entropy

MDPI

Article

Kernel-Based Approximation of the Koopman Generator and Schrödinger Operator

Stefan Klus [1,*,†], Feliks Nüske [2,†] and Boumediene Hamzi [3]

1 Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany
2 Department of Mathematics, Paderborn University, 33098 Paderborn, Germany; feliks.nueske@uni-paderborn.de
3 Department of Mathematics, Imperial College London, London SW7 2AZ, UK; b.hamzi@imperial.ac.uk
* Correspondence: stefan.klus@fu-berlin.de
† These authors contributed equally to this work.

Received: 27 May 2020; Accepted: 26 June 2020; Published: 30 June 2020



Abstract: Many dimensionality and model reduction techniques rely on estimating dominant eigenfunctions of associated dynamical operators from data. Important examples include the Koopman operator and its generator, but also the Schrödinger operator. We propose a kernel-based method for the approximation of differential operators in reproducing kernel Hilbert spaces and show how eigenfunctions can be estimated by solving auxiliary matrix eigenvalue problems. The resulting algorithms are applied to molecular dynamics and quantum chemistry examples. Furthermore, we exploit that, under certain conditions, the Schrödinger operator can be transformed into a Kolmogorov backward operator corresponding to a drift-diffusion process and vice versa. This allows us to apply methods developed for the analysis of high-dimensional stochastic differential equations to quantum mechanical systems.

Keywords: Koopman generator; Schrödinger operator; reproducing kernel Hilbert space

1. Introduction

The Koopman operator [1–4] plays a central role in the global analysis of complex dynamical systems. It is, for instance, used to find conformations of molecules and coherent patterns in fluid flows, but also for prediction, stability analysis, and control [5–10]. Instead of analyzing a given finite-dimensional, but highly nonlinear system directly, the underlying idea is to compute an associated infinite-dimensional, but linear operator [4]. By computing an approximation of this operator from measurement or simulation data, it is possible to extract Koopman eigenvalues, eigenfunctions, and modes. The most frequently used techniques are based on variants or generalizations of extended dynamic mode decomposition (EDMD) [11,12]. A reformulation of EDMD for the generator of the Koopman operator, called gEDMD, was recently proposed in [13]. It was shown that in addition to the previously mentioned applications, the generator contains valuable information about the governing equations of a system; see also [7,14]. System identification aims at learning a preferably parsimonious model from data. That is, the learned model should comprise as few terms as possible and still have predictive power, which is typically accomplished by utilizing sparse regression techniques. One drawback of gEDMD is that it requires a set of explicitly chosen basis functions and their first- and—if the system is non-deterministic and non-reversible—second-order derivatives. Moreover, the size of the resulting matrix eigenvalue problem that needs to be solved to compute eigenvalues, eigenfunctions, and modes of the generator depends on the size of the dictionary. The goal of this paper is to derive a kernel-based method to approximate the Koopman generator from data. A kernel-based variant of EDMD was proposed in [12] and generalized in [15]. We derive a

kernel-based variant of gEDMD. Employing the well-known kernel trick, a dual eigenvalue problem whose size depends on the number of snapshots can be constructed. The resulting methods allow for implicitly infinite-dimensional feature spaces and only require partial derivatives of the kernel function. This enables us to apply the methods to high-dimensional systems for which conventional techniques would be prohibitively expensive due to the curse of dimensionality, provided the number of snapshots is such that the eigenvalue problem can still be solved numerically or can be downsampled without losing essential information. Since we aim at approximating differential operators, we need to be able to represent derivatives in reproducing kernel Hilbert spaces. This requires the notion of derivative reproducing properties. Derivative reproducing kernels [16] were used to approximate Lyapunov functions for ordinary differential equations in [17] and to approximate center manifolds for ordinary differential equations in [18]. Reproducing kernel Hilbert spaces with derivative reproducing properties are related to the native spaces introduced in a different context in [19].

Similar operators are also used for manifold learning and understanding the geometry of high-dimensional data [20–23]. Methods like diffusion maps construct graph Laplacians with the aid of diffusion kernels, effectively approximating transition probabilities between data points. In the infinite data limit and letting the kernel bandwidth go to zero, it has been shown that these methods, depending on the normalization, essentially compute eigenfunctions of certain differential operators, e.g., the Laplace–Beltrami operator, the Kolmogorov backward operator, or the Fokker–Planck operator.

Another related differential operator that is of utmost importance in quantum mechanics is the Schrödinger operator. Solutions of the time-independent Schrödinger equation describe stationary states and associated energy levels. We will illustrate how kernel-based methods developed for the Koopman generator can be applied to these related problems. The main contributions of this paper are:

- We show how the derivative reproducing properties of kernels can be used to approximate differential operators such as the Koopman generator and the Schrödinger operator, as well as their eigenvalues and eigenfunctions from data. Additionally, we derive a kernel-based method tailored to reversible dynamics, which does not require estimating drift and diffusion terms, but only an equilibrated trajectory.
- Furthermore, we exploit the fact that, under certain conditions, the Schrödinger operator can be turned into a Kolmogorov backward operator (see, e.g., [24]), which allows for the interpretation of a quantum-mechanical system as a drift-diffusion process and, as a consequence, the application of methods developed for the analysis of stochastic differential equations or their generators.
- We demonstrate potential applications in molecular dynamics, using the example of a quadruple-well problem, and quantum mechanics, describing how to apply the proposed methods directly to the Schrödinger equation or the associated stochastic process. This will be illustrated with two well-known examples, the quantum harmonic oscillator and the hydrogen atom.

The remainder of the manuscript is structured as follows: We first introduce the necessary tools, namely the Koopman operator, its generator, and (derivative) reproducing kernel Hilbert spaces in Section 2. Additionally, relationships with the Schrödinger equation will be explored. The derivation of the kernel-based formulation of gEDMD will be detailed in Section 3. In Section 4, we will show how the derived methods can be applied to molecular dynamics and quantum mechanics problems. Concluding remarks and future work will be discussed in Section 5.

2. Koopman Theory and Reproducing Kernel Hilbert Spaces

We start directly with the non-deterministic setting; the Koopman operator and its generator for ordinary differential equations can then be regarded as a special case; see also [13] for a detailed comparison. The notation used below is summarized in Table 1.

Table 1. Overview of notation.

X_t	stochastic process
$\mathbb{X}$	state space
k, ϕ	kernel and associated feature map
$\mathbb{H}$	reproducing kernel Hilbert space induced by k
$\mathcal{K}^t$	Koopman operator with lag time t
$\mathcal{L}$	generator of the Koopman operator
$\mathcal{H}$	Schrödinger operator
$\mathcal{T}$	general differential operator
$\mathcal{T}_{\mathbb{H}}$	kernel-based differential operator
$\mathcal{C}_{00}$	covariance operator
$\widehat{\mathcal{A}}$	empirical estimate of operator $\mathcal{A}$
G_0, G_1, G_2	(generalizations of) Gram matrices

2.1. The Koopman Operator and Its Generator

In what follows, let $\mathbb{X} \subset \mathbb{R}^d$ be the state space and $f\colon \mathbb{X} \to \mathbb{R}$ a real-valued observable of the system. Furthermore, let $\mathbb{E}[\,\cdot\,]$ denote the expected value and Θ^t the flow map associated with a dynamical system, i.e., $\Theta^t(X_0) = X_t$. Given a stochastic differential equation of the form:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \tag{1}$$

where $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is called the drift term, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ the diffusion term, and B_t is d-dimensional Brownian motion, the stochastic Koopman operator is defined by:

$$(\mathcal{K}^t f)(x) = \mathbb{E}[f(\Theta^t(x))].$$

The infinitesimal generator $\mathcal{L}$ of the semigroup of Koopman operators is given by:

$$\mathcal{L}f = \sum_{i=1}^{d} b_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j} \tag{2}$$

and its adjoint, the generator of the Perron–Frobenius operator, by:

$$\mathcal{L}^* f = -\sum_{i=1}^{d} \frac{\partial (b_i f)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 (a_{ij} f)}{\partial x_i \partial x_j},$$

with $a = \sigma\sigma^\top$. We assume from now on that a is uniformly positive definite on $\mathbb{X}$. The second-order partial differential equation $\frac{\partial u}{\partial t} = \mathcal{L}u$ is also called the Kolmogorov backward equation and $\frac{\partial u}{\partial t} = \mathcal{L}^* u$ the Fokker–Planck equation [2].

Remark 1. *As in [13], we will often consider systems of the form:*

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}B_t,$$

where V is a given potential and β the inverse temperature. In this case, the operators can be written as:

$$\mathcal{L}f = -\nabla V \cdot \nabla f + \beta^{-1}\Delta f \quad \textit{and} \quad \mathcal{L}^* f = \nabla V \cdot \nabla f + \Delta V f + \beta^{-1}\Delta f.$$

2.2. Generator EDMD

A data-driven method for the approximation of the generator of the Koopman operator and Perron–Frobenius operator called generator extended dynamic mode decomposition (gEDMD) was derived in [13]. While standard EDMD requires a training dataset $\{\, x_m \,\}_{m=1}^M$ and the corresponding data points $\{\, y_m \,\}_{m=1}^M$, where $y_m = \Theta^\tau(x_m)$ for a fixed lag time τ, gEDMD assumes that we can evaluate

or estimate (using, for instance, Kramers–Moyal formulae) $\{b(x_m)\}_{m=1}^M$ and $\{\sigma(x_m)\}_{m=1}^M$. Choosing a dictionary of basis functions $\{\phi_n\}_{n=1}^N$, where $\phi_n\colon \mathbb{R}^d \to \mathbb{R}$, and defining $\phi(x) = [\phi_1(x), \ldots, \phi_N(x)]^\top$, we compute the matrices $\Phi_X, \mathrm{d}\Phi_X \in \mathbb{R}^{N\times M}$, with:

$$\Phi_X = \begin{bmatrix} \phi_1(x_1) & \ldots & \phi_1(x_M) \\ \vdots & \ddots & \vdots \\ \phi_N(x_1) & \ldots & \phi_N(x_M) \end{bmatrix} \quad \text{and} \quad \mathrm{d}\Phi_X = \begin{bmatrix} \mathrm{d}\phi_1(x_1) & \ldots & \mathrm{d}\phi_1(x_M) \\ \vdots & \ddots & \vdots \\ \mathrm{d}\phi_N(x_1) & \ldots & \mathrm{d}\phi_N(x_M) \end{bmatrix},$$

where:

$$\mathrm{d}\phi_n(x) = \sum_{i=1}^d b_i(x)\,\frac{\partial \phi_n}{\partial x_i}(x) + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d a_{ij}(x)\,\frac{\partial^2 \phi_n}{\partial x_i \partial x_j}(x).$$

The matrix representation of the least-squares approximation of the Koopman generator $\mathcal{L}$ is then given by:

$$\widehat{L}^\top = \mathrm{d}\Phi_X \Phi_X^+ = \widehat{A}\,\widehat{G}^+,$$

with:

$$\widehat{A} = \frac{1}{M}\sum_{m=1}^M \mathrm{d}\phi(x_m)\,\phi(x_m)^\top \quad \text{and} \quad \widehat{G} = \frac{1}{M}\sum_{m=1}^M \phi(x_m)\,\phi(x_m)^\top.$$

It was shown that gEDMD, in the infinite data limit, converges to a Galerkin projection of the generator onto the space spanned by the basis functions $\{\phi_n\}_{n=1}^N$ and that $\widehat{L}$ is an empirical estimate of the projected generator [13]. Approximations of eigenfunctions of $\mathcal{L}$ are then given by:

$$\varphi_\ell(x) = \langle \xi_\ell,\, \phi(x) \rangle\,,$$

where ξ_ℓ is an eigenvector of $\widehat{L}$ corresponding to the eigenvalue λ_ℓ and $\langle \cdot,\, \cdot \rangle$ denotes the standard Euclidean inner product. Analogously, the generator of the Perron–Frobenius operator is given by $(\widehat{L}^*)^\top = \widehat{A}^\top \widehat{G}^+$. Further details, examples, and different applications including system identification, coarse graining, and control can also be found in [13].

2.3. Second-Order Differential Operators

Consider the generator $\mathcal{L}$ in (2), and assume there is a unique strictly positive invariant density ρ_0, which we can write as $\rho_0(x) \propto \exp(-F(x))$. The function F is called a generalized potential (with $F = \beta V$ for the stochastic differential equation in Remark 1). The measure corresponding to ρ_0 is denoted by $\mathrm{d}\mu = \rho_0\,\mathrm{d}x$. The negative generator can be decomposed into a symmetric and an anti-symmetric part as:

$$-\mathcal{L} = -\frac{1}{2}e^F \nabla \cdot \left(e^{-F} a \nabla \cdot\right) + J \cdot \nabla = \mathcal{S} + \mathcal{A}, \tag{3}$$

$$J = \frac{1}{2}e^F \nabla \cdot (e^{-F} a) - b; \tag{4}$$

see [24]. The vector field J is called stationary probability flow. In the form of (3), $-\mathcal{L}$ is a special case of an elliptic second-order differential operator on L^2_μ, given by:

$$\mathcal{T} = -\frac{1}{2}e^F \nabla \cdot \left(e^{-F} a \nabla \cdot\right) + J \cdot \nabla + W, \tag{5}$$

for scalar functions F, W, a uniformly positive definite matrix field a, and a vector field J.

Remark 2. *Because of the general form of* (5)*, we avoid making too many assumptions about the coefficients of* $\mathcal{T}$ *or its domain of definition. The goal is to derive numerical algorithms using a minimal set of assumptions.*

A detailed analysis of the interplay between the domains and properties of the reproducing kernel Hilbert space (RKHS) will be carried out in future publications.

If $F \equiv 0$, we obtain generalized Schrödinger operators as another special case, i.e.,

$$\mathcal{H} = -\frac{1}{2}\nabla \cdot (a\nabla \cdot) + J \cdot \nabla + W, \tag{6}$$

with W called the potential energy in quantum mechanics. In particular, with the reduced Planck constant $\hbar$ and the mass $\mathbf{m}$, setting $a \equiv \frac{\hbar^2}{\mathbf{m}} I$ and $J \equiv 0$ leads to the Hamiltonian $\mathcal{H} = -\frac{\hbar^2}{2\mathbf{m}}\Delta + W$ of the time-independent Schrödinger equation in quantum mechanics:

$$\mathcal{H}\psi = E\psi. \tag{7}$$

We note for later use that, under certain conditions, Schrödinger operators and Koopman generators are equivalent; see, e.g., ([24] Chapter 4.9). For the sake of completeness, the proof is shown in Appendix A.

Lemma 1. *The ergodic generator $-\mathcal{L}$ with unique positive invariant density $\rho_0 \propto \exp(-F)$ is unitarily equivalent to the Schrödinger operator $\mathcal{H}$ in (6) on L^2, with J remaining unchanged and W given by:*

$$W = -\frac{1}{4}\nabla \cdot (a\nabla F) + \frac{1}{8}\nabla F^\top a \nabla F + \frac{1}{2} J \cdot \nabla F.$$

The function $e^{-\frac{1}{2}F}$ is an eigenfunction of $\mathcal{H}$ with eigenvalue zero. Conversely, let $\mathcal{H}$ be as in (6), and assume there is a non-degenerate smallest eigenvalue E_0 with strictly positive real eigenfunction $\psi_0 = \exp(-\eta)$. Then, $\mathcal{H}$ is unitarily equivalent to a negative ergodic generator $-\mathcal{L}$ on L^2_μ, where $\rho_0 \propto \exp(-2\eta)$ is the density associated with μ and ρ_0 is invariant for the corresponding SDE. The explicit form of $-\mathcal{L}$ is given by:

$$-\mathcal{L} = e^{\eta}\left[\mathcal{H} - E_0\right](e^{-\eta}\cdot) = -\frac{1}{2}e^{2\eta}\nabla \cdot \left(e^{-2\eta} a \nabla \cdot\right) + J \cdot \nabla.$$

Corollary 1. *Applying Lemma 1 to (7), we have:*

$$\frac{1}{\psi_0}(\mathcal{H} - E_0)(\psi_0 f) = -\left(-\frac{\hbar^2}{\mathbf{m}}\nabla\eta \cdot \nabla f + \frac{\hbar^2}{2\mathbf{m}}\Delta f\right) = -\mathcal{L}f,$$

where $\mathcal{L}$ is the Koopman generator of a drift-diffusion process (see Remark 1) with potential (up to an additive constant):

$$V(x) = \frac{\hbar^2}{\mathbf{m}}\eta(x),$$

and temperature $\beta^{-1} = \frac{\hbar^2}{2\mathbf{m}}$.

We will exploit this duality below to apply methods developed for the Koopman operator or generator to the Schrödinger operator. More details on quantum chemistry in general and also the quantum harmonic oscillator and the hydrogen atom studied in Section 4 can be found, e.g., in [25].

2.4. Reproducing Kernel Hilbert Spaces and Derivative Reproducing Properties

We aim at representing the differential operators introduced above in reproducing kernel Hilbert spaces.

Definition 1. *Let $\mathbb{X}$ be a set and $\mathbb{H}$ a space of functions $f\colon \mathbb{X} \to \mathbb{R}$. Then, $\mathbb{H}$ is called an RKHS with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ if a function $k\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ exists such that:*

(i) $\langle f, k(x,\cdot)\rangle_{\mathbb{H}} = f(x)$ *for all* $f \in \mathbb{H}$ *and*
(ii) $\mathbb{H} = \overline{\operatorname{span}\{k(x,\cdot) \mid x \in \mathbb{X}\}}$.

The function k is called a kernel. It was shown that every RKHS has a unique symmetric positive definite reproducing kernel and that, conversely, every symmetric positive definite kernel spans a unique RKHS; see [26–28]. Here, we use the terms positive definite and strictly positive definite, i.e., positive definite means that $\sum_{r=1}^{M}\sum_{s=1}^{M}\gamma_r\gamma_s k(x_r,x_s) \geq 0$ for all $M \in \mathbb{N}$, $\gamma_1,\ldots,\gamma_M \in \mathbb{R}$, and $x_1,\ldots,x_M \in \mathbb{X}$. Frequently used kernels include the polynomial kernel and the Gaussian kernel, given by:

$$k(x,x') = (c + x^\top x')^q \quad \text{and} \quad k(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\varsigma^2}\right),$$

respectively. Here, $q \in \mathbb{N}$ is the degree of the polynomial kernel, $c \geq 0$ a parameter, and ς the bandwidth of the Gaussian kernel. We now introduce the partial derivative reproducing properties of RKHSs [16]. Let $\alpha = (\alpha_1,\ldots,\alpha_d) \in \mathbb{N}_0^d$ be a multi-index and $|\alpha| = \sum_{i=1}^{d}\alpha_i$. Furthermore, for a fixed $p \in \mathbb{N}_0$, we define the index set $I_p = \{\alpha \in \mathbb{N}_0^d : |\alpha| \leq p\}$. Given $f\colon \mathbb{X} \to \mathbb{R}$, let D^α denote the partial derivative (assuming it exists):

$$D^\alpha f = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}} f.$$

Thus, the i^{th} entry of the gradient is given by $D^{e_i}f$ and the $(i,j)^{\text{th}}$ entry of the Hessian by $D^{e_i+e_j}$, where e_i and e_j are the i^{th} and j^{th} unit vectors, respectively. When we apply the differential operator D^α to the kernel k, the multi-index α is assumed to be embedded into $\mathbb{N}_0^{2d}$ by adding zeros, i.e., the derivatives are computed with respect to the first argument of the kernel. Furthermore, when we write $\nabla k(x,x')$, the gradient is computed with respect to x. In what follows, let $k(x,\cdot) = \phi(x)$, where ϕ is the canonical feature space mapping.

Theorem 1 ([16]). *Given* $p \in \mathbb{N}_0$ *and a positive definite kernel* $k\colon \mathbb{X}\times\mathbb{X} \to \mathbb{R}$ *with* $k \in C^{2p}(\mathbb{X}\times\mathbb{X})$, *the following holds:*

(i) $D^\alpha k(x,\cdot) \in \mathbb{H}$ *for any* $x \in \mathbb{X}$ *and* $\alpha \in I_p$.
(ii) $(D^\alpha f)(x) = \langle D^\alpha k(x,\cdot), f\rangle_{\mathbb{H}}$ *for any* $x \in \mathbb{X}$, $f \in \mathbb{H}$, *and* $\alpha \in I_p$.

The second property is called the derivative reproducing property. For $p = 0$, this reduces to the standard reproducing property of RKHSs.

Example 1. *Let us consider the two aforementioned kernels:*

1. *For the polynomial kernel, we obtain:*

$$D^{e_i}k(x,x') = q\,x_i'(c + x^\top x')^{q-1} \quad \text{and} \quad D^{e_i+e_j}k(x,x') = q\,(q-1)\,x_i'x_j'(c + x^\top x')^{q-2}.$$

Thus, $\nabla k(x,x') = q\,x'(c + x^\top x')^{q-1}$ *and* $\nabla^2 k(x,x') = q\,(q-1)\,x'x'^\top(c + x^\top x')^{q-2}$.

2. *Similarly, for the Gaussian kernel, this results in:*

$$D^{e_i}k(x,x') = -\frac{1}{\varsigma^2}(x_i - x_i')\,k(x,x'),$$

$$D^{e_i+e_j}k(x,x') = \begin{cases} \left[\frac{1}{\varsigma^4}(x_i - x_i')^2 - \frac{1}{\varsigma^2}\right]k(x,x'), & i = j, \\ \frac{1}{\varsigma^4}(x_i - x_i')(x_j - x_j')\,k(x,x'), & i \neq j, \end{cases}$$

$\nabla k(x,x') = -\frac{1}{\varsigma^2}(x - x')\,k(x,x')$, *and* $\nabla^2 k(x,x') = \left[\frac{1}{\varsigma^4}(x - x')(x - x')^\top - \frac{1}{\varsigma^2}I\right]k(x,x')$.

For the numerical experiments below, we will mainly use the Gaussian kernel. (To get error estimates, it might be more convenient to use Wendland functions [19]. We leave the formal analysis of the methods developed in this paper for future work.)

3. Kernel-Based Representation of Differential Operators

In this section, we introduce the Galerkin projection of the differential operators discussed above onto the RKHS, including the Koopman generator and Schrödinger operator. We then move on to show how these projected operators can be estimated from data.

3.1. Galerkin Projection of Operators

Let μ denote a probability measure on the state space $\mathbb{X}$, with density $\rho_0 \propto e^{-F}$ for a generalized potential F.

Definition 2. *We define the covariance operator* $\mathcal{C}_{00} \colon \mathbb{H} \to \mathbb{H}$ *by:*

$$\mathcal{C}_{00} = \int \phi(x) \otimes \phi(x) \, \mathrm{d}\mu(x), \tag{8}$$

and an operator $\mathcal{T}_{\mathbb{H}} \colon \mathbb{H} \to \mathbb{H}$ *by:*

$$\begin{aligned} \mathcal{T}_{\mathbb{H}} = & \int \phi(x) \otimes \left[-\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}(x) D^{e_i+e_j} \phi(x) \right] \mathrm{d}\mu(x) \\ & + \int \phi(x) \otimes \left[\sum_{i=1}^{d} \left(J_i(x) - \frac{1}{2} e^{F(x)} \nabla \cdot (e^{-F(x)} a_{:,i}(x)) \right) D^{e_i} \phi(x) \right] \mathrm{d}\mu(x) \\ & + \int W(x) \phi(x) \otimes \phi(x) \, \mathrm{d}\mu(x). \end{aligned} \tag{9}$$

If $J \equiv 0$, *we define* $\mathcal{T}_{\mathbb{H}}$ *by:*

$$\mathcal{T}_{\mathbb{H}} = \int \left[\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}(x) \left(D^{e_i} \phi(x) \otimes D^{e_j} \phi(x) \right) \right] + W(x) \phi(x) \otimes \phi(x) \, \mathrm{d}\mu(x). \tag{10}$$

The operator $\mathcal{C}_{00}$ is the standard covariance operator $\mathcal{C}_{XX}$; see [29,30]. The operator $\mathcal{T}_{\mathbb{H}}$ mimics the action of the bilinear form $\langle \mathcal{T} f, g \rangle_{\mu}$ on the RKHS. It plays the same role as the cross-covariance operator $\mathcal{C}_{XY}$ for the Koopman operator in [15]. The form of the symmetric operator for $J \equiv 0$ is motivated by the symmetry of $\mathcal{T}$, and that, at least formally:

$$\langle \mathcal{T} f, g \rangle_{\mu} = \int \left[\frac{1}{2} \nabla f(x)^{\top} a(x) \nabla g(x) \right] + W(x) f(x) g(x) \, \mathrm{d}\mu(x);$$

see also [31].

Lemma 2. *Assume that* $\mathbb{H} \subset \mathcal{D}(\mathcal{T})$ *and that all terms appearing under the integral signs in* (8) *and* (9) *(or* (10)*) are in* L^1_{μ} *as bounded operators on* $\mathbb{H}$, *that is:*

$$\int |a_{ij}(x)| \| D^{e_i+e_j} \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) < \infty, \tag{11}$$

$$\int \left(|J_i(x)| + \frac{1}{2} e^{F(x)} |\nabla \cdot (e^{-F(x)} a_{:,i}(x))| \right) \| D^{e_i} \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) < \infty, \tag{12}$$

$$\int |W(x)| \| \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) < \infty, \tag{13}$$

$$\int \| \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) < \infty. \tag{14}$$

Then, for all $f, g \in \mathbb{H}$,

$$\langle \mathcal{T} f, g \rangle_\mu = \langle \mathcal{T}_\mathbb{H} f, g \rangle_\mathbb{H}, \qquad \langle f, g \rangle_\mu = \langle \mathcal{C}_{00} f, g \rangle_\mathbb{H}.$$

The proof can be found in Appendix A. It uses the derivative reproducing properties and the definition of rank-one operators. Note that:

$$\begin{aligned} D^{e_i} f(x) g(x) &= \langle D^{e_i} \phi(x), f \rangle_\mathbb{H} \langle \phi(x), g \rangle_\mathbb{H} \\ &= \langle D^{e_i} \phi(x) \otimes \phi(x), f \otimes g \rangle_{\mathbb{H} \otimes \mathbb{H}} \\ &= \langle (\phi(x) \otimes D^{e_i} \phi(x)) f, g \rangle_\mathbb{H}. \end{aligned}$$

Lemma 3. *Assume that $\mathcal{T} f \in \mathbb{H}$ for all $f \in \mathbb{H}$, then $\mathcal{T}_\mathbb{H} f = \mathcal{C}_{00} \mathcal{T} f$.*

Proof. The proof is similar to the one for the corresponding result for kernel transfer operators; see [15]. With the previous lemma, we obtain:

$$\begin{aligned} \langle \mathcal{C}_{00} \mathcal{T} f, g \rangle_\mathbb{H} &= \mathbb{E}^\mu [(\mathcal{T} f)(x) g(x)] \\ &= \int (\mathcal{T} f)(x) g(x) \, \mathrm{d}\mu(x) \\ &= \langle \mathcal{T}_\mathbb{H} f, g \rangle_\mathbb{H} \end{aligned}$$

for arbitrary $g \in \mathbb{H}$. □

If the assumptions of Lemma 3 are satisfied and the operator $\mathcal{C}_{00}$ is invertible, the RKHS operators defined above can be used to compute exact eigenfunctions of $\mathcal{T}$. Indeed, if φ is a solution of:

$$\mathcal{T}_\mathbb{H} \varphi = \mathcal{C}_{00} \mathcal{T} \varphi = \lambda \mathcal{C}_{00} \varphi,$$

then multiplying this equation by $\mathcal{C}_{00}^{-1}$ shows that φ is also an eigenfunction for $\mathcal{T}$. A typical approach to circumvent the potential nonexistence of the inverse of the covariance operator is to consider a regularized version $\mathcal{T}_\varepsilon = (\mathcal{C}_{00} + \varepsilon I)^{-1} \mathcal{T}_\mathbb{H}$ for a regularization parameter ε. However, the assumptions of Lemma 3 are strong and may be hard to verify in practice. However, in any case, Lemma 2 shows that the operators defined in Definition 2 provide a Galerkin approximation of the full operator in the RKHS $\mathbb{H}$.

3.2. Empirical Estimates

The next step is to derive empirical estimates of the operators defined above. Given training data $\{ x_m \}_{m=1}^M$, sampling the probability distribution μ, we define $\Phi = [\phi(x_1), \ldots, \phi(x_M)]$ and $\mathrm{d}\Phi = [\mathrm{d}\phi(x_1), \ldots, \mathrm{d}\phi(x_M)]$, where:

$$\begin{aligned} \mathrm{d}\phi(x_m) = &-\frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij}(x_m) D^{e_i + e_j} \phi(x_m) \\ &+ \sum_{i=1}^d \left[J_i(x_m) - \frac{1}{2} \sum_{j=1}^d e^{F(x_m)} \frac{\partial}{\partial x_j} (e^{-F(x_m)} a_{ji}(x_m)) \right] D^{e_i} \phi(x_m) \\ &+ W(x_m) \phi(x_m). \end{aligned}$$

If $\mathcal{T}$ is the generator of an SDE with invariant measure μ, the data can also be obtained by integrating the stochastic dynamics with the initial condition drawn from μ. We see that Φ is the

standard feature map and $\mathrm{d}\Phi$ contains the action of the differential operator. The empirical estimates of the operators $\mathcal{C}_{00}$ and $\mathcal{T}_{\mathbb{H}}$ are then given by the following expressions:

$$\widehat{\mathcal{C}}_{00} = \frac{1}{M}\sum_{m=1}^{M} \phi(x_m) \otimes \phi(x_m) \quad = \frac{1}{M}\Phi\Phi^\top,$$
$$\widehat{\mathcal{T}}_{\mathbb{H}} = \frac{1}{M}\sum_{m=1}^{M} \phi(x_m) \otimes \mathrm{d}\phi(x_m) = \frac{1}{M}\Phi\,\mathrm{d}\Phi^\top.$$

Note that these are still finite-rank operators on the full RKHS $\mathbb{H}$. For the symmetric RKHS operator $\mathcal{T}_{\mathbb{H}}$, we need to define the empirical estimate in a slightly different way. Decompose the positive definite matrix $a(x_m) = \sigma(x_m)\sigma(x_m)^\top$. With:

$$\mathrm{d}\phi_l(x_m) = \sum_{i=1}^{d} \sigma_{il}(x_m) D^{e_i}\phi(x_m) = \left[\nabla\phi(x_m)^\top \sigma_l(x_m)\right],$$

where σ_l is the l^{th} column of σ, the empirical RKHS operator becomes:

$$\widehat{\mathcal{T}}_{\mathbb{H}} = \frac{1}{2M}\sum_{m=1}^{M}\sum_{l=1}^{d} \mathrm{d}\phi_l(x_m) \otimes \mathrm{d}\phi_l(x_m) + \frac{1}{M}\sum_{m=1}^{M} W(x_m)\phi(x_m) \otimes \phi(x_m).$$

Remark 3. *If the feature space associated with the kernel k is finite-dimensional and known explicitly, i.e., $\phi(x) = [\phi_1(x), \ldots, \phi_N(x)]^\top$ and $k(x, x') = \langle \phi(x), \phi(x')\rangle$, then for the Koopman generator, we obtain gEDMD as a special case, with $\widehat{\mathcal{C}}_{00} = \widehat{G}$ and $\widehat{\mathcal{T}}_{\mathbb{H}} = -\widehat{A}^\top$. However, the goal is to rewrite gEDMD in such a way that only kernel evaluations are required since ϕ can potentially be infinite-dimensional and might only be defined implicitly.*

3.3. Weak Formulation and Numerical Algorithm

With Lemma 2 in mind, we now proceed to the weak formulation of the eigenvalue problem for the operator $\mathcal{T}$. We then define the quadratic forms:

$$\begin{aligned}
\mathcal{Q}(f,g) &= \langle \mathcal{T}f, g\rangle_\mu, & f,g &\in \mathcal{D}_{\mathcal{Q}}, & \mathcal{S}(f,g) &= \langle f, g\rangle_\mu, & f,g &\in L^2_\mu,\\
\mathcal{Q}_{\mathbb{H}}(f,g) &= \langle \mathcal{T}_{\mathbb{H}}f, g\rangle_{\mathbb{H}}, & f,g &\in \mathbb{H}, & \mathcal{S}_{\mathbb{H}}(f,g) &= \langle \mathcal{C}_{00}f, g\rangle_{\mathbb{H}}, & f,g &\in \mathbb{H},\\
\widehat{\mathcal{Q}}_{\mathbb{H}}(f,g) &= \left\langle \widehat{\mathcal{T}}_{\mathbb{H}}f, g\right\rangle_{\mathbb{H}}, & f,g &\in \mathbb{H}, & \widehat{\mathcal{S}}_{\mathbb{H}}(f,g) &= \left\langle \widehat{\mathcal{C}}_{00}f, g\right\rangle_{\mathbb{H}}, & f,g &\in \mathbb{H},
\end{aligned}$$

where $\mathcal{D}_{\mathcal{Q}}$ is the domain of the quadratic form $\mathcal{Q}$. We consider the weak eigenvalue problems:

$$\mathcal{Q}(f_n, g) = \lambda_n \mathcal{S}(f_n, g) \quad \forall g \in \mathcal{D}_{\mathcal{Q}}, \tag{15}$$
$$\mathcal{Q}_{\mathbb{H}}(\widetilde{f}_n, g) = \widetilde{\lambda}_n \mathcal{S}_{\mathbb{H}}(\widetilde{f}_n, g) \quad \forall g \in \mathbb{H}, \tag{16}$$
$$\widehat{\mathcal{Q}}_{\mathbb{H}}(\widehat{f}_n, g) = \widehat{\lambda}_n \widehat{\mathcal{S}}_{\mathbb{H}}(\widehat{f}_n, g) \quad \forall g \in \mathbb{H}. \tag{17}$$

We will now rewrite (17) in such a way that only kernel evaluations—in the form of Gram matrices—are required. The derivation is similar to the kernel transfer operator counterpart in [15], but we now need to consider derivatives at the training data points instead of the time-lagged variables. We start by restricting (17) to the finite-dimensional space $\mathbb{H}^M = \operatorname{span}\{\phi(x_m)\}_{m=1}^M$, which we assume to be M-dimensional. Elements of this space are of the form $f = \Phi u$ for some vector $u \in \mathbb{R}^m$. We examine the quadratic forms $\widehat{\mathcal{Q}}_{\mathbb{H}}$ and $\widehat{\mathcal{S}}_{\mathbb{H}}$ on this space.

Lemma 4. *A solution of the problem $\widehat{\mathcal{Q}}_{\mathbb{H}}(f,g) = \widehat{\lambda}\,\widehat{\mathcal{S}}_{\mathbb{H}}(f,g)$ is given by $f = \Phi u$, where u is a solution of one of the following generalized eigenvalue problems:*

(i) *In the general case, u solves $G_2 u = \widehat{\lambda} G_0 u$, where the entries of the matrices G_2 and G_0 are given by:*

$$\left[G_2\right]_{mr} = \left[\mathrm{d}\phi(x_m)\right](x_r), \qquad \left[G_0\right]_{mr} = \left[\phi(x_m)\right](x_r).$$

(ii) *Analogously, for the symmetric case, we obtain $\frac{1}{2}\sum_{l=1}^{d} G_1^{(l)} G_1^{(l)} u = \widehat{\lambda} G_0 G_0 u$, where we define:*

$$\left[G_1^{(l)}\right]_{mr} = \sigma_l(x_m)^\top \nabla k(x_m, x_r)$$

and $\sigma_l(x_m)$ is the l^{th} column of the matrix $\sigma(x_m)$.

The proofs are shown in Appendix A. Since $[\phi(x_m)](x_r) = k(x_m, x_r)$, G_0 is the standard Gram matrix. The reversible case requires only first-order derivatives of the kernel. Furthermore, only trajectory data sampled from the invariant distribution μ and estimates of the diffusion term σ are needed. For typical problems, σ is constant and not position-dependent. As a result, the diffusion term needs to be estimated only once or might even be known. For molecular dynamics problems, for instance, it is proportional to the square root of the temperature. The overall approach is summarized in the following algorithm. Note that it is not a direct kernelization of gEDMD, but an extension that approximates the Koopman generator as a special case.

Algorithm 1. *The final numerical algorithm can be summarized as follows:*

1. *Choose a kernel k and compute all its required derivatives, either analytically or with the aid of automatic differentiation.*
2. *Assemble the Gram matrices G_2 and G_0 or, if the system is symmetric, $G_1^{(l)}$, for $l = 1, \ldots, d$, and G_0.*
3. *Solve the corresponding eigenvalue problem described in Lemma 4 to obtain an eigenvector u.*
4. *An eigenfunction is then given by $\varphi = \Phi u$.*

The two main steps of the algorithm are assembling the Gram matrices and solving the generalized eigenvalue problem. Since the size of the eigenvalue problem depends on the number of data points, the cost is cubic in M. This is a drawback of many kernel-based methods. The efficient approximation of solutions to this eigenvalue problem for large datasets will be considered in future work.

3.4. Analysis

In this section, we provide some preliminary analysis of the methods introduced above. The first result concerns the convergence of the empirical estimates.

Lemma 5. *As $M \to \infty$, the empirical estimates defined in Section 3.2 converge to the corresponding RKHS operators in Definition 2 with respect to the operator norm for almost all data sequences $\{x_m\}_{m=1}^{M}$, if the data were generated either as i.i.d. samples from μ or by integrating a stochastic dynamics, which is ergodic with respect to μ.*

Proof. The statement follows from ergodicity of the underlying dynamics, the integrability conditions in Lemma 2, and the Birkhoff individual ergodic theorem for Banach space valued functions [32]. □

Next, we generalize ([33] Theorem 7) to obtain convergence rates on the empirical estimates for i.i.d. data:

Lemma 6. *Assume that (11– 14) hold. Then:*

(i) *The operators C_{00}, $\widehat{C}_{00}$, $\mathcal{T}_{\mathbb{H}}$, and $\hat{\mathcal{T}}_{\mathbb{H}}$ are Hilbert–Schmidt.*

(ii) *Let $\delta \in (0,1]$. Assume the coefficients of the operator $\mathcal{T}$ are all globally bounded, and let $\sup_{x\in\mathbb{X}} D^\alpha k(x,x) < \infty$ for all $|\alpha| \leq 4$ ($|\alpha| \leq 2$ in the symmetric case). If the data are drawn i.i.d. from the distribution μ, then there are constants κ_0, κ_1 such that with probability at least $1-\delta$,*

$$\|\mathcal{C}_{00} - \widehat{\mathcal{C}}_{00}\|_{HS} \leq \frac{2\kappa_0\sqrt{2}}{\sqrt{M}} \log^{1/2}\frac{2}{\delta}, \qquad \|\mathcal{T}_{\mathbb{H}} - \widehat{\mathcal{T}}_{\mathbb{H}}\|_{HS} \leq \frac{2\kappa_1\sqrt{2}}{\sqrt{M}} \log^{1/2}\frac{2}{\delta},$$

where the $\|\cdot\|_{HS}$ is the Hilbert–Schmidt norm.

Proof. (i) The empirical estimates are all finite rank and, therefore, Hilbert-Schmidt. For $\mathcal{C}_{00}$ and $\mathcal{T}_{\mathbb{H}}$, this follows from the integrability conditions and the first part of the proof of Lemma 2; see Appendix A. (ii) For $\mathcal{C}_{00}$, the bound was already proven in [33] with $\kappa_0 = \sup_{x\in\mathbb{X}} k(x,x)^2$. We can employ the same strategy to obtain the bound for $\mathcal{T}_{\mathbb{H}}$. Consider the operator $\hat{\mathcal{T}}^m_{\mathbb{H}} = \phi(x_m) \otimes \mathrm{d}\phi(x_m) - \mathcal{T}_{\mathbb{H}}$, which satisfies $\mathbb{E}^\mu[\hat{\mathcal{T}}^m_{\mathbb{H}}] = 0$. By global boundedness of the coefficients of $\mathcal{T}$ and by:

$$\begin{aligned}\|\phi(x) \otimes D^\alpha\phi(x)\|_{HS} &= \|\phi(x)\|_{\mathbb{H}}\|D^\alpha\phi(x)\|_{\mathbb{H}} \\ &= \langle k(x,\cdot),\, k(x,\cdot)\rangle^{1/2}_{\mathbb{H}} \langle D^\alpha k(x,\cdot),\, D^\alpha k(x,\cdot)\rangle^{1/2}_{\mathbb{H}} \\ &= k(x,x)^{1/2} D^{2\alpha} k(x,x)^{1/2},\end{aligned}$$

we can find a κ_1 such that $\|\phi(x) \otimes \mathrm{d}\phi(x)\|_{HS} \leq \kappa_1$ for all $x \in \mathbb{X}$. We then have $\|\hat{\mathcal{T}}^m_{\mathbb{H}}\|_{HS} \leq 2\kappa_1$, and the result follows from the concentration bound ([33] Equation (3)). □

Finally, we show that solutions of (16) are also eigenvalues of the full operator $\mathcal{T}$ if the RKHS is dense in $\mathcal{D}_{\mathcal{Q}}$:

Proposition 1. *Let $\mathbb{H}$ be dense in $\mathcal{D}_{\mathcal{Q}}$ with respect to the norm in L^2_μ. If $\widetilde{\psi}_\ell \in \mathbb{H}$ is an eigenfunction of (16), it is also an eigenfunction of $\mathcal{T}$ with the same eigenvalue.*

Proof. Let $\widetilde{\psi}_\ell$ solve the variational problem (16). The definition of the operators $\mathcal{C}_{00}$, $\mathcal{T}_{\mathbb{H}}$ implies that for all $\phi \in \mathbb{H}$:

$$\langle \mathcal{T}\widetilde{\psi}_\ell,\, \phi\rangle_\mu = \langle \mathcal{T}_{\mathbb{H}}\widetilde{\psi}_\ell,\, \phi\rangle_{\mathbb{H}} = \widetilde{\lambda}_\ell \langle \mathcal{C}_{00}\widetilde{\psi}_\ell,\, \phi\rangle_{\mathbb{H}} = \widetilde{\lambda}_\ell \langle \widetilde{\psi}_\ell,\, \phi\rangle_\mu .$$

By the density of the RKHS, this also holds for all $\phi \in \mathcal{D}_{\mathcal{Q}}$, and consequently, $\widetilde{\psi}_\ell$ is an eigenfunction of $\mathcal{T}$. □

Note that even if the RKHS is dense, there might be additional eigenfunctions that are not contained in $\mathbb{H}$ and that will not appear as solutions of (16).

4. Applications

The methods described above have important applications in molecular dynamics and quantum physics, which we will show in an exemplary way, but can in principle be applied to data generated by arbitrary dynamical systems and also other differential operators. The code and select examples are available online [34]. Note that this is just a proof-of-concept implementation and that the methods could be sped up significantly by vectorizing and parallelizing the code and by tailoring the implementation to specific kernels.

4.1. Molecular Dynamics

Eigenvalues and eigenfunctions of transfer operators associated with molecular dynamics problems are often used to understand protein folding or binding/unbinding processes and their implied time scales. Conformations correspond to metastable sets and transitions between different

conformations to crossing energy barriers. The slowest dynamical processes are encoded in eigenfunctions whose eigenvalues are close to zero. Large-scale molecular dynamics examples, analyzed using kernel EDMD, can also be found in [35]. In this paper, we want to focus more on new applications.

Example 2. *Let us consider the simple quadruple-well problem whose potential V is visualized in Figure 1a; see also [13]. We first generate an equilibrated trajectory so that the training dataset of size $M = 5000$ is sampled from the invariant distribution and then apply kernel gEDMD for reversible processes, choosing a Gaussian kernel with bandwidth $\varsigma = 0.5$. The operator $-\mathcal{L}$ has four dominant eigenvalues $\lambda_0 = 0.009$, $\lambda_1 = 0.400$, $\lambda_2 = 1.011$, and $\lambda_3 = 1.55$, followed by a spectral gap. We then apply SEBA (sparse eigenbasis approximation; see [36]) to cluster the dominant eigenfunctions into four metastable sets. The results are shown in Figure 1b. As expected, the sets correspond to the wells of the potential. The computation and clustering of the eigenfunctions took approximately four minutes on a standard laptop (8 cores, 1.80 GHz, 16 GB of RAM). For comparison, we estimated the generator eigenvalues using a Markov state model. Applying both methods to 20 different trajectories, we computed the average of the eigenvalues and the standard deviation, see Figure 1c. The results were in excellent agreement. Clearly, the standard deviation increased for higher eigenvalues.*

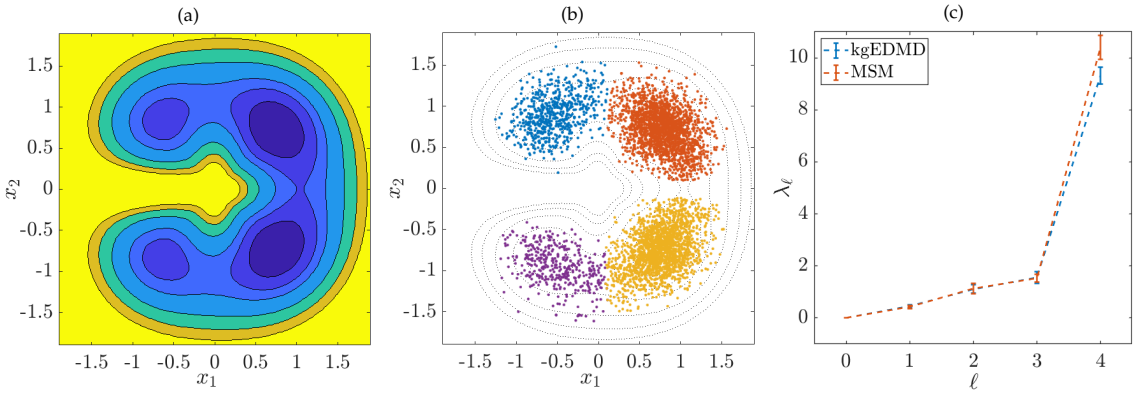


Figure 1. (**a**) Quadruple-well potential. The color blue corresponds to small values and yellow to large values. (**b**) Clustering into four metastable sets based on sparse eigenbasis approximation (SEBA). (**c**) Eigenvalues computed using kernel generator extended dynamic mode decomposition (gEDMD) and a Markov state model. The bars indicate the estimated standard deviation.

4.2. Quantum Mechanics

The goal now is to apply data-driven methods to simple quantum mechanics problems of the form (7) with $\mathcal{H} = -\frac{\hbar^2}{2\mathbf{m}}\Delta + W$.

4.2.1. Generator EDMD for the Schrödinger Equation

Let us consider two systems for which the eigenfunctions are well known.

Example 3. *For the quantum harmonic oscillator with angular frequency ω, the potential can be written as $W(x) = \frac{1}{2}\mathbf{m}\omega^2 x^2$. The eigenfunctions ψ_ℓ and corresponding energy levels E_ℓ of this system can be computed analytically, and we obtain:*

$$\psi_\ell(x) = \frac{1}{\sqrt{2^\ell\,\ell!}}\left(\frac{\mathbf{m}\omega}{\pi\hbar}\right)^{1/4}\exp\left(-\frac{\mathbf{m}\omega}{2\hbar}x^2\right)H_\ell\left(\sqrt{\frac{\mathbf{m}\omega}{\hbar}}x\right)$$

and $E_\ell = \hbar\omega\left(\ell + \frac{1}{2}\right)$, for $\ell = 0, 1, 2, \ldots$. Here, H_ℓ denotes the ℓ^{th} physicists' Hermite polynomial. For the numerical experiments, we set $\hbar = \mathbf{m} = \omega = 1$. Furthermore, the bandwidth of the kernel is set to $\varsigma = 1$. Computing the Gram matrices G_2 and G_0 for 100 uniformly distributed points in $[-5, 5]$ and solving the

corresponding eigenvalue problem, this results in the eigenfunctions shown in Figure 2. The probability densities p_ℓ *are defined by* $p_\ell(x) = |\psi_\ell(x)|^2$.

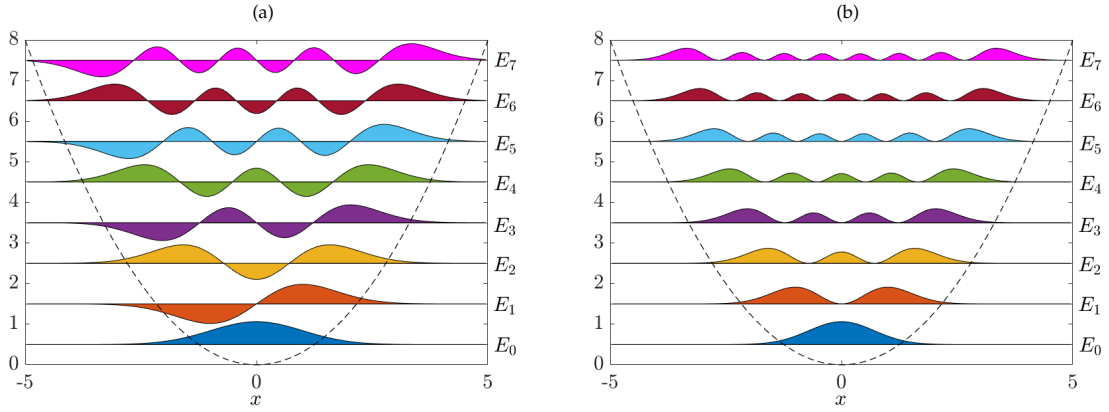


Figure 2. (**a**) Numerically computed eigenfunctions ψ_ℓ and associated energy levels E_ℓ of the quantum harmonic oscillator. The results are virtually indistinguishable from the analytical results. (**b**) Corresponding probability densities p_ℓ.

Example 4. *As a second example, let us analyze the Schrödinger equation for the hydrogen atom, where* $W(x) = -\frac{e^2}{4\pi\varepsilon_0\|x\|}$, *with* $x \in \mathbb{R}^3$. *Here, e is the electron charge and* ε_0 *the vacuum permittivity. Note that the parameter* **m** *in front of the Laplacian is the reduced mass of the system. As before, we define the physical constants to be one and use the Gaussian kernel, now with bandwidth* $\varsigma = 2$. *We then generate 5000 uniformly distributed test points in the ball with radius* 20 *and compute the Gram matrices* G_2 *and* G_0. *Solving the resulting eigenvalue problem, we obtain the eigenfunctions shown in Figure 3. As expected, there are several repeated eigenvalues (up to small perturbations due to the randomly sampled test points and numerical errors) for the higher energy states.*

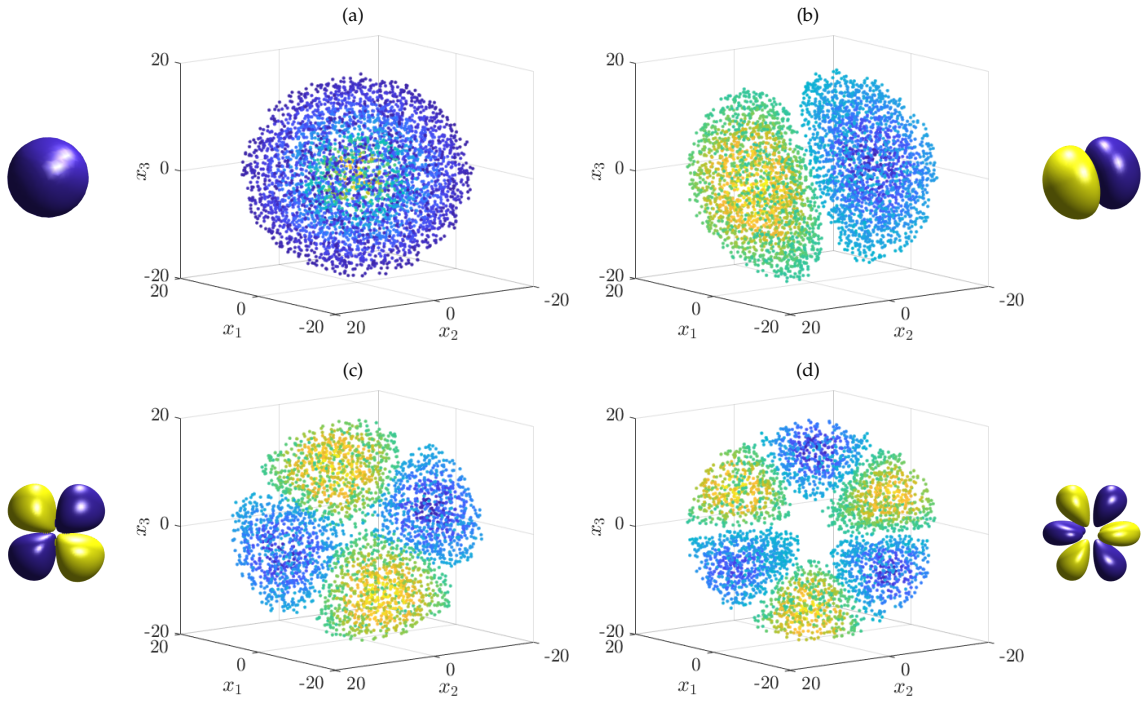


Figure 3. Numerically computed eigenfunctions of the Schrödinger equation associated with the hydrogen atom. Only points where the absolute value of the eigenfunction is larger than a given threshold are plotted. The shapes clearly resemble the well-known hydrogen atom orbitals shown next to the scatter plots. The eigenfunctions (or rotations thereof) correspond to the following quantum numbers $(\underline{n}, \underline{\ell}, \underline{m})$: (**a**) $(1,0,0)$, (**b**) $(2,1,1)$, (**c**) $(3,2,1)$, and (**d**) $(4,3,1)$.

4.2.2. SDE Formulation of the Schrödinger Equation

In order to derive gEDMD, we went from the stochastic differential equation to the Kolmogorov backward equation, which is the generator of the Koopman operator, or the adjoint Fokker–Planck equation, which is the generator of the Perron–Frobenius operator. Exploiting the resemblance between these two equations and the Schrödinger equation, we illustrated how data-driven methods can, in the same way, be used to compute wavefunctions. We now want to go in the opposite direction and find a stochastic differential equation whose eigenfunctions correspond to the wavefunctions. Formal similarities between quantum mechanics and the theory of stochastic processes have been investigated since the beginning of quantum mechanics by Schrödinger and others (see, for example, [37] and the references therein). The necessary transformations were already introduced in Section 2.3; we now want to exploit these relationships. Let us consider the two aforementioned examples again.

Example 5. *Using Corollary 1, the quantum harmonic oscillator can be transformed into an Ornstein–Uhlenbeck process:*

$$\mathrm{d}X_t = -\alpha\, X_t\, \mathrm{d}t + \sqrt{2\beta^{-1}}\, \mathrm{d}B_t,$$

with friction coefficient $\alpha = \hbar\omega$ *and temperature* $\beta^{-1} = \frac{\hbar^2}{2\mathbf{m}}$. *Since the eigenvalues of the Ornstein–Uhlenbeck process are* $\lambda_\ell = -\alpha\ell = -\hbar\omega\ell$, *the resulting eigenvalues of the quantum harmonic oscillator are* $E_\ell = E_0 - \lambda_\ell = \hbar\omega\left(\ell + \frac{1}{2}\right)$. *Correspondingly, the (unnormalized) eigenfunctions of the Ornstein–Uhlenbeck process are* $\varphi_\ell(x) = \widetilde{H}_\ell(\sqrt{\alpha\beta}\, x)$, *where* $\widetilde{H}_\ell$ *is the* ℓ^{th} *probabilists' Hermite polynomial. Thus,*

$$\psi_\ell(x) = \psi_0(x)\widetilde{H}_\ell\left(\sqrt{\frac{2\mathbf{m}\omega}{\hbar}}\, x\right) = \exp\left(-\frac{\mathbf{m}\omega}{2\hbar}x^2\right) H_\ell\left(\sqrt{\frac{\mathbf{m}\omega}{\hbar}}\, x\right),$$

which is consistent with the results obtained above. In the last step, we transformed the probabilists' Hermite polynomials into the physicists' Hermite polynomials.

Example 6. *Similarly, for the hydrogen atom, whose ground state is given by:*

$$\psi_0(x) = \frac{1}{\sqrt{\pi a_0^3}} \exp\left(-\frac{1}{a_0}\,\|x\|\right),$$

where $a_0 = \frac{4\pi\varepsilon_0\hbar^2}{\mathbf{m}e^2}$, *we obtain* $V(x) = \frac{\hbar^2}{\mathbf{m}\, a_0}\,\|x\|$, *and thus:*

$$\nabla V(x) = \frac{\hbar^2}{\mathbf{m}\, a_0}\frac{x}{\|x\|}.$$

There are now two options to compute the eigenfunctions numerically: we can either directly apply kernel gEDMD to the Koopman generator or generate time-series data by integrating the stochastic differential equation and then applying kernel EDMD or simply Ulam's method. We proceed with the former, but the latter leads to comparable results (although typically, more data points are required to achieve the same accuracy due to the stochasticity). We again generate uniformly distributed test points x_m *in the ball with radius* 20, *this time* $m = 10,000$, *and use the Gaussian kernel with bandwidth* $\varsigma = 2$. *This results in the same eigenfunctions as the ones shown in Figure 3. Due to the larger number of test points, even higher energy states can be well approximated. Two additional eigenfunctions are shown in Figure 4.*

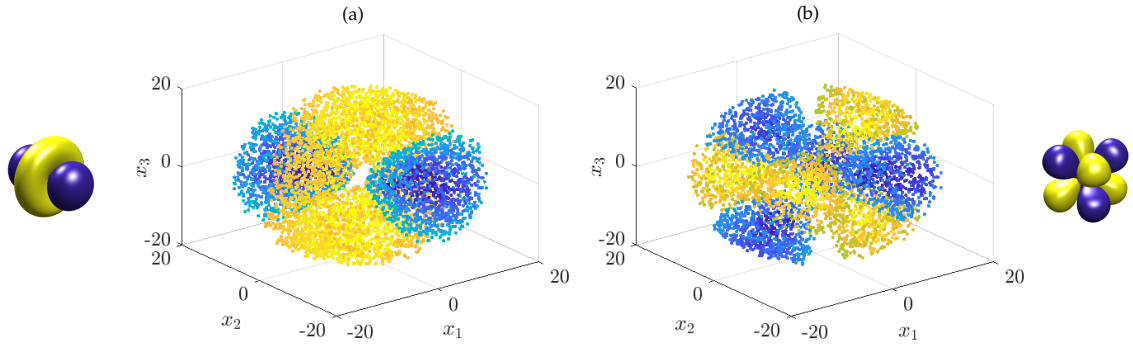


Figure 4. Eigenfunctions of the Schrödinger equation associated with the hydrogen atom computed by applying kernel gEDMD to the corresponding Koopman generator. The quantum numbers $(\underline{n}, \underline{\ell}, \underline{m})$ are: (**a**) $(3, 2, 0)$ and (**b**) $(4, 3, 2)$.

The examples illustrate that instead of solving partial differential equations, we can also compute eigenfunctions by approximating the Koopman operator from time-series data. The question under which conditions a non-degenerate strictly positive ground state exists needs to be addressed separately. One important theorem can be found in [38]:

Theorem 2. *Let $L^2_{loc}(\mathbb{X})$ be the space of locally square-integrable functions and $W \in L^2_{loc}(\mathbb{X})$ positive. Suppose $\lim_{|x|\to\infty} W(X) = \infty$, then $-\Delta + W$ has a non-degenerate strictly positive ground state.*

There are other results concerning the existence of such states; see [38] for details. Furthermore, diffusion Monte Carlo methods, which simultaneously compute the ground state energy and wavefunction, rely on similar assumptions [39]. However, in many cases of interest, the ground state of fermionic systems will have nodes so that these methods are not applicable [39]. The work presented here aims mainly at linking different operators describing the evolution of dynamical systems; more detailed relationships—in particular with the aforementioned diffusion Monte Carlo methods—and practical implications will be studied in future work.

4.3. Manifold Learning

So far, we assumed that the data were generated by a dynamical system. There is, however, a second scenario without any notion of time, where the Kolmogorov backward equation and Fokker–Planck equation are used for dimensionality reduction and manifold learning [21]; see also [20,22,23] and the references therein.

Let the data points $\{x_m\}_{m=1}^M$ be sampled from an arbitrary probability density ρ, then we can define the associated potential by:

$$U(x) = -\log \rho(x).$$

It was shown in [21] that, depending on some normalization parameter α, anisotropic diffusion maps approximate operators of the form:

$$\mathcal{L}_\alpha f = -2(1-\alpha)\nabla U \cdot \nabla f + \Delta f.$$

That is, for $\alpha = \frac{1}{2}$, we obtain the standard Kolmogorov backward equation with $\beta = 1$. Thus, the algorithms described above could also potentially be used for manifold learning purposes. We will illustrate this with a simple example.

Example 7. *We consider the well-known Swiss roll; see, for instance, [23]. The goal is to parametrize the two-dimensional manifold. We use kernel density estimation, cf. [40], and a Gaussian kernel with bandwidth* $\varsigma = 0.22$ *to learn* $U(x)$*, i.e.,*

$$U(x) = \frac{1}{M(\sqrt{2\pi}\varsigma)^d} \sum_{m=1}^{M} k(x, x_m)$$

and approximate the backward Kolmogorov operator by applying kernel gEDMD. Here, $M = 2000$ *and* $d = 2$*. The results are shown in Figure 5. The first eigenfunction parametrizes the angular direction, followed by higher order modes, and only the sixth eigenfunction corresponds to the* x_3 *direction. Considering these eigenfunctions as new coordinates, we obtain an unfolding of the roll. Note that also diffusion maps do not yield perfect rectangles in the embedded space due to the non-uniform density of points on the manifold [23].*

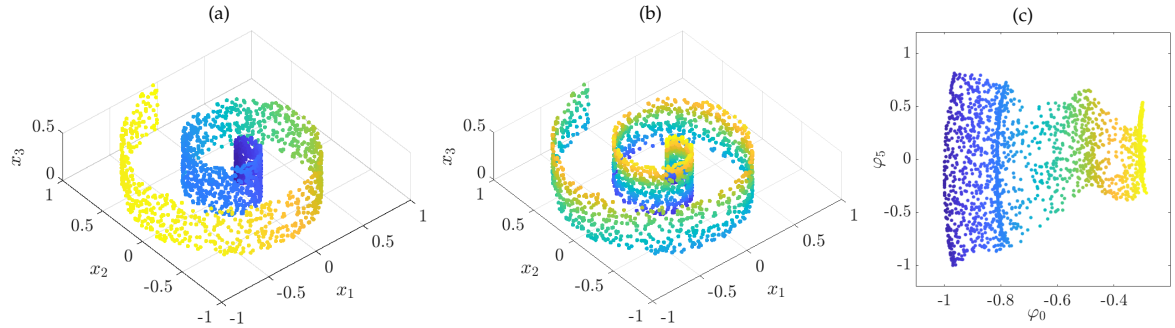


Figure 5. Swiss roll colored with respect to the eigenfunctions (**a**) φ_0 and (**b**) φ_5, which parametrize the angular and vertical direction, respectively. (**c**) Resulting two-dimensional embedding.

These results demonstrate that the eigenfunctions of certain differential operators capture geometrical properties of the data. However, the assumption that a strictly positive density in the ambient space exists will in general not be satisfied if the data are supported only on a lower dimensional manifold. This problem was circumvented by using kernel density estimation and a kernel with global support. Carrying over the definition of the differential operators involved and of their kernel-based analogues to the manifold case are beyond the scope of this paper and will be studied in future work. The same applies to the investigation of detailed relationships with diffusion maps or other manifold learning techniques. Concepts like neighborhood and sparsity will then need to be carried over to gEDMD to make this method amenable to large datasets. Furthermore, heuristics to find the optimal bandwidth ς are required since the results often strongly depend on the kernel hyperparameters.

5. Conclusions

Using the theory of derivative reproducing kernel Hilbert spaces, we derived a kernel-based formulation of gEDMD for approximating the Koopman generator, which allowed for the computation of eigenfunctions of potentially high-dimensional stochastic dynamical systems. If the system is reversible, the generator can be approximated from equilibrated time-series data, without having to estimate the drift and diffusion terms at the training data points. Furthermore, we showed that data-driven methods developed for the analysis of stochastic dynamical systems (kernel EDMD) can be carried over to their generators (kernel gEDMD) and, in turn, to the Schrödinger operator. Conversely, under certain assumptions on the ground state, the Schrödinger equation can be turned into a Kolmogorov backward equation corresponding to a drift-diffusion process. These results are summarized in Figure 6. Similar transformations also exist for the Fokker–Planck operator; see [24]. All derived approaches were illustrated with numerical results ranging from molecular dynamics to quantum mechanics.

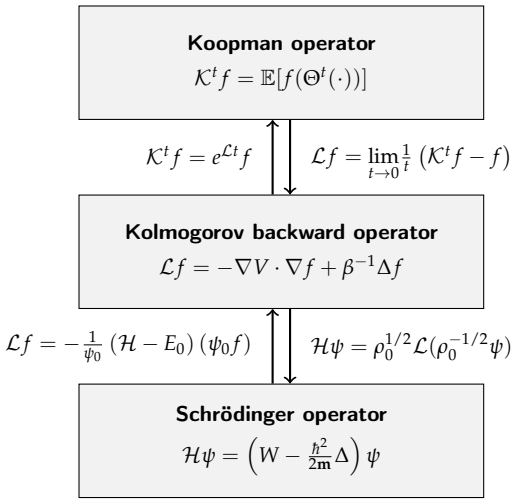


Figure 6. Relationships between the Koopman, Kolmogorov, and Schrödinger operators for a drift-diffusion process of the form $\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}B_t$. Here, ρ_0 denotes the invariant density, i.e., $\mathcal{L}^*\rho_0 = 0$. In our setting, the transformation of the Schrödinger operator requires a strictly positive real-valued ground state ψ_0.

Although we focused mainly on the Kolmogorov backward equation, the Fokker–Planck equation, and the Schrödinger equation, these methods can be applied to approximate other differential operators as well. An interesting open question is whether such algorithms can also be used for manifold learning. Some preliminary results were presented in Section 4, but a rigorous mathematical justification would require significant additional research. Analyzing connections with diffusion maps [20] or generalizations thereof in detail could be a potential direction for future work.

Another interesting avenue for future research could be to improve the efficiency and stability of the presented algorithms. Exploiting the properties of the given kernels, it might be possible to speed up computations significantly. The definition of a cutoff radius for the kernel or considering only a certain number of neighbors of data points, for instance, would—for suitable problems—result in sparse matrices. Moreover, the results sensitively depend on the hyperparameters such as the bandwidth of the Gaussian kernel. If the bandwidth is too small, this leads to overfitting and noisy eigenfunctions. If it is, on the other hand, too large, then the kernel is not able to capture the properties of the dynamical system accurately anymore. As a result, the Gram matrix G_0 has (numerically) essentially a low rank structure, and we obtain many zero eigenvalues. The question is then how to compute the smallest nonzero eigenvalues and corresponding eigenvectors efficiently.

Potential solutions for the hyperparameter tuning problem are techniques based on cross-validation [41] or so-called kernel flows [42]. By defining an optimization problem for the parameters of the kernel, e.g., based on a variational principle [43], gradient descent methods can help find suitable parameter values.

Author Contributions: Conceptualization, S.K., F.N., and B.H.; methodology, S.K. and F.N.; software, S.K. and F.N.; formal analysis, S.K. and F.N.; writing, original draft preparation, S.K., F.N., and B.H. All authors read and agreed to the published version of the manuscript.

Funding: S.K. was funded by Deutsche Forschungsgemeinschaft (DFG) through the grant CRC 1114 (Scaling Cascades in Complex Systems, project ID: 235221301) and through Germany's Excellence Strategy (MATH+: The Berlin Mathematics Research Center, EXC-2046/1, project ID: 390685689). B.H. thanks the European Commission for funding through the Marie Curie fellowships scheme.

Acknowledgments: The publication of this article was funded by Freie Universität Berlin. We would like to thank Luigi Delle Site for many helpful discussions regarding quantum chemistry and Amel Durakovic for useful discussions on the correspondence between the Schrödinger equation and complex Langevin dynamics.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Proof of Lemma 1. The unitary transformation is $\mathcal{H} = -e^{-\frac{1}{2}F}\mathcal{L}\left(e^{\frac{1}{2}F}\cdot\right)$. We obtain:

$$\begin{aligned}\mathcal{H}f &= e^{-\frac{1}{2}F}\mathcal{S}\left(e^{\frac{1}{2}F}f\right) + e^{-\frac{1}{2}F}J\cdot\left(e^{\frac{1}{2}F}\nabla f + \frac{1}{2}e^{\frac{1}{2}F}f\nabla F\right)\\ &= e^{-\frac{1}{2}F}\mathcal{S}\left(e^{\frac{1}{2}F}f\right) + J\cdot\nabla f + \frac{1}{2}J\cdot\nabla F f,\end{aligned}$$

which establishes the first-order term in $\mathcal{H}$ and the third term in the definition of W. For the symmetric part, we find that:

$$\begin{aligned}e^{-\frac{1}{2}F}\mathcal{S}\left(e^{\frac{1}{2}F}f\right) &= -\frac{1}{2}e^{\frac{1}{2}F}\nabla\cdot(e^{-F}a\nabla(e^{\frac{1}{2}F}f))\\ &= -\frac{1}{2}e^{\frac{1}{2}F}\nabla\cdot(e^{-\frac{1}{2}F}a[\nabla f + \frac{1}{2}\nabla F f])\\ &= -\frac{1}{2}\nabla\cdot(a\nabla f) + \frac{1}{4}\nabla F^{\top}a\nabla f - \frac{1}{4}e^{\frac{1}{2}F}\nabla\cdot(e^{-\frac{1}{2}F}fa\nabla F),\end{aligned}$$

which establishes the second-order term in the definition of $\mathcal{H}$. Expanding the third term above, we get:

$$-\frac{1}{4}e^{\frac{1}{2}F}\nabla\cdot(e^{-\frac{1}{2}F}fa\nabla F) = -\frac{1}{4}\nabla\cdot(a\nabla F)f - \frac{1}{4}\nabla f^{\top}a\nabla F + \frac{1}{8}\nabla F^{\top}a\nabla F f,$$

which cancels out the second term of the previous equation and establishes the remaining terms for W.

For the converse direction, we first translate the eigenvalue equation for ψ_0 into an equation for η:

$$\begin{aligned}0 = (\mathcal{H}-E_0)\psi_0 &= -\frac{1}{2}\nabla\cdot\left(a\nabla e^{-\eta}\right) + J\cdot\nabla e^{-\eta} + (W-E_0)e^{-\eta}\\ &= -\frac{1}{2}\nabla\cdot\left(-e^{-\eta}a\nabla\eta\right) - J\cdot\nabla\eta e^{-\eta} + (W-E_0)e^{-\eta}\\ &= e^{-\eta}\left[\frac{1}{2}\nabla\cdot(a\nabla\eta) - \frac{1}{2}\nabla\eta^{\top}a\nabla\eta - J\cdot\nabla\eta + W - E_0\right],\end{aligned}$$

implying that the term in brackets is also vanishing. Now, we define the negative generator by the transformation $-\mathcal{L} = e^{\eta}\left[\mathcal{H}-E_0\right](e^{-\eta}\cdot)$. Expanding the action of $-\mathcal{L}$, we find:

$$\begin{aligned}-\mathcal{L}f &= e^{\eta}\left[-\frac{1}{2}\nabla\cdot\left(a\nabla(e^{-\eta}f)\right) + J\cdot\nabla(e^{-\eta}f) + (W-E_0)(e^{-\eta}f)\right]\\ &= e^{\eta}\left[-\frac{1}{2}\nabla\cdot\left(e^{-\eta}a[\nabla f - \nabla\eta f]\right) + e^{-\eta}J\cdot\nabla f + (W-E_0-J\cdot\nabla\eta)(e^{-\eta}f)\right]\\ &= -\frac{1}{2}\nabla\cdot(a\nabla f) + \frac{1}{2}\nabla\eta^{\top}a\nabla f + \frac{1}{2}\nabla\cdot(a\nabla\eta)f + \frac{1}{2}\nabla\eta^{\top}a\nabla f - \frac{1}{2}\nabla\eta^{\top}a\nabla\eta f +\\ &\quad J\cdot\nabla f + (W-E_0-J\cdot\nabla\eta)f\\ &= -\frac{1}{2}\nabla\cdot(a\nabla f) + \nabla\eta^{\top}a\nabla f + J\cdot\nabla f\\ &= -\frac{1}{2}e^{2\eta}\nabla\cdot\left(e^{-2\eta}a\nabla f\right) + J\cdot\nabla f. \quad \square\end{aligned}$$

Proof of Lemma 2. We only show the proof for $\mathcal{T}_{\mathbb{H}}$. Similar to the argument in [44], $\mathcal{T}_{\mathbb{H}}$ is a bounded linear operator on $\mathbb{H}$ because of:

$$\left\| \int \phi(x)\otimes\left[-\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}a_{ij}(x)D^{e_i+e_j}\phi(x)\right]\mathrm{d}\mu(x)\right.$$

$$
\begin{aligned}
&+ \int \phi(x) \otimes \left[\sum_{i=1}^{d} \left(J_i(x) - \frac{1}{2} e^{F(x)} \nabla \cdot (e^{-F(x)} a_{:,i}(x)) \right) D^{e_i} \phi(x) \right] \mathrm{d}\mu(x) \\
&+ \int W(x) \phi(x) \otimes \phi(x) \, \mathrm{d}\mu(x) \Big\|_{HS} \\
\leq & \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \int |a_{ij}(x)| \| D^{e_i+e_j} \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
&+ \sum_{i=1}^{d} \int \left(|J_i(x)| + \frac{1}{2} e^{F(x)} |\nabla \cdot (e^{-F(x)} a_{:,i}(x))| \right) \| D^{e_i} \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
&+ \int |W(x)| \| \phi(x) \|_{\mathbb{H}} \| \phi(x) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
< & \infty.
\end{aligned}
$$

Using the derivative reproducing property, we obtain:

$$
\begin{aligned}
\langle \mathcal{T} f, g \rangle_{\mu} &= \int (\mathcal{T} f)(x) g(x) \, \mathrm{d}\mu(x) \\
&= \int \left[-\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] g(x) \, \mathrm{d}\mu(x) \\
&\quad + \int \left[\sum_{i=1}^{d} \left(J_i(x) - \frac{1}{2} e^{F(x)} \nabla \cdot (e^{-F(x)} a_{:,i}(x)) \right) \frac{\partial f}{\partial x_i} \right] g(x) \, \mathrm{d}\mu(x) \\
&\quad + \int W(x) f(x) g(x) \, \mathrm{d}\mu(x) \\
&= \int \left[-\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}(x) \left\langle D^{e_i+e_j} \phi(x), f \right\rangle_{\mathbb{H}} \right] \langle \phi(x), g \rangle_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
&\quad + \int \left[\sum_{i=1}^{d} \left(J_i(x) - \frac{1}{2} e^{F(x)} \nabla \cdot (e^{-F(x)} a_{:,i}(x)) \right) \langle D^{e_i} \phi(x), f \rangle_{\mathbb{H}} \right] \langle \phi(x), g \rangle_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
&\quad + \int W(x) \langle \phi(x), f \rangle_{\mathbb{H}} \langle \phi(x), g \rangle_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
&= - \int \left[\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}(x) \left\langle D^{e_i+e_j} \phi(x) \otimes \phi(x), f \otimes g \right\rangle_{\mathbb{H} \otimes \mathbb{H}} \right] \mathrm{d}\mu(x) \\
&\quad + \int \left[\sum_{i=1}^{d} \left(J_i(x) - \frac{1}{2} e^{F(x)} \nabla \cdot (e^{-F(x)} a_{:,i}(x)) \right) \langle D^{e_i} \phi(x) \otimes \phi(x), f \otimes g \rangle_{\mathbb{H} \otimes \mathbb{H}} \right] \mathrm{d}\mu(x) \\
&\quad + \int W(x) \langle \phi(x) \otimes \phi(x), f \otimes g \rangle_{\mathbb{H} \otimes \mathbb{H}} \, \mathrm{d}\mu(x) \\
&= \langle \mathcal{T}_{\mathbb{H}} f, g \rangle_{\mathbb{H}}.
\end{aligned}
$$

The same argument can be used to prove the statement about the symmetric case. □

Proof of Lemma 4. Let $f = \Phi u$ and $g = \Phi v$. Then:

$$
\begin{aligned}
\widehat{\mathcal{S}}_{\mathbb{H}}(f, g) &= \left\langle \left[\frac{1}{M} \sum_{m=1}^{M} \phi(x_m) \otimes \phi(x_m) \right] \sum_{r=1}^{M} u_r \phi(x_r), \sum_{s=1}^{M} v_s \phi(x_s) \right\rangle_{\mathbb{H}} \\
&= \frac{1}{M} \sum_{m=1}^{M} \sum_{r=1}^{M} \sum_{s=1}^{M} u_r v_s \langle \phi(x_m), \phi(x_r) \rangle_{\mathbb{H}} \langle \phi(x_m), \phi(x_s) \rangle_{\mathbb{H}} \\
&= \frac{1}{M} \sum_{m=1}^{M} \sum_{r=1}^{M} \sum_{s=1}^{M} u_r v_s k(x_m, x_r) k(x_m, x_s)
\end{aligned}
$$

$$= \frac{1}{M} \langle G_0 u,\, G_0 v \rangle .$$

Similarly,

$$\begin{aligned}
\widehat{\mathcal{Q}}_{\mathbb{H}}(f,g) &= \left\langle \left[\frac{1}{M} \sum_{m=1}^{M} \phi(x_m) \otimes \mathrm{d}\phi(x_m) \right] \sum_{r=1}^{M} u_r \phi(x_r),\, \sum_{s=1}^{M} v_s \phi(x_s) \right\rangle_{\mathbb{H}} \\
&= \frac{1}{M} \sum_{m=1}^{M} \sum_{r=1}^{M} \sum_{s=1}^{M} u_r v_s \, \langle \mathrm{d}\phi(x_m),\, \phi(x_r) \rangle_{\mathbb{H}} \, \langle \phi(x_m),\, \phi(x_s) \rangle_{\mathbb{H}} \\
&= \frac{1}{M} \langle G_2 u,\, G_0 v \rangle .
\end{aligned}$$

If the kernel functions at the training points are linearly independent, then G_0 is invertible, and it suffices to compute eigenvectors u of the generalized matrix eigenvalue problem $G_2 u = \lambda G_0 u$. In the symmetric case, the expression for the quadratic form $\widehat{\mathcal{Q}}_{\mathbb{H}}(f,g)$ changes to:

$$\begin{aligned}
\widehat{\mathcal{Q}}_{\mathbb{H}}(f,g) &= \frac{1}{2M} \sum_{l=1}^{d} \sum_{m=1}^{M} \sum_{r=1}^{M} \sum_{s=1}^{M} u_r v_s \left(\sigma_l(x_m)^\top \nabla k(x_m, x_r) \right) \left(\sigma_l(x_m)^\top \nabla k(x_m, x_s) \right) \\
&= \frac{1}{2M} \sum_{l=1}^{d} \left\langle G_1^{(l)} u,\, G_1^{(l)} v \right\rangle . \quad \square
\end{aligned}$$

References

1. Koopman, B. Hamiltonian systems and transformations in Hilbert space. *Proc. Natl. Acad. Sci. USA* **1931**, *17*, 315. doi:10.1073/pnas.17.5.315. [CrossRef]
2. Lasota, A.; Mackey, M.C. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics, 2nd ed. *Probab. Eng. Inf. Sci.*, **1994**, *10*, 311.
3. Mezić, I. Spectral Properties of Dynamical Systems, Model Reduction and Decompositions. *Nonlinear Dyn.* **2005**, *41*, 309–325. doi:10.1007/s11071-005-2824-x. [CrossRef]
4. Budišić, M.; Mohr, R.; Mezić, I. Applied Koopmanism. *Chaos: Interdiscip. J. Nonlinear Sci.* **2012**, *22*. doi:10.1063/1.4772195. [CrossRef]
5. Mauroy, A.; Mezić, I. Global stability analysis using the eigenfunctions of the Koopman operator. *IEEE Trans. Autom. Control* **2016**, *61*, 3356–3369. doi:10.1109/TAC.2016.2518918. [CrossRef]
6. Klus, S.; Koltai, P.; Schütte, C. On the numerical approximation of the Perron–Frobenius and Koopman operator. *J. Comput. Dyn.* **2016**, *3*, 51–79. doi:10.3934/jcd.2016003. [CrossRef]
7. Kaiser, E.; Kutz, J.N.; Brunton, S.L. Data-driven discovery of Koopman eigenfunctions for control. *arXiv* **2017**, arXiv:1707.01146.
8. Korda, M.; Mezić, I. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica* **2018**, *93*, 149–160. doi:10.1016/j.automatica.2018.03.046. [CrossRef]
9. Peitz, S.; Klus, S. Koopman operator-based model reduction for switched-system control of PDEs. *Automatica* **2019**, *106*, 184–191. doi:10.1016/j.automatica.2019.05.016. [CrossRef]
10. Klus, S.; Husic, B.E.; Mollenhauer, M.; Noé, F. Kernel methods for detecting coherent structures in dynamical data. *Chaos* **2019**. doi:10.1063/1.5100267. [CrossRef]
11. Williams, M.O.; Kevrekidis, I.G.; Rowley, C.W. A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *J. Nonlinear Sci.* **2015**, *25*, 1307–1346. doi:10.1007/s00332-015-9258-5. [CrossRef]
12. Williams, M.O.; Rowley, C.W.; Kevrekidis, I.G. A Kernel-Based Method for Data-Driven Koopman Spectral Analysis. *J. Comput. Dyn.* **2015**, *2*, 247–265. doi:10.3934/jcd.2015005. [CrossRef]
13. Klus, S.; Nüske, F.; Peitz, S.; Niemann, J.H.; Clementi, C.; Schütte, C. Data-driven approximation of the Koopman generator: Model reduction, system identification, and control. *Physica D* **2020**, *406*, 132416. doi:10.1016/j.physd.2020.132416. [CrossRef]

14. Mauroy, A.; Goncalves, J. Linear identification of nonlinear systems: A lifting technique based on the Koopman operator. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 6500–6505.
15. Klus, S.; Schuster, I.; Muandet, K. Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces. *J. Nonlinear Sci.* **2019**. doi:10.1007/s00332-019-09574-z. [CrossRef]
16. Zhou, D.X. Derivative reproducing properties for kernel methods in learning theory. *J. Comput. Appl. Math.* **2008**, *220*, 456–463. doi:10.1016/j.cam.2007.08.023. [CrossRef]
17. Giesl, P.; Hamzi, B.; Rasmussen, M.; Webster, K. Approximation of Lyapunov functions from noisy data. *J. Comput. Dyn.* **2019**. doi:10.3934/jcd.2020003. [CrossRef]
18. Haasdonk, B.; Hamzi, B.; Santin, G.; Witwar, D. Greedy Kernel Methods for Center Manifold Approximation. *arXiv* **2018**, arXiv:1810.11329.
19. Wendland, H. *Scattered Data Approximation*; Cambridge University Press: Cambridge, UK, 2004. doi:10.1017/CBO9780511617539. [CrossRef]
20. Coifman, R.R.; Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 5–30. [CrossRef]
21. Nadler, B.; Lafon, S.; Coifman, R.R.; Kevrekidis, I.G. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 113–127. doi:10.1016/j.acha.2005.07.004. [CrossRef]
22. Coifman, R.R.; Kevrekidis, I.G.; Lafon, S.; Maggioni, M.; Nadler, B. Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems. *Multiscale Model. Simul.* **2008**, *7*, 842–864. doi:10.1137/070696325. [CrossRef]
23. Nadler, B.; Lafon, S.; Coifman, R.R.; Kevrekidis, I.G. Diffusion Maps—A Probabilistic Interpretation for Spectral Embedding and Clustering Algorithms. In *Principal Manifolds for Data Visualization and Dimension Reduction*; Gorban, A., Kégl, B., Wunsch, D., Zinovyev, A., Eds.; Springer: Heidelberg, Germany, 2008; pp. 238–260.
24. Pavliotis, G.A. *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*; Springer: New York, NY, USA, 2014.
25. Levine, I.N. *Quantum Chemistry*; Prentice Hall: Upper Saddle River, NJ, USA, 2000.
26. Aronszajn, N. Theory of Reproducing Kernels. *Trans. Am. Math. Soc.* **1950**, *68*, 337–404. [CrossRef]
27. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*; MIT press: Cambridge, MA, USA, 2001.
28. Steinwart, I.; Christmann, A. *Support Vector Machines*, 1st ed.; Springer: New York, NY, USA, 2008.
29. Baker, C. Mutual Information for Gaussian Processes. *SIAM J. Appl. Math.* **1970**, *19*, 451–458. [CrossRef]
30. Baker, C. Joint Measures and Cross-Covariance Operators. *Trans. Am. Math. Soc.* **1973**, *186*, 273–289. [CrossRef]
31. Davies, E.B. *Spectral Theory and Differential Operators*; Cambridge University Press: Cambridge, UK, 1996; Volume 42.
32. Chacon, R.V. An ergodic theorem for operators satisfying norm conditions. *J. Math. Mech.* **1962**, *11*, 165–172.
33. Rosasco, L.; Belkin, M.; Vito, E.D. On Learning with Integral Operators. *J. Mach. Learn. Res.* **2010**, *11*, 905–934.
34. Klus, S. Data-Driven Dynamical Systems Toolbox. Available online: https://github.com/sklus/d3s/ (accessed on 1 May 2020).
35. Klus, S.; Bittracher, A.; Schuster, I.; Schütte, C. A kernel-based approach to molecular conformation analysis. *J. Chem. Phys.* **2018**, *149*, 244109. [CrossRef]
36. Froyland, G.; Rock, C.P.; Sakellariou, K. Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification. *Commun. Nonlinear Sci. Numer. Simul.* **2019**, *77*, 81–107. doi:10.1016/j.cnsns.2019.04.012. [CrossRef]
37. Okamoto, H. Stochastic formulation of quantum mechanics based on a complex Langevin equation. *J. Phys. A: Math. Gen.* **1990**, *23*, 5535–5545. doi:10.1088/0305-4470/23/23/025. [CrossRef]
38. Reed, M.; Simon, B. *Methods of Modern Mathematical Physics. IV Analysis of Operators*; Academic Press: San Diego, CA, USA, 1978.
39. Kosztin, I.; Faber, B.; Schulten, K. Introduction to the diffusion Monte Carlo method. *Am. J. Phys.* **1996**, *64*, 633–644. doi:10.1119/1.18168. [CrossRef]
40. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. doi:10.1214/aoms/1177704472. [CrossRef]

41. McGibbon, R.T.; Pande, V.S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **2015**, *142*, 03B621_1. [CrossRef]
42. Owhadi, H.; Yoo, G.R. Kernel Flows: From learning kernels from data into the abyss. *J. Comput. Phys.* **2019**, *389*, 22–47. doi:10.1016/j.jcp.2019.03.040. [CrossRef]
43. Wu, H.; Noé, F. Variational approach for learning Markov processes from time series data. *arXiv* **2017**, arXiv:1707.04659.
44. Muandet, K.; Fukumizu, K.; Sriperumbudur, B.; Schölkopf, B. Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.* **2017**, *10*, 1–141. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).