

Data-driven approximation of the Koopman generator: Model reduction, system identification, and control

Stefan Klus^{a,*}, Feliks Nüske^{b,c}, Sebastian Peitz^c, Jan-Hendrik Niemann^d,
Cecilia Clementi^b, Christof Schütte^{a,d}

^a Department of Mathematics and Computer Science, Freie Universität Berlin, Germany

^b Center for Theoretical Biological Physics and Department of Chemistry, Rice University, USA

^c Department of Mathematics, Paderborn University, Germany

^d Zuse Institute Berlin, Germany

ARTICLE INFO

Article history:

Received 23 September 2019

Received in revised form 7 February 2020

Accepted 12 February 2020

Available online 15 February 2020

Keywords:

Data-driven methods

Koopman operator

Infinitesimal generator

System identification

Coarse graining

Control

ABSTRACT

We derive a data-driven method for the approximation of the Koopman generator called gEDMD, which can be regarded as a straightforward extension of EDMD (*extended dynamic mode decomposition*). This approach is applicable to deterministic and stochastic dynamical systems. It can be used for computing eigenvalues, eigenfunctions, and modes of the generator and for system identification. In addition to learning the governing equations of deterministic systems, which then reduces to SINDy (*sparse identification of nonlinear dynamics*), it is possible to identify the drift and diffusion terms of stochastic differential equations from data. Moreover, we apply gEDMD to derive coarse-grained models of high-dimensional systems, and also to determine efficient model predictive control strategies. We highlight relationships with other methods and demonstrate the efficacy of the proposed methods using several guiding examples and prototypical molecular dynamics problems.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Data-driven approaches for the analysis of complex dynamical systems – be it methods to approximate transfer operators for computing metastable or coherent sets, methods to learn physical laws, or methods for optimization and control – have been steadily gaining popularity over the last years. Algorithms such as DMD [1,2], EDMD [3,4], SINDy [5], and their various kernel- [3,6,7], tensor- [8–10], or neural network-based [11–13] extensions and generalizations have been successfully applied to a plethora of different problems, including molecular and fluid dynamics, meteorology, finance, as well as mechanical and electrical engineering. An overview of different applications can be found, e.g., in [14]. Similar methods, developed mainly for reversible molecular dynamics problems, have been proposed in [15]. Most of the aforementioned techniques turn out to be strongly related, with the unifying concept being Koopman operator theory [16–18]. In what follows, we will focus mainly on the generator of the Koopman operator and its properties and applications.

SINDy [5] constitutes a milestone for data-driven discovery of dynamical systems. Because of the close relationship between the vector field of a deterministic dynamical system and its Koopman

generator, SINDy is a special case of the framework we will introduce in this study. In [19,20], an extension of SINDy to determine eigenfunctions of the Koopman generator was presented. The discovered eigenfunctions are then used for control, resulting in the so-called KRONIC framework. Another extension of SINDy was derived in [21], allowing for the identification of parameters of a stochastic system using Kramers–Moyal formulae.

A different avenue towards system identification was taken in [22,23]. Here, the Koopman operator is first approximated with the aid of EDMD, and then its generator is determined using the matrix logarithm. Subsequently, the right-hand side of the differential equation is extracted from the matrix representation of the generator. The relationship between the Koopman operator and its generator was also exploited in [24] for parameter estimation of stochastic differential equations.

A method for computing eigenfunctions of the Koopman generator was proposed in [25], where the diffusion maps algorithm is used to set up a Galerkin-projected eigenvalue problem with orthogonal basis elements. Two efficient methods for computing the generator of the adjoint Perron–Frobenius operator based on Ulam’s method and spectral collocation were presented in [26]. Provided that a model of the system dynamics is available, the computation of trajectories can be replaced by evaluations of the right-hand side of the system, which is often orders of magnitude faster.

* Corresponding author.

E-mail address: stefan.klus@fu-berlin.de (S. Klus).

The purpose of this study is to present a general framework to compute a matrix approximation of the Koopman generator, both for deterministic and stochastic systems, and to explore a range of applications. The main contributions of this work are:

1. We reformulate standard EDMD in such a way that it can be used to approximate the generator of the Koopman operator – as well as its eigenvalues, eigenfunctions, and modes – from data without resorting to trajectory integration. Exploiting duality, this can be extended naturally to the generator of the Perron–Frobenius operator.
2. We illustrate that the governing equations of deterministic as well as stochastic dynamical systems can be obtained from empirical estimates of the generator. Furthermore, we highlight relationships with related system identification techniques such as the Koopman lifting approach [22], SINDy [5], and KRONIC [19], which focus mainly on identifying ordinary differential equations.
3. Lastly, we explore two powerful applications of the approximated Koopman generator. We show that gEDMD can be used to identify coarse-grained models based on data of the full system, which is a highly relevant topic across different research fields, like molecular dynamics simulations for instance. Moreover, we apply the Koopman generator to control dynamical systems, providing flexible and efficient model predictive control strategies.

The efficacy of the resulting methods will be demonstrated with the aid of guiding examples and illustrative benchmark problems.

The remainder of this paper is structured as follows: In Section 2, we introduce the Koopman operator and its generator for different kinds of dynamical systems. We then derive an extension of EDMD for the approximation of the Koopman generator, named gEDMD, in Section 3. Furthermore, relationships with other methods are described. Section 4 explores additional applications of the proposed methods, namely coarse-graining and the application to control problems. Open questions and future work are discussed in Section 5.

2. The Koopman operator and its generator

In what follows, let $\mathbb{X}$ be the state space, e.g., $\mathbb{X} \subset \mathbb{R}^d$, and $f \in L^\infty(\mathbb{X})$ a real-valued observable of the system.

2.1. Deterministic dynamical systems

Given an ordinary differential equation of the form $\dot{x} = b(x)$, where $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$, the so-called *Koopman semigroup* of operators $\{\mathcal{K}^t\}$ is defined as

$$(\mathcal{K}^t f)(x) = f(\Phi^t(x)),$$

where Φ^t is the flow map, see [4,17,18]. That is, if $x(t)$ is a solution of the initial value problem with initial condition $x(0) = x_0$, then $\Phi^t(x_0) = x(t)$. The infinitesimal generator $\mathcal{L}$ of the semigroup, defined as

$$\mathcal{L}f = \lim_{t \rightarrow 0} \frac{1}{t} (\mathcal{K}^t f - f),$$

is given by

$$\mathcal{L}f = \frac{d}{dt}f = b \cdot \nabla_x f = \sum_{i=1}^d b_i \frac{\partial f}{\partial x_i},$$

see, e.g., [17]. Thus, if f is continuously differentiable, then $u(t, x) = \mathcal{K}^t f(x)$ satisfies the first-order partial differential equation $\frac{\partial u}{\partial t} = \mathcal{L}u$. The adjoint operator $\mathcal{L}^*$, i.e., the generator of the

Perron–Frobenius operator, is given by

$$\mathcal{L}^* f = - \sum_{i=1}^d \frac{\partial (b_i f)}{\partial x_i}.$$

Example 2.1. Throughout the paper, we will use the simple system

$$\begin{aligned} \dot{x}_1 &= \gamma x_1, \\ \dot{x}_2 &= \delta (x_2 - x_1^2), \end{aligned}$$

taken from [27], as a guiding example. In addition to the trivial eigenfunction $\varphi_1(x) = 1$ with corresponding generator eigenvalue $\lambda_1 = 0$, we obtain $\varphi_2(x) = x_1$ and $\varphi_3(x) = \frac{2\gamma-\delta}{\delta}x_2 + x_1^2$ with corresponding generator eigenvalues $\lambda_2 = \gamma$ and $\lambda_3 = \delta$, respectively. Moreover, products of eigenfunctions are again eigenfunctions. $\triangle$

2.2. Non-deterministic dynamical systems

Similarly, the definition of the Koopman operator can be generalized to stochastic differential equations

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t \quad (1)$$

as described, e.g., in [28], resulting in

$$(\mathcal{K}^t f)(x) = \mathbb{E}[f(\Phi^t(x))]. \quad (2)$$

Here, $\mathbb{E}[\cdot]$ denotes the expected value, $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the drift term, $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times s}$ the diffusion term, and W_t an s -dimensional Wiener process. Given a twice continuously differentiable function f , it can be shown using Itô's lemma that the infinitesimal generator of the stochastic Koopman operator is then characterized by

$$\mathcal{L}f = b \cdot \nabla_x f + \frac{1}{2} a : \nabla_x^2 f = \sum_{i=1}^d b_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}, \quad (3)$$

where $a = \sigma \sigma^\top$ and ∇_x^2 denotes the Hessian. Properties of the generator associated with non-deterministic dynamical systems are studied in [29]. The function $u(t, x) = \mathcal{K}^t f(x)$ satisfies the second-order partial differential equation $\frac{\partial u}{\partial t} = \mathcal{L}u$, which is called the *Kolmogorov backward equation* [30]. The adjoint operator in this case is

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

so that $\frac{\partial u}{\partial t} = \mathcal{L}^* u$ becomes the *Fokker–Planck equation* or *Kolmogorov forward equation* [17].

If μ is a stationary measure for the process X_t , the Koopman operator can be extended from $L^\infty_\mu(\mathbb{X})$ to the Hilbert space $L^2_\mu(\mathbb{X})$ with inner product $\langle f, g \rangle_\mu = \int_\mathbb{X} f(x)g(x)d\mu(x)$ [31]. We will frequently consider this situation in what follows. An important class of stochastic differential equations are those which are reversible with respect to a measure μ , which is necessarily a stationary measure in this case. The Koopman operator becomes self-adjoint on $L^2_\mu(\mathbb{X})$ in the reversible setting. Reversible systems can be characterized by the diffusion σ and a scalar potential $F: \mathbb{R}^d \rightarrow \mathbb{R}$, from which the drift is then obtained by

$$b = -\frac{1}{2} a \nabla F + \frac{1}{2} \nabla \cdot a,$$

where the divergence in the second term is applied to each column of a [32]. The generator of a reversible stochastic differential equation is a self-adjoint and typically unbounded operator on a suitable dense subspace of $L^2_\mu(\mathbb{X})$.

Remark 2.2. For systems of the form $dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t$, which play an important role in molecular dynamics, we obtain

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

Here, V describes the potential and β is the inverse temperature. The resulting dynamics are reversible with invariant measure $\mu(x) \sim \exp(-\beta V(x))$. The generator $\mathcal{L}$ is self-adjoint on $L^2_\mu(\mathbb{X})$ and it can be shown that, assuming suitable growth conditions on the potential, the spectrum of $\mathcal{L}$ is discrete [33].

Example 2.3. We will use the one-dimensional Ornstein–Uhlenbeck process, given by the stochastic differential equation

$$dX_t = -\alpha X_t dt + \sqrt{2\beta^{-1}}dW_t,$$

which is of the above form with $V(x) = \frac{1}{2}\alpha x^2$, as a second guiding example. The parameter α is the friction coefficient. The generator becomes self-adjoint in the space $L^2(\rho)$ weighted by the invariant density

$$\rho(x) = \frac{1}{\sqrt{2\pi\alpha^{-1}\beta^{-1}}} \exp\left(-\alpha\beta \frac{x^2}{2}\right)$$

and the eigenvalues λ_ℓ and eigenfunctions φ_ℓ are given by

$$\lambda_\ell = -\alpha(\ell - 1), \quad \varphi_\ell(x) = \frac{1}{\sqrt{(\ell - 1)!}} H_{\ell-1}\left(\sqrt{\alpha\beta}x\right), \quad \ell = 1, 2, \dots,$$

where H_ℓ denotes the ℓ th probabilists' Hermite polynomial [32]. That these functions are indeed eigenfunctions can be verified easily using recurrence relations for the Hermite polynomials, i.e., $H_{\ell+1}(x) = xH_\ell(x) - H'_\ell(x)$. $\triangle$

2.3. Galerkin approximation

Given a set of basis functions $\{\psi_i\}_{i=1}^n$, where $\psi_i: \mathbb{R}^d \rightarrow \mathbb{R}$, a Galerkin approximation $\mathbf{L}$ of the generator $\mathcal{L}$ can be obtained by computing the matrices $A, G \in \mathbb{R}^{n \times n}$ with

$$\begin{aligned} A_{ij} &= \langle \mathcal{L} \psi_i, \psi_j \rangle_\mu, \\ G_{ij} &= \langle \psi_i, \psi_j \rangle_\mu, \end{aligned} \quad (4)$$

where μ is a given measure. The matrix representation L of the projected operator $\mathbf{L}$ is then given by $L^\top = AG^{-1}$. We define $\psi(x) = [\psi_1(x), \dots, \psi_n(x)]^\top$. That is, for a function $f(x) = \sum_{i=1}^n c_i \psi_i(x) = c^\top \psi(x)$, it holds that $(\mathbf{L}f)(x) = (Lc)^\top \psi(x)$, where $c = [c_1, \dots, c_n]^\top \in \mathbb{R}^n$. It follows that an eigenvector ξ_ℓ of L corresponding to the eigenvalue λ_ℓ contains the coefficients for the eigenfunctions of $\mathbf{L}$ since defining $\varphi_\ell(x) = \xi_\ell^\top \psi(x)$ yields

$$(\mathbf{L}\varphi_\ell)(x) = (L\xi_\ell)^\top \psi(x) = \lambda_\ell \xi_\ell^\top \psi(x) = \lambda_\ell \varphi_\ell(x).$$

In many applications, the reciprocals of the generator eigenvalues (or their approximations) are also of interest, as they can be interpreted as decay time scales of dynamical processes in the system. We will refer to them as *implied time scales*

$$t_\ell := \frac{1}{\lambda_\ell}.$$

Example 2.4. For the Ornstein–Uhlenbeck process and a basis comprising monomials of order up to $n - 1$, i.e., $\psi(x) = [1, x, \dots, x^{n-1}]^\top$, we can compute the matrix L analytically. Note that $\mathcal{L}\psi_k$ is again in the subspace spanned by $\{\psi_i\}_{i=1}^n$. In particular, for $k \geq 3$, we have

$$(\mathcal{L}\psi_k)(x) = -\alpha(k-1)x^{k-1} + \beta^{-1}(k-1)(k-2)x^{k-3}$$

and the matrix $L \in \mathbb{R}^{n \times n}$ is of the form

$$L = \begin{matrix} & \begin{matrix} 1 & x & x^2 & x^3 & x^4 & x^5 & x^6 & \dots \end{matrix} \\ \begin{matrix} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \\ x^5 \\ x^6 \\ \vdots \end{matrix} & \begin{bmatrix} 0 & & & & & & & \\ & -\alpha & & & & & & \\ & & -2\alpha & & & & & \\ & & & -3\alpha & & & & \\ & & & & -4\alpha & & & \\ & & & & & -5\alpha & & \\ & & & & & & -6\alpha & \\ & & & & & & & \ddots \end{bmatrix} \end{matrix},$$

where the row and column labels correspond to the respective basis functions. The eigenvalues of the generator are given by $\lambda_\ell = -\alpha(\ell - 1)$, for $\ell = 1, \dots, n$, and the resulting eigenfunctions whose coefficients are given by the eigenvectors are the (transformed) probabilists' Hermite polynomials as described above. An approach to compute Hermite polynomials by solving an eigenvalue problem, resulting in a similar matrix representation, is also described in [34]. $\triangle$

Since we in general cannot compute the required integrals analytically, the aim is to estimate them from data using, e.g., Monte Carlo integration. More details regarding different types of Galerkin approximations and other methods for the approximation of transfer operators from data can be found in [4,35].

Remark 2.5. Issues pertaining to non-compactness or continuous spectra of Koopman operators associated with systems of high complexity are beyond the scope of this paper. Although such cases can theoretically be handled, the numerical analysis is often challenging and typically requires regularization, which is, for instance, implicitly given by Galerkin projections [25]. This is discussed in detail in the aforementioned work by Giannakis. Moreover, the projected generator does in general not result in a rate matrix, see [36,37] for details on Galerkin discretizations of transfer operators and their properties.

3. Infinitesimal generator EDMD

EDMD [4,38] was developed for the approximation of the Koopman or Perron–Frobenius operator from data. However, it can be reformulated to compute also the associated infinitesimal generators. We will call the resulting method gEDMD.

3.1. Deterministic dynamical systems

Let us first consider the deterministic case, which – albeit derived in another way and with different applications in mind – has already been studied in [19,20] so that we only briefly summarize and extend these results and then generalize them to the non-deterministic setting. Detailed relationships with other methods can be found in Section 3.3. We now assume that we have m measurements of the states of the system, given by $\{x_i\}_{i=1}^m$, and the corresponding time derivatives, given by $\{\dot{x}_i\}_{i=1}^m$. The derivatives might also be estimated from data, cf. [5].

3.1.1. Generator approximation

Similar to the Galerkin projection described above, we then choose a set of basis functions, also sometimes called *dictionary*, defined by $\{\psi_i\}_{i=1}^n$, and write this again in vector form as $\psi(x) = [\psi_1(x), \dots, \psi_n(x)]^\top$. Additionally, we define

$$\dot{\psi}_k(x) = (\mathcal{L}\psi_k)(x) = \sum_{i=1}^d b_i(x) \frac{\partial \psi_k}{\partial x_i}(x).$$

For all data points and basis functions, this can be written in matrix form as

$$\Psi_X = \begin{bmatrix} \psi_1(x_1) & \dots & \psi_1(x_m) \\ \vdots & \ddots & \vdots \\ \psi_n(x_1) & \dots & \psi_n(x_m) \end{bmatrix} \quad \text{and} \quad \dot{\Psi}_X = \begin{bmatrix} \dot{\psi}_1(x_1) & \dots & \dot{\psi}_1(x_m) \\ \vdots & \ddots & \vdots \\ \dot{\psi}_n(x_1) & \dots & \dot{\psi}_n(x_m) \end{bmatrix},$$

where $\Psi_X, \dot{\Psi}_X \in \mathbb{R}^{n \times m}$. The partial derivatives of the basis functions required for $\dot{\psi}_k(x_i)$ can be precomputed analytically.¹ Note that we additionally need $b(x_i)$ which is simply $\dot{x}_i$. If the time derivatives cannot be measured directly, they can be approximated using, e.g., finite differences. We now assume there exists a matrix M such that $\dot{\Psi}_X = M\Psi_X$. Since this equation in general cannot be satisfied exactly, we solve it in the least squares sense – analogously to the derivation of EDMD – by minimizing $\|\dot{\Psi}_X - M\Psi_X\|_F$, resulting in

$$M = \dot{\Psi}_X \Psi_X^+ = (\dot{\Psi}_X \Psi_X^T)(\Psi_X \Psi_X^T)^+ = \widehat{A} \widehat{G}^+,$$

with

$$\widehat{A} = \frac{1}{m} \sum_{i=1}^m \dot{\psi}(x_i) \psi(x_i)^T \quad \text{and} \quad \widehat{G} = \frac{1}{m} \sum_{i=1}^m \psi(x_i) \psi(x_i)^T.$$

We call this approach gEDMD. The advantage is that the generator might be sparse even when the Koopman operator for the time- t map is not.

Remark 3.1. The sparsification approach proposed for SINDy, see [5], can be added in the same way to gEDMD in order to minimize the number of spurious nonzero entries caused, for instance, by the numerical approximation of the time derivatives or by noisy data.

The convergence to the Galerkin approximation in the infinite data limit will be shown for the non-deterministic case, the deterministic counterpart follows as a special case. The matrix M is thus an empirical estimate of L^T and we write $M = \widehat{L}^T = \widehat{A} \widehat{G}^+$. Accordingly, exploiting duality, the matrix representation of the adjoint operator $\mathcal{L}^*$, the generator of the Perron-Frobenius operator, is given by $M^* = (\widehat{L}^*)^T = \widehat{A}^T \widehat{G}^+$. A detailed derivation for standard EDMD, which can be carried over to gEDMD, can be found in [4]. The convergence of the standard EDMD approximation to the Koopman operator as the number of basis functions goes to infinity is discussed in [39]. Whether the results can be extended to gEDMD will be studied in future work.

Example 3.2. Let us again consider the system defined in Example 2.1 using monomials up to order 8. We set $\gamma = -0.8$ and $\delta = -0.7$ and generate 1000 uniformly distributed test points in $[-2, 2] \times [-2, 2]$. Then gEDMD results in eigenvalues and (rescaled) eigenfunctions

$$\begin{aligned} \lambda_1 &\approx 0, & \varphi_1(x) &\approx 1, \\ \lambda_2 &\approx -0.7 = \delta, & \varphi_2(x) &= 1.286 x_2 + 1.000 x_1^2 \approx \frac{2\gamma-\delta}{\delta} x_2 + x_1^2, \\ \lambda_3 &\approx -0.8 = \gamma, & \varphi_3(x) &\approx x_1. \end{aligned}$$

The subsequent eigenfunctions are products of the above eigenfunctions, we obtain, for instance, $\lambda_6 \approx -1.6 = 2\gamma$ with $\varphi_6(x) = 1.000 x_1^2 \approx \varphi_3(x)^2$. Note that the ordering of the eigenvalues, which are typically sorted by decreasing values, and associated eigenfunctions depends on the values of γ and δ . $\triangle$

¹ Alternatively, automatic differentiation or symbolic computing toolboxes could be utilized.

3.1.2. System identification

With the aid of the full-state observable $g(x) = x$, it is possible to reconstruct the governing equations of the underlying dynamical system. Note that $\mathbb{X}$ needs to be bounded here – and for the identification of stochastic differential equations introduced below – so that g is (component-wise) contained in $L^\infty(\mathbb{X})$. Let ξ_ℓ be the ℓ th eigenvector of $\widehat{L}$ and $\mathcal{E} = [\xi_1, \dots, \xi_n]$. Furthermore, assume that $B \in \mathbb{R}^{n \times d}$ is the matrix such that $g(x) = B^T \psi(x)$. This can be easily accomplished by adding the observables $\{x_i\}_{i=1}^d$ to the dictionary. In order to obtain the Koopman modes for the full-state observable, define $\varphi(x) = [\varphi_1(x), \dots, \varphi_n(x)]^T = \mathcal{E}^T \psi(x)$. Then

$$g(x) = B^T \psi(x) = B^T \mathcal{E}^{-T} \varphi(x).$$

The column vectors of the matrix $V = B^T \mathcal{E}^{-T}$ are the Koopman modes v_ℓ . We obtain

$$(\mathcal{L}g)(x) = b(x) \approx \sum_{\ell=1}^n \lambda_\ell \varphi_\ell(x) v_\ell,$$

where the generator is applied component-wise. This allows us to decompose a system into different frequencies. The derivation of the modes is equivalent to the standard EDMD case, see [4,38] for more details. Instead of representing the system in terms of the eigenvalues, eigenfunctions, and modes of the generator, we can also express it directly in terms of the basis functions, i.e.,

$$(\mathcal{L}g)(x) = b(x) \approx (LB)^T \psi(x),$$

which is then equivalent to SINDy, see Section 3.3.

Example 3.3. Using the eigenvalues λ_ℓ and corresponding eigenfunctions $\varphi_\ell(x)$ as determined in Example 3.2, we can reconstruct the dynamical system from Example 2.1. Only the Koopman modes $v_2 = [0, 0.778]^T \approx [0, \frac{\delta}{2\gamma-\delta}]^T$, $v_3 = [1, 0]^T$, and $v_6 = [0, -0.778]^T \approx [0, -\frac{\delta}{2\gamma-\delta}]^T$ are required for the reconstruction, the other modes are numerically zero. That is,

$$b(x) \approx \lambda_2 \varphi_2(x) v_2 + \lambda_3 \varphi_3(x) v_3 + \lambda_6 \varphi_6(x) v_6 \approx \begin{bmatrix} \gamma x_1 \\ \delta (x_2 - x_1^2) \end{bmatrix}.$$

Expressing the system directly in terms of the basis functions results in the same representation, the governing equations are hence identified correctly in both cases. $\triangle$

Remark 3.4. In the above example, we assumed that the derivatives for the training data are known or can be computed with sufficient accuracy. If the derivatives, however, are noisy or inaccurate, the resulting matrix representations of the operators often become nonsparse and additional techniques such as denoising, total-variation regularization, or iterative hard thresholding might be required to eliminate spurious nonzero entries, see also [5] and references therein. In order to model the presence of noise, we replace $b(x_i)$ by $b(x_i) + \eta$, where η is sampled from a Gaussian distribution with standard deviation ς . By adding the iterative hard thresholding procedure proposed in [5] to gEDMD, which step by step removes entries larger than a given threshold δ and then recomputes the coefficients, we can eliminate unwanted entries. The results, however, depend strongly on the chosen threshold as shown in Fig. 1. The smaller the signal-to-noise ratio, the larger the threshold needs to be to eliminate spurious nonzero entries, but a too large threshold will also eliminate the actual coefficients. The error here is defined to be the average difference between the true and the estimated coefficients after 10 iterations of the hard thresholding algorithm.

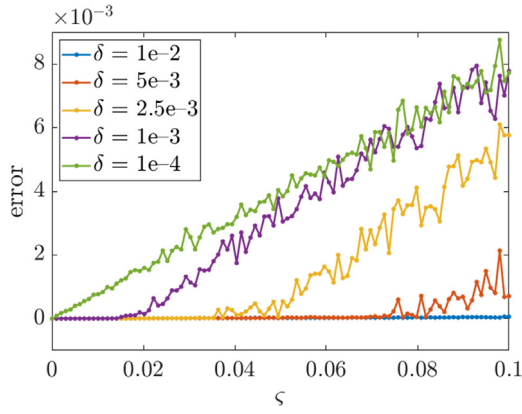


Fig. 1. Recovery error as a function of the standard deviation ς for different thresholds δ . If no thresholding is used, the results coincide with the $\delta = 1e-4$ case. The results show that for inaccurate estimates of the derivatives additional techniques are required to obtain suitable representations of the system. Provided that the cut-off value is chosen judiciously, the hard thresholding approach enables us to recover the correct dynamics even in the presence of noise.

3.1.3. Conservation laws

A function $E: \mathbb{R}^d \rightarrow \mathbb{R}$ is said to be a *conserved quantity* if it remains constant for all t and all initial values, i.e., $\frac{d}{dt}E = \nabla E \cdot b = 0$, which immediately implies that E is an eigenfunction of the Koopman generator corresponding to the eigenvalue $\lambda = 0$; such invariants have already been considered in Koopman's original paper [16]. Similarly, eigenfunctions of the Perron–Frobenius generator associated with $\lambda = 0$ represent invariant densities. Conservation laws play an important role in physics and engineering, but are in principle hard to discover. The relationship between conservation laws and Koopman eigenfunctions has recently been exploited in [19,20], where conserved quantities are learned from data. In the same way, we can apply gEDMD to find non-trivial eigenfunctions corresponding to $\lambda = 0$.

3.2. Non-deterministic dynamical systems

Let us now extend these results to stochastic differential equations of the form (1). Given a set of training data $\{x_i\}_{i=1}^m$ as above, we assume that $\{b(x_i)\}_{i=1}^m$ and $\{\sigma(x_i)\}_{i=1}^m$ are known or can be estimated.

3.2.1. Generator approximation

Let

$$d\psi_k(x) = (\mathcal{L}\psi_k)(x) = \sum_{i=1}^d b_i(x) \frac{\partial \psi_k}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij}(x) \frac{\partial^2 \psi_k}{\partial x_i \partial x_j}(x) \quad (5)$$

and

$$d\psi_X = \begin{bmatrix} d\psi_1(x_1) & \dots & d\psi_1(x_m) \\ \vdots & \ddots & \vdots \\ d\psi_k(x_1) & \dots & d\psi_k(x_m) \end{bmatrix}.$$

That is, in addition to the first derivatives of the basis functions, we now also need the second derivatives, which can again be precomputed analytically. Solving the resulting minimization problem, this leads to the least-squares approximation

$$M = d\psi_X \psi_X^+ = (d\psi_X \psi_X^T)(\psi_X \psi_X^T)^+ = \widehat{A} \widehat{G}^+,$$

with

$$\widehat{A} = \frac{1}{m} \sum_{i=1}^m d\psi(x_i) \psi(x_i)^T \quad \text{and} \quad \widehat{G} = \frac{1}{m} \sum_{i=1}^m \psi(x_i) \psi(x_i)^T.$$

As above, we obtain $M = \widehat{L}^T = \widehat{A} \widehat{G}^+$ as an empirical estimate of the generator and $M^* = (\widehat{L}^*)^T = \widehat{A}^T \widehat{G}^+$ as an estimate of the adjoint operator.

Proposition 3.5. *In the infinite data limit, gEDMD converges to the Galerkin projection of the generator onto the space spanned by the basis functions $\{\psi_i\}_{i=1}^n$.*

Proof. The proof is equivalent to the counterpart for standard EDMD, see [4,38]. Letting m go to infinity, we obtain

$$\widehat{A}_{ij} = \frac{1}{m} \sum_{l=1}^m d\psi_i(x_l) \psi_j(x_l) \xrightarrow{m \rightarrow \infty} \int (\mathcal{L}\psi_i)(x) \psi_j(x) d\mu(x) = \langle \mathcal{L}\psi_i, \psi_j \rangle_\mu = A_{ij},$$

$$\widehat{G}_{ij} = \frac{1}{m} \sum_{l=1}^m \psi_i(x_l) \psi_j(x_l) \xrightarrow{m \rightarrow \infty} \int \psi_i(x) \psi_j(x) d\mu(x) = \langle \psi_i, \psi_j \rangle_\mu = G_{ij},$$

where $x_l \sim \mu$. That is, the matrices $\widehat{A}$ and $\widehat{G}$ are empirical estimates of the matrices A and G , respectively. $\square$

Remark 3.6. If the drift and diffusion coefficients of the stochastic differential equation (1) are not known, they can be approximated via finite differences. In fact, by the Kramers–Moyal formulae,

$$b(x) = \lim_{t \rightarrow 0} b^t(x) := \lim_{t \rightarrow 0} \mathbb{E} \left[\frac{1}{t} (X_t - x) \mid X_0 = x \right],$$

$$a(x) = \lim_{t \rightarrow 0} a^t(x) := \lim_{t \rightarrow 0} \mathbb{E} \left[\frac{1}{t} (X_t - x)(X_t - x)^T \mid X_0 = x \right].$$

These expressions can be evaluated pointwise by spawning multiple short trajectories from each data point x_i , and then estimating the expectations above via Monte Carlo. Alternatively, if a single ergodic simulation at time step t is available, we can also replace the definition of $d\psi_k$ in (5) by

$$d\psi_k(x_i) = \frac{1}{t} (x_{i+1} - x_i) \cdot \nabla \psi_k(x_i) + \frac{1}{2t} [(x_{i+1} - x_i)(x_{i+1} - x_i)^T] : \nabla^2 \psi_k(x_i).$$

It was shown in [21] that in the infinite data limit

$$\lim_{m \rightarrow \infty} \widehat{A}_{ij} = \left\langle b^t \cdot \nabla \psi_i + \frac{1}{2} a^t : \nabla^2 \psi_i, \psi_j \right\rangle_\mu.$$

In this case, gEDMD converges to a Galerkin approximation of the differential operator with drift and diffusion coefficients b^t and a^t .

Remark 3.7. If the stochastic dynamics (1) are reversible with respect to the measure μ , we only require first-order derivatives of the basis. In this case, the Galerkin matrix A in (4) can be expressed as

$$A_{ij} = \langle \mathcal{L}\psi_i, \psi_j \rangle_\mu = -\frac{1}{2} \int \nabla \psi_i \sigma \sigma^T \nabla \psi_j^T d\mu,$$

where the drift coefficient enters only implicitly via the invariant measure μ , see [40]. Using the gradient matrix $\nabla \Psi \in \mathbb{R}^{n \times d}$, where each row corresponds to the gradient of a basis function, the empirical estimator $\widehat{A}$ for A is then defined as follows:

$$\widehat{A} = -\frac{1}{2m} \sum_{i=1}^m d\psi(x_i) d\psi(x_i)^T,$$

with $d\psi(x_i) = \nabla \Psi(x_i) \sigma(x_i)$.

Example 3.8. Let us first compute eigenfunctions of the generator. We assume that $\{b(x_i)\}_{i=1}^m$ and $\{\sigma(x_i)\}_{i=1}^m$ are known and not estimated from data.

1. We consider again the Ornstein–Uhlenbeck process defined in Example 2.3. For the numerical experiments, we set $\alpha = 1$ and $\beta = 4$ and select a basis comprising

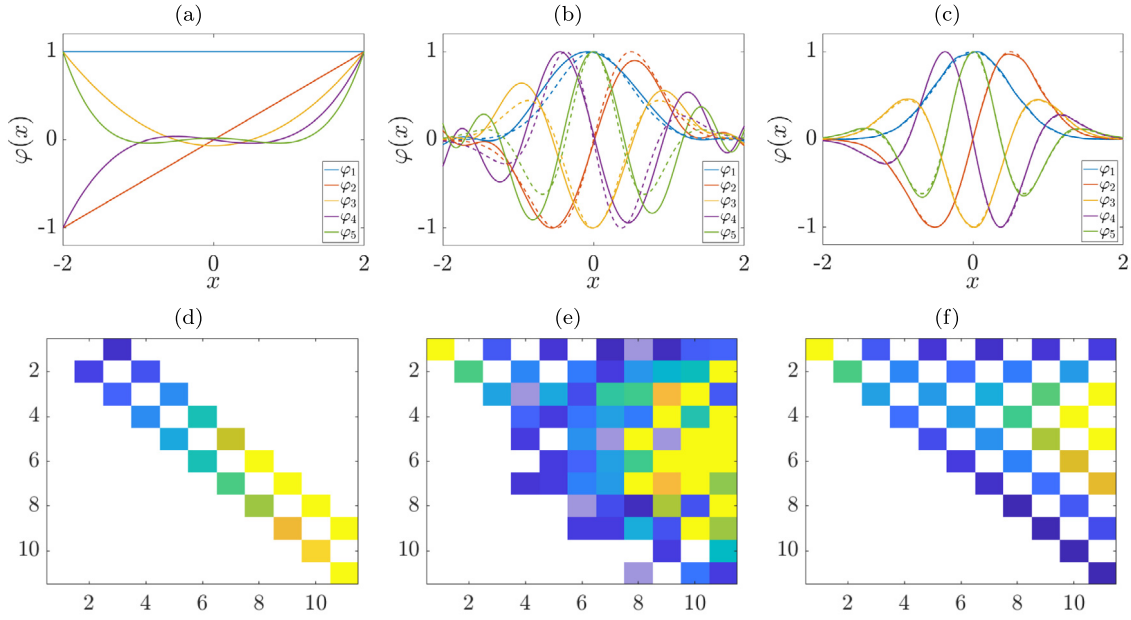


Fig. 2. Eigenfunctions of (a) the Koopman generator and (b) the Perron-Frobenius generator associated with the Ornstein-Uhlenbeck process computed using gEDMD with monomials of order up to ten. The dashed lines represent the analytically computed eigenfunctions. (c) Eigenfunctions of the Perron-Frobenius generator, where the basis now comprises 30 Gaussian functions. (d) Sparsity pattern of L computed with gEDMD, (e) sparsity pattern of $\tilde{K}_\tau$ computed with EDMD, and (f) sparsity pattern of $\exp(\tau L)$, where τ is the lag time used for EDMD.

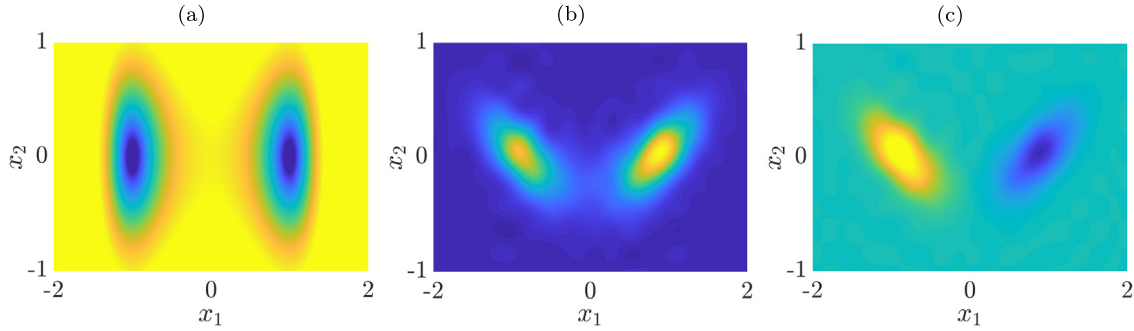


Fig. 3. (a) Double-well potential. (b) First and (c) second eigenfunctions of the Perron-Frobenius generator. Due to the non-isotropic noise the wells are tilted. The second eigenfunction clearly separates the two wells. In all plots, blue corresponds to small and yellow to large values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

monomials of order up to and including ten. Using only 100 uniformly generated test points in $\mathbb{X} = [-2, 2]$, we obtain the Koopman eigenfunctions shown in Fig. 2(a), which are virtually indistinguishable from the analytical solution. Standard EDMD would typically need more test points for such an accurate approximation of the dominant eigenfunctions, see [35] for details.² The results for the Perron-Frobenius generator using monomials are not as good, see Fig. 2(b). Replacing monomials by a basis containing Gaussian functions the results improve considerably as shown in Fig. 2(c). This illustrates that it is crucial to select suitable basis functions, which are, however, generally not known in advance. The sparsity patterns of the generator approximation using EDMD and gEDMD are compared in Fig. 2(d-f), showing that EDMD leads to less sparse matrices with additional spurious nonzero entries. Δ

2. We construct a more complicated example by defining $V(x) = (x_1^2 - 1)^2 + x_2^2$, which represents the renowned double-well potential, but then, instead of using isotropic

noise, add a state-dependent diffusion term to obtain a stochastic differential equation of the form (1), with

$$b(x) = -\nabla V(x) = \begin{bmatrix} 4x_1 - 4x_1^3 \\ -2x_2 \end{bmatrix} \quad \text{and} \quad \sigma(x) = \begin{bmatrix} 0.7 & x_1 \\ 0 & 0.5 \end{bmatrix}.$$

The system exhibits metastable behavior, where the rare transitions are the jumps between the two wells. The potential and the two dominant eigenfunctions of the Perron-Frobenius generator computed with the aid of gEDMD are shown in Fig. 3. Here, we generated 30 000 test points in $\mathbb{X} = [-2, 2] \times [-1, 1]$ and selected a basis comprising 300 radial basis functions (whose centers are the midpoints of a regular box discretization) with bandwidth $\sigma = 0.2$. Δ

3.2.2. System identification

As for deterministic systems, we can utilize the generator approximation also for system identification. In order to determine b , we simply plug in the full-state observable g again. In addition to the drift term, we need to identify the diffusion term. This can be accomplished as follows: Note that for $\psi_k(x) = x_i x_j$, it holds that

$$(\mathcal{L}\psi_k)(x) = b_i(x)x_j + b_j(x)x_i + a_{ij}(x).$$

² Note that although the definition of the Ornstein-Uhlenbeck process is slightly different in [35], the systems are in fact identical.

Since we already obtained a representation of b in the previous step, we can subtract the first two terms to obtain a_{ij} . Here, we have to assume that both b_i and b_j can be written in terms of the basis functions and that, furthermore, also the functions multiplied by x_j or x_i , respectively, are contained in the space spanned by $\{\psi_i\}_{i=1}^n$. For instance, if b contains monomials of degree p , then the dictionary must also contain monomials of degree $p+1$. For other types of basis functions, we have to make sure that the aforementioned requirement is satisfied as well.

Example 3.9. Let us illustrate the recovery of b and a from the generator representation.

1. For the Ornstein–Uhlenbeck process, we immediately obtain $b(x) = (\mathcal{L}\psi_2)(x) = -\alpha x$ and $a(x) = (\mathcal{L}\psi_3)(x) - 2b(x)x = 2\beta^{-1}$, see the matrix representation of the generator in Example 2.4, which implies $\sigma(x) = \sqrt{2\beta^{-1}}$. Thus, the system is identified correctly.
2. For the double-well problem, we generate 8000 random points in $\mathbb{X} = [-2, 2] \times [-1, 1]$ and use the exact values for $b(x)$ and $\sigma(x)$. We then obtain an approximation of the generator whose first six columns for a dictionary comprising monomials up to order four are given by

$$\begin{array}{c} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \\ x_1^3 \\ x_1^2 x_2 \\ x_1 x_2^2 \\ x_2^3 \\ x_1^4 \\ x_1^3 x_2 \\ x_1^2 x_2^2 \\ x_1 x_2^3 \\ x_2^4 \end{array} \begin{bmatrix} 1 & x_1 & x_2 & x_1^2 & x_1 x_2 & x_2^2 \\ 0 & 0 & 0 & 0.49 & 0 & 0.25 \\ 0 & 4 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & -4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -8 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We can see that b is recovered correctly by the columns two and three. Furthermore, for the entries of the matrix a , we obtain

$$\begin{aligned} a_{11}(x) &= (\mathcal{L}\psi_4)(x) - 2b_1(x)x_1 = 0.49 + x_1^2, \\ a_{12}(x) &= (\mathcal{L}\psi_5)(x) - b_1(x)x_2 - b_2(x)x_1 = 0.5x_1, \\ a_{22}(x) &= (\mathcal{L}\psi_6)(x) - 2b_2(x)x_2 = 0.25, \end{aligned}$$

which is indeed $\sigma\sigma^\top$. Note that using only monomials of order up to three would allow us to recover b but not a . $\triangle$

Remark 3.10. It is worth noting that:

1. Although we presented only systems composed of monomials (mainly for the sake of illustration), the proposed method allows for arbitrary dictionaries containing twice continuously differentiable functions.
2. We identify $a = \sigma\sigma^\top$ and not σ itself. If it is necessary to evaluate σ , e.g., when using the identified system to generate new trajectories, we can obtain it, for instance, by a Cholesky decomposition of a , see also [40]. Note, however, that σ is not uniquely defined.
3. The method relies on accurate estimates of the drift and diffusion terms. Noisy data will lead to nonsparse solutions, which can then be improved by applying iterative hard thresholding again, see Remark 3.4. We now add noise with variance $\varsigma = 0.1$ to the drift and diffusion terms.

After sparsifying the estimated matrix approximation of the Koopman generator with a threshold $\delta = 0.1$, we obtain

$$b(x) = \begin{bmatrix} 4.00057x_1 - 4.00012x_1^3 \\ -1.99998x_2 \end{bmatrix}$$

and

$$\begin{aligned} a_{11}(x) &= 0.50035 + 0.99901x_1^2 - 0.00016x_1^4, \\ a_{12}(x) &= 0.49729x_1 - 0.00250x_1x_2 + 0.00097x_1^3x_2, \\ a_{22}(x) &= 0.25648 + 0.00720x_2^2. \end{aligned}$$

Note that the noise is picked up by the diffusion term which might thus be overestimated. Nevertheless, the coefficients are still close to the exact solution. Instead of eliminating small coefficients of $(\mathcal{L}\psi_k)(x)$, we could apply iterative hard thresholding to the coefficients of $a_{ij}(x)$ to find a parsimonious representation of $a(x)$.

This method to discover the drift and diffusion terms of stochastic differential equations suffers from the same shortcomings as SINDy: The validity of the learned model depends crucially on whether or not both b and a can be expressed in terms of the basis functions and also on the availability of accurate estimates of the derivatives. Ideally, the resulting model is parsimonious, minimizing model complexity while simultaneously enabling accurate predictions without overfitting. Nonsparse solutions typically indicate that the expressivity of the dictionary is not sufficient or that the data is too noisy. Adding more basis functions or increasing the size of the data set might alleviate such problems. However, positing that the model comprises only a few simple terms, the method presented here allows for the identification of the governing equations of stochastic dynamical systems. Additionally, the approximation of the generator is an important problem in itself. The eigenvalues and eigenfunctions contain information about time scales and metastable sets and can be used for model reduction and control. This will be described in more detail in Section 4.

3.2.3. Conservation laws

If E is a conserved quantity of a non-deterministic system, then the definition of the Koopman operator (2) and the partial differential equation $\frac{\partial u}{\partial t} = \mathcal{L}u$ imply that $\mathcal{L}E = 0$, just as in the deterministic case. Hence, conserved quantities can also be approximated by extracting non-trivial eigenfunctions associated with $\lambda = 0$ using gEDMD. The same precautions as discussed in Section 3.1.3 apply.

Remark 3.11. For a stochastic dynamical system in the sense of Stratonovich, i.e.,

$$dX_t = b(X_t)dt + \sigma(X_t) \circ dW_t,$$

a sufficient condition for E to be conserved is

$$\nabla E^\top \left[b + \sum_{i=1}^s \sigma_i \right] = 0,$$

which is similar to the deterministic case. Here, σ_i denotes the i th column of σ . This result follows directly from the chain rule of Stratonovich calculus, see [41,42].

Example 3.12. Consider the noisy Duffing oscillator, i.e., for $\alpha, \beta, \varepsilon \in \mathbb{R}$ we have a Stratonovich stochastic differential equation with

$$b(x) = \begin{bmatrix} x_2 \\ -\alpha x_1 - \beta x_1^3 \end{bmatrix} \quad \text{and} \quad \sigma(x) = \varepsilon b(x).$$

To apply gEDMD, we convert it to an Itô stochastic differential equation using the drift correction formula to correct the noise-induced drift, which is defined componentwise as

$$c_i(x) = \sum_{j=1}^d \sum_{k=1}^s \frac{\partial \sigma_{ik}}{\partial x_j}(x) \sigma_{jk}(x), \quad i = 1, \dots, d,$$

see [43]. We obtain the Itô stochastic differential equation

$$dX_t = (b(X_t) + \frac{1}{2} c(X_t)) dt + \sigma(X_t) dW_t \quad \text{with} \\ c(x) = \varepsilon^2 \begin{bmatrix} b_2(x) \\ (-\alpha - 3\beta x_1^2) b_1(x) \end{bmatrix}.$$

Setting $\alpha = -1.1$, $\beta = 1.1$, $\varepsilon = 0.05$, choosing a dictionary that contains monomials, and applying gEDMD, the multiplicity of the eigenvalue $\lambda = 0$ is two and we obtain a conserved quantity of the form

$$E(x) \approx \frac{\alpha}{2} x_1^2 + \frac{\beta}{4} x_1^4 + \frac{1}{2} x_2^2 + c,$$

where $c \in \mathbb{R}$ is an arbitrary constant. $\triangle$

3.3. Relationships with other methods

We will now point out similarities and differences between the methods presented above and other well-known approaches for systems identification and generator approximation.

3.3.1. SINDy

SINDy [5] was designed to learn ordinary differential equations from simulation or measurement data. Just like gEDMD, it requires a set of states and the corresponding time derivatives. Defining $\dot{X} = [\dot{x}_1, \dot{x}_2, \dots, \dot{x}_m]$, SINDy minimizes the cost function $\|\dot{X} - M_s \Psi_X\|_F$, i.e., $M_s = \dot{X} \Psi_X^+$. Here, we omit the sparsification constraints, which can be added in the same way to gEDMD as described above. Recall that we assume that the full-state observable is given by $g(x) = B^\top \psi(x)$. SINDy can thus be seen as a special case of gEDMD since

$$\dot{x} = B^\top \dot{\psi}(x) \approx B^\top M \psi(x) = \underbrace{B^\top \dot{\psi}_X}_{\dot{X}} \underbrace{\psi_X^+}_{M_s} \psi(x) = \underbrace{\dot{X} \psi_X^+}_{M_s} \psi(x) = M_s \psi(x).$$

3.3.2. Koopman lifting technique

The Koopman lifting technique [22,23] uses the infinitesimal generator $\mathcal{L}$ for system identification. While tailored mainly to ordinary differential equations, extensions to stochastic differential equations with isotropic noise are also considered. First, the Koopman operator for a fixed lag time τ is estimated from trajectory data with the aid of standard EDMD. Then an approximation of the generator is obtained by taking the matrix logarithm, i.e.,

$$\widehat{L} = \frac{1}{\tau} \log \widehat{K}_\tau,$$

where $\widehat{K}_\tau$ is the matrix representation of the Koopman operator with respect to the chosen basis ψ (and lag time τ). The last step is to estimate the governing equations in the same way as illustrated in Example 3.3 for gEDMD. The Koopman lifting technique does not require the time-derivatives of the states or the partial derivatives of the basis functions, but only pairs of time-lagged data. However, the non-uniqueness of the matrix logarithm can cause problems and a sufficiently small sampling time τ is needed to ensure that the (possibly complex) eigenvalues lie in the strip $\{z \in \mathbb{C} : |\Im(z)| < \pi\}$, where $\Im$ denotes the imaginary part. Roughly speaking, only an infinite sampling rate allows us to capture the entire spectrum of frequencies [23]. Our approach generalizes to arbitrary systems of the form (1), but the estimation of the diffusion term can be carried over to the lifting technique as well. This could be a valuable alternative, e.g., when

only trajectory data is available. If the exact derivatives for the training data are known, then gEDMD is in general more accurate than the lifting approach. If, on the other hand, the derivatives for gEDMD have to be approximated from trajectory data, then the accuracy depends on the order of the finite-difference approximation and the step size, while the accuracy of the lifting approach depends on the lag time and the matrix logarithm implementation.

3.3.3. KRONIC

KRONIC [19,20], which stands for *Koopman reduced order non-linear identification and control*, is a data-driven method for discovering Koopman eigenfunctions, which are then used for control and the detection of conservation laws. The approach is based on SINDy and assumes that an eigenvalue is known a priori (or simultaneously learns the eigenvalue and corresponding eigenfunction). In our notation, the resulting problem can be written as

$$(\lambda_\ell \Psi_X^\top - \dot{\Psi}_X^\top) \xi_\ell = 0,$$

which, multiplying from the left by Ψ_X and assuming that $\Psi_X \Psi_X^\top$ is regular, becomes the gEDMD eigenvalue problem. This operator formulation is briefly mentioned in [19] as well. Thus, for deterministic systems, despite their different derivations, gEDMD and KRONIC are strongly related.

4. Further applications

In addition to identifying fast and slow modes, governing equations, or conservation laws, the Koopman generator has further applications that we will briefly demonstrate.

4.1. Coarse-grained dynamics and gEDMD

4.1.1. Galerkin approximation

In what follows, we describe how models of the Koopman generator can be used to identify reduced order models of a (possibly high-dimensional) stochastic dynamical system. To get started, we recapitulate the model reduction formalism introduced by [40,44]. Assume the stochastic process given by (1) possesses a unique invariant density μ , and let $\xi: \mathbb{R}^d \rightarrow \mathbb{R}^p$ be a coarse-graining function which maps $\mathbb{R}^d$ to a lower-dimensional space $\mathbb{R}^p$. The coarse-graining map induces a reduced probability measure with density ν on $\mathbb{R}^p$. Consider the space L_ν^2 of square-integrable functions of the reduced variables z . In fact, L_ν^2 is an infinite-dimensional subspace of L_μ^2 , if each function $f \in L_\nu^2$ is identified with the function $f \circ \xi \in L_\mu^2$. Let $\mathcal{P}$ be the orthogonal projection onto L_ν^2 . Define a coarse-grained generator as

$$\mathcal{L}^\xi = \mathcal{P} \mathcal{L} \mathcal{P}. \quad (6)$$

Given suitable assumptions on the original process (1), $\mathcal{L}^\xi$ is again the infinitesimal generator of a stochastic dynamics on $\mathbb{R}^p$, with invariant density ν and effective drift and diffusion coefficient b^ξ , a^ξ [40,44].

First, we show that for a basis set of functions defined only on $\mathbb{R}^p$, gEDMD converges to a Galerkin approximation of the coarse-grained generator $\mathcal{L}^\xi$:

Proposition 4.1. *Let $\mathbb{V} = \text{span}\{\psi_k\}_{k=1}^n$ be a subspace of L_ν^2 . Then gEDMD applied to the functions $\psi_k = \psi_k \circ \xi$ converges to the Galerkin projection of $\mathcal{L}^\xi$ onto $\mathbb{V}$. Here, (5) needs to be updated by*

$$d\tilde{\psi}_k(x) = b(x) \nabla_x \xi(x) \nabla_z \psi_k^\top(\xi(x)) + \frac{1}{2} (a(x) : H^\xi(x)) \nabla_z \psi_k^\top(\xi(x)) \\ + \frac{1}{2} \nabla_z^2 \psi_k(\xi(x)) : [\nabla \xi(x)^\top a(x) \nabla \xi(x)],$$

where $\nabla_x \xi \in \mathbb{R}^{d \times p}$ is the Jacobian of ξ , and $H^\xi \in \mathbb{R}^{d \times d \times p}$ is the tensor of Hessian matrices for each component of ξ . The Frobenius inner product between a and H^ξ is applied to the first two dimensions of H^ξ .

Proof. The expression for $d\tilde{\psi}_k(x)$ follows from the chain rule. It was already shown in [40] that

$$\begin{aligned} \langle \psi_i, \psi_j \rangle_\nu &= \langle \psi_i \circ \xi, \psi_j \circ \xi \rangle_\mu, \\ \langle \mathcal{L}^\xi \psi_i, \psi_j \rangle_\nu &= \langle \mathcal{L}(\psi_i \circ \xi), \psi_j \circ \xi \rangle_\mu. \end{aligned}$$

Thus, Proposition 3.5 implies that

$$\begin{aligned} \hat{A}_{ij} &\rightarrow \langle \mathcal{L}\tilde{\psi}_i, \tilde{\psi}_j \rangle_\mu = \langle \mathcal{L}^\xi \psi_i, \psi_j \rangle_\nu, \\ \hat{G}_{ij} &\rightarrow \langle \tilde{\psi}_i, \tilde{\psi}_j \rangle_\mu = \langle \psi_i, \psi_j \rangle_\nu. \quad \square \end{aligned}$$

In summary, data of the original process, sampling the distribution μ , can be used to learn a matrix representation of the coarse-grained generator (6). This matrix approximation can then be used to perform system identification, simulation, and control of the coarse-grained system the same way as described in Section 3.2.

4.1.2. Separate identification

For a reversible stochastic differential equation (1), we present an alternative approach to identify the parameters of the corresponding coarse-grained system. The method is related to spectral matching as introduced in [45]. The authors of [40] have shown that reversibility of the full process implies the dynamics generated by (6) are also reversible. Recalling that reversible dynamics are characterized by a scalar potential and the diffusion field, the basic idea is simply to estimate these two terms separately. The resulting framework, which we will call *separate identification*, consists of four steps, which are only partially dependent on each other:

Force matching. The scalar potential F^ξ of the coarse-grained dynamics can be estimated by an established technique called *force matching* [46,47]. It is based on the fact that the gradient of F^ξ solves the following minimization problem [48]:

$$\nabla_z F^\xi = \operatorname{argmin}_{g \in (L^2_\nu)^p} \int_{\mathbb{R}^d} \|g(\xi(x)) - f_{\text{lmf}}^\xi(x)\|^2 d\mu(x), \quad (7)$$

$$f_{\text{lmf}}^\xi = -\nabla_x F \cdot G^\xi + \nabla_x \cdot G^\xi, \quad (8)$$

$$G^\xi = \nabla_x \xi [(\nabla_x \xi)^T \nabla_x \xi]^{-1}, \quad (9)$$

where the minimization is over all square-integrable vector fields g of the reduced variables z , and the divergence is applied separately to each column of G^ξ in (8). The vector field f_{lmf}^ξ is called *local mean force*, while F is the scalar potential of the full process.

Application of gEDMD. The second step consists of applying gEDMD to estimate a finite-dimensional model of the coarse-grained generator $\mathcal{L}^\xi$ as described in Section 4.1.1, using a basis of functions $\{\psi_i\}_{i=1}^n$ defined on the reduced space $\mathbb{R}^p$. In particular, we obtain an estimate of the Galerkin matrix

$$\hat{A}_{ij} = \langle \mathcal{L}^\xi \psi_i, \psi_j \rangle_\nu.$$

Learning the diffusion field. As already discussed in Remark 3.7, matrix elements of the reduced generator are given by

$$\langle \mathcal{L}^\xi \psi_i, \psi_j \rangle_\nu = -\frac{1}{2} \int \nabla_z \psi_i a^\xi \nabla_z \psi_j d\nu. \quad (10)$$

It follows that the effective diffusion can be learned by matching it to the generator matrix $\hat{A}$ via (10). Let $a^\xi(\theta)$ be a parametric model for the effective diffusion. Then the optimal set of

parameters can be found by minimizing the Frobenius norm error

$$E(\theta) = \|\hat{A} - A(\theta)\|_F^2, \quad (11)$$

$$A(\theta)_{ij} = -\frac{1}{2} \int \nabla_z \psi_i a^\xi(\theta) \nabla_z \psi_j d\nu. \quad (12)$$

Determination of the drift. Using the relationship between drift and diffusion of a reversible system, inserting estimates for F^ξ and a^ξ into (13) completes the definition of the reduced model

$$b^\xi = -\frac{1}{2} a^\xi \nabla F^\xi + \frac{1}{2} \nabla \cdot a^\xi. \quad (13)$$

The above formulation seems advantageous compared to the direct system identification described in Section 3.2.2 for several reasons:

- Separate basis sets can be used to calculate the Galerkin matrix, the potential, and the diffusion. Specifically, constraints on each of these (such as positive definiteness of the diffusion) can be incorporated into each basis individually. Moreover, learning of the potential and the diffusion can also be accomplished using nonlinear models.
- The coordinate functions z_i and $z_i z_j$, as well as the products of the coordinate functions with the effective drift, are no longer required to be contained in the basis set.
- Both force matching and (11) are regression problems, allowing for the use of model validation techniques like cross-validation.
- The dynamics obtained by combining the learned potential and diffusion via (13) are automatically reversible.
- By diagonalizing the generator matrix corresponding to $A(\theta)$ above, the spectrum of the learned dynamics can be calculated directly and compared to the spectrum of the generator matrix corresponding to A , providing a further means of model validation.

On the other hand, the direct system identification is more general since the reconstruction via the local mean force may fail to yield good approximations of the effective drift in cases where some parts of the dynamics orthogonal to the low-dimensional manifold defined by the reaction coordinate are slow.

4.1.3. Example 1: Lemon-slice potential

We consider overdamped Langevin dynamics (see Remark 2.2) at inverse temperature $\beta = 1$ in the following two-dimensional potential V , expressed in polar coordinates:

$$V(r, \varphi) = \cos(k\varphi) + \sec(0.5\varphi) + 10(r-1)^2 + \frac{1}{r}.$$

For $k = 4$, a contour of the potential is shown in Fig. 4(a). Because of the two singular terms, the system's state space does not include the set $\{(x_1, x_2) : x_1 \leq 0, x_2 = 0\}$, enabling us to map the two-dimensional state space to polar coordinates unambiguously.

The polar angle φ is a suitable reaction coordinate, so we choose $\xi(x_1, x_2) = \varphi(x_1, x_2)$. Due to the simplicity of the system, all relevant quantities can be calculated analytically. Using the full-state partition function Z and two numerical constants C_1, C_2 , see [49], the invariant distribution, the effective drift and the effective diffusion along φ are given by

$$\nu(\varphi) = \frac{C_2}{Z} \exp(-[\cos(k\varphi) + \sec(0.5\varphi)]),$$

$$b^\varphi(\varphi) = \frac{C_1}{C_2} [k \sin(k\varphi) - 0.5 \tan(0.5\varphi) \sec(0.5\varphi)],$$

$$a^\varphi(\varphi) = \frac{2C_1}{C_2}.$$

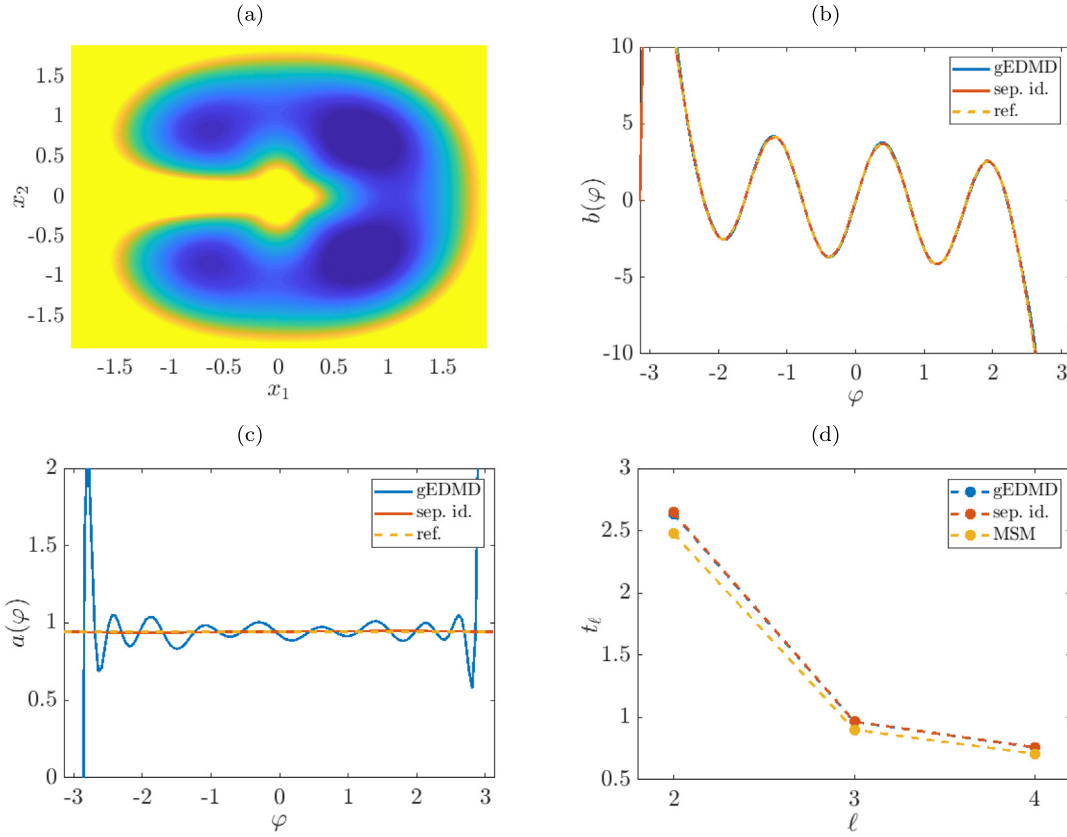


Fig. 4. Application of gEDMD with 21 Legendre polynomials to one-dimensional coarse-graining of the two-dimensional lemon-slice potential. (a) Visualization of the potential. (b) Estimates of the effective drift along the polar angle φ obtained directly from the generator matrix (blue), and by combining the solutions of (11) and (7) via (13) (red). The analytical reference is shown in yellow. (c) Estimates of the effective diffusion along the polar angle φ obtained directly from the generator matrix (blue), and by learning the diffusion via (11) using a Gaussian basis set (red), compared to the analytical reference in yellow. (d) Estimates of the three slowest implied time scales using a Markov state model (yellow), diagonalization of the generator matrix (blue), and diagonalization of the generator matrix corresponding to the optimal diffusion (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We apply coarse-grained gEDMD with a basis set of Legendre polynomials up to degree 20, scaled to fit the domain $[-\pi, \pi]$. In this example and the next, we use exact expressions for the drift and diffusion coefficient, as the parameters of the full system are usually known in the context of model reduction. From the generator matrix, we obtain estimates of the effective drift and diffusion as described in Section 3.2.2. Moreover, we also apply separate identification to learn the scalar potential and the diffusion. To this end, we use a basis set of periodic Gaussian functions centered at equidistant points between $\varphi = -2.8$ and $\varphi = 2.8$. The bandwidth of these Gaussians is determined by cross-validation, and is found to be 0.1 for force matching and 2.0 for the diffusion. We also enforce positivity of the diffusion by applying positivity constraints to the regression problem (11). We see in Fig. 4(b) and (c) that both methods provide accurate representations of the effective parameters. However, the diffusion estimated from (11) is virtually indistinguishable from the analytical solution, while the representation obtained from the polynomial basis is more oscillatory.

We also verify that gEDMD correctly captures the slow dynamics in this example. We diagonalize the generator matrix obtained from the polynomial basis, and compute the first three implied time scales by taking reciprocals of the first three non-trivial eigenvalues (leaving out the zero eigenvalue). We compare these time scales to those extracted from a Markov state model (MSM) [50] inferred directly from the data. We find in Fig. 4(d)

that the time scales are in very good agreement. As described above, we also use the generator matrix corresponding to the optimal $A(\theta)$ to estimate the first three implied time scales, and find them to match almost perfectly as well.

4.1.4. Example 2: Alanine dipeptide

As a more complex example, we derive a coarse-grained model from molecular dynamics simulations of alanine dipeptide, which has been used as a test case in numerous previous studies. The data set is the same as in reference [51] and comprises one million snapshots of Langevin dynamics saved every 1 ps. As is well-known, the positional component of Langevin dynamics behaves approximately like an overdamped process, see Remark 2.2, up to a re-scaling of time. Hence, we apply gEDMD assuming the original process is overdamped, and we extract this effective unit of time by comparing the first two implied time scales obtained from gEDMD and from a reference Markov state model.

The slowest dynamics of alanine dipeptide are captured by a single internal molecular coordinate, called ϕ -dihedral angle, which we choose to be the coarse-graining coordinate. Fig. 5(a) shows the empirical coarse-grained energy F^ϕ , and an approximation obtained by applying force matching. The basis set for force matching consists of 57 periodic Gaussians of bandwidth 1.2, centered at equidistant points between -2.8 and 2.8 . The slowest dynamical process corresponds to the transition across the highest barrier in this energy landscape.

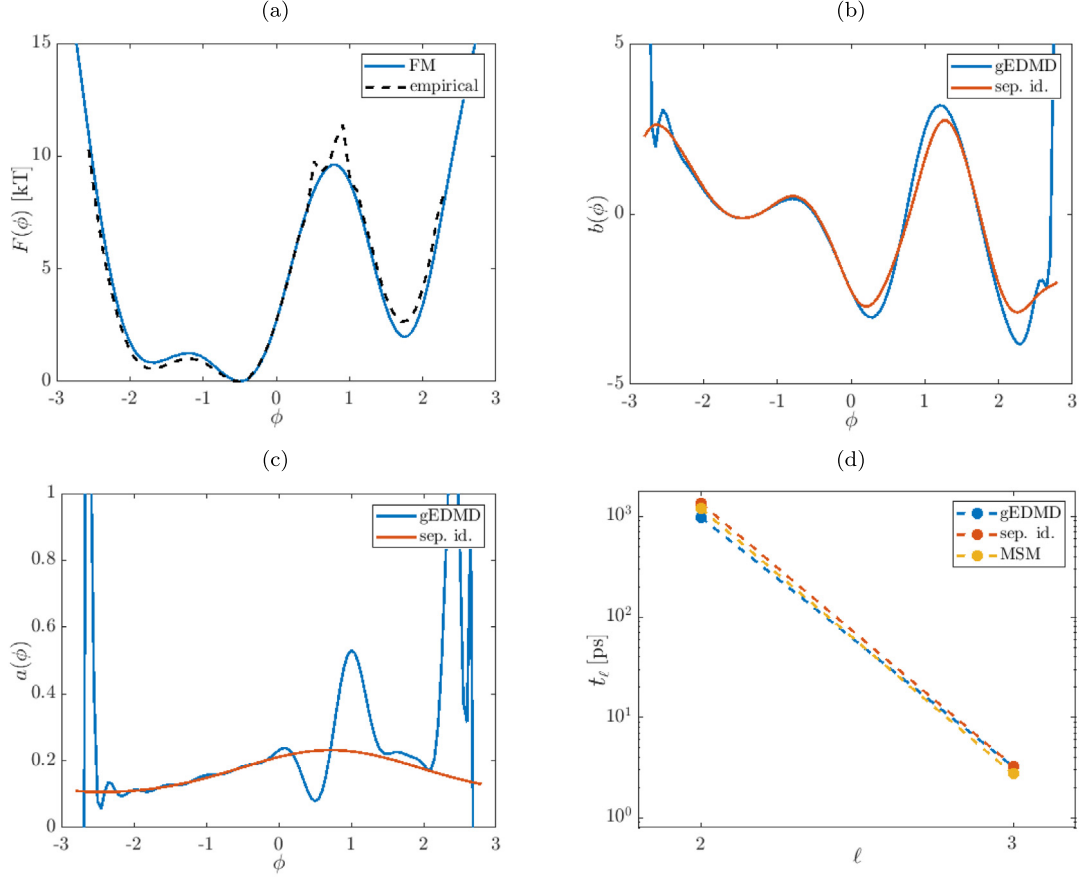


Fig. 5. Coarse-grained gEDMD along the ϕ -angle coordinate of alanine dipeptide, using a basis set of 26 Legendre polynomials (a) Effective energy F^ϕ , as estimated by histogramming the molecular dynamics simulation data (black), and by applying force matching with a Gaussian basis set (blue). (b) Estimates of the effective drift obtained directly from the generator matrix (blue), and by combining the solutions of (11) and (7) via (13) (red). (c) Estimates of the effective diffusion obtained directly from the generator matrix (blue), and by learning the diffusion via (11) using a Gaussian basis set (red). (d) Estimates of the two slowest implied time scales using a Markov state model (yellow), diagonalization of the generator matrix (blue), and diagonalization of the generator matrix corresponding to the optimal diffusion (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We apply gEDMD with the first 26 Legendre polynomials scaled to fit the domain $[-2.7, 2.7]$. From the generator matrix, we extract the effective drift and diffusion, which are depicted by blue lines in Fig. 5(b) and (c). As a comparison, we also compute an estimate of the effective diffusion by minimization of (11), including positivity constraints, with a set of 29 Gaussians of bandwidth 0.8, where the optimal bandwidth was determined by cross-validation. The resulting estimate of the diffusion is far less oscillatory than the direct estimate using the generator matrix, while the corresponding drift obtained from (13) is similar to the direct estimate.

Finally, we verify that gEDMD accurately reproduces the spectral properties of the original dynamics. As we can see in Fig. 5(d), after re-scaling the first two time scales provided by gEDMD by the effective time unit described above, they agree well with the results of an MSM analysis. The same is true for the time scales calculated based on the generator matrix corresponding to the optimal diffusion obtained by solving (11).

4.2. Control

The predictive capabilities of the Koopman operator have also raised interest in the control community, where the aim is to determine a system input u such that the non-autonomous control system $\dot{x} = b(x, u)$ behaves in a desired way, which results in the

following control problem:

$$\begin{aligned} \min_{u \in L^2([t_0, t_e], \mathbb{R})} J(x, u) &= \min_{u \in L^2([t_0, t_e], \mathbb{R})} \int_{t_0}^{t_e} \|x(t) - x^{\text{ref}}(t)\|_2^2 + \alpha \|u(t)\|_2^2 dt \\ \text{s.t.} \quad \dot{x}(t) &= b(x(t), u(t)), \\ x(t_0) &= x_0. \end{aligned} \quad (14)$$

In this formulation, the goal is to track a desired state over the control horizon $[t_0, t_e]$, and $\alpha \in \mathbb{R}^{>0}$ is a small number penalizing the control cost. In order to achieve a feedback behavior, problem (14) is embedded into a model predictive control (MPC) [52] scheme, where it has to be solved repeatedly over a relatively short horizon while the system (the plant) is running at the same time. The first part $[t_0, t_0 + h]$ of the optimal control u is then applied to the plant, and (14) has to be solved again on a shifted horizon $[t_0 + h, t_e + h]$.

Since the real-time requirements in MPC are often very hard to satisfy, a promising approach is to replace the system dynamics by a surrogate model, and one possibility is to use the Koopman operator or its generator for prediction. Introducing the variable $z = \psi(f(x))$, we obtain a linear system via the approximation $\mathbf{L}$ of the generator:

$$\dot{z}(t) \approx \mathbf{L}z(t).$$

However, as we see above, the Koopman operator is only defined for autonomous systems. Hence, a transformation has to be used

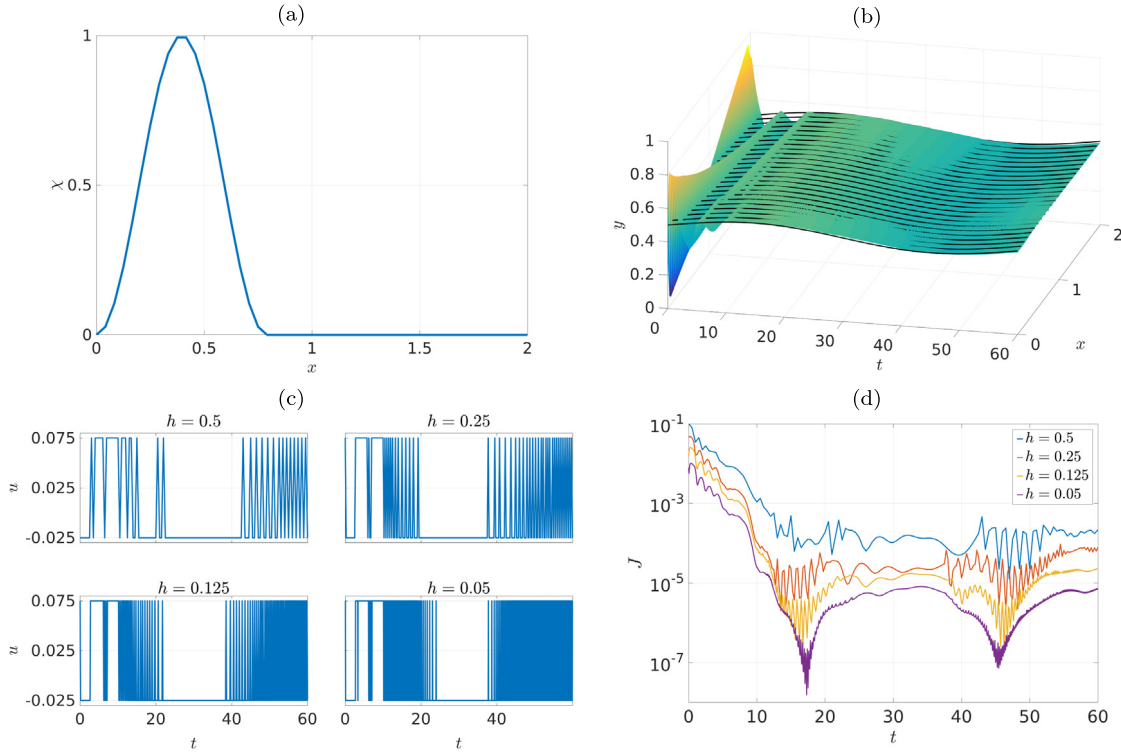


Fig. 6. Control of the Burgers equation using the Koopman generator and switching control. (a) The shape function used for the distributed control term. (b) The optimal state (colored) and the reference trajectories (black) for $h = 0.05$. (c) The optimal switching sequence as a function of the time step h . (d) The tracking error as a function of the time step h . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(the exception being control-affine systems, where only the autonomous part needs to be modeled). In [53], the control system was autonomized by introducing an augmented state $\hat{x} = (x, u)^\top$, and DMD was performed on the augmented system. The same approach was also used in combination with MPC in [54]. This state augmentation significantly increases the data requirements (all combinations of states and control inputs should be covered), such that an alternative transformation was proposed in [55,56] by restricting $u(t)$ to a finite set of inputs $\{u^1, \dots, u^{n_c}\}$. This way, the control system can be replaced by a finite set of autonomous systems $b_{u^i}(x) = b(x, u^i)$ for which the corresponding generators $\{L_{u^1}, \dots, L_{u^{n_c}}\}$ can be approximated. The control task is thus to determine the optimal right-hand side in each time step instead of computing a continuous input u :

$$\begin{aligned} \min_{u \in L^2([t_0, t_e], \{u^1, \dots, u^{n_c}\})} \int_{t_0}^{t_e} \|z(t) - z^{\text{ref}}(t)\|_2^2 + \alpha \|u(t)\|_2^2 dt \\ \text{s.t.} \quad \dot{z}(t) = \mathbf{L}_{u(t)} z(t), \\ z(t_0) = \psi(f(x_0)). \end{aligned} \quad (15)$$

Note that the quantization (i.e., the switching control) is encoded in the function space the control u lives in. For a more detailed description, the reader is referred to [55].

Regardless of the approach, a drawback of Koopman operator based surrogate models is that the control freedom is limited by the finite lag time. While larger lag times are often beneficial for the approximation of the dynamics, this is counterproductive for control, as the control frequency is strongly limited. This issue is overcome by the generator approach (15) since we can choose arbitrary time steps here, and results on mixed integer optimal control problems (see, e.g., [57]) suggest that fast switches allow for solutions of any desired accuracy. Moreover, the continuous-time generator model is much better suited for switching time optimization approaches. Therein, the combinatorial problem of

selecting the optimal right-hand side is replaced by a continuous optimization problem for the time instances τ_j at which the right-hand side is switched from the input u^i to u^{i+1} :

$$\begin{aligned} \min_{\tau \in \mathbb{R}^p} \int_{t_0}^{t_e} \|z(t) - z^{\text{ref}}(t)\|_2^2 + \alpha \|u(t)\|_2^2 dt \\ \text{s.t.} \quad \dot{z}(t) = \mathbf{L}_{u^i} z(t), \quad \text{for } t \in [\tau_{j-1}, \tau_j], \\ t_0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_p \leq t_e, \\ i = 1 + j \bmod n_c, \\ z(t_0) = \psi(f(x_0)). \end{aligned} \quad (16)$$

By fixing the number p of switches, this reformulation of (15) is now a continuous, finite-dimensional optimization problem for the switching times with a given switching sequence (cf. [58] for details), and both open and closed-loop control schemes (using MPC in the latter case) can be constructed.

Problem (16) was also used in combination with the Koopman operator in [55], but the discrete-time system prohibits arbitrary switching points which results in a reduced performance. Using the generator solves this problem and additionally, there even exist efficient second-order methods for this problem class [59].

In what follows, we present examples for the two extensions for MPC based on the Koopman generator. For the deterministic case, we use the 1D viscous Burgers equation with varying lag times in an MPC framework (Problem (15)) and for the non-deterministic case, we control the expected value of an Ornstein–Uhlenbeck process using both MPC (Problem (15)) and open loop switching time control (Problem (16)).

4.2.1. Partial differential equations

Consider the 1D Burgers equation with distributed control

$$\dot{y}(t, x) - \nu \Delta y(t, x) + y(t, x) \nabla y(t, x) = u(t) \chi(x).$$

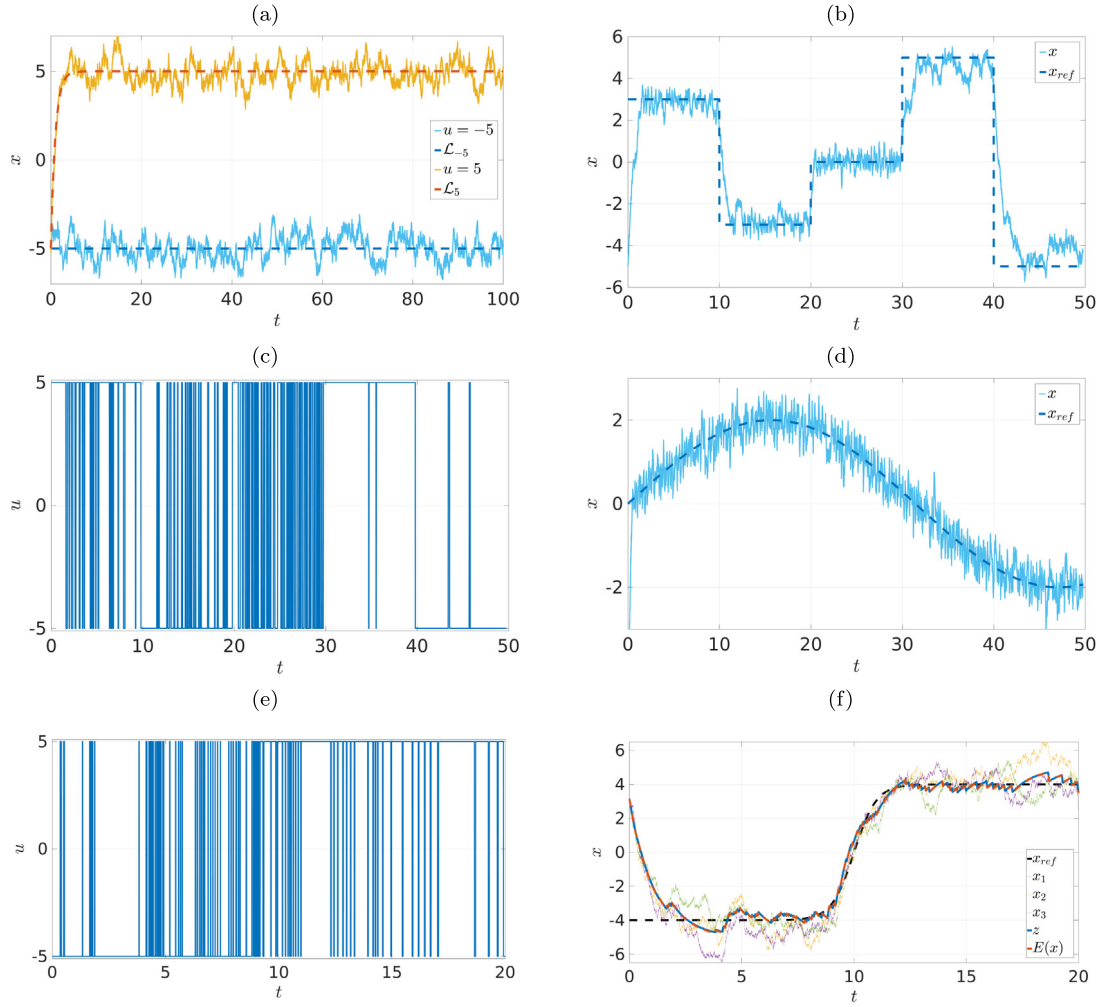


Fig. 7. Control of the expected value of the Ornstein–Uhlenbeck process. (a) Simulation and generator prediction for $u = 5$ and $u = -5$, respectively. (b) Tracking of a piecewise constant reference trajectory using MPC based on Problem (15). (c) The corresponding optimal input signal. (d) MPC based tracking of a continuous reference trajectory. (e) Solution of Problem (16) with $p = 200$ for the reference trajectory $x_{ref} = \tanh(t - 10)$, see (f). (f) The corresponding optimal trajectories of the generator model (z , blue line), of three realizations of the Ornstein–Uhlenbeck process (x_1 to x_3 , dotted lines), and of the expected value of the controlled process ($E(x)$, dashed orange line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Here, y denotes the state depending on space x and time t , and the system is controlled by a shape function χ (see Fig. 6(a)) that can be scaled by the input $u \in \{-0.025, 0.075\}$. The objective is to track a sinusoidal reference trajectory (shown in black in (b)), and we do this by solving problem (15) in an MPC framework. To this end, we approximate the Koopman generator using a relatively coarse “full state observable” (a grid of 25 equidistantly distributed points in space) and monomials up to order two. The data is collected from one trajectory with a piecewise constant input signal $u(t) \in \{-0.025, 0.075\}$. It is then divided into two data sets corresponding to the constant inputs 0.025 and -0.075 , respectively. The time derivative $\dot{y}$ is computed via finite differences.

We see in (c) and (d) that with decreasing time steps h (over which the input u is constant) that the control performance increases significantly. While the time step $h = 0.5$ corresponds to a solution that can be obtained by a Koopman operator approximation as well, the generator framework allows us to decrease the time steps and thereby the error by two orders of magnitude. Note that we can formally also decrease the lag time for the Koopman operator to increase the performance. In our experiments, the results were of comparable quality, and this is likely due to the high robustness of the MPC algorithm, which can cope well with small model inaccuracies. However,

an advantage of the generator approach is that we can choose the time step adaptively—in contrast to the Koopman operator approach, where a change in the lag time requires a different data set and new computations. This can be beneficial in terms of computational efficiency and is thus particularly important for long control horizons (see, e.g., [60]), due to which real-time capability may otherwise be jeopardized.

4.2.2. Stochastic differential equations

In the case of non-deterministic systems, the generator approach allows for a very elegant solution of *stochastic control problems*. In stochastic (or *robust*) control (see [61,62] for introductions), the goal is very often to steer the expected value to some desired value. In many situations, determining this expected value (e.g., via Monte Carlo methods) is numerically challenging. As the Koopman generator for stochastic systems describes the evolution of the expected value, see (2), problem (15) can be used to solve a control problem for the expected value using a deterministic linear system. To this end, we replace the computation of the initial value by an average over the recent past:

$$z_0 = \frac{1}{h} \int_{t_0-h}^{t_0} z(t) dt.$$

We again consider the Ornstein–Uhlenbeck process from [Example 2.3](#), with the only difference that we now add a control input:

$$dX_t = -\alpha(X_t - u)dt + \sqrt{2\beta^{-1}}dW_t,$$

with $\alpha = 1$ and $\beta = 2$. We compute two generator approximations corresponding to $u = -5$ and $u = 5$ using monomials up to order 12. [Fig. 7\(a\)](#) shows the trajectories of the two systems and the predictions using the corresponding generators, and we see that the expected values are accurately predicted. We set $h = 0.05$ as a discretization for the control u as well as the length of the input that is applied to the plant in each loop. The MPC controller based on [\(15\)](#) with the modified initial condition z_0 yields very good performance, as is shown for a tracking problem with a piecewise constant reference value in [Fig. 7\(b\)](#). The corresponding optimal control is shown in [Fig. 7\(c\)](#), and (d) shows that continuously varying inputs can be approximated equally well.

Finally, we use the switching time reformulation [\(16\)](#) in an open loop fashion in order to track a tanh profile over 20 s. The results are shown in [Fig. 7\(e\)](#) and (f), where the optimal input with $p = 200$ switches is shown in (e) and the corresponding dynamics are shown in (f). We observe a remarkable performance, as the optimal trajectory of the generator model and the expected value of the controlled Ornstein–Uhlenbeck process (computed as the mean over 1000 simulations) are almost indistinguishable.

To summarize, the generator approach yields highly efficient control schemes both for open and closed-loop control, as the linear system for the prediction of the expected value requires no further sampling of multiple noisy trajectories.

5. Conclusion

We presented an extension of standard EDMD to approximate the generator of the Koopman or Perron–Frobenius operator from data and highlighted several important applications pertaining to model reduction, system identification, and control. We illustrated that this approach can be used to obtain a decomposition into eigenvalues, eigenfunctions, and modes and, furthermore, that SINDy emerges as a special case. The proposed methods were implemented in Python, the gEDMD code and some of the above examples are available at <https://github.com/sklus/d3s/>.

Open questions include the convergence of gEDMD if not only the number of data points but also the number of basis functions tends to infinity. It is also unclear which part of the spectrum is approximated if the generator does not possess a pure point spectrum. Furthermore, is it possible to learn coarse-grained dynamics by only considering the dominant terms of the decomposition of the system's equations into eigenvalues, eigenfunctions, and modes (cf. [Example 3.3](#) and also [\[45\]](#))? Another interesting application of gEDMD would be to compute committor functions or hitting times. Extensions to non-autonomous systems will be considered in future work.

CRedit authorship contribution statement

Stefan Klus: Conceptualization, Methodology, Software, Investigation. **Felix Nüske:** Conceptualization, Methodology, Software, Investigation. **Sebastian Peitz:** Conceptualization, Methodology, Software, Investigation. **Jan-Hendrik Niemann:** Methodology, Software, Investigation. **Cecilia Clementi:** Conceptualization. **Christof Schütte:** Conceptualization.

Acknowledgments

S. K., J. N., and C. S. were funded by Deutsche Forschungsgemeinschaft (DFG), Germany through 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). F. N. was partially funded by the Rice University Academy of Fellows (USA). F. N. and C. C. were supported by the National Science Foundation, USA (CHE-1265929, CHE-1738990, CHE-1900374, PHY-1427654) and the Welch Foundation, USA (C-1570). C. C. also acknowledges funding from the Einstein Foundation Berlin (Germany). S. P. acknowledges support by the DFG Priority Programme 1962 (Germany).

References

- [1] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28, <http://dx.doi.org/10.1017/S0022112010001217>.
- [2] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, J.N. Kutz, On dynamic mode decomposition: Theory and applications, *J. Comput. Dyn.* 1 (2) (2014) <http://dx.doi.org/10.3934/jcd.2014.1.391>.
- [3] M.O. Williams, C.W. Rowley, I.G. Kevrekidis, A kernel-based method for data-driven Koopman spectral analysis, *J. Comput. Dyn.* 2 (2) (2015) 247–265, <http://dx.doi.org/10.3934/jcd.2015005>.
- [4] S. Klus, P. Koltai, C. Schütte, On the numerical approximation of the Perron–Frobenius and Koopman operator, *J. Comput. Dyn.* 3 (1) (2016) 51–79, <http://dx.doi.org/10.3934/jcd.2016003>.
- [5] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci.* 113 (15) (2016) 3932–3937, <http://dx.doi.org/10.1073/pnas.1517384113>.
- [6] C.R. Schwantes, V.S. Pande, Modeling molecular kinetics with tICA and the kernel trick, *J. Chem. Theory Comput.* 11 (2) (2015) 600–608, <http://dx.doi.org/10.1021/ct5007357>.
- [7] S. Klus, I. Schuster, K. Muandet, Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces, *J. Nonlinear Sci.* (2019) <http://dx.doi.org/10.1007/s00332-019-09574-z>.
- [8] S. Klus, P. Gelß, S. Peitz, C. Schütte, Tensor-based dynamic mode decomposition, *Nonlinearity* 31 (7) (2018) <http://dx.doi.org/10.1088/1361-6544/aabc8f>.
- [9] P. Gelß, S. Klus, J. Eisert, C. Schütte, Multidimensional approximation of nonlinear dynamical systems, *J. Comput. Nonlinear Dyn.* 14 (2019) 061006, <http://dx.doi.org/10.1115/1.4043148>.
- [10] C. Chen, A. Surana, A. Bloch, I. Rajapakse, Multilinear time invariant systems theory, 2019, arXiv e-prints.
- [11] Q. Li, F. Dietrich, E.M. Bollt, I.G. Kevrekidis, Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator, *Chaos* 27 (10) (2017) 103111, <http://dx.doi.org/10.1063/1.4993854>.
- [12] B. Lusch, J.N. Kutz, S. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, *Nature Commun.* 9 (2017) <http://dx.doi.org/10.1038/s41467-018-07210-0>.
- [13] A. Mardt, L. Pasquali, H. Wu, F. Noé, VAMPnets for deep learning of molecular kinetics, *Nature Commun.* 9 (2018) <http://dx.doi.org/10.1038/s41467-017-02388-1>.
- [14] J.N. Kutz, S.L. Brunton, B.W. Brunton, J.L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, SIAM, Philadelphia, 2016.
- [15] F. Nüske, B.G. Keller, G. Pérez-Hernández, A.S.J.S. Mey, F. Noé, Variational approach to molecular kinetics, *J. Chem. Theory Comput.* 10 (4) (2014) 1739–1752.
- [16] B. Koopman, Hamiltonian systems and transformations in Hilbert space, *Proc. Natl. Acad. Sci.* 17 (5) (1931) 315, <http://dx.doi.org/10.1073/pnas.17.5.315>.
- [17] A. Lasota, M.C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, second ed., in: *Applied Mathematical Sciences*, vol. 97, Springer, New York, 1994.
- [18] M. Budišić, R. Mohr, I. Mezić, Applied Koopmanism, *Chaos* 22 (4) (2012) <http://dx.doi.org/10.1063/1.4772195>.
- [19] E. Kaiser, J.N. Kutz, S.L. Brunton, Data-driven discovery of Koopman eigenfunctions for control, 2017, ArXiv e-prints.
- [20] E. Kaiser, J.N. Kutz, S.L. Brunton, Discovering conservation laws from data for control, 2018, arXiv e-prints.
- [21] L. Boninsegna, F. Nüske, C. Clementi, Sparse learning of stochastic dynamical equations, *J. Chem. Phys.* 148 (24) (2018) 241723, <http://dx.doi.org/10.1063/1.5018409>.

- [22] A. Mauroy, J. Goncalves, Linear identification of nonlinear systems: A lifting technique based on the Koopman operator, in: 2016 IEEE 55th Conference on Decision and Control, CDC, 2016, pp. 6500–6505, <http://dx.doi.org/10.1109/CDC.2016.7799269>.
- [23] A. Mauroy, J. Goncalves, Koopman-based lifting techniques for nonlinear systems identification, 2017, ArXiv e-prints.
- [24] A.N. Riseth, J.P. Taylor-King, Operator fitting for parameter estimation of stochastic differential equations, 2017, ArXiv e-prints.
- [25] D. Giannakis, Data-driven spectral decomposition and forecasting of ergodic dynamical systems, *Appl. Comput. Harmon. Anal.* 47 (2) (2019) 338–396, <http://dx.doi.org/10.1016/j.acha.2017.09.001>.
- [26] G. Froyland, O. Junge, P. Koltai, Estimating long term behavior of flows without trajectory integration: The infinitesimal generator approach, *SIAM J. Numer. Anal.* 51 (1) (2013) 223–247, <http://dx.doi.org/10.1137/110819986>.
- [27] S.L. Brunton, B.W. Brunton, J.L. Proctor, J.N. Kutz, Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control, *PLoS One* 11 (2) (2016) <http://dx.doi.org/10.1371/journal.pone.0150171>.
- [28] B.J. Hollingsworth, *Stochastic Differential Equations: A Dynamical Systems Approach* (Ph.D. thesis), Auburn University, 2008.
- [29] N. Črnjarić-Žić, S. Maćešić, I. Mezić, Koopman operator spectrum for random dynamical systems, 2019, arXiv e-prints.
- [30] P. Metzner, *Transition Path Theory for Markov Processes: Application to Molecular Dynamics* (Ph.D. thesis), Freie Universität Berlin, 2007.
- [31] J.R. Baxter, J.S. Rosenthal, Rates of convergence for everywhere-positive markov chains, *Statist. Probab. Lett.* 22 (4) (1995) 333–338, [http://dx.doi.org/10.1016/0167-7152\(94\)00085-M](http://dx.doi.org/10.1016/0167-7152(94)00085-M).
- [32] G.A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*, in: *Texts in Applied Mathematics*, vol. 60, Springer, New York, 2014.
- [33] D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Vol. 348, Springer Science & Business Media, 2013.
- [34] V. Aboites, Hermite polynomials through linear algebra, *Int. J. Pure Appl. Math.* 114 (05) (2017) 401–406, <http://dx.doi.org/10.12732/ijpam.v114i2.19>.
- [35] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, F. Noé, Data-driven model reduction and transfer operator approximation, *J. Nonlinear Sci.* 28 (2018) 985–1010, <http://dx.doi.org/10.1007/s00332-017-9437-7>.
- [36] C. Schütte, M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, No. 24, in: *Courant Lecture Notes*, American Mathematical Society, 2013.
- [37] C. Schütte, M. Sarich, A critical appraisal of Markov state models, *Eur. Phys. J. Spec. Top.* 224 (12) (2015) 2445–2462, <http://dx.doi.org/10.1140/epjst/e2015-02421-0>.
- [38] M.O. Williams, I.G. Kevrekidis, C.W. Rowley, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition, *J. Nonlinear Sci.* 25 (6) (2015) 1307–1346, <http://dx.doi.org/10.1007/s00332-015-9258-5>.
- [39] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the Koopman operator, *J. Nonlinear Sci.* 28 (2) (2018) 687–710, <http://dx.doi.org/10.1007/s00332-017-9423-0>.
- [40] W. Zhang, C. Hartmann, C. Schütte, Effective dynamics along given reaction coordinates, and reaction rate theory, *Faraday Discuss.* 195 (2016) 365–394, <http://dx.doi.org/10.1039/C6FD00147E>.
- [41] E. Faou, T. Lelièvre, Conservative stochastic differential equations: Mathematical and numerical analysis, *Math. Comp.* 78 (268) (2009) 2047–2074, <http://dx.doi.org/10.1090/S0025-5718-09-02220-0>.
- [42] W. Zhou, L. Zhang, J. Hong, S. Song, Projection methods for stochastic differential equations with conserved quantities, *BIT Numer. Math.* 56 (4) (2016) 1497–1518, <http://dx.doi.org/10.1007/s10543-016-0614-0>.
- [43] R.L. Stratonovich, A new representation for stochastic integrals and equations, *SIAM J. Comput.* 4 (2) (1966) 362–371, <http://dx.doi.org/10.1137/0304028>.
- [44] F. Legoll, T. Lelièvre, Effective dynamics using conditional expectations, *Nonlinearity* 23 (9) (2010) 2131, <http://dx.doi.org/10.1088/0951-7715/23/9/006>.
- [45] F. Nüske, L. Boninsegna, C. Clementi, Coarse-graining molecular systems by spectral matching, *J. Chem. Phys.* 151 (4) (2019) 044116.
- [46] S. Izvekov, G.A. Voth, A multiscale coarse-graining method for biomolecular systems, *J. Phys. Chem. B* 109 (7) (2005) 2469–2473.
- [47] W.G. Noid, J.-W. Chu, G.S. Ayton, V. Krishna, S. Izvekov, G.A. Voth, A. Das, H.C. Andersen, The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models, *J. Chem. Phys.* 128 (24) (2008) 244114, <http://dx.doi.org/10.1063/1.2938860>.
- [48] G. Ciccotti, T. Lelièvre, E. Vanden-Eijnden, Projection of diffusions on submanifolds: Application to mean force computation, *Comm. Pure Appl. Math.* 61 (3) (2008) 371–408, <http://dx.doi.org/10.1002/cpa.20210>.
- [49] F. Nüske, P. Koltai, L. Boninsegna, C. Clementi, Spectral properties of effective dynamics from conditional expectations, 2019, [arXiv:1901.01557](https://arxiv.org/abs/1901.01557).
- [50] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Generation and validation, *J. Chem. Phys.* 134 (17) (2011) 174105, <http://dx.doi.org/10.1063/1.3565032>.
- [51] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N.E. Charron, G. De Fabritiis, F. Noé, C. Clementi, Machine learning of coarse-grained molecular dynamics force fields, *ACS Cent. Sci.* (2019) <http://dx.doi.org/10.1021/acscentsci.8b00913>.
- [52] L. Grüne, J. Pannek, *Nonlinear Model Predictive Control*, second ed., Springer International Publishing, 2017.
- [53] J.L. Proctor, S.L. Brunton, J.N. Kutz, Dynamic mode decomposition with control, *SIAM J. Appl. Dyn. Syst.* 15 (1) (2015) 142–161, <http://dx.doi.org/10.1137/15M1013857>.
- [54] M. Korda, I. Mezić, Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control, *Automatica* 93 (2018) 149–160, <http://dx.doi.org/10.1016/j.automatica.2018.03.046>.
- [55] S. Peitz, S. Klus, Koopman operator-based model reduction for switched-system control of PDEs, *Automatica* 106 (2019) 184–191, <http://dx.doi.org/10.1016/j.automatica.2019.05.016>.
- [56] S. Peitz, S. Klus, Feedback control of nonlinear PDEs using data-efficient reduced order models based on the Koopman operator, in: A. Mauroy, Y. Suzuki, I. Mezić (Eds.), *The Koopman Operator in Systems and Control: Theory, Numerics, and Applications*, Springer, 2020.
- [57] S. Sager, H.G. Bock, M. Diehl, The integer approximation error in mixed-integer optimal control, *Math. Program.* 133 (1–2) (2012) 1–23, <http://dx.doi.org/10.1007/s10107-010-0405-3>.
- [58] M. Egerstedt, Y. Wardi, F. Delmotte, Optimal control of switching times in switched dynamical systems, in: 42nd IEEE International Conference on Decision and Control, CDC, 2003, pp. 2138–2143.
- [59] B. Stellato, S. Ober-Blöbaum, P.J. Goulart, Second-order switching time optimization for switched dynamical systems, *IEEE Trans. Automat. Control* 62 (10) (2017) 5407–5414.
- [60] C. Kirches, L. Wirsching, H.G. Bock, J.P. Schlöder, Efficient direct multiple shooting for nonlinear model predictive control on long horizons, *J. Process Control* 22 (3) (2012) 540–550.
- [61] A. Bemporad, M. Morari, Robust model predictive control: A survey, in: A. Garulli, A. Tesi (Eds.), *Robustness in Identification and Control*, Springer, London, 2007, pp. 207–226.
- [62] A. Mesbah, Stochastic model predictive control: An overview and perspectives for future research, *IEEE Control Syst.* 36 (6) (2016) 30–44, <http://dx.doi.org/10.1109/MCS.2016.2602087>.