure on $\Sigma_\xi(z)$, defined by

$$\int_{\Sigma_\xi(z)} f(x)\, d\mu_z(x) = \int_{\Sigma_\xi(z)} f(x)\pi(x) \det\left(\nabla\xi(x)^\intercal \nabla\xi(x)\right)^{-1/2} dH_{n-r}(x),$$

where H_d denotes the d-dimensional Hausdorff measure. By $|\mathbb{Z}|$, we denote the Lebesgue-measure of $\mathbb{Z}$.

Lumpability

The first condition is as follows:

Definition 3.1 (Lumpability). *If there exists a domain $\mathbb{Z} \subset \mathbb{R}^r$, a continuous function $\xi : \mathbb{X} \to \mathbb{Z}$, and a family of time-parametrized functions $p_L^t : \mathbb{Z} \times \mathbb{X} \to \mathbb{R}^+$ and a lag time $\tau > 0$ such that*

$$\frac{1}{|\mathbb{Z}|} \left\| p^t(*, \cdot) - p_L^t(\xi(*), \cdot) \right\|_{\mathbb{K}} \leq \varepsilon \tag{L}$$

*is fulfilled for $t \geq \tau$, we say the system is ε-*lumpable *with respect to ξ.*

In words, lumpability means that for sufficiently large t, the transition densities $p^t(x, \cdot)$, i.e., the probabilities to transition out of x, depend essentially only on the value $\xi(x)$ of the reaction coordinate at x, and not on the precise location of x on the $\xi(x)$-level set of ξ. Again, this is a reasonable property for a reaction coordinate that is supposed to describe the effective dynamics of X_t.

Remark 3.2. Our definition of lumpability can be seen as a continuous version of the lumpability condition for discrete Markov chains, originally formulated by Kemeny and Snell [22]. There it has been shown that the discrete version of lumpability is a necessary and sufficient condition for the Markov chain to be "compressible" into a Markov chain between certain *aggregates* of the original chain. In general, this compression however leads to a loss of information, i.e., restoration of the original chain is in general not possible under the lumpability assumption alone.

Deflatability

The second condition we want to discuss is:

Definition 3.3 (Deflatability). *If there exists a domain $\mathbb{Z} \subset \mathbb{R}^r$, a continuous function $\xi : \mathbb{X} \to \mathbb{Z}$, a family of time-parametrized functions $p_D^t : \mathbb{X} \times \mathbb{Z} \to \mathbb{R}^+$ and a lag time $\tau > 0$ such that*

$$\frac{1}{|\mathbb{Z}|} \left\| p^t(*, \cdot) - p_D^t(*, \xi(\cdot))\pi(\cdot) \right\|_{\mathbb{K}} \leq \varepsilon \tag{D}$$

*is fulfilled for $t \geq \tau$, we say the system is ε-*deflatable *with respect to ξ.*

Here the intuition is that for t large enough, the transition observable $p^t(*, y)$, i.e., the probabilities to transition *to* y, is effectively defined by 1) the term $p_D^t(*, \xi(y))$, i.e. the probability density to transition to the $\xi(y)$-level set of ξ, and 2) the value $\pi(y)$, which crucially does not depend on the starting point.

Remark 3.4. The above notion of deflatability also has a time- and space-discrete equivalent for discrete Markov chains, which was defined by some of the authors in [6]. There it was shown that for Markov chains fulfilling both the lumpability and deflatability condition, a "compressed" Markov chain between certain aggregates of the original states exists, and that the full Markov chain can be (approximately) restored from the compressed chain.

Remark 3.5. It should be noted that, for large enough lag time, every uniformly ergodic system is trivially lumpable and deflatable since

$$\sup_{x \in \mathbb{X}} \left\| P^t(x, \cdot) - \mu(\cdot) \right\|_{TV} = \sup_{x \in \mathbb{X}} \frac{1}{2} \left\| p^t(x, \cdot) - \pi(\cdot) \right\|_{L^1} \to 0 \tag{4}$$

as $t \to 0$ (see for example [39, Section 3.3 together with Proposition 3(f)]. Hence, choosing $p_L^t(z, \cdot) = \pi$ and $p_D^t(z, \cdot) = 1$ in the above definitions will give lumpability and deflatability with respect to any constant reaction coordinate since then p_L^t and p_D^t are independent of x. Likewise, every system is trivially lumpable and deflatable, for any tolerance and lag time, with respect to the trivial n-dimensional reaction coordinate $\xi(x) = x$, since choosing $p_L^t(x, \cdot) = p^t(x, \cdot)$ and $p_D^t(\cdot, y) = p^t(\cdot, y)/\pi(y)$ fulfils (L) and (D) with tolerance zero. We emphasize that in this paper we specifically care about systems which are lumpable/deflatable with respect to intermediate lag

times τ that are much smaller than the equilibration time scale of the system, as well as small dimensions r and tolerances ε.

Moreover, a system may be lumpable/deflatable with respect to more than one non-trivial combination of ε, τ and r. In cases where no clear time scale separation exists in the full system, a balance has to be struck between the achievable approximation error of a reduced model built using ξ (acceptable ε), the time scale above which the reduced model is valid (choice of τ), and the dimension of the reduced model (choice of r).

Optimizing the choice of r and τ are however not subject of this paper. We will later consider r and τ to be fixed, and search for corresponding "optimal" reaction coordinates, i.e. an r-dimensional ξ for which (L) and/or (D) are fulfilled for the smallest possible ε (more on that in Section 4).

Connection to Reversibility

As mentioned above, the two conditions (L) and (D) are equivalent in reversible systems:

Proposition 3.6. *Let the system be reversible, i.e., let (2) hold. Then the system is ε-lumpable if and only if it is ε-deflatable.*

Proof. Let (L) hold for some family of functions $p_L^t : \mathbb{Z} \times \mathbb{X} \to \mathbb{R}$. Define the family of functions $p_D^t : \mathbb{X} \times \mathbb{Z} \to \mathbb{R}$ by

$$p_D^t(x, z) := \frac{p_L^t(z, x)}{\pi(x)}.$$

Then

$$\begin{aligned}
\|p_D^\tau(x, \xi(\cdot))\pi(\cdot) - p^\tau(x, \cdot)\|_{L^1} &= \|p_L^\tau(\xi(\cdot), x)\frac{\pi(\cdot)}{\pi(x)} - p^\tau(x, \cdot)\|_{L^1} \\
&\overset{(2)}{=} \|p_L^\tau(\xi(\cdot), x)\frac{\pi(\cdot)}{\pi(x)} - p^\tau(x, \cdot)\frac{\pi(\cdot)}{\pi(x)}\|_{L^1} \\
&= \|p_L^\tau(\xi(x), \cdot) - p^\tau(x, \cdot)\|_{L^1_\mu}\pi(x)^{-1}.
\end{aligned}$$

Hence, with this p_D^t, it holds

$$\begin{aligned}
\frac{1}{|\mathbb{Z}|}\left\|p^\tau(*, \cdot) - p_D^\tau(*, \xi(\cdot))\pi(\cdot)\right\|_{\mathbb{K}} &= \frac{1}{|\mathbb{Z}|}\int_{\mathbb{X}} \left\|p^\tau(x, \cdot) - p_D^\tau(x, \xi(\cdot))\pi(\cdot)\right\|_{L^1} d\mu(x) \\
&= \frac{1}{|\mathbb{Z}|}\int_{\mathbb{X}} \left\|p_L^\tau(\xi(x), \cdot) - p^\tau(x, \cdot)\right\|_{L^1_\mu}\pi(x)^{-1} d\mu(x) \\
&= \frac{1}{|\mathbb{Z}|}\left\|p_L^\tau(\xi(*), \cdot) - p^\tau(*, \cdot)\right\|_{\mathbb{K}} \leq \varepsilon.
\end{aligned}$$

For the reverse direction, let (D) hold for some family of functions $p_D^t : \mathbb{X} \times \mathbb{Z} \to \mathbb{R}$ and define $p_L^t : \mathbb{Z} \times \mathbb{X} \to \mathbb{R}$ by

$$p_L^t(z, y) := p_D^t(y, z)\pi(y).$$

We then obtain ε-lumpability with this p_L^t by performing the above transformations in reverse. □

Remark 3.7. Proposition 3.6 implies that, whenever a system is ε-lumpable or ε-deflatable, there exists a reduced transition kernel $\tilde{p}^t : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}^+$, such that

$$\left\|p^t(*, \cdot) - \tilde{p}^t(\xi(*), \xi(\cdot))\pi(\cdot)\right\|_{\mathbb{K}} \leq \varepsilon$$

for $t \geq \tau$. Under this condition, knowing the triple $(\xi, \tilde{p}^t, \pi)$ allows us to effectively re-construct the long-term dynamics of the full system. This specifies our understanding of a *good reaction coordinate*. Furthermore, we call a system for which such a good reaction coordiante exists, a *reducible* system.

3.3. Examples of reducible systems

While the conditions defined in Section 3.2 certainly can be seen as a sensible quality measure for reaction coordinates, an obvious question is whether they are consistent with established concepts of reducible systems. We hence now present several systems with known good reaction coordinates and show that they are indeed either lumpable or deflatable.

3.3.1. Existence of a transition manifold

Lumpability is strongly connected to the concept of the so-called *transition manifold*, which was introduced a few years ago by some of the authors in order to formulate a geometrical approach to the computation of optimal reaction coordinates [4].

Definition 3.8 (Reducibility and Transition Manifold)**.** *For $\varepsilon > 0, r \leq n, \tau \in \mathbb{R}_0^+$, we call the system* (ε, r, τ)-reducible *if there exists an r-dimensionally parametrizable manifold $\mathbb{M} \subset \mathbb{P}_\tau$ so that for all $x \in \mathbb{X}$*

$$\left\| \mathcal{Q}\left(p^\tau(x,\cdot)\right) - p^\tau(x,\cdot) \right\|_{L^2_{1/\pi}} \leq \varepsilon, \tag{5}$$

where $\mathcal{Q} : \mathbb{P}_\tau \to \mathbb{M}$ is the nearest point projection onto $\mathbb{M}$,

$$\mathcal{Q}\left(p^\tau(x',\cdot)\right) := \operatorname*{arg\,min}_{p \in \mathbb{M}} \left\| p^\tau(x,\cdot) - p \right\|_{L^2_{1/\pi}}.$$

We call any $\mathbb{M}$ that fulfills (5) a transition manifold *of the system.*

The intuition behind (7) is that the set of all transition densities $\mathbb{P}_\tau$ clusters ε-closely around an r-dimensional manifold $\mathbb{M}$ with respect to the $L^2_{1/\pi}$ norm. Let $\mathcal{E} : \mathbb{M} \to \mathbb{R}^r$ be any parametrization of $\mathbb{M}$. It can be shown that the *transition manifold reaction coordinate*

$$\xi(x) := \mathcal{E}\left(\mathcal{Q}(p^\tau(x,\cdot))\right) \tag{6}$$

is a good reaction coordinate, in the sense that the projection error of the leading transfer operator eigenfunctions onto ξ is at most ε [4].

The computational strategy behind the transition manifold approach now is to sample the set $\{p^t(x,\cdot),\ x \in \mathbb{X}\}$ (for example by randomly selecting starting points $x_m \in \mathbb{X},\ m = 1, 2, \ldots$ and estimating the $p^\tau(x_m,\cdot)$ by parallel simulation), and applying an *unsupervised manifold learning method* (such as diffusion maps [8]) to the samples. This strategy has been successfully applied to multiple high-dimensional molecular systems and was confirmed to produce physically interpretable reaction coordinates [2, 3].

The concept of the transition manifold was recently re-visited and extended to a broader class of dynamical systems [5] (the central object now being called *weak transition manifold*), by requiring the "closeness" to the manifold now only averaged over the level sets of $\mathcal{Q}$:

Definition 3.9 (Weak reducibility and weak transition manifold)**.** *For $\varepsilon > 0, r \leq n, \tau \in \mathbb{R}_0^+$, we call the system* weakly (ε, r, τ)-reducible *if there exists an r-dimensionally parametrizable manifold $\mathbb{M} \subset \mathbb{P}_\tau$ so that for all $x \in \mathbb{X}$*

$$\int_{\Sigma_\mathcal{Q}(\mathcal{Q}(x))} \left\| \mathcal{Q}\left(p^\tau(x',\cdot)\right) - p^\tau(x',\cdot) \right\|_{L^2_{1/\pi}} d\mu_{\mathcal{Q}(x)}(x') \leq \varepsilon, \tag{7}$$

where $\mathcal{Q} : \mathbb{P}_\tau \to \mathbb{M}$ is the nearest point projection onto $\mathbb{M}$,

$$\mathcal{Q}\left(p^\tau(x',\cdot)\right) := \operatorname*{arg\,min}_{p \in \mathbb{M}} \left\| p^\tau(x,\cdot) - p \right\|_{L^2_{1/\pi}},$$

and $\Sigma_\mathcal{Q}(q)$ is the q-level set of $\mathcal{Q}$,

$$\Sigma_\mathcal{Q}(p) = \left\{ p \in \mathbb{P}_\tau \mid \mathcal{Q}(p) = q \right\}.$$

We call any manifold $\mathbb{M}$ that fulfills (7) a weak transition manifold.

It is easy to confirm that every reducible system is weakly reducible.

The careful reader might have noticed that the closeness condition for the weak transition manifold (7) resembles the lumpability condition (L). Indeed, we will now show that a system which is weakly (ε, r, τ)-reducible is ε-lumpable with respect to the transition manifold reaction coordinate ξ.

Proposition 3.10. *Let the system be weakly (ε, r, τ)-reducible. Then there exists a domain $\mathbb{Z} \subset \mathbb{R}^r$, and a family of functions $p_L^t : \mathbb{Z} \times \mathbb{X} \to \mathbb{R}^+$ such that (L) is fulfilled with respect to ξ defined by (6).*

Proof. Let $\mathcal{E} : \mathbb{M} \to \mathbb{R}^r$ be any parametrization of the transition manifold $\mathbb{M}$ and let $\mathbb{Z} := \mathcal{E}(\mathbb{M})$. Note that $\mathcal{E} : \mathbb{M} \to \mathbb{Z}$ is one-to-one. Define the reaction coordinate ξ by (6) and the reduced density p_L^τ by

$$p_L^\tau(z, \cdot) := \mathcal{E}^{-1}(z).$$

Then for any $z \in \mathbb{Z}$ there exists a $x \in \mathbb{X}$ such that $z = \mathcal{E}\big(\mathcal{Q}(x)\big)$ and hence, due to $\mathcal{E}$ being one-to-one, $\Sigma_\xi(z) = \Sigma_\mathcal{Q}\big(\mathcal{Q}(x)\big)$. For this x it holds

$$\begin{aligned}
\int_{\Sigma_\xi(z)} \big\|p_L^\tau(z,\cdot) - p^\tau(x',\cdot)\big\|_{L^1} d\mu_z(x') &= \int_{\Sigma_\mathcal{Q}(\mathcal{Q}(x))} \big\|p_L^\tau\big(\mathcal{E}(\mathcal{Q}(x)),\cdot\big) - p^\tau(x',\cdot)\big\|_{L^1} d\mu_{\mathcal{Q}(x)}(x') \\
&= \int_{\Sigma_\mathcal{Q}(\mathcal{Q}(x))} \big\|\mathcal{E}^{-1}\big(\mathcal{E}(\mathcal{Q}(x))\big) - p^\tau(x',\cdot)\big\|_{L^1} d\mu_{\mathcal{Q}(x)}(x') \\
&= \int_{\Sigma_\mathcal{Q}(\mathcal{Q}(x))} \big\|\mathcal{Q}(x) - p^\tau(x',\cdot)\big\|_{L^1} d\mu_{\mathcal{Q}(x)}(x').
\end{aligned}$$

Finally, with $\|f\|_{L^1} = \|f/\pi\|_{L^1_\mu} \le \|f/\pi\|_{L^2_\mu} = \|f\|_{L^2_{1/\pi}}$, we get

$$\begin{aligned}
\frac{1}{|\mathbb{Z}|}\int_\mathbb{Z}\int_{\Sigma_\xi(z)} \big\|p_L^\tau(z,\cdot) - p^\tau(x',\cdot)\big\|_{L^1} d\mu_z(x')\,dz &\le \frac{1}{|\mathbb{Z}|}\int_\mathbb{Z}\int_{\Sigma_\mathcal{Q}(\mathcal{Q}(x))} \big\|\mathcal{Q}(x) - p^\tau(x',\cdot)\big\|_{L^2_{1/\pi}} d\mu_{\mathcal{Q}(x)}(x')\,dz \\
&\le \sup_{z\in\mathbb{Z}} \int_{\Sigma_\mathcal{Q}(\mathcal{Q}(x))} \big\|\mathcal{Q}(x) - p^\tau(x',\cdot)\big\|_{L^2_{1/\pi}} d\mu_{\mathcal{Q}(x)}(x') \\
&\overset{(7)}{\le} \varepsilon.
\end{aligned}$$

□

3.3.2. Slow- and fast components

Next, we show that a process defined by an SDE with slow and fast components is deflatable with respect to the slow component.

We utilize a multiscale decomposition of the corresponding infinitesimal generator, together with a multiscale ansatz for the transition densities $p^t(x,\cdot)$. In that we utilize well-known averaging and homogenization techniques from [36].

It will prove advantageous to consider the transition densities $p^t(x,\cdot)$ as densities $q^t(x,\cdot)$ with respect to the stationary density π. That is, we define for each $x \in \mathbb{X}$ $q^t(x,\cdot)$ by

$$p^t(x,\cdot) = q^t(x,\cdot)\pi(\cdot).$$

Let the components of (X_t) be dividable into two processes (Y_t) on $\mathbb{Y}$, (Z_t) on $\mathbb{Z}$, such that $\mathbb{X} = \mathbb{Y} \oplus \mathbb{Z}$,

$$(X_t) = \begin{pmatrix} Y_t \\ Z_t \end{pmatrix},$$

and the components fulfil the system of SDEs

$$\begin{aligned}
\varepsilon dY_t &= -\nabla_y V(Y_t, Z_t)\,dt + \sqrt{\varepsilon}\sigma dW_t^\mathbb{Y} \\
dZ_t &= -\nabla_z V(Y_t, Z_t)\,dt + \sigma dW_t^\mathbb{Z},
\end{aligned} \tag{8}$$

with potential $V : \mathbb{X} \to \mathbb{R}$, scalar diffusion parameter $\sigma > 0$ and $(W_t^{\mathbb{Y}})$, $(W_t^{\mathbb{Z}})$ standard Wiener processes on $\mathbb{Y}$ and $\mathbb{Z}$, respectively. A timescale separation parameter $0 < \varepsilon \ll 1$ ensures that (Y_t) evolves "fast" compared to (Z_t).

The evolution of q^t under (X_t) is now governed by the Fokker Planck equation

$$\partial_t q^t(x, \cdot) = \mathcal{L}_\varepsilon q^t(x, \cdot), \quad q^0(x, \cdot) = \delta_x(\cdot),$$

where the infinitesimal generator $\mathcal{L}_\varepsilon$ has the multiscale structure

$$\mathcal{L}_\varepsilon = \frac{1}{\varepsilon}\mathcal{L}_z + \mathcal{L}_y, \tag{9}$$

with the two compontents

$$\mathcal{L}_z = \frac{\sigma^2}{2}\Delta_y - \nabla_y V \cdot \nabla_y,$$
$$\mathcal{L}_y = \frac{\sigma^2}{2}\Delta_z - \nabla_z V \cdot \nabla_z.$$

We want to investigate to what extent $q^t(x, \cdot)$, and by extension $p^t(x, \cdot)$, can be approximated by an "essential transition density" that depends only on the slow variable z in the situation $t \gg \varepsilon$. To this end, we make the multiscale ansatz for q^t

$$q^t(x, \cdot) = q_0^t(x, \cdot) + \varepsilon q_1^t(x, \cdot) + \mathcal{O}(\varepsilon^2). \tag{10}$$

Inserting (10) into (9) gives

$$\partial_t q_0^t(x, \cdot) + \varepsilon \partial_t q_1^t(x, \cdot) + \mathcal{O}(\varepsilon^2) = \frac{1}{\varepsilon}\mathcal{L}_z q_0^t(x, \cdot) + \mathcal{L}_z q_1^t(x, \cdot) + \mathcal{L}_y q_0^t(x, \cdot) + \mathcal{O}(\varepsilon). \tag{11}$$

Comparing the terms of order ε^{-1} gives

$$\mathcal{L}_z q_0^t(x, \cdot) = 0.$$

By [36, Sec. 10.4], this implies that for each fixed z, $\mathcal{L}_z$ has a one-dimensional null space, namely the constant functions in y. Hence, $q_0^t(x, \cdot)$ is independent of y in its second argument, and so

$$q^t\left((y, z), (y', z')\right) = q_0^t\left((y, z), z'\right) + \mathcal{O}(\varepsilon).$$

Therefore, for $t = \mathcal{O}(1)$, the transition density $p^t(x, \cdot)$ takes the form

$$p^t\left((y, z), (y', z')\right) = q_0^t((y, z), z')\,\pi\left((y', z')\right) + g\left((y, z), (y', z')\right)\,\pi\left((y', z')\right).$$

for some function $g \in \mathcal{O}(\varepsilon)$. Applying the $\|\cdot\|_{\mathbb{K}}$-norm, it follows that the system is deflatable with respect to the reaction coordinate $\xi(x) = z$, i.e.,

$$\frac{1}{|\mathbb{Z}|}\left\|q_0^t(*, \xi(\cdot))\pi(\cdot) - p^t(*, \cdot)\right\|_{\mathbb{K}} = \mathcal{O}(\varepsilon).$$

Remark 3.11. While not necessarily in the scope of this paper, we can continue the multiscale analysis in order to derive an evolution equation for q_0^t. See Appendix A for details.

3.3.3. Generator spectral gap

Finally, we show that systems whose infinitesimal generator exhibits a spectral gap of sufficient size, such as metastable systems, are deflatable with respect to some non-trivial reaction coordinate.

Consider again the transition densities q^t with respect to the stationary measure, and its Fokker Planck equation

$$\partial_t q^t(x, \cdot) = \mathcal{L} q^t(x, \cdot), \quad q^0(x, \cdot) = \delta_x(\cdot).$$

We assume that the spectrum of the generator $\mathcal{L}$ is real, non-positive and discrete, and denote the eigenvalues of $\mathcal{L}$ in descending order:

$$0 = \theta_0 > \theta_1 \geq \theta_2 \geq \dots,$$

repeated by geometric multiplicity. Let furthermore φ_i denote the eigenfunction belonging to θ_i. The φ_i then form an orthonormal basis of $L^2_\mu(\mathbb{X})$ [35]. As $q^t(x,\cdot) \in L^1_\mu(\mathbb{X}) \cap L^\infty_\mu(\mathbb{X})$ for any x, $q^t(x,\cdot) \in L^2_\mu(\mathbb{X})$. We can then describe the evolution of the density $q^t(x,\cdot)$ by

$$q^t(x,\cdot) = \sum_{i=0}^{\infty} e^{\theta_i t} c_i(x) \varphi_i(\cdot) \tag{12}$$

where $c_i : \mathbb{X} \to \mathbb{R}$ are some functions that do not depend on the θ_i.

Now, we additionally assume that the eigenvalues can be separated into a dominant and a non-dominant part. Specifically, we assume there exists an index $K > 0$, so that the ratio

$$\rho := \frac{\theta_K}{\theta_{K+1}} \tag{13}$$

is small. This situation for example occurs if the system is *metastable*, i.e., there exists a partition $\mathbb{X} = \mathbb{X}_1 \cup \ldots \cup \mathbb{X}_K$ of state space into disjoint regions, and the system is almost invariant on each $\mathbb{X}_i$ on relatively long time scales. Each For a precise introduction of metastability and its connection to the dominant spectrum see [44].

Now suppose that there exists an integer $r \leq K$ and a reaction coordinate $\xi : \mathbb{X} \to \mathbb{R}^r$ such that ξ parametrizes the dominant φ_i, i.e., there exist some functions $\tilde{\varphi}_i : \mathbb{R}^r \to \mathbb{R}$ such that

$$\varphi_i = \tilde{\varphi}_i \circ \xi \quad i = 1, \dots, K. \tag{14}$$

Such a ξ always exists, as one can always choose

$$r := K, \quad \xi_i := \varphi_i \quad \text{and} \quad \tilde{\varphi}_i(z) := z_i.$$

Often, however, the dominant eigenfunctions possess some common lower-dimensional, non-linear parametrization. For metastable systems, this is for example the case if the metastable sets $\mathbb{X}_1, \dots, \mathbb{X}_K$ are connected by just a small number of transition pathways. An example system with five metastable sets, but one common transition pathway, hence a one-dimensional reaction coordinate, can be found in Section 6.2.

Let $\varepsilon > 0$ be some small constant. We now show that, if $t = t(\varepsilon)$ is large enough, and the metastability ratio $\rho = \rho(\varepsilon)$ is small enough, then the system is ε-deflatable with respect to any ξ fulfilling (14). To see this, split the right hand side of (12) into the dominant and the non-dominant part:

$$q^t(x,\cdot) = \sum_{i=0}^{K} e^{\theta_i t} c_i(x) \varphi_i(\cdot) + \sum_{i=K+1}^{\infty} e^{\theta_i t} c_i(x) \varphi_i(\cdot).$$

Due to (14), the first summand depends only on ξ:

$$\sum_{i=0}^{K} e^{\theta_i t} c_i(x) \varphi_i(\cdot) = \underbrace{\sum_{i=0}^{K} e^{\theta_i t} c_i(x) \tilde{\varphi}_i\left(\xi(\cdot)\right)}_{=: p_D^t(x, \xi(\cdot))}.$$

The system hence is ε-deflatable with respect to ξ and p_D^t, if

$$\frac{1}{|\mathbb{Z}|} \left\| \sum_{i=K+1}^{\infty} e^{\theta_i t} c_i(*) \varphi_i(\cdot) \pi(\cdot) \right\|_{\mathbb{K}} \leq \varepsilon. \tag{15}$$

Since the θ_i are decreasing, it holds

$$\left\| \sum_{i=K+1}^{\infty} e^{\theta_i t} c_i(*)\varphi_i(\cdot)\pi(\cdot) \right\|_{\mathbb{K}} \leq e^{\theta_{K+1} t} \left\| \sum_{i=K+1}^{\infty} c_i(*)\varphi_i(\cdot)\pi(\cdot) \right\|_{\mathbb{K}},$$

and further

$$\leq e^{\theta_{K+1} t} \left(\underbrace{\left\| \sum_{i=0}^{K} c_i(*)\varphi_i(\cdot)\pi(\cdot) \right\|_{\mathbb{K}}}_{=:\tilde{C}} + \left\| \sum_{i=0}^{\infty} c_i(*)\varphi_i(\cdot)\pi(\cdot) \right\|_{\mathbb{K}} \right).$$

The first summand, denoted $\tilde{C}$, is finite as a finite sum. As

$$p^t(*,\cdot) = q^t(*,\cdot)\pi(\cdot) = \sum_{i=0}^{\infty} e^{\theta_i t} c_i(*)\varphi_i(\cdot)\pi(\cdot),$$

and $1 = e^{\theta_i 0}$, the second summand can be interpreted as the $\mathbb{K}$-norm of $p^t\big|_{t=0}$:

$$\left\| \sum_{i=0}^{\infty} c_i(*)\varphi_i(\cdot)\pi(\cdot) \right\|_{\mathbb{K}} = \lim_{t\to 0} \left\| p^t(*,\cdot) \right\|_{\mathbb{K}}.$$

By writing out the $\mathbb{K}$-norm, it can easily be seen that $\|p^t(*,\cdot)\|_{\mathbb{K}} = 1$ for all $t > 0$, and thus

$$\lim_{t\to 0} \left\| p^t(*,\cdot) \right\|_{\mathbb{K}} = 1.$$

With $C := \tilde{C} + 1$, we therefore get

$$\frac{1}{|\mathbb{Z}|} \left\| \sum_{i=K+1}^{\infty} e^{\theta_i t} c_i(*)\varphi_i(\cdot)\pi(\cdot) \right\|_{\mathbb{K}} \leq \frac{C}{|\mathbb{Z}|} e^{\theta_{K+1} t}.$$

Hence, in the for us relevant situation $\varepsilon < \frac{C}{|\mathbb{Z}|}$, if we choose

$$t \geq t(\varepsilon) := \frac{1}{\theta_{K+1}} \log\left(\frac{|Z|}{C} \varepsilon \right),$$

then (15) is fulfilled, and the system is ε-deflatable.

Now, of course, every system is arbitrarily deflatable if only the lag time is chosen large enough (see Remark 3.5). In order to claim non-trivial deflatability for the lag time $t(\varepsilon)$, we have to ensure that $p_D^t(x,\cdot)$ is not close to the identity, i.e., the factors $e^{\theta_i t(\varepsilon)}, i = 1,\dots,K$, do not decay for $\varepsilon \to 0$. This is indeed the case if the metastability ratio $\rho = \rho(\varepsilon)$ is sufficiently small, since

$$\begin{aligned} e^{\theta_i t(\varepsilon)} &= e^{\frac{\theta_i}{\theta_{K+1}} \log\left(\frac{|Z|}{C}\varepsilon\right)} \\ &\geq e^{\frac{\theta_K}{\theta_{K+1}} \log\left(\frac{|Z|}{C}\varepsilon\right)} \quad \text{for} \quad i = 1,\dots,K. \end{aligned}$$

Hence, for

$$\frac{\theta_K}{\theta_{K+1}} = \rho(\varepsilon) = \mathcal{O}\left(\log\left(\frac{|Z|}{C}\varepsilon \right)^{-1} \right) \quad (\varepsilon \to 0),$$

we have

$$e^{\theta_i t(\varepsilon)} = \mathcal{O}(1) \quad (\varepsilon \to 0),$$

and thus also, for all $x \in \mathbb{X}$,

$$p_D^{t(\varepsilon)}(x,\cdot) = \mathcal{O}(1) \quad (\varepsilon \to 0).$$

In words: if the metastability ratio falls as $\mathcal{O}\left(\log\left(\frac{|Z|}{C}\varepsilon \right)^{-1} \right)$ as $\varepsilon \to 0$, then the slow modes have not yet decayed on timescales $t \approx t(\varepsilon)$, and the effective density p_D^t describes their further evolution.

4. A variational principle for optimal reaction coordinates

From here on, we consider the reduced dimension $r \leq n$ as well the lag time τ to be predetermined and fixed. For simplicity of notation, we will omit the lag time as parameter for the transition kernel, i.e., define $p(\cdot,\cdot) := p^\tau(\cdot,\cdot)$, $p_L(\cdot,\cdot) := p_L^\tau(\cdot,\cdot)$, $p_D(\cdot,\cdot) := p_D^\tau(\cdot,\cdot)$.

Our goal is now to find an optimal reaction coordinate, i.e., a function $\xi : \mathbb{X} \to \mathbb{Z} \subset \mathbb{R}^r$ that fulfills (L) or equivalently (D) for the smallest possible $\varepsilon \geq 0$. Hence formally, we seek the minimizers of the following loss functions:

Definition 4.1 (Lumpability and deflatability loss function)**.** *The nonlinear functional* $\mathscr{L}_L : C(\mathbb{X},\mathbb{Z}) \to \mathbb{R}^+$*, defined by*

$$\mathscr{L}_L(\vartheta) := \frac{1}{|\mathbb{Z}|} \min_{p_L:\mathbb{Z}\times\mathbb{X}\to\mathbb{R}^+} \|p(*,\cdot) - p_L(\vartheta(*),\cdot)\|_{\mathbb{K}} \tag{16}$$

is called the lumpability loss function *of the system.*

The nonlinear functional $\mathscr{L}_D : C(\mathbb{X},\mathbb{Z}) \to \mathbb{R}^+$*, defined by*

$$\mathscr{L}_D(\vartheta) := \frac{1}{|\mathbb{Z}|} \min_{p_D:\mathbb{X}\times\mathbb{Z}\to\mathbb{R}^+} \|p(*,\cdot) - p_D(*,\vartheta(\cdot))\pi(\cdot)\|_{\mathbb{K}} \tag{17}$$

is called the deflatability loss function *of the system.*

From the equivalence of lumpability and deflatability (Proposition 3.6) it follows that for every $\vartheta \in C(\mathbb{X},\mathbb{Z})$ holds

$$\mathscr{L}_{\mathrm{L}}(\vartheta) = \mathscr{L}_{\mathrm{D}}(\vartheta), \tag{18}$$

hence we can find the optimal reaction coordinate with respect to both (L) and (D) by solving

$$\xi := \operatorname*{arg\,min}_{\vartheta\in C(\mathbb{X},\mathbb{Z})} \mathscr{L}_{\mathrm{L}}(\vartheta), \tag{19}$$

at least in theory. Evaluating, let alone minimizing $\mathscr{L}_{\mathrm{L}}$ (or $\mathscr{L}_{\mathrm{D}}$) would however prove difficult, due to the minimization over an infinite-dimensional function space involved in its definition (the search for the functions p_L or p_D in every step). In the following section, we therefore derive *essentially* equivalent reformulations of $\mathscr{L}_{\mathrm{L}}$ and $\mathscr{L}_{\mathrm{D}}$ that do not involve this minimization.

4.1. Differential formulation of lumpability and deflatability

The condition (L) can be interpreted as the closeness of the transition densities $p(x,\cdot)$ to some reduced reference density $p_L(\xi(x),\cdot)$. This implies that all densities $p(x,\cdot)$ whose starting points x lie on one level set of ξ are close *to each other*. Likewise, condition (D) can be seen as the closeness of the transition observables $p(\cdot,y)$ to some reduced reference observable $p_D(\cdot,\xi(y))$, which implies that all observables $p(\cdot,y)$ whose end points y lie on one level set of ξ are close *to each other*. This observation motivates the following "differential" characterization of lumpability and deflatability:

Definition 4.2 (Differential lumpability)**.** *If there exists a domain* $\mathbb{Z} \subset \mathbb{R}^r$ *and a continuous function* $\xi : \mathbb{X} \to \mathbb{Z}$ *such that*

$$\frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\xi(z)} \int_{\Sigma_\xi(z)} \left\| p(x^{(1)},\cdot) - p(x^{(2)},\cdot) \right\|_{L^1} d\mu_z\big(x^{(1)}\big)\, d\mu_z\big(x^{(2)}\big)\, dz \leq \varepsilon \tag{L'}$$

is fulfilled, we say the system is differentially ε-lumpable *with respect to* ξ*.*

Definition 4.3 (Differential deflatability)**.** *If there exists a domain* $\mathbb{Z} \subset \mathbb{R}^r$ *and a continuous function* $\xi : \mathbb{X} \to \mathbb{Z}$ *such that*

$$\frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\xi(z)} \int_{\Sigma_\xi(z)} \left\| p(\cdot,y^{(1)})/\pi(y^{(1)}) - p(\cdot,y^{(2)})/\pi(y^{(2)}) \right\|_{L^1_\mu} d\mu_z\big(x^{(1)}\big)\, d\mu_z\big(x^{(2)}\big)\, dz \leq \varepsilon. \tag{D'}$$

is fulfilled, we say the system is differentially ε deflatable *with respect to* ξ*.*

As one easily sees using a triangle inequality argument, the differential conditions imply the original conditions, and the original conditions *almost* imply the differential conditions:

Lemma 4.4. *If the system is ε-lumpable with respect to ξ, then the system is differentially 2ε-lumpable with respect to ξ.*

Conversely, if the system is differentially ε-lumpable with respect to ξ, then the system is ε-lumpable with respect to ξ.

Proof. Assume that the system is ε-lumpable, i.e., (L) holds for some function p_L. Then we have

$$\begin{aligned}
&\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\Sigma_\xi(z)} \left\|p(x^{(1)},\cdot)-p(x^{(2)},\cdot)\right\|_{L^1} d\mu_z\big(x^{(1)}\big)\,d\mu_z\big(x^{(2)}\big)\,dz \\
&\qquad= \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\Sigma_\xi(z)} \Big\|p(x^{(1)},\cdot)-\underbrace{p_L(z,\cdot)}_{=p_L(\xi(x^{(1)}),\cdot)}+\underbrace{p_L(z,\cdot)}_{=p_L(\xi(x^{(2)}),\cdot)}-p(x^{(2)},\cdot)\Big\|_{L^1} d\mu_z\big(x^{(1)}\big)\,d\mu_z\big(x^{(2)}\big)\,dz \\
&\qquad\le \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)} \big\|p(x^{(1)},\cdot)-p_L(\xi(x^{(1)}),\cdot)\big\|_{L^1} d\mu_z(x^{(1)})\,dz + \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)} \big\|p(x^{(2)},\cdot)-p_L(\xi(x^{(2)}),\cdot)\big\| \\
&\qquad\le 2\varepsilon.
\end{aligned}$$

For the reverse statement, assume that (L') holds. Define

$$p_L(z,\cdot) := \int_{\Sigma_\xi(z)} p(x',\cdot)\,d\mu_z(x').$$

This p_L exists because $p(\cdot,y)\in L^1_\mu$ for all $y\in\mathbb{X}$.

Recall that all μ_z are probability measures on the respective $\Sigma_\xi(z)$. Then

$$\begin{aligned}
&\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)} \big\|p_L(z,\cdot)-p(x,\cdot)\big\|_{L^1}\,d\mu_z(x)\,dz \\
&\qquad= \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)} \Big\|\int_{\Sigma_\xi(z)} p(x',y)-p(x,y)\,d\mu_z(x')\Big\|_{L^1} d\mu_z(x)\,dz \\
&\qquad\le \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\Sigma_\xi(z)} \big\|p(x',\cdot)-p(x,\cdot)\big\|_{L^1}\,d\mu_z(x')\,d\mu_z(x)\,dz \\
&\qquad\overset{\text{(L')}}{\le} \varepsilon.
\end{aligned}$$

□

Lemma 4.5. *If the system is ε-deflatable with respect to ξ, then the system is differentially 2ε-deflatable with respect to ξ.*

Conversely, if the system is differentially 2ε-deflatable with respect to ξ, then the system is ε-deflatable with respect to ξ.

Proof. Assume that the system is ε-deflatable, i.e., (D) holds for some function p_D. Then we have

$$\begin{aligned}
&\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\Sigma_\xi(z)}\left\|p(\cdot,y^{(1)})/\pi(y^{(1)})-p(\cdot,y^{(2)})/\pi(y^{(2)})\right\|_{L^1_\mu}d\mu_z\big(y^{(1)}\big)\,d\mu_z\big(y^{(2)}\big)\,dz\\
&\quad=\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\Sigma_\xi(z)}\left\|p(\cdot,y^{(1)})/\pi(y^{(1)})-p_D(\cdot,z)+p_D(\cdot,z)-p(\cdot,y^{(2)})/\pi(y^{(2)})\right\|_{L^1_\mu}d\mu_z\big(y^{(1)}\big)\,d\mu_z\big(y^{(2)}\big)\,d\\
&\quad\le\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\left\|p(\cdot,y^{(1)})/\pi(y^{(1)})-p_D(\cdot,\xi(y^{(1)}))\right\|_{L^1_\mu}d\mu_z\big(y^{(1)}\big)\,dz\\
&\qquad+\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\left\|p(\cdot,y^{(2)})/\pi(y^{(2)})-p_D(\cdot,\xi(y^{(2)}))\right\|_{L^1_\mu}d\mu_z\big(y^{(2)}\big)\,dz\\
&\quad=\frac{2}{|\mathbb{Z}|}\int_{\mathbb{X}}\|p(\cdot,y)-p_D(\cdot,\xi(y))\pi(y)\|_{L^1_\mu}\,dy\\
&\quad=\frac{2}{|\mathbb{Z}|}\big\|\|p(*,\cdot)-p_D(*,\cdot)\pi(\cdot)\|_{L^1}\big\|_{L^1_\mu}\overset{\text{(D)}}{\le}2\varepsilon.
\end{aligned}$$

For the reverse statement, define

$$p_D(\cdot,z):=\int_{\Sigma_\xi(z)}p(\cdot,y')/\pi(y')\,d\mu_z(y').$$

This p_D exists because $p(x,\cdot)\in L^1$ for all $x\in\mathbb{X}$. Then

$$\begin{aligned}
\frac{1}{|\mathbb{Z}|}\big\|p(*,\cdot)-p_D(*,\xi(\cdot))\pi(\cdot)\big\|_{\mathbb{K}}&=\frac{1}{|\mathbb{Z}|}\big\|\|p(*,\cdot)-p_D(*,\xi(\cdot))\pi(\cdot)\|_{L^1}\big\|_{L^1_\mu}\\
&=\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\left\|p_D(\cdot,z)-p(\cdot,y)/\pi(y)\right\|_{L^1_\mu}d\mu_z(y)\,dz\\
&=\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\mathbb{X}}\Big|\int_{\Sigma_\xi(z)}p(x,y')/\pi(y')-p(x,y)/\pi(y)d\mu_z(y')\Big|\,d\mu(x)\,d\mu_z(y)\,dz\\
&\le\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\xi(z)}\int_{\Sigma_\xi(z)}\|p(\cdot,y')/\pi(y')-p(\cdot,y)/\pi(y)\|_{L^1_\mu}d\mu_z(y')d\mu_z(y)\,dz\\
&\overset{\text{(D')}}{\le}\varepsilon.
\end{aligned}$$

□

4.2. Differential loss functions

We can now define *differential* versions of the loss functions $\mathscr{L}_{\mathrm{L}}$ and $\mathscr{L}_{\mathrm{D}}$ in which the hard-to-identify terms p_L and p_D no longer appear:

Definition 4.6 (differential lumpability and differential deflatability loss function)**.** *The nonlinear functional $\widetilde{\mathscr{L}_L}:C(\mathbb{X},\mathbb{Z})\to\mathbb{R}^+$, defined by*

$$\widetilde{\mathscr{L}_L}(\vartheta):=\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\vartheta(z)}\int_{\Sigma_\vartheta(z)}\left\|p(x^{(1),\cdot})-p(x^{(2)},\cdot)\right\|_{L^1}d\mu_z\big(x^{(1)}\big)\,d\mu_z\big(x^{(2)}\big)\,dz \tag{20}$$

is called the differential lumpability loss function *of the system.*

The nonlinear functional $\widetilde{\mathscr{L}_D}:C(\mathbb{X},\mathbb{Z})\to\mathbb{R}^+$, defined by

$$\widetilde{\mathscr{L}_D}(\vartheta):=\frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\vartheta(z)}\int_{\Sigma_\vartheta(z)}\left\|p(\cdot,y^{(1)})/\pi(y^{(1)})-p(\cdot,y^{(2)})/\pi(y^{(2)})\right\|_{L^1_\mu}d\mu_z\big(x^{(1)}\big)\,d\mu_z\big(x^{(2)}\big)\,dz \tag{21}$$

is called the differential deflatability loss function *of the system.*

Note that, unlike $\mathscr{L}_\mathrm{L}$ and $\mathscr{L}_\mathrm{D}$, $\widetilde{\mathscr{L}_\mathrm{L}}$ and $\widetilde{\mathscr{L}_\mathrm{D}}$ are in general neither identical to each other, nor to $\mathscr{L}_\mathrm{L}$. The following result characterizes the relationship between the different loss functions:

Corollary 4.7. *For any $\vartheta \in C(\mathbb{X}, \mathbb{Z})$ holds*

$$\mathscr{L}_L(\vartheta) \leq \widetilde{\mathscr{L}_L}(\vartheta) \leq 2\mathscr{L}_L(\vartheta),$$

and

$$\mathscr{L}_L(\vartheta) \leq \widetilde{\mathscr{L}_D}(\vartheta) \leq 2\mathscr{L}_L(\vartheta).$$

Proof. The first pair of inequalities follows directly from Lemma 4.4. The second pair of inequalities follows from Lemma 4.5 and (18). □

Hence, for arbitrary (non-optimal) reaction coordinates ϑ, we cannot expect $\widetilde{\mathscr{L}_\mathrm{L}}(\vartheta)$ or $\widetilde{\mathscr{L}_\mathrm{D}}(\vartheta)$ to be similar to $\mathscr{L}_\mathrm{L}(\vartheta)$. However, under the assumption that the system is indeed ε-lumpable for small ε, we can expect their *minima* to be similar:

Corollary 4.8. *Let ξ be an optimal reaction coordinate, defined by (19), and set $\varepsilon := \mathscr{L}_L(\xi)$. Let ξ_L and ξ_D be minimizers of $\widetilde{\mathscr{L}_L}$ and $\widetilde{\mathscr{L}_D}$, respectively. Then*

$$\widetilde{\mathscr{L}_L}(\xi) \leq 2\varepsilon \qquad \text{and} \qquad \widetilde{\mathscr{L}_D}(\xi) \leq 2\varepsilon.$$

Moreover,

$$\mathscr{L}_L(\xi_L) \leq 2\varepsilon \qquad \text{and} \qquad \mathscr{L}_L(\xi_D) \leq 2\varepsilon.$$

Proof. We show the assertions for $\widetilde{\mathscr{L}_\mathrm{L}}$. For $\widetilde{\mathscr{L}_\mathrm{D}}$, the proof is identical.

The first inequality follows from applying Corollary 4.7 to $\vartheta = \xi$. The second inequality then follows from

$$\mathscr{L}_\mathrm{L}(\xi_L) \overset{\text{Cor. } 4.7}{\leq} \widetilde{\mathscr{L}_\mathrm{L}}(\xi_L) \overset{\text{Def. } \xi_L}{\leq} \widetilde{\mathscr{L}_\mathrm{L}}(\xi) \leq 2\varepsilon.$$

□

Remark 4.9. The above variational principle implies that a minimizer η to $\widetilde{\mathscr{L}_\mathrm{L}}$ or $\widetilde{\mathscr{L}_\mathrm{D}}$ is not necessarily a strict minimizer of $\mathscr{L}_\mathrm{L}$. However, the difference between the minima will be at most ε. In other words, the system will be 2ε-lumpable and -deflatable with respect to η. Thus, for practical purposes, we can expect

$$\xi_L := \operatorname*{arg\,min}_{\vartheta \in C(\mathbb{X},\mathbb{Z})} \widetilde{\mathscr{L}_\mathrm{L}}(\vartheta) \tag{22}$$

and

$$\xi_D := \operatorname*{arg\,min}_{\vartheta \in C(\mathbb{X},\mathbb{Z})} \widetilde{\mathscr{L}_\mathrm{D}}(\vartheta) \tag{23}$$

to be "quasi-optimal" reaction coordinates.

5. Numerical approximation of the loss function

In order to solve the above optimization problems, the loss functions $\widetilde{\mathscr{L}_\mathrm{L}}$ and $\widetilde{\mathscr{L}_\mathrm{D}}$, and, depending on the optimization scheme, also their gradients, need to be evaluated numerically for candidate reaction coordinates ϑ. The presumed high dimension n of the state space however presents several challenges, which are discussed in the following.

5.1. Sampling requirements of the deflatability loss function

Note that at the heart of both $\widetilde{\mathscr{L}_{\mathrm{L}}}$ and $\widetilde{\mathscr{L}_{\mathrm{D}}}$ stands the evaluation of the transition density p^t at certain points. The function p^t is however not known analytically in practice, and must be estimated empirically by simulations of the dynamics. A critical question is thus how many of these estimates are required in order to approximate $\widetilde{\mathscr{L}_{\mathrm{L}}}$ or $\widetilde{\mathscr{L}_{\mathrm{D}}}$ up to a given tolerance. In particular, an exponential dependency of that number on the dimension n would be fatal. We will now show that, if the system is indeed highly lumpable with respect to some reaction coordinate ξ (which in general is different from the candidate reaction coordinate ϑ), the differential deflatability loss function $\widetilde{\mathscr{L}_{\mathrm{D}}}$ can be approximated very efficiently. Crucially, ξ does not have to be known, its implied existence is sufficient to guarantee the efficiency.

By expanding the $\|\cdot\|_{L^1_\mu}$ norm in (21), we can write $\widetilde{\mathscr{L}_{\mathrm{D}}}$ as

$$\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta) = \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\vartheta(z)}\int_{\Sigma_\vartheta(z)}\int_{\mathbb{X}}\left|\frac{p(x,y^{(1)})}{\pi(y^{(1)})} - \frac{p(x,y^{(2)})}{\pi(y^{(2)})}\right| \pi(x)\,dx\,d\mu_z(y^{(1)})\,d\mu_z(y^{(2)})\,dz. \tag{24}$$

or, by changing the integration order, as

$$= \frac{1}{|\mathbb{Z}|}\int_{\mathbb{X}}\left(\int_{\mathbb{Z}}\int_{\Sigma_\vartheta(z)}\int_{\Sigma_\vartheta(z)}\left|\frac{p(x,y^{(1)})}{\pi(y^{(1)})} - \frac{p(x,y^{(2)})}{\pi(y^{(2)})}\right| d\mu_z(y^{(1)})\,d\mu_z(y^{(2)})\,dz\right)\pi(x)\,dx. \tag{25}$$

Exact evaluation of the outermost integral in (25) would require analytical knowledge of the transition kernel p. Specifically, it would require knowledge of *all* transition densities, i.e., of $p(x,\cdot)$ for *all* starting points $x \in \mathbb{X}$, which we cannot assume in practice. We can however assume that a sufficiently precise approximation of a certain *limited* number of transition densities $p(x^{(k)},\cdot)$, $k = 1,\dots,M$ can be obtained. This approximation would typically be realized by parallel simulation of the stochastic dynamics starting from $x^{(k)}$, thus creating a point cloud that is distributed according to $p(x^{(k)},\cdot)$. We can then apply density estimation techniques such as kernel- or spectral density estimation. In any realistic scenario, we would however not to be able to approximate $p(x^{(k)},\cdot)$ for a uniform covering of $\mathbb{X}$ by a sufficient amount of points, due to the typically high dimension of $\mathbb{X}$. For example, a regular grid-based covering of the unit cube $[0,1]^n$ with spacing $\delta > 0$ would require $M = \left(\frac{1}{\delta}\right)^n$ grid points. This exponential dependence of M on the dimension renders classical numerical quadrature schemes infeasible, leaving Monte Carlo (MC) quadrature as the only viable option for solving the integral over $\mathbb{X}$.

To quantify the MC error in M for this integral, define

$$f(x) := \frac{1}{|\mathbb{Z}|}\int_{\mathbb{Z}}\int_{\Sigma_\vartheta(z)}\int_{\Sigma_\vartheta(z)}\left|\frac{p(x,y^{(1)})}{\pi(y^{(1)})} - \frac{p(x,y^{(2)})}{\pi(y^{(2)})}\right| d\mu_z(y^{(1)})\,d\mu_z(y^{(2)})\,dz, \tag{26}$$

$$I(f) := \int_{\mathbb{X}} f(x)\,d\mu(x), \tag{27}$$

$$I_M(f) := \frac{1}{M}\sum_{i=1}^{M} f(x^{(k)}), \quad x^{(k)} \sim \mu \text{ i.i.d.}. \tag{28}$$

For the expected approximation error then holds [21, Chap. 4]

$$\mathbb{E}\left[|I(f) - I_M(f)|\right] = \frac{\mathrm{Var}_\mu(f)}{\sqrt{M}}, \tag{29}$$

where $\mathrm{Var}_\mu(f)$ denotes the variance of f with respect to μ, defined by

$$\mathrm{Var}_\mu(f) = \mathbb{E}_\mu\left[f^2\right] - \left(\mathbb{E}_\mu[f]\right)^2.$$

The independence of the convergence rate $1/\sqrt{M}$ from the dimension is what gives MC methods an edge above conventional methods, at least in theory. However, the *effective* convergence speed, and hence the required number of start points $x^{(k)}$, is highly influenced by the prefactor $\mathrm{Var}_\mu(f)$.

We will show now that for highly lumpable systems, $\mathrm{Var}_\mu(f)$ tends to be substantially smaller than for non-lumpable systems. The intuitive explanation is that, for systems that are lumpable with respect to some reaction coordinate ξ, f is essentially constant along every level set of ξ, hence only the variation of f along ξ contributes to $\mathrm{Var}_\mu(f)$. This holds even in the candidate reaction coordinate ϑ that appears in the definition of f is not close to ξ. The intuition is formalized by the following theorem:

Theorem 5.1. *Assume that the system is ε-lumpable with respect to $\xi : \mathbb{X} \to \mathbb{Z}$ and the effective density $p_L : \mathbb{Z} \times \mathbb{X} \to \mathbb{R}$. Define $f_L : \mathbb{Z} \to \mathbb{R}$ by*

$$f_L(z) := \frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z')} \int_{\Sigma_\vartheta(z')} \left| \frac{p_L(z, y^{(1)})}{\pi(y^{(1)})} - \frac{p_L(z, y^{(2)})}{\pi(y^{(2)})} \right| d\mu_z(y^{(1)})\, d\mu_z(y^{(2)})\, dz'. \tag{30}$$

Furthermore, define the effective invariant density *by*

$$\bar{\pi}(z) := \int_{\Sigma_\xi(z)} \pi(x)\, d\sigma_z(x) \tag{31}$$

where σ_z denotes the surface measure on $\Sigma_\xi(z)$. Then there exists a constant $C > 0$ such that

$$|\mathrm{Var}_\mu(f) - \mathrm{Var}_\mu(f_L \circ \xi)| \leq 2(1 + C^2)\|f_L \circ \xi\|_{L^1_\mu} \|\bar{\pi}\|_\infty \varepsilon + \mathcal{O}(\varepsilon^2).$$

In words, the variance of f is ε-close to the variance of $f_L \circ \xi$. Also, the variance of $f_L \circ \xi$ is equal to the variance of f_L:

Lemma 5.2. *Assume that the system is pointwise ε-lumpable with effective density p_L, and let f_L and $\bar{\mu}$ be defined as in Lemma 5.1. Then*

$$\mathrm{Var}_\mu(f_L \circ \xi) = \mathrm{Var}_{\bar{\mu}}(f_L).$$

The proof of Theorem 5.1 and Lemma 5.2 can be found in Appendix B.

Remark 5.3. To summarize, the variance of f is ε-close to the variance of f_L. As f_L is defined on $\mathbb{Z}$, the variance of f cannot depend on the full phase space dimension n through the dimension of its argument.

Looking at the definition of f_L, it is clear (and also intutitive) that the variance of the effective transition density $p_L(z, y)$ in its first argument influences $\mathrm{Var}_{\bar{\mu}}(f_L)$. Note, however, that the dimension n also indirectly appears in the definition of f_L, through the integrations over the $(n - r)$-dimensional level sets $\Sigma_\vartheta(z)$. As the candidate reaction coordinate ϑ can be arbitrarily complex, it is very hard to give any general estimates on the influence of ϑ and n on $\mathrm{Var}_{\bar{\mu}}(f_L)$. However, we found no argument why $\mathrm{Var}_{\bar{\mu}}(f_L)$ should *increase* with increasing n in general. Indeed, in Section 6.1, we present an example system for which $\mathrm{Var}_{\bar{\mu}}(f_L)$ is inversely correlated to n. There, we also numerically confirm the predicted dependence of the Monte Carlo error on $\mathrm{Var}_\mu(f)$, and the predicted dependence of $\mathrm{Var}_\mu(f)$ on ε.

5.2. Further numerical challenges

In order to solve the optimization problem (23) in practice, several more numerical challenges will need to be overcome. These challenges do however not involve the collection of dynamical data, i.e., the expensive realization of the dynamical system, and hence can be considered peripheral. We will therefore only briefly sketch possible solutions to these challenges, and leave the exact elaboration to future work.

Discretization of the solution space

The solution space $C(\mathbb{X}, \mathbb{Z})$ for ξ_D in (22) needs to be replaced by a finite-dimensional parametric ansatz space. While in principle classical approaches like Galerkin or grid-based finite

element discretization methods could be explored, these methods suffer heavily from the curse of dimensionality, that is, the exponential dependence of the number of parameters on the system dimension [34].

Due to its prevalence in scientific computing, many approaches have been suggested to tackle the curse of dimensionality, including sparse grids [18], mesh-free methods [17], and non-parametric approaches [42]. All of these methods are in principle compatible with the task of evaluating $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta)$, provided the level sets $\Sigma_\vartheta(z)$ can be sampled for the discretized ϑ.

However, one particular discretization method practically suggests itself for our overall task of finding the minimizer of $\widetilde{\mathscr{L}_{\mathrm{D}}}$, and that is the representation of ξ_D via a multilayer neural network [28]. These models have demonstrated impressive performance for problems of high input dimensions, such as image recognition [27], protein structure prediction [45] or, quite relevant to our task, Markov model construction via the approximation of transfer operator eigenfunctions [30]. Minimization of the loss function $\widetilde{\mathscr{L}_{\mathrm{D}}}$ can then be accomplished by gradient descent methods, of which there exist numerous efficient variants and implementations [41].

Level set integrals

Besides the integral over $\mathbb{X}$, computation of $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta)$ requires the numerical solution of an integral over $\mathbb{Z}$ and, for each $z \in \mathbb{Z}$, two surface integrals over the z-level set $\Sigma_\vartheta(z)$ of ϑ, see (25).As the dimension r of Z was assumed to be small, classical grid-based quadrature methods are sufficient for the former. The level sets $\Sigma_\vartheta(z)$, however, are $n-r$ dimensional nonlinear sub-manifolds of $\mathbb{X}$, and thus too high-dimensional for grid based methods.

The canonical approach would be again to use Monte Carlo quadrature, like for the integral over $\mathbb{X}$. This approach was used for the low-dimensional examples presented in Section 6. Sampling of the surface measure μ_z can be accomplished by restricted simulation of the dynamics, for which numerical schemes were suggested in [7]. Moreover, if ϑ is modeled by a neural network, sampling $\Sigma_\vartheta(z)$ could be alternated with the optimization step, leading to an optimization scheme similar to stochastic gradient descent.

6. Numerical examples

6.1. Timescale-separated Brownian motion

We consider a process $(X_t) = (X_t^{(1)}, \ldots, X_t^{(n)})$ on the n-dimensional torus $\mathbb{T}^n := [-\pi, \pi]^n$ for $n \geq 2$ with slow diffusion in the first coordinate direction (standard Brownian motion) and instantaneous and pairwise independent equilibrations to the uniform distribution in the $n-1$ remaining coordinate components, i.e.,

$$\begin{aligned} \mathrm{d}X_t^{(1)} &= \mathrm{d}W_t, \\ X_t^{(i)} &\sim U([-\pi, \pi]), \quad 2 \leq i \leq n, \end{aligned}$$

such that

$$\begin{aligned} X_{t_1}^{(i)} &\perp X_{t_2}^{(i)}, \quad 1 \leq i \leq n, t_1 \neq t_2, \\ X_{t_1}^{(i)} &\perp X_{t_2}^{(j)} \quad 1 \leq i \neq j \leq n, \forall t_1, t_2, \end{aligned}$$

with a standard Brownian motion W_t. We introduce a second process $(\widetilde{X}_t)$ on $\mathbb{T}^n$ which is parametrized by a variance $\sigma^2 > 0$ by the global diffusion

$$\begin{aligned} \mathrm{d}\widetilde{X}_t^{(1)} &= \mathrm{d}\widetilde{W}_t^{(1)} \\ \mathrm{d}\widetilde{X}_t^{(i)} &= \sigma \mathrm{d}\widetilde{W}_t^{(i)}, \quad 2 \leq i \leq n, \end{aligned}$$

where the $\widetilde{W}_t^{(i)}, 1 \leq i \leq n$ are pairwise independent standard Brownian motions.

6.1.1. Transition densities

The transition densities of (X_t) and $(\widetilde{X}_t)$ can be conveniently described in terms of the one-dimensional *wrapped normal distribution* [20] on the circle $[-\pi, \pi]$ with variance $\sigma^2 \geq 0$, which is defined by the density $g^\sigma : [-\pi, \pi] \times [-\pi, \pi] \to \mathbb{R}_+$ given by

$$g^\sigma(x, y) := \frac{1}{2\pi} \left(1 + 2 \sum_{k=1}^{\infty} \rho^{k^2} \cos(k(y - x)) \right), \text{ where } \rho = \exp(-\sigma^2/2).$$

In particular, for $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{T}^n$ we straightforwardly obtain the transition density of (X_t) as

$$p^{\tau,\infty}(x, y) := \frac{1}{(2\pi)^{n-1}} \, g^\tau(x_1, y_1)$$

as well as the transition density of $(\widetilde{X}_t)$ as

$$p^{\tau,\sigma}(x, y) := g^\tau(x^{(1)}, y^{(1)}) \prod_{i=2}^{n} g^{\tau\sigma}(x_i, y_i). \tag{32}$$

The transition densities are illustrated in Figure 1 for the case $n = 2$.

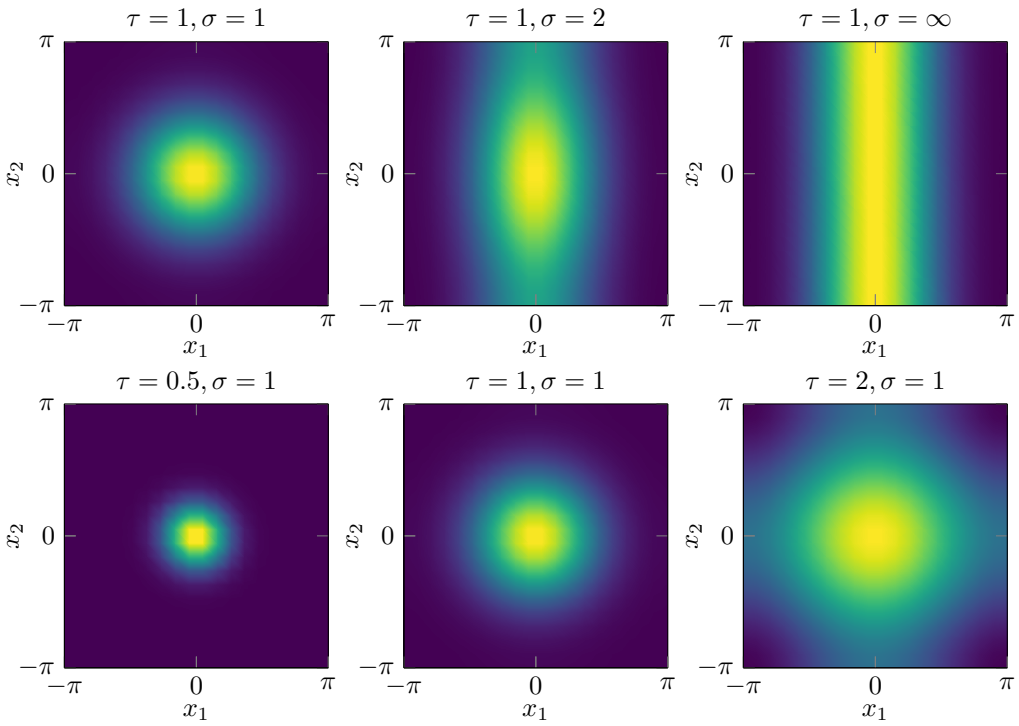

Figure 1.: Illustration of transition densities $p^{\tau,\sigma}(0, \cdot)$ for different lag times τ and standard deviations σ in the case $n = 2$. The top row illustrates how the transition densities $p^{\tau,\sigma}(x, \cdot)$ approximate $p^{\tau,\infty}(x, \cdot)$ for an increasing standard deviation σ when τ is fixed. Note that for fixed x_1 coordinates, the limiting density $p^{\tau,\infty}(x, \cdot)$ is constant along the x_2-coordinate. The bottom row illustrates the larger degree of global dispersion for inceasing lag times when σ is fixed.

6.1.2. Lumpability

As σ increases, the transition density $p^{\tau,\sigma}(x, y)$ uniformly approximates $p^{\tau,\infty}(x, y)$, we may therefore understand (X_t) as the limiting process of $(\widetilde{X}_t)$ for large variances. Because the transition

density of $p^{\tau,\infty}(x,y)$ only depends on the first coordinate components x_1, y_1, we can verify lumpability of $(\widetilde{X}_t)$ for large enough σ, as we will briefly illustrate now.

Consider the reaction coordinate $\xi : [-\pi,\pi]^n \to [-\pi,\pi]$ given by the orthogonal projection onto the first coordinate component $\xi((x_1,x_2,\ldots,x_n) = x_1$. We call ξ the *optimal* reaction coordinate. With the effective transition density

$$p_L^\tau(z,y) := \frac{1}{(2\pi)^{n-1}}\, g^\tau(z,y_1),$$

one can then show that

$$\left\| p^{\tau,\sigma}(*,\cdot) - p_L^\tau(\xi(*),\cdot) \right\|_{\mathbb{K}} = \mathcal{O}\left(\exp(-(\tau\sigma)^2/2)\right). \tag{33}$$

which for increasing σ or τ becomes arbitrarily small. The derivation of (33) can be found in Appendix C. In particular, for any lag time $\tau > 0$, we will find a σ for which the system becomes arbitrarily lumpable, while in the limit $\sigma \to \infty$, the system is "perfectly" lumpable (i.e., $\left\| p^{\tau,\sigma}(*,\cdot) - p_L^\tau(\xi(*),\cdot) \right\|_{\mathbb{K}} = 0$) for every $\tau > 0$. Note that ε-lumpability of $(\tilde{X}_t)$ also implies ε-deflatability due to Proposition 3.6.

As our estimates of the distance $\left\| p^{\tau,\sigma}(*,\cdot) - p_L^\tau(\xi(*),\cdot) \right\|_{\mathbb{K}}$ in Appendix C are asymptotic in nature, we can give no precise formula for the dependence of the minimal ε in (L) on σ. In particular, we cannot predict the dependence of ε on the system dimension n. Nevertheless, when computing the distance numerically, its dependence on n appears to be moderate, and seems to diminish with growing n (Figure 2). Hence, we can expect a high degree of lumpability for moderately large σ and moderately high dimensions. Also, the computations confirm the convergence rate predicted in (33).

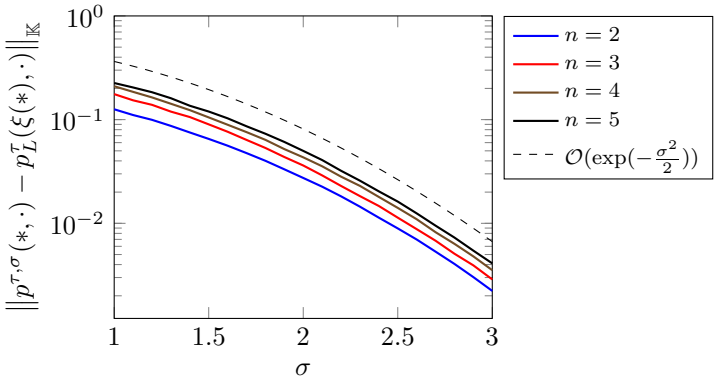

Figure 2.: $\mathbb{K}$-distance of the transition density $p^{\tau,\sigma}$ to the effective transition density p_L^τ, indicating the degree of lumpability. We observe the predicted decay in the diffusion constant σ, as well as a relative insensitivity to the system dimension n.

6.1.3. Loss function illustration

Next, we investigate the dependence of the loss function $\widetilde{\mathscr{L}_{\mathrm{D}}}$ on the reaction coordinate, for the two-dimensional process (one slow and one fast component). To be precise, we investigate the dependence of $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta_\alpha)$ on the parameter α of the family of "test" reaction coordiantes

$$\vartheta_\alpha(x) := \frac{1}{\pi(\cos|\alpha| + \sin|\alpha|)} \begin{pmatrix} \cos\alpha & \sin\alpha \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \alpha \in (-\pi,\pi). \tag{34}$$

for different values of the diffusion constant σ. The level sets of ϑ_α are one-dimensional hyperplanes that intersect the x_1-axis with angle $\pi/2 - \alpha$. The prefactor in (34) ensures that $\mathbb{Z} = \mathrm{range}(\vartheta_\alpha) = [-1,1]$ for all α. Figure 3 illustrates ϑ_α and its level sets.

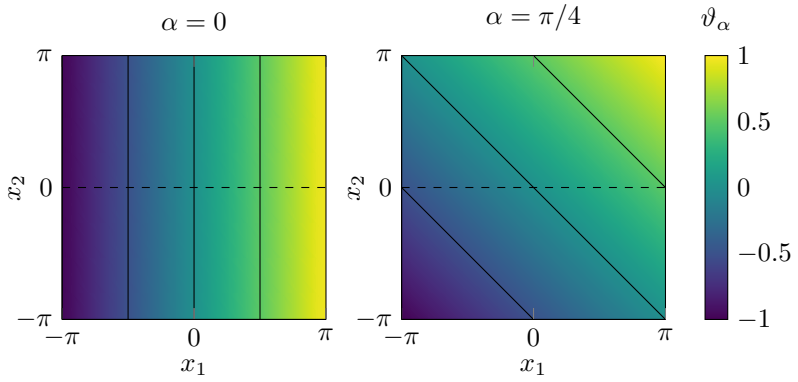

Figure 3.: The linear reaction coordinate ϑ_α for two different values of α.

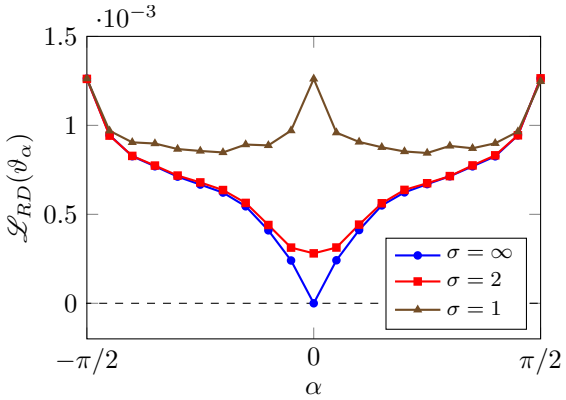

Figure 4.: The differential deflatability loss function of the reaction coordinate ϑ_α for $\alpha \in (-\pi/2, \pi/2)$ and different values of σ.

We use a combination of symbolic computation and numerical quadrature techniques for computing $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta_\alpha)$, using Mathematica and Matlab. In particular, for sampling points on the level sets $\Sigma_{\vartheta_\alpha}(z)$, required for numerically computing the two innermost integrals in (25), we utilized the Matlab function `randFixedLinearCombination`, written by John D'Errico and provided through MatlabCentral [10]. The scripts containing our computations are provided in the SI. We do however *not* suggest to use these scripts for the computation and minimization of $\widetilde{\mathscr{L}_{\mathrm{D}}}$ in real-world systems, as in particular the integration over the $(n-r)$-dimensional level sets quickly become infeasible. See the discussion in Section 5.2 for a first outlook on how we plan to solve the optimization problem in practice.

The differential deflatability loss function $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta_\alpha)$ in dependence of α is shown in Figure 4. We observe that, for $\sigma = 2$ and $\sigma = \infty$, $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta_\alpha)$ indeed takes its unique global minimum for $\alpha = 0$, i.e., the optimal reaction coordinate $\vartheta_0(x) = \xi(x) = x_1$. We also observe that for $\sigma = 2$, the minimum value is not zero, as the system is not "perfectly lumpable" with respect to ϑ_0, whereas for $\sigma = \infty$ it is. Further, the "worst" reaction coordinates $\vartheta_{\pm\pi/2}(x) = \pm x_2$, which project onto the system's fast instead of its slow coordinate, correctly get assigned the global maximum value. For $\alpha = 1$, on the other hand, the diffusion is isotropic, and hence the "slow directions" coincide with the directions of largest extent, which are the diagonals of the domain $[-\pi, \pi]^2$. Consequently, $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta_\alpha)$ becomes minimal for $\alpha = \pm\pi/4$, for which ϑ_α projects onto the diagonals $\{x_1 = \pm x_2\}$.

The behavior of $\widetilde{\mathscr{L}_{\mathrm{D}}}$ is therefore perfectly consistent with our intuition, and we can expect to identify the optimal reaction coordinate by minimizing $\widetilde{\mathscr{L}_{\mathrm{D}}}$.

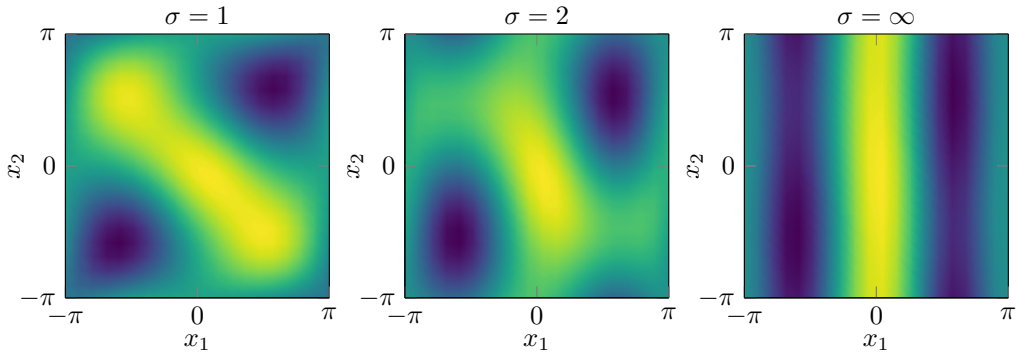

Figure 5.: The function f, i.e., the integrand of the integral over $\mathbb{X}$ in $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta)$ for $\vartheta(x) = x_1 + x_2$ and different values of σ.

6.1.4. Loss function Monte Carlo error

Finally we investigate the Monte Carlo quadrature error for the integral over $\mathbb{X}$ in $\widetilde{\mathscr{L}_{\mathrm{D}}}$. As discussed in Section 5.1, this error is, besides the number of sample points M, determined by the variance of the integrand f, which for the current system takes the form

$$f(x) = \frac{(2\pi)^n}{|\vartheta_{\max} - \vartheta_{\min}|} \int_{\vartheta_{\min}}^{\vartheta_{\max}} \int_{\Sigma_\vartheta(z)} \int_{\Sigma_\vartheta(z)} \left| p^{\tau,\sigma}(x, y^{(1)}) - p^{\tau,\sigma}(x, y^{(2)}) \right| d\sigma_z(y^{(1)})\, d\sigma_z(y^{(2)})\, dz,$$

where $\vartheta_{\max}$ and $\vartheta_{\min}$ describe the maximum and minimum of ϑ. We point out in particular that f depends on the diffusion coefficient σ and, through the reaction coordinate ϑ, on the system dimension n. The influence of these two parameters on the Monte Carlo error is the primary subject of investigation for this section.

We consider in dimension 2 and 3, respectively, the test reaction coordinates

$$\vartheta^2(x) = x_1 + x_2, \qquad \vartheta^3(x) = x_1 + x_2 + x_3.$$

Figure 5 shows f for the reaction coordinate ϑ^2 and different values of σ. We observe that with increasing σ, f indeed becomes increasingly constant on the level sets $\{x_1 = z\}$ of the optimal reaction coordinate ξ. This behavior was predicted in Section 5.1.

Figure 6 illustrates the variance of f. We first note that the variance indeed decreases for increasing σ, and converges towards the variance of the process with instantaneously equilibrating components $x_2, \ldots, x_n$ (equivalent to choosing "$\sigma = \infty$"). Moreover, we observe that the variance of the three-dimensional process is substantially smaller compared to the two-dimensional process. This observation still holds when considering the relative variance $\mathrm{Var}[f]/\mathbb{E}[f]$. This demonstrates that higher-dimensional function do not per se possess higher variance. In fact, $\mathrm{Var}[f]$ appears to be more dependent on the choice of the particular reaction coordinate ϑ than on n, although, to avoid digression, we will refrain from detailed analysis.

Finally, we estimate the relative expected MC error

$$\frac{\mathbb{E}\left[|I(f) - I_M(f)|\right]}{I(f)}$$

where $I(f)$ and $I_M(f)$ are the exact integral and Monte Carlo integral with M samples of f defined in (27) and (28). In practice, this error indicates how many "dynamical samples" $p^{\tau,\sigma}(x^{(k)}, \cdot)$ need to be created by numerical simulation in order to approximate $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta)$ up to a given accuracy. In the present example, however, $p^{\tau,\sigma}$ is known analytically by (32).

Figure 7 shows the relative error in dependence on the sample size M for ϑ^2 and ϑ^2, each for a finite value of σ as well as $\sigma = \infty$. In all four cases we observe the expected Monte Carlo

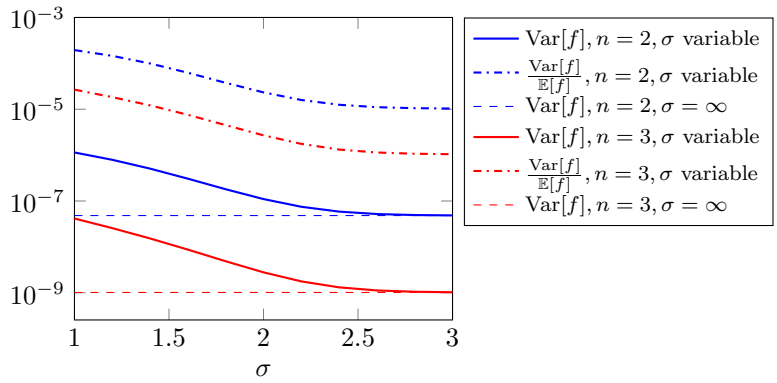

Figure 6.: (Relative) variance of f associated in dimensions 2 and 3 for various σ. We observe convergence of $\mathrm{Var}[f]$ with variable σ towards $\mathrm{Var}[f]$ associated with the σ-independent limit process X_t.

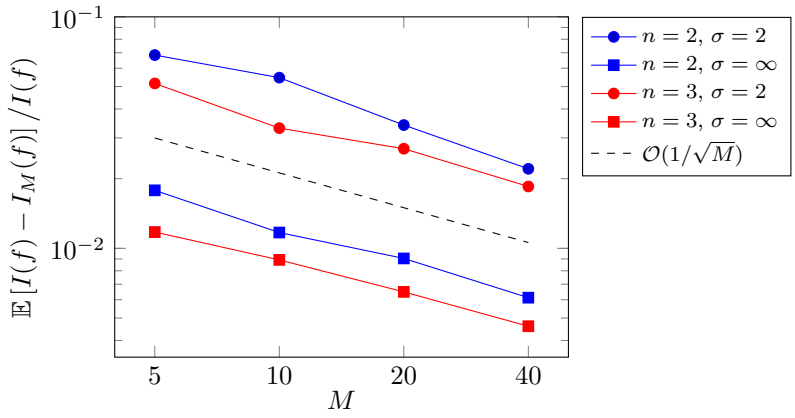

Figure 7.: The relative Monte Carlo quadrature error for the integral over $\mathbb{X}$ in $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta_w)$ for different dimensions n and different diffusion coefficients σ.

convergence rate of $\mathcal{O}(1/\sqrt{M})$. Moreover, the error is significantly smaller for higher σ, i.e., the more lumpable system. Finally, the expected error for the higher dimensional system is slightly smaller. All these phenomena are perfectly consistent with the preceding analysis of $\mathrm{Var}[f]$.

We conclude from this example that for lumpable systems, we can expect to find the optimal reaction coordinate by minimising $\widetilde{\mathscr{L}_{\mathrm{D}}}$, that the Monte Carlo approximation to $\widetilde{\mathscr{L}_{\mathrm{D}}}(\vartheta)$ requires only few dynamical samples for highly lumpable systems, and that a high dimension of the base system has no negative impact on the performance.

6.2. Metastable circular system

To demonstrate the behavior of the loss function for nonlinear optimal reaction coordinates, we consider as a second example a two-dimensional system governed by the overdamped Langevin dynamics

$$dX_t = -\nabla V(X_t) + \sqrt{2\beta^{-1}} dW_t, \tag{35}$$

where $V : \mathbb{R}^2 \to \mathbb{R}$ is the potential energy surface and $\beta > 0$ is the inverse temperature determining the strength of the Brownian motion W_t. Informally, movement of this system can be described as a random walk within the energy landscape, aiming in the direction of steepest descent of V but being disturbed by temperature-scaled white noise.

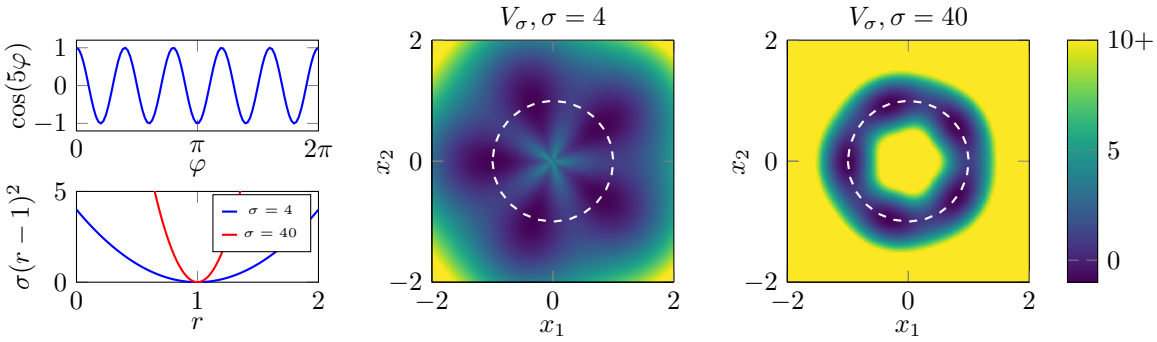

Figure 8.: Left: The two components of the potential, which depend on the angular and the radial part of the polar coordinates, respectively. Center & right: The circular potential and its components for different values of σ. The dashed white line indicates the unit circle, i.e., the minimal energy pathway connecting the local minima.

In particular, we here consider the family of potentials

$$V_\sigma(x) = \cos\left(5\varphi(x)\right) + \sigma\left(r(x) - 1\right)^2, \quad \sigma > 0,$$

where $(\varphi(x), r(x))$ describe the polar coordinates of x, i.e.,

$$\varphi(x) = \operatorname{atan2}(x), \qquad r(x) = \|x\|_2.$$

The potential consists of two components: a cosine term in the angular coordinate with five local minima of equal depth, and a quadratic term in the radial coordinate with a single minimum at $r = 1$. The full potential, shown in Figure 8, therefore possesses five local minima arranged along the unit circle. The two components φ and r as functions on $\mathbb{X}$ are shown in Figure 10 (left).

6.2.1. Metastability analysis

At moderate temperatures, the five local minima of V induce *metastability* in the system. This means that a typical trajectory will vibrate around a minimum for a long time, until suddenly, induced by sufficiently strong stochastic excitation, undergo a rapid transition to another local minimum. If additionally the equilibration in the radial direction is sufficiently fast, i.e., the parameter σ is sufficiently large, then these rare transitions will be the slowest sub-processes of the system. To confirm this, we compute the leading eigenvalues and associated eigenfunctions of the system's transfer operator $\mathcal{P}^t : L^2(\mathbb{X}) \to L^2(\mathbb{X})$, which is defined by

$$\mathcal{P}^t f(x) = \int_{\mathbb{X}} f(y) p^t(y, x)\, dy.$$

Going back to the late 90s, spectral analysis of the transfer operator, or its adjoint, the Koopman operator, forms the basis of many model reduction methods that aim to preserve the system's long timescales, for an overview see [24].

The eigenvalues of $\mathcal{P}^t$ for $t = 0.1$ and various values of σ are shown in Figure 9. We see that for $\sigma = 10$ and $\sigma = 100$, there is a significant *spectral gap* after λ_5, indicating a significant time scale separation between the associated sub-processes. Analysing the sign structure of the associated eigenfunctions confirms that the slowest processes are indeed associated with the transitions between the metastable sets (see the SI).

For small σ, on the other hand, no spectral gap can be observed, hence the metastable transitions are not considerably slower than the remaining processes. Indeed, sign analysis of the sixth eigenfunction for $\sigma = 1$ shows that the corresponding sub-process describes the equilibration in the radial direction (see the SI). As λ_6 is not significantly separated from λ_5, the radial equilibration occurs on roughly the same timescale as the metastable transitions.

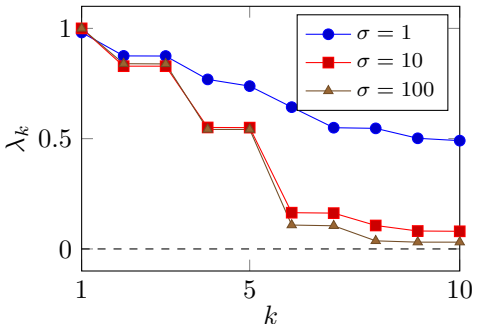

Figure 9.: Leading eigenvalues of the circular system's transfer operator for different values of σ. For larger σ, we observe a spectral gap, which indicates that the equilibration times of the metastable transitions are much slower than that of other sub-processes in the system.

6.2.2. Lossfunction computation

The preceding analysis confirms that, for high enough σ, a good reaction coordinate should resolve the transitions between the metastable sets. As these transitions occur predominantly within certain "transition channels" that surround the minimum energy pathways (MEPs) [32], and as in our system the MEPs between two neighbouring minima are segments of the unit circle, we would expect the angular component $\varphi(x)$ of the polar coordinates of x to be a good reaction coordinate. Conversely, we would expect the radial component $r(x)$ to be a bad reaction coordinate, especially for high values of σ.

To test this hypothesis, we now compute the differential lumpability loss function $\widetilde{\mathscr{L}_{\mathrm{D}}}$ for both φ and r. For the lag time parameter in $\widetilde{\mathscr{L}_{\mathrm{D}}}$, we use $\tau = 0.1$, as the observed spectral gap for this time indicates that the time scale separation between the slow and fast processes has already manifested itself. We again solve the various integrals in (25) by Monte Carlo quadrature. One complication over the simple slow-fast system is that the transition density functions $p^\tau(x, \cdot)$ are not known analytically, so they have to be approximated empirically. For that, we first draw an empirical sample of $p^t(x, \cdot)$ by simulating many trajectories with starting point x, and then apply the kernel density estimation algorithm to the end points. For details on the numerical implementation see the SI.

Figure 10 (right) shows the loss function for the two reaction coordinates in dependendce of σ. We see that, indeed, $\widetilde{\mathscr{L}_{\mathrm{D}}}(\varphi) < \widetilde{\mathscr{L}_{\mathrm{D}}}(r)$ for all values of σ, hence φ is the better reaction coordinate. Moreover, for increasing σ, $\widetilde{\mathscr{L}_{\mathrm{D}}}(\varphi)$ continually decreases[1], whereas $\widetilde{\mathscr{L}_{\mathrm{D}}}(r)$ increases. This agrees well with our intuitive understanding of the role of σ: with increasing σ, the radial component r equilibrates more quickly, so the long-term future evolution of X_t depends more and more only on $\varphi(X_t)$ (i.e., the degree of lumpability with respect to φ increases), and less and less on $r(X_t)$ (i.e., the degree of lumpability with respect to r decreases).

7. Discussion and outlook

In this paper, we formally characterized optimal reaction coordinates in continuous Markovian systems, that is, observables that optimally describe latent long-term mechanisms of the dynamics. We have seen, by both theoretical analysis (Section 3.3) and numerical examples (Section 6), that our definition is applicable to several common types of multiscale systems, such as slow-fast systems and metastable systems. To add further interpretability to our definition, it would be desireable to draw a formal connection to the well-established transition path theory [13, 32], which

[1] Note that the flattening of $\widetilde{\mathscr{L}_{\mathrm{D}}}(\varphi)$ towards the right side of the plot is a numerical artefact. For high degrees of lumpability, miniscule differences between transition densities need to be quantified, which presents a challenge to our fixed-bandwidth kernel density estimator.

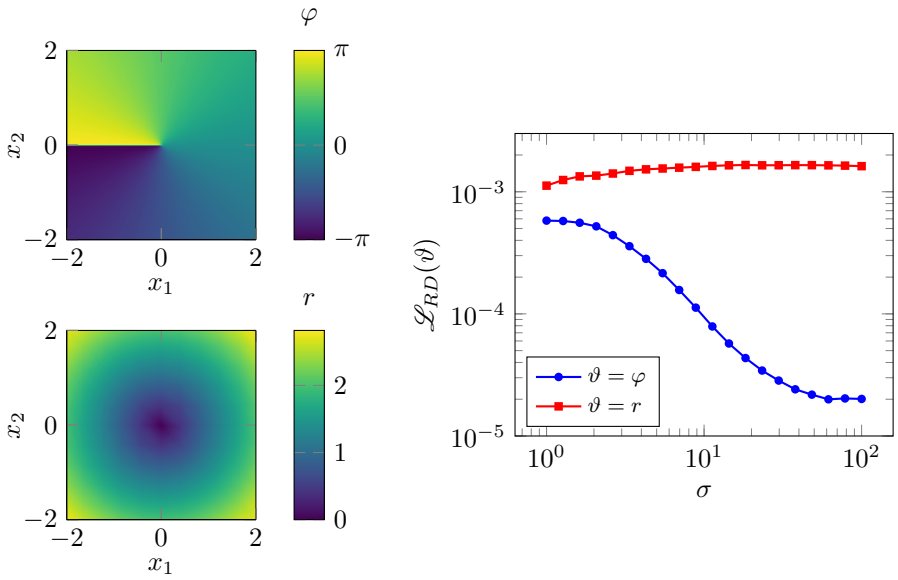

Figure 10.: Left: the two components φ and r as functions on $\mathbb{X}$. Right: Loss function $\widetilde{\mathscr{L}_{\mathrm{D}}}$ for φ and r interpreted as reaction coordinates.

characterizes reaction coordinates in terms of committor functions and minimum energy pathways on the potential energy surface. In [2], a connection between the transition manifold approach (a predecessor to the present characterization) and transition path theory has been discussed rather informally, but a rigorous investigation is still outstanding.

We then presented a variational formulation of this characterization as a computational strategy for the discovery of optimal reaction coordinates. In particular, that variational formulation provides a leverage point for modern machine learning methods, such as deep learning and stochastic gradient descent. The implementation of these methods and their demonstration of efficiency are subject of ongoing work. The reduced data requirement for the optimization problem that was demonstrated in this paper raises confidence that we will be able to apply our method to high dimensional problems such as the identification of reaction coordinates in large molecular systems.

Acknowledgements

This research has been funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 "Scaling Cascades in Complex Systems", Project Number 235221301, Project B03 "Multilevel coarse graining of multiscale problems", and by Deutsche Forschungsgemeinschaft (DFG) through grant EXC 2046 "MATH+", Project Number 390685689, Project AA1-2 "Learning Transition Manifolds and Effective Dynamics of Biomolecules".

References

[1] J. Berner, P. Grohs, G. Kutyniok, and P. Petersen. The Modern Mathematics of Deep Learning. *arXiv:2105.04026 [cs, stat]*, May 2021.

[2] A. Bittracher, R. Banisch, and C. Schütte. Data-driven computation of molecular reaction coordinates. *The Journal of Chemical Physics*, 149(15):154103, 2018.

[3] A. Bittracher, S. Klus, B. Hamzi, P. Koltai, and C. Schütte. Dimensionality Reduction of Complex Metastable Systems via Kernel Embeddings of Transition Manifolds. *J Nonlinear Sci*, 31(1):3, Dec. 2020.

[4] A. Bittracher, P. Koltai, S. Klus, R. Banisch, M. Dellnitz, and C. Schütte. Transition Manifolds of Complex Metastable Systems: Theory and Data-driven Computation of Effective Dynamics. *J. Nonlinear Sci.*, 28(2):471–512, 2017.

[5] A. Bittracher and C. Schütte. A Weak Characterization of Slow Variables in Stochastic Dynamical Systems. In O. Junge, O. Schütze, G. Froyland, S. Ober-Blöbaum, and K. Padberg-Gehle, editors, *Advances in Dynamics, Optimization and Computation*, Studies in Systems, Decision and Control, pages 132–150, Cham, 2020. Springer International Publishing.

[6] A. Bittracher and C. Schütte. A probabilistic algorithm for aggregating vastly undersampled large Markov chains. *Physica D: Nonlinear Phenomena*, 416:132799, Feb. 2021.

[7] G. Ciccotti, T. Lelievre, and E. Vanden-Eijnden. Projection of Diffusions on Submanifolds: Application to Mean Force Computation. *Communications on Pure and Applied Mathematics*, 61(3):371–408, 2007.

[8] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, July 2006.

[9] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behaviour. *SIAM J. Num. Anal.*, 36(2), 1999.

[10] J. D'Errico. randFixedLinearCombination. https://de.mathworks.com/matlabcentral/fileexchange/49795-randfixedlinearcombination, Retrieved June 16, 2021.

[11] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. *Linear algebra and its applications*, 398:161–184, 2005.

[12] P. Drineas and M. W. Mahoney. On the Nyström Method for Approximating a Gram Matrix for Improved Kernel-Based Learning. *J. Mach. Learn. Res.*, 6:2153–2175, Dec. 2005.

[13] W. E and E. Vanden-Eijnden. Towards a Theory of Transition Paths. *J. Stat. Phys.*, 123(3):503–523, 2006.

[14] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY, 2013.

[15] T. A. Frewen, G. Hummer, and I. G. Kevrekidis. Exploration of effective potential landscapes using coarse reverse integration. *J Chem Phys*, 131(13), Oct. 2009.

[16] G. Froyland, G. A. Gottwald, and A. Hammerlindl. A Computational Method to Extract Macroscopic Variables and Their Dynamics in Multiscale Systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1816–1846, Jan. 2014.

[17] S. Garg and M. Pant. Meshfree Methods: A Comprehensive Review of Applications. *Int. J. Comput. Methods*, 15(04):1830001, June 2018.

[18] M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numerische Mathematik*, 83(2):279–312, 1999.

[19] N. Halko, P. Martinsson, and J. Tropp. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, 53(2):217–288, 2011.

[20] S. R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*, volume 5 of *Series on Multivariate Analysis*. WORLD SCIENTIFIC, Apr. 2001.

[21] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods*. John Wiley & Sons, June 2009.

[22] J. Kemeny and J. L. Snell. *Finite Markov Chains*. Springer-Verlag New York, 1976.

[23] Y. Khoo, J. Lu, and L. Ying. Solving for high-dimensional committor functions using artificial neural networks. *Res Math Sci*, 6(1):1, Oct. 2018.

[24] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé. Data-Driven Model Reduction and Transfer Operator Approximation. *J Nonlinear Sci*, 28(3):985–1010, June 2018.

[25] S. Klus, I. Schuster, and K. Muandet. Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces. *Journal of Nonlinear Science*, 30(1):283–315, Feb. 2020.

[26] U. Krengel and A. Brunel. *Ergodic Theorems*. Number 6 in De Gruyter Studies in Mathematics. W. de Gruyter, Berlin ; New York, 1985.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017.

[28] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[29] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, 125(2):024106, 2006.

[30] A. Mardt, L. Pasquali, H. Wu, and F. Noé. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.*, 9(1):5, 2018.

[31] R. T. McGibbon, B. E. Husic, and V. S. Pande. Identification of simple reaction coordinates from complex dynamics. *The Journal of Chemical Physics*, 146(4):044109, 2017.

[32] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *The Journal of Chemical Physics*, 125(8):084110, Aug. 2006.

[33] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23):3634–3637, 1994.

[34] E. Novak and H. Woźniakowski. Approximation of infinitely differentiable multivariate functions is intractable. *Journal of Complexity*, 25(4):398–404, Aug. 2009.

[35] F. Nüske, P. Koltai, L. Boninsegna, and C. Clementi. Spectral Properties of Effective Dynamics from Conditional Expectations. *Entropy*, 23(2):134, Feb. 2021.

[36] G. A. Pavliotis and A. M. Stuart. *Multiscale Methods: Averaging and Homogenization*. Springer Science & Business Media, 2008.

[37] G. Perez-Hernandez, F. Paul, T. Giorgino, G. de Fabritiis, and F. Noé. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics*, 139(1):015102, 2013.

[38] A. Rahimi and B. Recht. Random Features for Large-Scale Kernel Machines. page 8.

[39] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1(none):20–71, Jan. 2004.

[40] M. Rosenblatt. Transition probability operators. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 2*, pages 473–483, Berkeley, Calif., 1967. University of California Press.

[41] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv:1609.04747 [cs]*, June 2017.

[42] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT press, Cambridge, USA, 2001.

[43] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.*, 151(1):146–168, 1999.

[44] C. Schütte and M. Sarich. *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*. Courant Lecture Notes in Mathematics. American Mathematical Society, 2014.

[45] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, Jan. 2020.

[46] A. Singer, R. Erban, I. G. Kevrekidis, and R. R. Coifman. Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38):16090–5, 2009.

[47] M. Udell and A. Townsend. Why Are Big Data Matrices Approximately Low Rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, Jan. 2019.

[48] W. Wang and A. Roberts. Slow manifold and averaging for slow–fast stochastic differential system. *Journal of Mathematical Analysis and Applications*, 398(2):822–839, Feb. 2013.

[49] C. Wehmeyer and F. Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.*, 148(24):241703, June 2018.

[50] R. Zwanzig, A. Szabo, and B. Bagchi. Levinthal's paradox. *PNAS*, 89(1):20–22, Jan. 1992.

A. Extended multiscale expansion of slow-fast systems

In this section, we continue the multiscale analysis for the slow-fast system from Section 3.3.2. That is, we derive an evolution equation for the dominant component q_0^t of the transition density. Comparing the terms of order ε^0 in (11) yields

$$\partial_t q_0^t(x,\cdot) = \mathcal{L}_z q_1^t(x,\cdot) + \mathcal{L}_y q_0^t(x,\cdot). \tag{36}$$

The middle term is zero, which can be seen as follows: Let the averaging operator $\Pi : L^1_\mu(\mathbb{X}) \to L^1_{\bar\mu}(\mathbb{Z})$ be defined by

$$\Pi g(z) = \frac{\int_{\mathbb{Y}} g(y,z)\pi(y,z)\,dy}{\bar\pi(z)} \quad \text{where} \quad \bar\pi(z) = \int_{\mathbb{Y}} \pi(y,z)\,dy \tag{37}$$

and $\bar\mu$ is the measure induced by $\bar\pi$. As $q_0^t(x,\cdot)$ is independent of y in its first argument, we have

$$\partial_t q_0^t(x,\cdot) = \frac{1}{\bar\pi(z)}\Pi\partial_t q_0^t(x,\cdot) \quad \text{and} \quad \mathcal{L}_y q_0^t(x,\cdot) = \frac{1}{\bar\pi(z)}\Pi\mathcal{L}_y q_0^t(x,\cdot).$$

Furthermore, $\Pi\mathcal{L}_z = 0$. It follows that $\mathcal{L}_z q_1^t(x,\cdot) = 0$.

Therefore, applying Π to (36) gives

$$\partial_t \Pi q_0^t(x,\cdot) = \Pi\mathcal{L}_y q_0^t(x,\cdot).$$

As both $q_0^t(x,\cdot)$ and $\mathcal{L}_y g$ (for all g) are independent of y, we finally get for the evolution equation for q_0^t

$$\partial_t q_0^t(x,\cdot) = \mathcal{L}_y q_0^t.$$

Hence, with $\overline{\mathcal{L}} := \Pi \mathcal{L}_y \Pi$, the evolution equation of $q_0^t(x,\cdot)$ up to order ε becomes

$$\partial_t q_0^t(x,\cdot) = \overline{\mathcal{L}} q_0^t(x,\cdot) + \mathcal{O}(\varepsilon) \quad \text{for all } x \in \mathbb{X}.$$

B. Proof of Theorem 5.1

The proof of consists of simple integral and norm estimations. The main argument, used multiple times in the final proof, is formulated in the following lemma:

Lemma B.1. *Assume that the system is ε-lumpable with respect to $\xi : \mathbb{X} \to \mathbb{Z}$ and the effective density $p_L : \mathbb{Z} \times \mathbb{X} \to \mathbb{R}$. Let f_L and $\operatorname{Var}\pi$ be defined as in Theorem 5.1. Then*

$$\|f - f_L \circ \xi\|_{L^1_\mu} \leq \|\bar{\pi}\|_\infty \, \varepsilon. \tag{38}$$

Proof. Writing out the left hand side of (38), we get

$$\begin{aligned}\|f - f_L \circ \xi\|_{L^1_\mu} = \int_{\mathbb{X}} \Bigg| &\left(\frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} \int_{\Sigma_\vartheta(z)} \left| \frac{p(x,y^{(1)})}{\pi(y^{(1)})} - \frac{p(x,y^{(2)})}{\pi(y^{(2)})} \right| d\mu_z(y^{(1)})\, d\mu_z(y^{(2)})\, dz \right) \\ &- \left(\frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} \int_{\Sigma_\vartheta(z)} \left| \frac{p_L(\xi(x),y^{(1)})}{\pi(y^{(1)})} - \frac{p_L(\xi(x),y^{(2)})}{\pi(y^{(2)})} \right| d\mu_z(y^{(1)})\, d\mu_z(y^{(2)})\, dz \right) \Bigg| \, d\mu(x).\end{aligned}$$

Applying the reverse triangle inequality, this becomes

$$\begin{aligned}\|f - f_L \circ \xi\|_{L^1_\mu} &\leq \int_{\mathbb{X}} \frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} \int_{\Sigma_\vartheta(z)} \left| \frac{p(x,y^{(1)})}{\pi(y^{(1)})} - \frac{p(x,y^{(2)})}{\pi(y^{(2)})} \right. \\ &\qquad \left. - \frac{p_L(\xi(x),y^{(1)})}{\pi(y^{(1)})} + \frac{p_L(\xi(x),y^{(2)})}{\pi(y^{(2)})} \right| d\mu_z(y^{(1)})\, d\mu_z(y^{(2)})\, d\mu(x)\, dz \\ &\leq \int_{\mathbb{X}} \frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} \int_{\Sigma_\vartheta(z)} \left| \frac{p(x,y^{(1)})}{\pi(y^{(1)})} - \frac{p_L(\xi(x),y^{(1)})}{\pi(y^{(1)})} \right| d\mu_z(y^{(1)})\, d\mu_z(y^{(2)})\, dz\, d\mu(x) \\ &\quad + \int_{\mathbb{X}} \frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} \int_{\Sigma_\vartheta(z)} \left| \frac{p(x,y^{(2)})}{\pi(y^{(2)})} - \frac{p_L(\xi(x),y^{(2)})}{\pi(y^{(2)})} \right| d\mu_z(y^{(1)})\, d\mu_z(y^{(2)})\, dz\, d\mu(x).\end{aligned}$$

In each of the two summands, the ingrand is independent of $y^{(2)}$ and $y^{(1)}$, respectively, hence one integral over $\Sigma_\vartheta(z)$ becomes the factor $\bar{\pi}(z)$:

$$\begin{aligned}\|f - f_L \circ \xi\|_{L^1_\mu} &\leq 2 \int_{\mathbb{X}} \frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} |p(x,y) - p_L(\xi(x),y)|\, \bar{\pi}(z)\, d\sigma_z(y)\, dz\, d\mu(x) \\ &\leq 2\, \|\bar{\pi}\|_\infty \int_{\mathbb{X}} \frac{1}{|\mathbb{Z}|} \int_{\mathbb{Z}} \int_{\Sigma_\vartheta(z)} |p(x,y) - p_L(\xi(x),y)|\, d\sigma_z(y)\, dz\, d\mu(x)\end{aligned}$$

By the coarea formula, the inner two integrals simply describe the integration over $\mathbb{X}$ with respect to the Lebesgue measure, i.e.,

$$\|f - f_L \circ \xi\|_{L^1_\mu} \leq 2\, \|\bar{\pi}\|_\infty \frac{1}{|\mathbb{Z}|} \int_{\mathbb{X}} \|p(x,\cdot) - p_L(\xi(x),\cdot)\|_{L^1}\, d\mu(x).$$

The integral over $\mathbb{X}$ is the L^1_μ norm, so the overall expression is the $\mathbb{K}$-norm (see (3)):

$$\|f - f_L \circ \xi\|_{L^1_\mu} \leq 2\, \|\bar{\pi}\|_\infty \frac{1}{|\mathbb{Z}|} \|p(*,\cdot) - p_L(\xi(*),\cdot)\|_{\mathbb{K}}$$

which by the ε-lumpability assumption (L) is

$$\leq 2 \left\|\bar{\pi}\right\|_\infty \varepsilon.$$

□

The proof of the main result now consists only of reducing the difference of the variances to the expression $\|f - f_L \circ \xi\|_{L^1_\mu}$.

Proof of Theorem 5.1. We have

$$\begin{aligned}
\left|\mathrm{Var}_\mu(f) - \mathrm{Var}_\mu(f_L \circ \xi)\right| &= \left|\mathbb{E}_\mu\left[f^2 - (f_L \circ \xi)^2\right] - \left(\mathbb{E}_\mu[f]^2 - \mathbb{E}_\mu[f_L \circ \xi]^2\right)\right| \\
&\leq \underbrace{\left|\mathbb{E}_\mu\left[f^2 - (f_L \circ \xi)^2\right]\right|}_{=:(\star)} + \underbrace{\left|\mathbb{E}_\mu[f]^2 - \mathbb{E}_\mu[f_L \circ \xi]^2\right|}_{=:(\star\star)}.
\end{aligned}$$

For the first summand, we get

$$\begin{aligned}
(\star) &= \left\|f^2 - (f_L \circ \xi)^2\right\|_{L^1_\mu} = \left\|(f - f_L \circ \xi)(f + f_L \circ \xi)\right\|_{L^1_\mu} \\
&\leq \left\|f - f_L \circ \xi\right\|_{L^2_\mu} \left\|f + f_L \circ \xi\right\|_{L^2_\mu}.
\end{aligned}$$

As μ is a finite measure on $\mathbb{X}$, we have $L^2_\mu(\mathbb{X}) \subset L^1_\mu(\mathbb{X})$, hence there exists a $C > 0$ such that

$$(\star) \leq C^2 \left\|f - f_L \circ \xi\right\|_{L^1_\mu} \left\|f + f_L \circ \xi\right\|_{L^1_\mu},$$

which by Lemma B.1 can be estimated as

$$\leq C^2 \|\bar{\pi}\|_\infty \varepsilon \left\|f + f_L \circ \xi\right\|_{L^1_\mu}.$$

Using the inverse triangle inequality, the remaining factor $\|f + f_L \circ \xi\|_{L^1_\mu}$ can be estimated as

$$\begin{aligned}
\left\|f + f_L \circ \xi\right\|_{L^1_\mu} &\leq \left\|f - f_L \circ \xi\right\|_{L^1_\mu} + 2\left\|f_L \circ \xi\right\|_{L^1_\mu} \\
&\leq \|\bar{\pi}\|_\infty \varepsilon + 2\left\|f_L \circ \xi\right\|_{L^1_\mu}.
\end{aligned}$$

Overall, we get for the first summand

$$(\star) \leq 2C^2 \|f_L \circ \xi\|_{L^1_\mu} \|\bar{\pi}\|_\infty \varepsilon + C^2 \|\bar{\pi}\|^2_\infty \varepsilon^2$$

For the second summand, we get

$$\begin{aligned}
(\star\star) &= \left|\left(\mathbb{E}_\mu[f] + \mathbb{E}_\mu[f_L \circ \xi]\right)\left(\mathbb{E}_\mu[f] - \mathbb{E}_\mu[f_L \circ \xi]\right)\right| \\
&\leq \left\|f + f_L \circ \xi\right\|_{L^1_\mu} \left\|f - f_L \circ \xi\right\|_{L^1_\mu}.
\end{aligned}$$

By using Lemma B.1, and the same estimation of $\|f + f_L \circ \xi\|_{L^1_\mu}$ as above, this becomes

$$(\star\star) \leq 2\|f_L \circ \xi\|_{L^1_\mu} \|\bar{\pi}\|_\infty \varepsilon + \|\bar{\pi}\|^2_\infty \varepsilon^2$$

Overall, we receive

$$\left|\mathrm{Var}_\mu(f) - \mathrm{Var}_\mu(f_L \circ \xi)\right| \leq 2(1 + C^2)\|f_L \circ \xi\|_{L^1_\mu} \|\bar{\pi}\|_\infty \varepsilon + (1 + C^2)\|\bar{\pi}\|^2_\infty \varepsilon^2.$$

Finally, we show Lemma 5.2:

Proof of Lemma 5.2. We have

$$\begin{aligned}
\mathrm{Var}_\mu(f_L \circ \xi) &= \mathbb{E}_\mu[(f_L \circ \xi)^2] - (\mathbb{E}_\mu[f_L])^2 \\
&= \int_{\mathbb{X}} \left(f_L(\xi(x))\right)^2 \pi(x)\, dx - \left(\int_{\mathbb{X}} f_L(\xi(x))\pi(x)\, dx\right)^2.
\end{aligned}$$

Using the coarea formula, this becomes

$$
\begin{aligned}
&= \int_{\mathbb{Z}} f_L^2(z) \int_{\Sigma_\xi(z)} \pi(x)\, d\sigma_z(x)\, dz + \left(\int_{\mathbb{Z}} f_L(z) \int_{\Sigma_\xi(z)} \pi(x)\, d\sigma_z(x)\, dz \right)^2 \\
&= \int_{\mathbb{Z}} f_L^2(z) \bar{\pi}(z)\, dz + \left(\int_{\mathbb{Z}} f_L(z) \bar{\pi}(z)\, dz \right)^2 \\
&= \mathrm{Var}_{\bar{\mu}}(f_L).
\end{aligned}
$$

□

□

C. Lumpability of Example 6.1

For $x, y \in \mathbb{T}$, we define the shorthand notation $h^\sigma(x,y) := \left(1 + 2\sum_{k=1}^\infty \rho^{k^2} \cos(k(y-x))\right)$, where $\rho = \exp(-\sigma^2/2)$. Note that since $g^\sigma(x,y) = h^\sigma(x,y)/(2\pi)$ is a density, the function h is cleary nonnegative. For all $x, y \in \mathbb{T}^n$, we have

$$
\begin{aligned}
|p^{\tau,\infty}(x,y) - p^{\tau,\sigma}(x,y)| &= \left| \frac{h^\tau(x^{(1)}, y^{(1)})}{(2\pi)^n} \left(1 - \prod_{i=2}^n h^{\tau\sigma}(x^{(i)}, y^{(i)})\right) \right| \\
&\leq \frac{C(\tau)}{(2\pi)^d} \max\left\{ \max_{x,y\in\mathbb{T}^n} \left|1 - \prod_{i=2}^n h^{\tau\sigma}(x^{(i)}, y^{(i)})\right|, \min_{x,y\in\mathbb{T}^n} \left|1 - \prod_{i=2}^n h^{\tau\sigma}(x^{(i)}, y^{(i)})\right| \right\} \\
&= \frac{C(\tau)}{(2\pi)^n} \max\left\{ \left|1 - \left(1 + 2\sum_{k=1}^\infty (\pm 1)^k \rho^{k^2}\right)^{n-1}\right| \right\} \qquad (39)
\end{aligned}
$$

with the constant $C(\tau) := \max_{x^{(1)}, y^{(1)} \in \mathbb{T}} h^\tau(x^{(1)}, y^{(1)})$. Here, we use the fact that for all $\tau, \sigma > 0$, the maximum of $\prod_{i=2}^n h^{\tau\sigma}(x^{(i)}, y^{(i)})$ is attained at $x^{(i)} = y^{(i)}$ and its minimum at $\left|x^{(i)} - y^{(i)}\right| = \pi$. When we apply the binomial theorem to the right hand side of (39), we obtain

$$
\begin{aligned}
\left|1 - \left(1 + 2\sum_{k=1}^\infty (\pm 1)^k \rho^{k^2}\right)^{n-1}\right| &= \left| \sum_{i=1}^{n-1} \binom{n-1}{i} \left(2\sum_{k=1}^\infty (\pm 1)^k \rho^{k^2}\right)^i \right| \\
&= 2(n-1)\left(\sum_{k=1}^\infty \rho^{k^2}\right) + \mathcal{O}\left(\left(\sum_{k=1}^\infty \rho^{k^2}\right)^2\right). \qquad (40)
\end{aligned}
$$

We now neglect the higher order terms in (40). Whenever σ is large enough such that $\rho < 1$, we have by the limit of the geometric series

$$
\left(\sum_{k=1}^\infty \rho^{k^2}\right) \leq \left(\sum_{k=1}^\infty \rho^k\right) = \frac{1}{1-\rho} - 1 = \mathcal{O}(\rho) = \mathcal{O}\left(\exp(-(\tau\sigma)^2/2)\right).
$$

Hence, we have shown

$$
\max_{x,y\in\mathbb{T}^n} |p^{\tau,\infty}(x,y) - p^{\tau,\sigma}(x,y)| = \left(\exp(-(\tau\sigma)^2/2)\right).
$$

As such, bounding both integrals in the norms in the lumpability condition (L) by the maximum above, we obtain the assertion (33).

Note that by neglecting the higher order terms in (40), we essentially ignore the impact of the prefactor depending on n in terms of the binomial coefficient, constituting the unknown dependence of the error for flexible n as mentioned in Section 6.1.2.