Journal of

Nonlinear Science

CrossMark

Data-Driven Model Reduction and Transfer Operator Approximation

Stefan Klus[1] · Feliks Nüske[1] · Péter Koltai[1] · Hao Wu[1] ·
Ioannis Kevrekidis[2,3] · Christof Schütte[1,3] · Frank Noé[1]

Abstract In this review paper, we will present different data-driven dimension reduction techniques for dynamical systems that are based on transfer operator theory as well as methods to approximate transfer operators and their eigenvalues, eigenfunctions, and eigenmodes. The goal is to point out similarities and differences between methods developed independently by the dynamical systems, fluid dynamics, and molecular dynamics communities such as *time-lagged independent component analysis*, *dynamic mode decomposition*, and their respective generalizations. As a result, extensions and best practices developed for one particular method can be carried over to other related methods.

Keywords Koopman operator · Perron-Frobenius operator · Model reduction · Data-driven methods

Mathematics Subject Classification 37M10 · 37M25 · 37L65 · 34L16

1 Introduction

The numerical solution of complex systems of differential equations plays an important role in many areas such as molecular dynamics, fluid dynamics, mechanical as

✉ Stefan Klus
   stefan.klus@fu-berlin.de

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

[2] Department of Chemical and Biological Engineering and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA

[3] Zuse Institute Berlin, Berlin, Germany

well as electrical engineering, and physics. These systems often exhibit multi-scale behavior which can be due to the coupling of subsystems with different time scales—for instance, fast electrical and slow mechanical components—or due to the intrinsic properties of the system itself—for instance, the fast vibrations and slow conformational changes of molecules. Analyzing such problems using transfer operator-based methods is often infeasible or prohibitively expensive from a computational point of view due to the so-called *curse of dimensionality*. One possibility to avoid this is to project the dynamics of the high-dimensional system onto a lower-dimensional space and to then analyze the reduced system representing, for instance, only the relevant slow dynamics, see, for example, Pérez-Hernández et al. 2013; Froyland et al. 2014.

In this paper, we will introduce different methods such as *time-lagged independent component analysis* (TICA) (Molgedey and Schuster 1994; Pérez-Hernández et al. 2013; Schwantes and Pande 2013) and *dynamic mode decomposition* (DMD) (Schmid 2010; Chen et al. 2012; Tu et al. 2014; Kutz et al. 2016) to identify the dominant dynamics using only simulation data or experimental data. It was shown that these methods are related to Koopman operator approximation techniques (Koopman 1931; Mezić 2005; Rowley et al. 2009; Budišić et al. 2012). Extensions of the aforementioned methods called the *variational approach of conformation dynamics* (VAC) (Noé and Nüske 2013; Nüske et al. 2014, 2016) developed mainly for reversible molecular dynamics problems and *extended dynamic mode decomposition* (EDMD) (Williams et al. 2015a, b; Klus et al. 2016) can be used to compute eigenfunctions, eigenvalues, and eigenmodes of the Koopman operator (and its adjoint, the Perron–Frobenius operator). Interestingly, although the underlying ideas, derivations, and intended applications of these methods differ, the resulting algorithms share a lot of similarities. The goal of this paper is to show the equivalence of different data-driven methods which have been widely used by the dynamical systems, fluid dynamics, and molecular dynamics communities, but under different names. Hence, extensions, generalizations, and algorithms developed for one method can be carried over to its counterparts, resulting in a unified theory and set of tools. An alternative approach to data-driven model reduction—also related to transfer operators and their generators—would be to use *diffusion maps* (Nadler et al. 2006; Coifman et al. 2008; Ferguson et al. 2011; Giannakis 2015). Manifold learning methods, however, are beyond the scope of this paper.

The outline of the paper is as follows: Sect. 2 briefly introduces transfer operators and the concept of reversibility. In Sect. 3, different data-driven methods for the approximation of the eigenvalues, eigenfunctions, and eigenmodes of transfer operators will be described. The theoretical background and the derivation of these methods will be outlined in Sect. 4. Section 5 addresses open problems and lists possible future work.

2 Transfer Operators and Reversibility

In the literature, the term *transfer operator* is sometimes used in different contexts. In this section, we will briefly introduce the Perron–Frobenius operator, the Perron–Frobenius operator with respect to the equilibrium density, and the Koopman operator. All these three operators are, according to our definition, transfer operators.

2.1 Guiding Example

Our paper will deal with data-driven methods to analyze both stochastic and deterministic dynamical systems. To illustrate the concepts of transfer operators and their spectral components, we first introduce a simple stochastic dynamical system that will be revisited throughout the paper.

Example 2.1 Consider the following one-dimensional Ornstein–Uhlenbeck process, given by an Itô stochastic differential equation[1] of the form:

$$dX_t = -\alpha D X_t \, dt + \sqrt{2D} \, dW_t.$$

Here, $\{W_t\}_{t \geq 0}$ is a one-dimensional standard Wiener process (Brownian motion), the parameter α is the friction coefficient, and $D = \beta^{-1}$ is the diffusion coefficient. The stochastic forcing usually models physical effects, most often thermal fluctuations and it is customary to call β the inverse temperature.

The transition density of the Ornstein–Uhlenbeck process, i.e., the conditional probability density to find the process near y a time τ after it had been at x, is given by

$$p_\tau(x, y) = \frac{1}{\sqrt{2\pi\,\sigma^2(\tau)}} \exp\left(-\frac{\left(y - x\,e^{-\alpha D\tau}\right)^2}{2\,\sigma^2(\tau)}\right), \tag{1}$$

where $\sigma^2(\tau) = \alpha^{-1}\left(1 - e^{-2\alpha D\tau}\right)$. Figure 1a shows the transition densities for different values of τ. More details can be found in Pavliotis (2014). For complex dynamical systems, the transition density is not known explicitly, but must be estimated from simulation or measurement data.

In this work we will describe the dynamics of a system in terms of dynamical operators such as the propagator $\mathcal{P}_\tau$, which is defined by the transition density $p_\tau(x, y)$ and propagates a probability density of Brownian walkers in time by

$$p_{t+\tau}(x) = \mathcal{P}_\tau\, p_t(x).$$

See Fig. 1b for the time evolution of the Ornstein–Uhlenbeck process initiated from a localized starting condition. It can be seen that the distribution spreads out and converges toward a Gaussian distribution, which is then invariant in time. For this simple dynamical system we can give the equation for the invariant density explicitly:

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

which is a Gaussian whose variance is decreasing with increasing friction and decreasing temperature. □

---

[1] A general time-homogeneous Itô stochastic differential equation is given by $dX_t = -\alpha(X_t)\,X_t\,dt + \sigma(X_t)\,dW_t$, where $\alpha : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are coefficient functions, and $\{W_t\}_{t \geq 0}$ is a d-dimensional standard Wiener process.
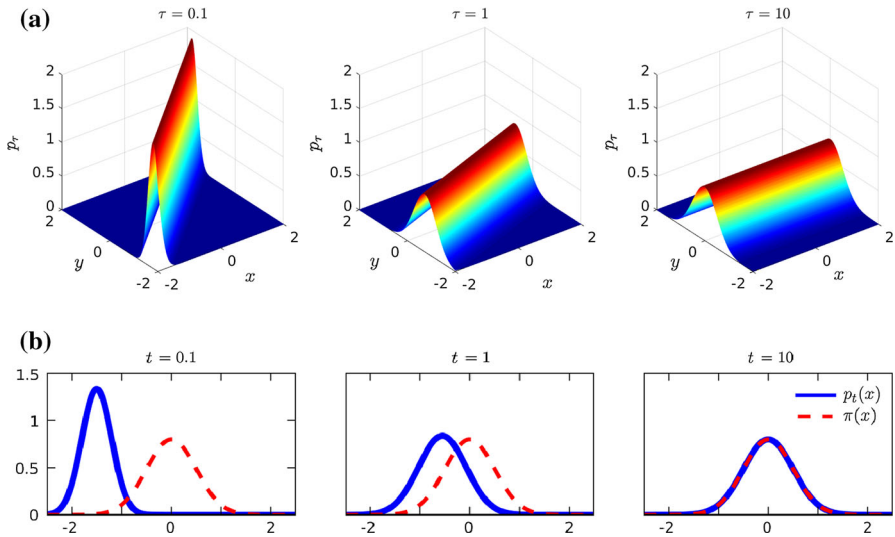
Fig. 1 **a** Transition density function of the Ornstein–Uhlenbeck process for different values of τ. If τ is small, values starting in x will stay close to it. For larger values of τ, the influence of the starting point x is negligible. Densities converge to the equilibrium density, denoted by π. Here, $\alpha = 4$ and $D = 0.25$. **b** Evolution of the probability to find the dynamical system at any point x over time t, after starting with a peaked distribution at $t = 0$. We show the resulting distributions at times $t = 0.1$, and $t = 1$, and $t = 10$. The system relaxes toward the stationary density $\pi(x)$

2.2 Transfer Operators

Let $\{\mathbf{X}_t\}_{t \geq 0}$ be a time-homogeneous[2] stochastic process defined on the bounded state space $\mathbb{X} \subset \mathbb{R}^d$. It can be genuinely stochastic or it might as well be deterministic, such that there is a flow map $\Phi_\tau : \mathbb{X} \to \mathbb{X}$ with $\Phi_\tau(\mathbf{X}_t) = \mathbf{X}_{t+\tau}$ for $\tau \geq 0$. Let the measure[3] $\mathbb{P}$ denote the law of the process $\{\mathbf{X}_t\}_{t \geq 0}$ that we will study in terms of its statistical transition properties. To this end, under some mild regularity assumptions[4] which are satisfied by Itô diffusions with smooth coefficients (Hopf 1954; Krengel 1985), we can give the following definition.

Definition 2.2 The transition density function $p_\tau : \mathbb{X} \times \mathbb{X} \to [0, \infty]$ of a process $\{\mathbf{X}_t\}_{t \geq 0}$ is defined by

$$\mathbb{P}[\mathbf{X}_{t+\tau} \in \mathbb{A} \mid \mathbf{X}_t = x] = \int_{\mathbb{A}} p_\tau(x, y) \, dy,$$

---

[2] We call a stochastic process $\{\mathbf{X}_t\}_{t \geq 0}$ *time-homogeneous*, or *autonomous*, if it holds for every $t \geq s \geq 0$ that the distribution of $\mathbf{X}_t$ conditional to $\mathbf{X}_s = x$ only depends on x and $(t - s)$. It is the stochastic analogue of the flow of an autonomous (time-independent) ordinary differential equation.

[3] For a measure-theoretic discussion of this construction, please refer to Klus et al. (2016). For our purposes, it is sufficient to equip $\mathbb{X}$ with the standard Lebesgue measure. In particular, if not stated otherwise, measurability of a set $\mathbb{A} \subset \mathbb{X}$ is meant with respect to the Borel σ-algebra.

[4] These conditions are called interchangeably *absolute continuity*, *μ-compatibility*, or *null preservingness*.

for every measurable set $\mathbb{A}$. Here and in what follows, $\mathbb{P}[\cdot \mid \mathfrak{E}]$ denotes probabilities conditioned on the event $\mathfrak{E}$. That is, $p_\tau(x, y)$ is the conditional probability density of $\mathbf{X}_{t+\tau} = y$ given that $\mathbf{X}_t = x$.

For $1 \le r \le \infty$, the spaces $L^r(\mathbb{X})$ denote the usual spaces of r-Lebesgue integrable functions, which is a Banach space with the corresponding norm $\| \cdot \|_{L^r}$.

Definition 2.3 Let $p_t \in L^1(\mathbb{X})$ be the probability density and $f_t \in L^\infty(\mathbb{X})$ an observable of the system. For a given lag time τ:

(a) The *Perron–Frobenius operator* or *propagator* $\mathcal{P}_\tau : L^1(\mathbb{X}) \to L^1(\mathbb{X})$ is defined by

$$\mathcal{P}_\tau p_t(x) = \int_{\mathbb{X}} p_\tau(y, x)\, p_t(y)\, \mathrm{d}y.$$

(b) The *Koopman operator* $\mathcal{K}_\tau : L^\infty(\mathbb{X}) \to L^\infty(\mathbb{X})$ is defined by

$$\mathcal{K}_\tau f_t(x) = \int_{\mathbb{X}} p_\tau(x, y)\, f_t(y)\, \mathrm{d}y = \mathbb{E}[f_t(\mathbf{X}_{t+\tau}) \mid \mathbf{X}_t = x].$$

Both $\mathcal{P}_\tau$ and $\mathcal{K}_\tau$ are linear but infinite-dimensional operators which are adjoint to each other with respect to the standard duality pairing $\langle \cdot, \cdot \rangle$, defined by $\langle f, g \rangle = \int_{\mathbb{X}} f(x)\, g(x)\, \mathrm{d}x$. The homogeneity of the stochastic process $\{\mathbf{X}_t\}_{t \ge 0}$ implies the so-called *semigroup property* of the operators, i.e., $\mathcal{P}_{\tau+\sigma} = \mathcal{P}_\tau \mathcal{P}_\sigma$ and $\mathcal{K}_{\tau+\sigma} = \mathcal{K}_\tau \mathcal{K}_\sigma$ for $\tau, \sigma \ge 0$. In other words, these operators describe time-stationary Markovian dynamics. While the Perron–Frobenius operator describes the evolution of densities, the Koopman operator describes the evolution of observables. For the analysis of the long-term behavior of dynamical systems, densities that remain unchanged by the dynamics play an important role (one can think of the concept of *ergodicity*).

Definition 2.4 A density π is called an *invariant density* or *equilibrium density* if $\mathcal{P}_\tau \pi = \pi$. That is, the equilibrium density π is an eigenfunction of the Perron–Frobenius operator $\mathcal{P}_\tau$ with corresponding eigenvalue 1.

In what follows, $L_\pi^r(\mathbb{X})$ denotes the weighted L^r-space of functions f such that $\|f\|_{L_\pi^r} := \int_{\mathbb{X}} |f(x)|^r \pi(x)\, \mathrm{d}x < \infty$. While one can consider the evolution of densities with respect to any density, we are particularly interested in the evolution with respect to the equilibrium density. From this point on, we assume there is a unique invariant density. This assumption is typically satisfied for molecular dynamics applications, where the invariant density is given by the Boltzmann distribution.

Definition 2.5 Let $L_\pi^1(\mathbb{X}) \ni u_t(x) = \pi(x)^{-1} p_t(x)$ be a probability density with respect to the equilibrium density π. Then the *Perron–Frobenius operator (propagator) with respect to the equilibrium density*, denoted by $\mathcal{T}_\tau$, is defined by

$$\mathcal{T}_\tau u_t(x) = \int_{\mathbb{X}} \frac{\pi(y)}{\pi(x)}\, p_\tau(y, x)\, u_t(y)\, \mathrm{d}y.$$

The operators $\mathcal{P}_\tau$ and $\mathcal{K}_\tau$ can be defined on other spaces L^r and $L^{r'}$, with $r \neq 1$ and $r' \neq \infty$, see (Baxter and Rosenthal 1995; Klus et al. 2016) for more details. By defining the weighted duality pairing $\langle f, g \rangle_\pi = \int_{\mathbb{X}} f(x) g(x) \pi(x) \, dx$ for $f \in L^r_\pi(\mathbb{X})$ and $g \in L^{r'}_\pi(\mathbb{X})$, where $\frac{1}{r} + \frac{1}{r'} = 1$, $\mathcal{T}_\tau$ defined on $L^{r'}_\pi(\mathbb{X})$ is the adjoint of $\mathcal{K}_\tau$ defined on $L^r_\pi(\mathbb{X})$ with respect to $\langle \cdot, \, \cdot \rangle_\pi$:

$$\langle \mathcal{K}_\tau f, \, g \rangle_\pi = \langle f, \, \mathcal{T}_\tau g \rangle_\pi.$$

For more details, see (Lasota and Mackey 1994; Schütte et al. 1999; Froyland et al. 2013; Noé and Nüske 2013; Nüske et al. 2014; Klus et al. 2016; Wu et al. 2017). The two operators $\mathcal{P}_\tau$ and $\mathcal{T}_\tau$ are often referred to as *forward operators*, whereas $\mathcal{K}_\tau$ is also called *backward operator*, as they are the solution operators of the forward (Fokker–Planck) and backward Kolmogorov equations (Lasota and Mackey 1994, Section 11), respectively.

2.3 Spectral Decomposition of Transfer Operators

In what follows, let $\mathcal{A}_\tau$ denote one of the transfer operators defined above, i.e., $\mathcal{P}_\tau$, $\mathcal{T}_\tau$, or $\mathcal{K}_\tau$. We are particularly interested in computing eigenvalues $\lambda_\ell(\tau) \in \mathbb{C}$ and eigenfunctions $\varphi_\ell : \mathbb{X} \to \mathbb{C}$ of transfer operators, i.e.,

$$\mathcal{A}_\tau \varphi_\ell = \lambda_\ell(\tau) \, \varphi_\ell.$$

Note that the eigenvalues depend on the lag time τ. For the sake of simplicity, we will often omit this dependency. The eigenvalues and eigenfunctions of transfer operators contain important information about the global properties of the system such as metastable sets or fast and slow processes and can also be used as reduced coordinates, see (Dellnitz and Junge 1999; Schütte et al. 1999; Mezić 2005; Schütte and Sarich 2013; Froyland et al. 2013, 2014) and references therein.

Example 2.6 The eigenvalues λ_ℓ and eigenfunctions φ_ℓ of $\mathcal{K}_\tau$ associated with the Ornstein–Uhlenbeck process introduced in Example 2.1 are given by

$$\lambda_\ell(\tau) = e^{-\alpha D (\ell-1) \tau}, \quad \varphi_\ell(x) = \frac{1}{\sqrt{(\ell-1)!}} \, H_{\ell-1}\left(\sqrt{\alpha}\, x\right), \quad \ell = 1, 2, \ldots,$$

where H_ℓ denotes the ℓth probabilists' Hermite polynomial (Pavliotis 2014). The eigenfunctions of $\mathcal{P}_\tau$ are given by the eigenfunctions of $\mathcal{K}_\tau$ multiplied by the equilibrium density π, see also Fig. 2.   □

In addition to the eigenvalues and eigenfunctions, an essential part of the Koopman operator analysis is the set of *Koopman modes* for the so-called *full-state observable* $g(x) = x$. The Koopman modes are vectors that, together with the eigenvalues and eigenfunctions, allow us to reconstruct and to propagate the system's state (Williams et al. 2015a). More precisely, assume that each component g_i of the full-state observable, i.e., $g_i(x) = x_i$ for $i = 1, \ldots, d$, can be written in terms of the eigenfunctions
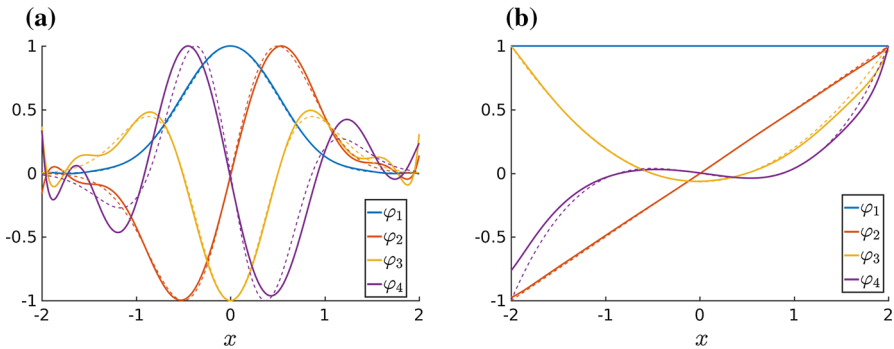
Fig. 2 Dominant eigenfunctions of the Ornstein–Uhlenbeck process computed analytically (dotted lines) and using VAC/EDMD (solid lines). **a** Eigenfunctions of the Perron–Frobenius operator $\mathcal{P}_\tau$. **b** Eigenfunctions of the Koopman operator $\mathcal{K}_\tau$

as $g_i(x) = \sum_\ell \varphi_\ell(x)\, \eta_{i\ell}$. Defining the Koopman modes by $\eta_\ell = [\eta_{1\ell}, \ldots, \eta_{d\ell}]^T$, we obtain $g(x) = x = \sum_\ell \varphi_\ell(x)\, \eta_\ell$ and thus

$$\mathcal{K}_\tau g(x) = \mathbb{E}[g(\mathbf{X}_\tau)\,|\,\mathbf{X}_0 = x] = \mathbb{E}[\mathbf{X}_\tau\,|\,\mathbf{X}_0 = x] = \sum_\ell \lambda_\ell(\tau)\, \varphi_\ell(x)\, \eta_\ell. \quad (3)$$

For vector-valued functions, the Koopman operator is defined to act componentwise. In order to be able to compute eigenvalues, eigenfunctions, and eigenmodes numerically, we project the infinite-dimensional operators onto finite-dimensional spaces spanned by a given set of basis functions. This will be described in detail in Sect. 4.

2.4 Reversibility

We briefly recapitulate the properties of reversible systems. For many applications, including commonly used molecular dynamics models, the dynamics in full phase space are known to be reversible.

Definition 2.7 A system is said to be reversible if the so-called detailed balance condition is fulfilled, i.e., it holds for all $x, y \in \mathbb{X}$ that

$$\pi(x)\, p_\tau(x, y) = \pi(y)\, p_\tau(y, x). \quad (4)$$

Example 2.8 The Ornstein–Uhlenbeck process is reversible. It is straightforward to verify that (4) is fulfilled by the transition density (1) and the stationary density (2) for all values of x, y and τ. Also general Smoluchowski equations of a d-dimensional system of the form

$$d\mathbf{X}_t = -D\nabla V(\mathbf{X}_t)\, dt + \sqrt{2dD}\, d\mathbf{W}_t$$

with dimensionless potential $V(x)$ are reversible. The stationary density is then given by $\pi \propto \exp(-V(x))$ (Leimkuhler et al. 2006). $\qquad \square$

As a result of the detailed balance condition, the Koopman operator $\mathcal{K}_\tau$ and the Perron–Frobenius operator with respect to the equilibrium density, $\mathcal{T}_\tau$, are identical (hence also self-adjoint):

$$\mathcal{K}_\tau f = \int_{\mathbb{X}} p_\tau(x, y) \, f(y) \, \mathrm{d}y = \int_{\mathbb{X}} \frac{\pi(y)}{\pi(x)} \, p_\tau(y, x) \, f(y) \, \mathrm{d}y = \mathcal{T}_\tau \, f.$$

Moreover, both $\mathcal{K}_\tau$ and $\mathcal{P}_\tau$ become self-adjoint with respect to the stationary density, i.e.,

$$\langle \mathcal{P}_\tau f, \, g \rangle_{\pi^{-1}} = \langle f, \, \mathcal{P}_\tau g \rangle_{\pi^{-1}},$$
$$\langle \mathcal{K}_\tau f, \, g \rangle_{\pi} = \langle f, \, \mathcal{K}_\tau g \rangle_{\pi}.$$

Hence, the eigenvalues λ_ℓ are real and the eigenfunctions φ_ℓ of $\mathcal{K}_\tau$ form an orthogonal basis with respect to $\langle \cdot, \, \cdot \rangle_\pi$. That is, the eigenfunctions can be scaled so that $\langle \varphi_\ell, \, \varphi_{\ell'} \rangle_\pi = \delta_{\ell\ell'}$. Furthermore, the leading eigenvalue λ_1 is the only eigenvalue with absolute value 1 and we obtain

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots,$$

see, for example, Nüske et al. (2014). We can then expand a function $f \in L_\pi^2(\mathbb{X})$ in terms of the eigenfunctions as $f = \sum_{\ell=1}^\infty \langle f, \, \varphi_\ell \rangle_\pi \, \varphi_\ell$ such that

$$\mathcal{K}_\tau f = \sum_{\ell=1}^\infty \lambda_\ell(\tau) \, \langle f, \, \varphi_\ell \rangle_\pi \, \varphi_\ell. \tag{5}$$

Furthermore, the eigenvalues decay exponentially with $\lambda_\ell(\tau) = \exp(-\kappa_\ell \tau)$ with relaxation rate κ_ℓ and relaxation timescale t_ℓ^{-1}. Thus, for a sufficiently large lag time τ, the fast relaxation processes have decayed and (5) can be approximated by finitely many terms. The propagator $\mathcal{P}_\tau$ has the same eigenvalues and the eigenfunctions $\tilde{\varphi}_\ell$ are given by $\tilde{\varphi}_\ell(x) = \pi(x) \, \varphi_\ell(x)$.

3 Data-Driven Approximation of Transfer Operators

In this section, we will describe different data-driven methods to identify the dominant dynamics of dynamical systems and to compute eigenfunctions of transfer operators associated with the system, namely TICA and DMD as well as VAC and EDMD. A formal derivation of methods to compute finite-dimensional approximations of transfer operators—resulting in the aforementioned methods – will be given in Sect. 4. Although TICA can be regarded as a special case of VAC, and DMD as a special case of EDMD, these methods are often used in different settings. With the aid of TICA, for instance, it is possible to identify the main slow coordinates and to project the dynamics onto the resulting reduced space, which can then be discretized using conventional Markov state models (a special case of VAC or EDMD, respectively, see Sect. 3.5). We

will introduce the original methods—TICA and DMD—first and then extend these methods to the more general case. Since in many publications a different notation is used, we will first start with the required basic definitions.

In what follows, let $x_i, y_i \in \mathbb{R}^d$, $i = 1, \ldots, m$, be a set of pairs of d-dimensional data vectors, where $x_i = \mathbf{X}_{t_i}$ and $y_i = \mathbf{X}_{t_i+\tau}$. Here, the underlying dynamical system is not necessarily known, the vectors x_i and y_i can simply be measurement data or data from a black box simulation. In matrix form, this can be written as

$$X = \begin{bmatrix} x_1 \ x_2 \ \cdots \ x_m \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \ y_2 \ \cdots \ y_m \end{bmatrix}, \tag{6}$$

with $X, Y \in \mathbb{R}^{d \times m}$. If one long trajectory $\{z_0, z_1, z_2, \ldots\}$ of a dynamical system is given, i.e., $z_i = \mathbf{X}_{t_0+h\,i}$, where h is the step size and $\tau = n_\tau h$ the lag time, we obtain

$$X = \begin{bmatrix} z_0 \ z_1 \ \cdots \ z_{m-1} \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} z_{n_\tau} \ z_{n_\tau+1} \ \cdots \ z_{n_\tau+m-1} \end{bmatrix}.$$

That is, in this case Y is simply X shifted by the lag time τ. Naturally, if more than one trajectory is given, the data matrices X and Y can be concatenated.

In addition to the data, VAC and EDMD require a set of uniformly bounded basis functions or observables, given by $\{\psi_1, \psi_2, \ldots, \psi_k\} \subset L^\infty(\mathbb{X})$. Since $\mathbb{X}$ is assumed to be bounded, we have $\psi_i \in L^r(\mathbb{X})$ for all $i = 1, \ldots, k$ and $1 \le r \le \infty$. The basis functions could, for instance, be monomials, indicator functions, radial basis functions, or trigonometric functions. The optimal choice of basis functions remains an open problem and depends strongly on the system. If the set of basis functions is not sufficient to represent the eigenfunctions, the results will be inaccurate. A too large set of basis functions, on the other hand, might lead to ill-conditioned matrices and overfitting. Cross-validation strategies have been developed to detect overfitting (McGibbon and Pande 2015).

For a basis ψ_i, $i = 1, \ldots, k$, define $\psi : \mathbb{R}^d \to \mathbb{R}^k$ to be the vector-valued function given by

$$\psi(x) = [\psi_1(x), \ \psi_2(x), \ \ldots, \ \psi_k(x)]^T. \tag{7}$$

The goal then is to find the best approximation of a given transfer operator in the space spanned by these basis functions. This will be explained in detail in Sect. 4. In addition to the data matrices X and Y, we will need the transformed data matrices

$$\Psi_X = \begin{bmatrix} \psi(x_1) \ \psi(x_2) \ \ldots \ \psi(x_m) \end{bmatrix} \quad \text{and} \quad \Psi_Y = \begin{bmatrix} \psi(y_1) \ \psi(y_2) \ \ldots \ \psi(y_m) \end{bmatrix}. \tag{8}$$

3.1 Time-Lagged Independent Component Analysis

Time-lagged independent component analysis (TICA) has been introduced in Molgedey and Schuster (1994) as a solution to the *blind source separation* problem, where the correlation matrix and the time-delayed correlation matrix are used to separate superimposed signals. The term TICA has been introduced later (Hyvärinen et al. 2001). TDSEP Ziehe and Müller (1998), an extension of TICA, is popular in the machine learning community. It was shown only recently that TICA is a special case of the VAC by computing the optimal linear projection for approximating the slowest

relaxation processes, and as such provides an approximation of the leading eigenvalues and eigenfunctions of transfer operators (Pérez-Hernández et al. 2013). TICA is now a popular dimension reduction technique in the field of molecular dynamics (Pérez-Hernández et al. 2013; Schwantes and Pande 2013). That is, TICA is used as a preprocessing step to reduce the size of the state space by projecting the dynamics onto the main coordinates. The time-lagged independent components are required (a) to be uncorrelated and (b) to maximize the autocovariances at lag time τ, see (Hyvärinen et al. 2001; Pérez-Hernández et al. 2013) for more details. Assuming that the system is reversible, the TICA coordinates are the eigenfunctions of $\mathcal{T}_\tau$ or $\mathcal{K}_\tau$, respectively, projected onto the space spanned by linear basis functions, i.e., $\psi(x) = x$.

Let $C(\tau)$ be the time-lagged covariance matrix defined by

$$C_{ij}(\tau) = \langle \mathbf{X}_{t,i}\, \mathbf{X}_{t+\tau,j} \rangle_t = \mathbb{E}_\pi \left[\mathbf{X}_{t,i}\, \mathbf{X}_{t+\tau,j} \right].$$

Given data X and Y as defined above, estimators C_0 and C_τ for the covariance matrices $C(0)$ and $C(\tau)$ can be computed as

$$
\begin{aligned}
C_0 &= \frac{1}{m-1} \sum_{k=1}^{m} x_k\, x_k^T = \frac{1}{m-1} X X^T, \\
C_\tau &= \frac{1}{m-1} \sum_{k=1}^{m} x_k\, y_k^T = \frac{1}{m-1} X Y^T.
\end{aligned}
\tag{9}
$$

The time-lagged independent components are defined to be solutions of the eigenvalue problem

$$C_\tau\, \xi_\ell = \lambda_\ell\, C_0\, \xi_\ell \quad \text{or} \quad C_0^+ C_\tau\, \xi_\ell = \lambda_\ell\, \xi_\ell, \tag{10}$$

respectively. In what follows, let $M_{\text{TICA}} = C_0^+ C_\tau$, where C_0^+ denotes the Moore–Penrose pseudo-inverse of C_0.

In applications, often the symmetrized estimators

$$C_0 = \frac{1}{2m-2}(X X^T + Y Y^T) \quad \text{and} \quad C_\tau = \frac{1}{2m-2}(X Y^T + Y X^T)$$

are used so that the resulting TICA coordinates become real-valued. This corresponds to averaging over the trajectory and the time-reversed trajectory. Note that this symmetrization can introduce a large estimator bias that affects the dominant spectrum of (10), if the process is non-stationary, or the distribution of the data is far from the equilibrium of the process. In the latter case, a reweighting procedure can be applied to obtain weighted versions of the estimators (9), to reduce that bias (Wu et al. 2017).

Example 3.1 Let us illustrate the idea behind TICA with a simple example. Consider the data shown in Fig. 3, which was generated by a stochastic process which will typically spend a long time in one of the two clusters before it jumps to the other. We are interested in finding these metastable sets. Standard *principal component analysis* (PCA) leads to the coordinate shown in red, whereas TICA—shown in black—takes
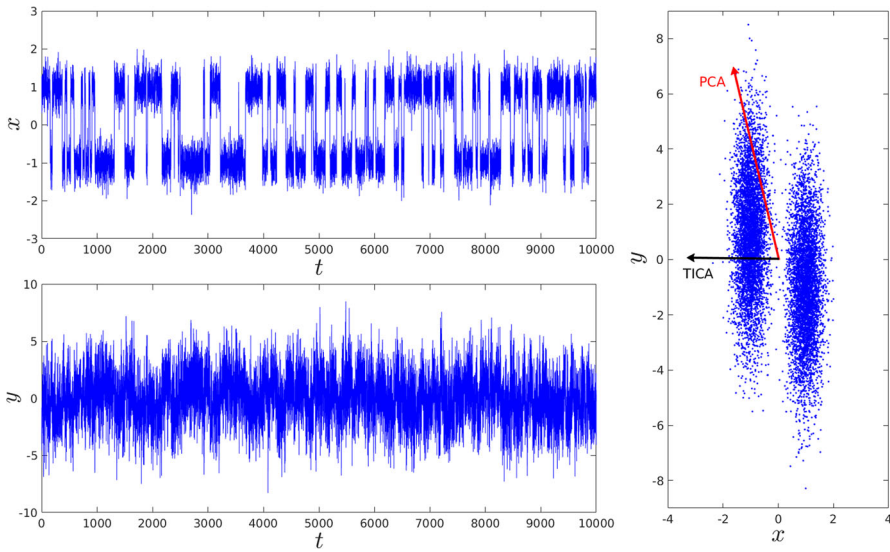
Fig. 3 The difference between PCA and TICA. The top and bottom plot show the x- and y-component of the system, respectively, the plot on the right the resulting main principal component vector and the main TICA coordinate

time information into account and is thus able to identify the slow direction of the system correctly. Projecting the system onto the x-coordinate will preserve the slow process while eliminating the fast stochastic noise. □

Algorithm 1 AMUSE algorithm to compute TICA.

1. Compute a reduced SVD of X, i.e., $X = U \Sigma V^T$.
2. Whiten data: $\tilde{X} = \Sigma^{-1} U^T X$ and $\tilde{Y} = \Sigma^{-1} U^T Y$.
3. Compute $\overline{M}_{\text{TICA}} = \tilde{X}\tilde{Y}^T = \Sigma^{-1} U^T X Y^T U \Sigma^{-1}$.
4. Solve the eigenvalue problem $\overline{M}_{\text{TICA}} w_\ell = \lambda_\ell w_\ell$.
5. The TICA coordinates are then given by $\xi_\ell = U \Sigma^{-1} w_\ell$.

The TICA coordinates can be computed using AMUSE[5] Tong et al. (1990) as shown in Algorithm 1. Instead of computing a singular value decomposition of the data matrix X in step 1, an eigenvalue decomposition of the covariance matrix XX^T could be computed, which is more efficient if $m \gg d$, but less accurate. The vectors ξ_ℓ computed by AMUSE are solutions of the eigenvalue problem (10), since

$$M_{\text{TICA}} \xi_\ell = \left(XX^T\right)^+ XY^T U \Sigma^{-1} w_\ell$$
$$= U \Sigma^{-1} \tilde{X}\tilde{Y}^T w_\ell$$

---

[5] Algorithm for Multiple Unknown Signals Extraction.

$$= \lambda_\ell \, U \Sigma^{-1} w_\ell$$
$$= \lambda_\ell \, \xi_\ell.$$

In the second line, we used the fact that $(XX^T)^+ = U\Sigma^{-2}U^T$ and in the third that w_ℓ is an eigenvector of $\overline{M}_{\mathrm{TICA}} = \tilde{X}\tilde{Y}^T$.

3.2 Dynamic Mode Decomposition

Dynamic mode decomposition (DMD) was developed by the fluid dynamics community as a tool to identify coherent structures in fluid flows (Schmid 2010). Since its introduction, several variants and extensions have been proposed, see (Chen et al. 2012; Tu et al. 2014; Jovanović et al. 2014; Brunton et al. 2015; Klus et al. 2016). A review of the applications of Koopman operator theory in fluid mechanics can be found in Mezić (2013). DMD can be viewed as a combination of a PCA in the spatial domain and a Fourier analysis in the frequency domain (Brunton et al. 2016). It can be shown that the DMD modes are the Koopman modes for the set of basis functions defined by $\psi(x) = x$. Given again data X and Y as above, the idea behind DMD is to assume that there exists a linear operator M_{DMD} such that $y_i = M_{\mathrm{DMD}}\, x_i$. Since the underlying dynamical system is in general nonlinear, this equation cannot be fulfilled exactly and we want to compute the matrix M_{DMD} in such a way that the Frobenius norm of the deviation is minimized, i.e.,

$$\min \| Y - M_{\mathrm{DMD}} X \|_F . \tag{11}$$

The solution of this minimization problem is given by

$$M_{\mathrm{DMD}} = Y X^+ = \left(Y X^T\right)\left(X X^T\right)^+ = C_\tau^T C_0^+ = M_{\mathrm{TICA}}^T. \tag{12}$$

The eigenvalues and eigenvectors of M_{DMD} are called DMD eigenvalues and modes, respectively. That is, we are solving

$$M_{\mathrm{DMD}} \, \xi_\ell = \lambda_\ell \, \xi_\ell.$$

The above equations already illustrate the close relationship with TICA, cf. (10). The DMD modes are the right eigenvectors of M_{DMD}, whereas the TICA coordinates are defined to be the right eigenvectors of the transposed matrix M_{TICA}. Hence, the TICA coordinates are the left eigenvectors of the DMD matrix and the DMD modes the left eigenvectors of the TICA matrix. This is consistent with the results that will be presented in the VAC and EDMD subsections below: The TICA coordinates represent the *Koopman eigenfunctions* projected onto the space spanned by linear basis functions, i.e., $\psi(x) = x$, while the DMD modes are the corresponding *Koopman modes*.

   Similar to AMUSE, the DMD modes and eigenvalues can be obtained without explicitly computing M_{DMD} by using a reduced singular value decomposition of X. Standard and exact DMD are presented in Algorithm 2. The standard DMD modes are

simply the exact DMD modes projected onto the range of the matrix X, see Tu et al. (2014).

Algorithm 2 Standard and exact DMD.

1. Compute compact SVD of X, given by $X = U \Sigma V^T$.
2. Define $\overline{M}_{\mathrm{DMD}} = U^T Y V \Sigma^{-1}$.
3. Compute eigenvalues and eigenvectors of $\overline{M}_{\mathrm{DMD}}$, i.e., $\overline{M}_{\mathrm{DMD}} w_\ell = \lambda_\ell w_\ell$.
4. The DMD mode corresponding to the eigenvalue λ_ℓ is defined as

$$\text{a)} \; \xi_\ell = U w_\ell. \qquad \text{(Standard DMD)}$$

$$\text{b)} \; \xi_\ell = \frac{1}{\lambda} Y V \Sigma^{-1} w_\ell. \quad \text{(Exact DMD)}$$

Remark 3.2 TICA and standard DMD are closely related. When comparing with the AMUSE formulation, we obtain

$$\overline{M}_{\mathrm{TICA}} = \tilde{X}\tilde{Y}^T = \Sigma^{-1} U^T X Y^T U \Sigma^{-1} = \Sigma\, U^T M_{\mathrm{TICA}}\, U \Sigma^{-1} =: W \Lambda W^{-1}$$

and

$$\overline{M}_{\mathrm{DMD}} = U^T Y V \Sigma^{-1} = U^T M_{\mathrm{DMD}}\, U = U^T M_{\mathrm{TICA}}^T U =: \tilde{W}\tilde{\Lambda}\tilde{W}^{-1}.$$

The TICA coordinates are given by $\Xi = U \Sigma^{-1} W$ and the standard DMD modes by $\tilde{\Xi} = U \tilde{W}$ so that—except for the scaling Σ^{-1}—AMUSE and standard DMD use the same projection, the main difference is that the former computes the eigenvectors of M_{TICA} and the latter the eigenvectors of the transposed matrix M_{TICA}^T. As a result, AMUSE could be rewritten to compute the DMD modes if we define $\overline{M}'_{\mathrm{DMD}} = \tilde{Y}\tilde{X}^T = \Sigma^{-1} U^T Y X^T U \Sigma^{-1}$ in step 3 of the algorithm and $\xi_\ell = U \Sigma w_\ell$ in step 5, where w_ℓ now denotes the eigenvectors of $\overline{M}'_{\mathrm{DMD}}$.

3.3 Variational Approach of Conformation Dynamics

The *variational approach of conformation dynamics* (VAC) (Noé and Nüske 2013; Nüske et al. 2014, 2016) is a generalization of the frequently used Markov state modeling framework that allows arbitrary basis functions and is similar to the variational approach in quantum mechanics (Nüske et al. 2014). As described above, VAC and EDMD (see below) require—in addition to the data—a set of basis functions (also called *dictionary*), given by ψ. The variational approach is defined only for reversible systems—EDMD does not require this restriction—and computes eigenfunctions of $\mathcal{T}_\tau$ or $\mathcal{K}_\tau$, respectively. Using the data matrices Ψ_X and Ψ_Y defined in (8), C_0 and C_τ defined in (9) for the transformed data can be estimated as

$$C_0 = \tfrac{1}{m-1} \sum_{k=1}^{m} \psi(x_k)\,\psi(x_k)^T = \tfrac{1}{m-1} \Psi_X \Psi_X^T,$$

$$C_\tau = \tfrac{1}{m-1} \sum_{k=1}^{m} \psi(x_k)\,\psi(y_k)^T = \tfrac{1}{m-1} \Psi_X \Psi_Y^T.$$

In what follows, let $M_{\text{VAC}} = C_0^+ C_\tau$ for the transformed data matrices Ψ_X and Ψ_Y. The matrix M_{VAC} can be regarded as a finite-dimensional approximation of $\mathcal{K}_\tau$ (or $\mathcal{T}_\tau$, since the system is assumed to be reversible; the derivation is shown in Sect. 4), respectively. Eigenfunctions of the operator can then be approximated by the eigenvectors of the matrix M_{VAC}. Let ξ_ℓ be an eigenvector of M_{VAC}, i.e.,

$$M_{\text{VAC}}\,\xi_\ell = \lambda_\ell\,\xi_\ell,$$

and $\varphi_\ell(x) = \xi_\ell^* \psi(x)$, where $*$ denotes the conjugate transpose. Since

$$\mathcal{K}_\tau \varphi_\ell(x) \approx (M_{\text{VAC}}\,\xi_\ell)^* \,\psi(x) = \lambda_\ell\,\xi_\ell^*\,\psi(x) = \lambda_\ell\,\varphi_\ell(x),$$

we obtain an approximation of the eigenfunctions of $\mathcal{K}_\tau$. The derivation will be described in detail in Sect. 4.

3.4 Extended Dynamic Mode Decomposition

Extended dynamic mode decomposition (EDMD), a generalization of DMD, can be used to compute finite-dimensional approximations of the Koopman operator, its eigenvalues, eigenfunctions, and eigenmodes (Williams et al. 2015a, b). It was shown in Klus et al. (2016) that EDMD can be extended to approximate also eigenfunction of the Perron–Frobenius operator with respect to the density underlying the data points. With the notation introduced above, the minimization problem (11) for the transformed data matrices Ψ_X and Ψ_Y can be written as

$$\min \|\Psi_Y - M_{\text{EDMD}}\Psi_X\|_F. \tag{13}$$

The solution—see also (12)—is given by

$$M_{\text{EDMD}} = \Psi_Y \Psi_X^+ = (\Psi_Y \Psi_X^T)(\Psi_X \Psi_X^T)^+ = C_\tau^T\, C_0^+ = M_{\text{VAC}}^T.$$

That is, instead of assuming a linear relationship between the data matrices X and Y, EDMD aims at finding a linear relationship between the transformed data matrices Ψ_X and Ψ_Y. Eigenfunctions of the Koopman operator are then given by the left eigenvectors ξ_ℓ of M_{EDMD}, i.e.,

$$\varphi_\ell(x) = \xi_\ell^*\,\psi(x).$$

The derivation of EDMD can be found in Sect. 4. Since the left eigenvectors of M_{EDMD} are the right eigenvectors of M_{VAC}, VAC and EDMD are equivalent as they compute exactly the same eigenvalue and eigenfunction approximations for a data and basis set.

As shown in Klus et al. (2016), EDMD can also be used to approximate the Perron–Frobenius operator as follows:

$$\tilde{M}_{\text{EDMD}} = \left(\Psi_X \Psi_Y^T\right)\left(\Psi_X \Psi_X^T\right)^+ = C_\tau \, C_0^+.$$

It is important to note that the Perron–Frobenius operator is computed with respect to the density underlying the data matrices. That is, if X is sampled from a uniform distribution, we obtain the eigenfunctions of the Perron–Frobenius operator $\mathcal{P}_\tau$. If we, on the other hand, use one long trajectory, the underlying density converges to the equilibrium density π and we obtain the eigenfunctions of the Perron–Frobenius operator with respect to the equilibrium density, denoted by $\mathcal{T}_\tau$. An approach to compute the equilibrium density from off-equilibrium data is proposed in Wu et al. (2017).

Example 3.3 Let us consider the Ornstein–Uhlenbeck process introduced in Example 2.1. Here, $\alpha = 4$ and $D = 0.25$. The lag time is defined to be $\tau = 1$. We generated 10^5 uniformly distributed test points in $[-2, \, 2]$ and used a basis comprising monomials of order up to 10. With the aid of EDMD, we computed the dominant eigenfunctions of the Perron–Frobenius operator $\mathcal{P}_\tau$ and the Koopman operator $\mathcal{K}_\tau$ (which is identical to $\mathcal{T}_\tau$ here due to reversibility). The results are shown in Fig. 2. The corresponding eigenvalues are given by

$$\lambda_1(\tau) = 1.00, \quad \lambda_2(\tau) = 0.37, \quad \lambda_3(\tau) = 0.13, \quad \lambda_4(\tau) = 0.049,$$

which is a good approximation of the analytically computed eigenvalues (Example 2.6). □

In order to approximate the Koopman modes, let $\varphi(x) = [\varphi_1(x), \, \ldots, \, \varphi_k(x)]^T$ be the vector of eigenfunctions and

$$\Xi = \begin{bmatrix} \xi_1 & \xi_2 & \ldots & \xi_k \end{bmatrix}$$

the matrix that contains all left eigenvectors of M_{EDMD}. Furthermore, define $B \in \mathbb{R}^{d \times k}$ such that $g(x) = B \, \psi(x)$. That is, the full-state observable is written in terms of the basis functions[6]. Since $\varphi(x) = \Xi^* \psi(x)$, this leads to $g(x) = B \, \psi(x) = B \, (\Xi^*)^{-1} \varphi(x)$. Thus, the ℓth column vector of the matrix $\boldsymbol{\eta} = B \, (\Xi^*)^{-1}$ represents the Koopman mode η_ℓ required for the reconstruction of the dynamical system, see (3).

---

[6] The easiest way to accomplish this is by adding the observables x_i, $i = 1, \ldots, d$, to the set of basis functions.
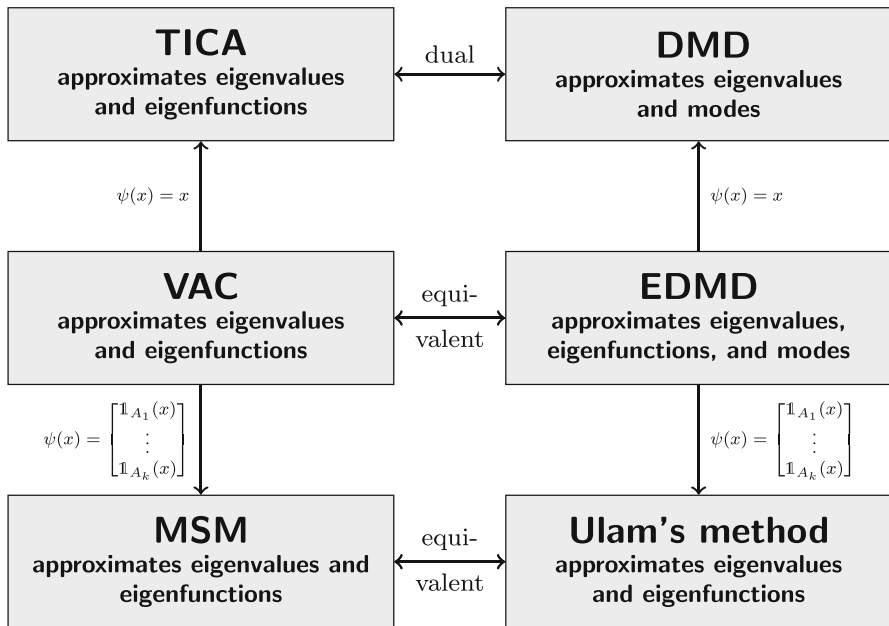
Fig. 4 Relationships between data-driven methods. While VAC was derived for reversible dynamical systems, the derivation of EDMD covers non-reversible dynamics as well

3.5 Relationships with Other Methods

For particular choices of basis functions, VAC and EDMD are equivalent to other methods (see also Nüske et al. 2016; Klus et al. 2016):

1. If we choose $\psi(x) = x$, we obtain TICA and DMD, respectively. That is, the TICA coordinates are the eigenfunctions of the Koopman operator projected onto the space spanned by linear basis functions and the DMD modes are the corresponding Koopman modes. (Note that in this case $\boldsymbol{B} = \boldsymbol{I}$ and the matrix $\boldsymbol{\eta} = (\Xi^*)^{-1}$ contains the right eigenvectors of M_{EDMD}.) In many applications of TICA, the basis functions are modified to have zero mean. For reversible processes, this eliminates the stationary eigenvalue $\lambda_1 = 1$ and its eigenfunction $\varphi_1 \equiv 1$. The largest eigenpair then approximates the slowest dynamical eigenvalue and eigenfunction, respectively.

2. If the set of basis functions comprises indicator functions $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k}$ for a given decomposition of the state space into disjoint sets $A_1, \ldots, A_k$, VAC and EDMD result in Ulam's method Ulam (1960) and thus a Markov state model (MSM).

These relationships are shown in Fig. 4. Detailed examples illustrating the use of VAC and EDMD can be found in Nüske et al. (2014), Williams et al. (2015a), Klus et al. (2016).

3.6 Examples

3.6.1 Double Gyre

Let us consider the autonomous double gyre, which was introduced in Shadden et al. (2005), given by the SDE

$$d\mathbf{X}_t = -\pi \, A \, \sin(\pi \, \mathbf{X}_t) \, \cos(\pi \, \mathbf{Y}_t) + \varepsilon \, d\mathbf{W}_{t,1},$$
$$d\mathbf{Y}_t = \quad \pi \, A \, \cos(\pi \, \mathbf{X}_t) \, \sin(\pi \, \mathbf{Y}_t) + \varepsilon \, d\mathbf{W}_{t,2}$$

on the domain $\mathbb{X} = [0, 2] \times [0, 1]$ with reflecting boundary. For $\varepsilon = 0$, there is no transport between the left half and the right half of the domain and both subdomains are invariant sets with measure $\frac{1}{2}$, cf. (Froyland and Padberg 2009; Froyland and Padberg-Gehle 2014). For $\varepsilon > 0$, there is a small amount of transport due to diffusion and the subdomains are almost invariant. For the Koopman operator $\mathcal{K}_\tau$, this means that for $\varepsilon = 0$ the characteristic functions $\tilde{\varphi}_1 = \mathbb{1}_{[0,1] \times [0,1]}$ and $\tilde{\varphi}_2 = \mathbb{1}_{[1,2] \times [0,1]}$ are both eigenfunctions with corresponding eigenvalue 1. If, on the other hand, $\varepsilon > 0$, then the two-dimensional eigenspace subdivides into two one-dimensional eigenspaces with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1 - \mathcal{O}(\varepsilon)$ and eigenfunctions $\varphi_1 = \mathbb{1}_{[0,2] \times [0,1]}$ and $\varphi_2 \approx \tilde{\varphi}_1 - \tilde{\varphi}_2$. Typical trajectories of the system are shown in Fig. 5a. Using the parameters $A = 0.25$ and $\varepsilon = 0.05$, we integrated 10^5 randomly generated test points using the Euler–Maruyama scheme with step size $h = 10^{-3}$.

For the computation of the eigenfunctions, we choose a set of radial basis functions whose centers were given by the midpoints of an equidistant 50×25 box discretization, and a lag time $\tau = 3$. The resulting nontrivial leading eigenfunctions of the Koopman operator computed with EDMD are shown in Fig. 5b. The two almost invariant sets are clearly visible. The eigenfunctions of the Perron–Frobenius operator exhibit similar patterns (but "rotating" in the opposite direction).

3.6.2 Deca Alanine

As a second example, we illustrate what has become a typical workflow for the application of VAC/EDMD in molecular dynamics, using deca alanine as a model system. Deca alanine is a small peptide comprised of ten alanine residues, and it has been used as a test system many times before. Here, we analyze equilibrium simulations of $3\,\mu\text{s}$ total simulation time using the Amber03 force field, see (Nüske et al. 2014, 2016) for the detailed simulation setup. A set of important quantities for our analysis are the leading *implied timescales*

$$t_m = -\frac{\tau}{\log |\lambda_m|}, \tag{14}$$

for $m = 2, 3, \ldots$. Implied timescales are independent of the lag time (Pazy 1983, Theorem 2.2.4). However, if they are estimated using (14) and an approximation to the eigenvalues λ_m obtained from VAC/EDMD, the timescales will be underestimated (see Sect. 4.2.1) and the error will decrease as a function of the lag time (Djurdjevac et al. 2012). Approximate convergence of implied timescales with increasing lag time
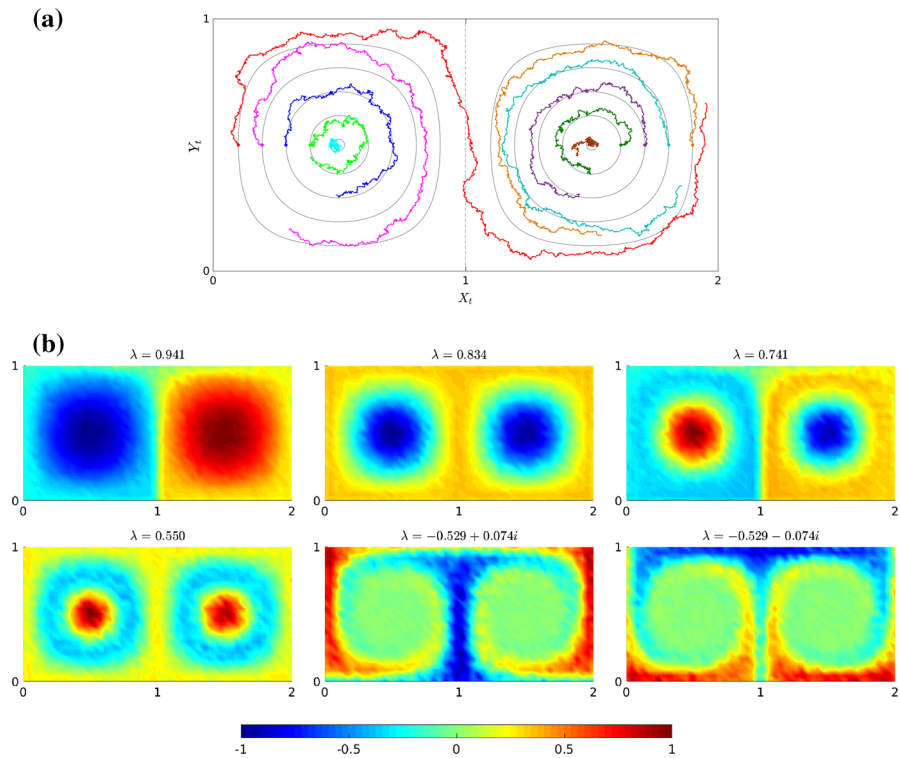
Fig. 5 **a** Typical trajectories (of different lengths) of the double gyre system for $\varepsilon = 0.05$. The initial states are marked by dots. Due to the diffusion term, particles can cross the separatrix (dashed line). The gray lines show the trajectories with the same initial conditions for $\varepsilon = 0$. **b** Leading eigenfunctions of the Koopman operator associated with the double gyre system computed using EDMD

has become a standard model validation criterion in molecular dynamics (Prinz et al. 2011).

In the first step, a set of internal molecular coordinates is extracted from the simulation data, to which TICA is applied. In our example, we select all 16 backbone dihedral angles as internal molecular coordinates. Figure 6a shows the first five implied timescales estimated by TICA as a function of the lag time τ.

Next, a first dimension reduction is performed, where the data are projected onto the leading M TICA eigenvectors. The number M is selected by the criterion of total kinetic variance, that is, M is the smallest number such that the cumulative sum of the first M squared eigenvalues exceeds 95 per cent of the total sum of squared eigenvalues (Noé and Clementi 2015). Figure 6c shows the resulting dimension M as a function of the lag time.

As a third step, the reduced data set is discretized by application of a clustering method. In our case, we use k-means clustering to assign the data to 50 discrete states. A Markov state model (MSM, equivalent to Ulam's method, see above) is estimated from the discretized time series. We show the first five implied timescales from the MSM in Fig. 6c and observe that estimates improve compared to the TICA approximations. Also, timescale estimates converge for lag times $\tau \geq 4$ ns.
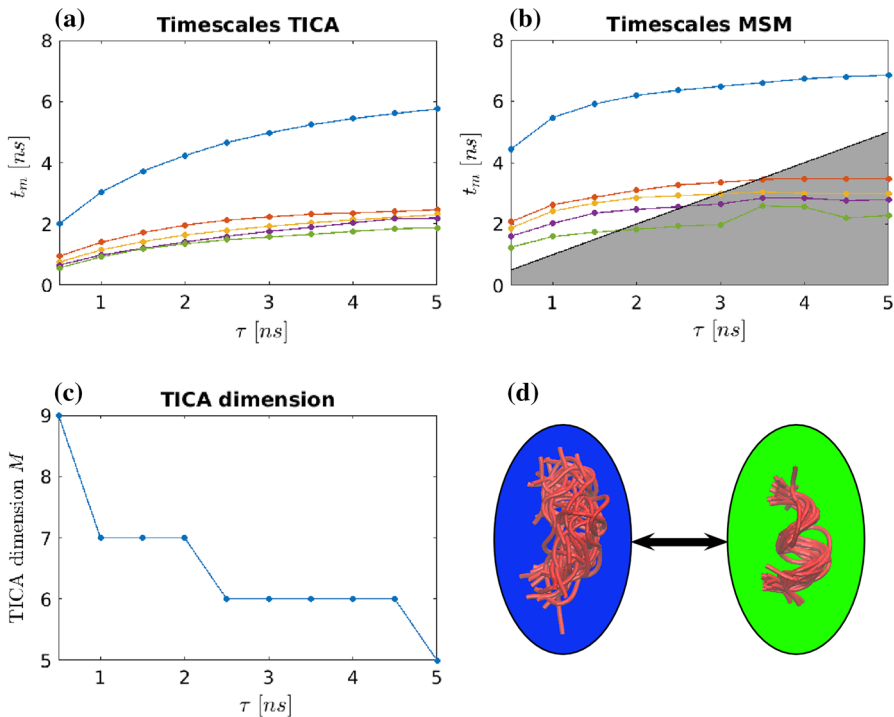
Fig. 6 Illustration of standard EDMD workflow in molecular dynamics using the deca alanine model system. **a** Leading implied timescales t_m (in nanoseconds) as estimated by TICA as a function of the lag time. **b** Effective dimension M selected by applying the criterion of total kinetic variance to the TICA eigenvalues. **c** Leading implied timescales t_m estimated by a Markov state model after projecting the data onto the first M TICA eigenvectors and discretizing this data set into 50 states using k-means. **d** Simple visualization of effective coarse grained dynamics. All MSM states are assigned to two macrostates using the PCCA algorithm. An overlay of representative structures from both macrostates shows that the dynamics between them corresponds to helix formation. Macrostates are drawn proportionally to their stationary probability

Finally, we use the converged model at lag time $\tau = 4$ ns for further analysis. As the slowest implied timescale t_2 dominates all others, and as it is the only one which is larger than the lag time used for analysis (indicated by the gray line in Fig. 6c), we attempt to extract a two-state model that captures the essential dynamics. We employ the PCCA + algorithm (Deuflhard and Weber 2005; Röblitz and Weber 2013) to coarse grain all MSM states into two macrostates. Inspection of randomly selected trajectory frames belonging to each macrostate reveals that the slow dynamical process in the data corresponds to the formation of a helix, see Fig. 6d. It should be noted that this coarse graining works well for visualization purposes, but some details need to be taken into account. In fact, PCCA performs a fuzzy assignment of MSM states to macrostates, where each MSM state belongs to each macrostate with a certain *membership* in [0, 1]. We simply assign each MSM state to the macrostate with maximal membership here. Alternatively, we could also use a *hidden Markov model* (HMM) to perform the coarse graining (Noé et al. 2013).

4 Derivations

In this section, we will show how VAC and EDMD as well as their respective special cases TICA and DMD can be derived and how these methods are related to eigenfunctions and eigenmodes of transfer operators.

4.1 General Dynamical Systems

Let us begin with general, not necessarily reversible dynamical systems. In order to be able to compute eigenfunctions of transfer operators numerically, the infinite-dimensional operators are projected onto a finite-dimensional space. We will briefly outline how the EDMD minimization problem (13) leads to an approximation of the Koopman operator.

Theorem 4.1 *Let the process $\{\mathbf{X}_t\}_{t \geq 0}$ be Feller-continuous[7]. Let ψ_i, $i = 1, \ldots, k$, be the set of at least piecewise continuous basis functions of the finite-dimensional linear space $\mathbb{V}$. Let the empirical distribution of the data points $x_1, x_2, \ldots$ converge weakly to the density ρ. Then the minimization problem*

$$\min_{K \in \mathbb{R}^{k \times k}} \frac{1}{m} \sum_{j=1}^{m} \left\| \psi(y_j) - K^T \psi(x_j) \right\|_2^2$$

converges, as $m \to \infty$, almost surely to $\min_{\hat{K}} \sum_{i=1}^{k} \|\mathcal{K}_\tau \psi_i - \hat{K} \psi_i\|_\rho^2$, where the minimization is over all linear mappings $\hat{K} : \mathbb{V} \to \mathbb{V}$.

Proof Let $f = \sum_{i=1}^{k} a_i \psi_i = a^T \psi \in \mathbb{V}$ be an arbitrary function, where $a = [a_1, \ldots, a_k]^T$. For a single data point x_j, we have for a linear mapping $\hat{K} : \mathbb{V} \to \mathbb{V}$ with matrix representation $K \in \mathbb{R}^{k \times k}$ that

$$\hat{K} f(x_j) = \sum_{i=1}^{k} (K a)_i \psi_i(x_j) = a^T K^T \psi(x_j).$$

Here, the ith column of the matrix K corresponds to $\hat{K} \psi_i$. Thus, we obtain

$$\frac{1}{m} \sum_{j=1}^{m} \left\| \psi(y_j) - K^T \psi(x_j) \right\|_2^2$$

$$= \sum_{i=1}^{k} \frac{1}{m} \sum_{j=1}^{m} (\psi_i(y_j) - \hat{K} \psi_i(x_j))^2$$

---

[7] A process $\{\mathbf{X}_t\}_{t \geq 0}$ is called Feller-continuous if the mapping $x \mapsto \mathbb{E}[g(\mathbf{X}_t)|\mathbf{X}_0 = x]$ is continuous for any fixed continuous function g. This implies, that the Koopman operator of a Feller-continuous process has a well-defined restriction from $L^\infty(\mathbb{X})$ to the set of continuous functions. Any stochastic process generated by an Itô stochastic differential equation with Lipschitz-continuous coefficients is Feller-continuous (Øksendal 2003, Lemma 8.1.4).

$$\overset{m \to \infty}{\longrightarrow} \sum_{i=1}^{k} \int_{\mathbb{X}} (\mathbb{E}[\psi_i(\mathbf{X}_\tau) \mid \mathbf{X}_0 = x] - \hat{K}\psi_i(x))^2 \rho(x)\, dx$$

$$= \sum_{i=1}^{k} \|\mathcal{K}_\tau \psi_i - \hat{K}\psi_i\|_\rho^2,$$

where the convergence for $m \to \infty$ is almost sure. From the first line to the second we used that the y_j are realizations of the random variables $\mathbf{X}_\tau$ given $\mathbf{X}_0 = x_j$, that $\mathbf{X}_\tau$ is a Feller-continuous process, that the ψ_i are (piecewise) continuous functions, and that the sampling process of x_j is independent of the noise process that decides over $\mathbf{X}_\tau$ given $\mathbf{X}_0 = x_j$. □

With the aid of the data matrices Ψ_X and Ψ_Y defined in (8), this minimization problem can be written as

$$\min \left\| \Psi_Y - K^T \Psi_X \right\|_F^2,$$

which is identical to (13), where now $K^T = M_{\text{EDMD}}$. Thus, the transposed EDMD matrix M_{EDMD} is an approximation of the Koopman operator. A similar setup allows for the approximation of the Perron–Frobenius operator with respect to the data point density ρ. For details, we refer to (Klus et al. 2016, Appendix A). Note, however, that although the Perron–Frobenius and Koopman operators are adjoint, the matrix representation of the discrete Perron–Frobenius operator will in general not just be the transposed of the matrix K, unless the ansatz functions ψ_i are orthonormal with respect to $\langle \cdot, \cdot \rangle_\rho$.

If the dynamical system is deterministic, we can already interpret the minimization (13) for finite values of m. As shown, e.g., in Klus et al. (2016), Korda and Mezić (2017), the solution of (13) is a Petrov–Galerkin projection of the Koopman operator on the ansatz space $\mathbb{V}$.

4.2 Reversible Dynamical Systems

Let us now assume that the system is reversible. That is, it holds that $\pi(x)\, p_\tau(x, y) = \pi(y)\, p_\tau(y, x)$ for all x and y.

4.2.1 Variational Principle for the Rayleigh Trace

We can also derive a variational formulation for the first M eigenvalues of the Koopman operator $\mathcal{K}_\tau$ in the reversible setting. It is a standard result for self-adjoint operators on a Hilbert space with bounded eigenvalue spectrum, see, for example, Bandle (1980):

Proposition 4.2 *Assume that $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_M$ are the dominant eigenvalues of the Koopman operator $\mathcal{K}_\tau$ on L_π^2. Then*

$$\sum_{\ell=1}^{M} \lambda_\ell = \sup \sum_{\ell=1}^{M} \langle \mathcal{K}_\tau v_\ell, v_\ell \rangle_\pi,$$

$$\langle v_\ell, v_{\ell'} \rangle_\pi = \delta_{\ell\ell'}$$

(15)

The sum of the first M eigenvalues maximizes the Rayleigh trace, which is the sum on the right-hand side of (15) over all selections of M orthonormal functions v_ℓ. The maximum is attained for the first M eigenfunctions $\varphi_1, \ldots, \varphi_M$.

Proof The M-dimensional space $\mathbb{V}$ spanned by the functions v_ℓ must contain an element u_M which is orthonormal to the first $M-1$ eigenfunctions φ_ℓ, i.e., $\langle u_M, \varphi_\ell \rangle_\pi = 0, \ell = 1, \ldots, M-1$, and $\|u_M\|_\pi = 1$. By the standard Rayleigh principle for self-adjoint operators

$$\langle K_\tau u_M, u_M \rangle_\pi \le \lambda_M.$$

Next, determine a normalized element u_{M-1} of the orthogonal complement of u_M in $\mathbb{V}$ with $\langle u_{M-1}, \varphi_\ell \rangle_\pi = 0, \ell = 1, \ldots, M-2$. Again, we can invoke the Rayleigh principle to find

$$\langle \mathcal{K}_\tau u_{M-1}, u_{M-1} \rangle_\pi \le \lambda_{M-1}.$$

Repeating this argument another $M-2$ times provides an orthonormal basis $u_1, \ldots, u_M$ of the space $\mathbb{V}$ such that

$$\sum_{\ell=1}^{M} \langle \mathcal{K}_\tau u_\ell, u_\ell \rangle_\pi \le \sum_{\ell=1}^{M} \lambda_\ell.$$

As the Rayleigh trace is independent of the choice of orthonormal basis for the subspace $\mathbb{V}$, and the space itself was arbitrary, this proves (15). Clearly, the maximum is attained for the first M eigenfunctions. $\qquad\square$

Proposition 4.2 motivates the variational approach developed in Noé and Nüske (2013), Nüske et al. (2014) to maximize the Rayleigh trace restricted to some fixed space of ansatz functions:

Proposition 4.3 *Let $\mathbb{V}$ be a space of k linearly independent ansatz functions ψ_i given by a dictionary as above. The set of $M \le k$ mutually orthonormal functions $f^\ell = \sum_{i=1}^{k} a_i^\ell \psi_i$ which maximize the Rayleigh trace of the Koopman operator restricted to $\mathbb{V}$ is given by the first M eigenvectors of the generalized eigenvalue problem*

$$C_\tau \, a^\ell = \hat{\lambda}_\ell \, C_0 \, a^\ell,$$

(16)

where $a^\ell = (a_i^\ell)_{i=1}^{k}$, and the matrices C_τ, C_0 are given by

$$(C_\tau)_{ij} = \langle \mathcal{K}_\tau \psi_i, \psi_j \rangle_\pi,$$
$$(C_0)_{ij} = \langle \psi_i, \psi_j \rangle_\pi.$$

Proof First, note that for any functions $f = \sum_{i=1}^{k} a_i \, \psi_i$ and $g = \sum_{i=1}^{k} b_i \, \psi_i$, we have that

$$\langle \mathcal{K}_\tau f, g \rangle_\pi = a^T C_\tau \, b,$$

$$\langle f, g \rangle_\pi = a^T C_0 \, b.$$

Let us assume that the ansatz functions are mutually orthonormal, i.e., $C_0 = I$. Then, maximization of the Rayleigh trace is equivalent to finding M vectors a^ℓ, such that $(a^\ell)^T a^{\ell'} = \delta_{\ell \ell'}$ and

$$\sum_{\ell=1}^{M} \left(a^\ell \right)^T C_\tau \, a^\ell = \sum_{\ell=1}^{M} \langle C_\tau \, a^\ell, a^\ell \rangle$$

is maximal. By Proposition 4.2 applied to the operator C_τ on $\mathbb{R}^N$, the vectors a^ℓ are given by the first M eigenvectors of C_τ. In the general case, transform the basis functions into a set of mutually orthonormal functions $\tilde{\psi}_i$ via $\tilde{\psi}_i = \sum_{j=1}^{k} C_0^{-1/2}(j, i) \, \psi_j$. For the transformed basis, we need to compute the eigenvectors $\tilde{a}^\ell$ of

$$C_0^{-1/2} C_\tau C_0^{-1/2} \, \tilde{a}^\ell = \hat{\lambda}_\ell \, \tilde{a}^\ell.$$

This is equivalent to the generalized eigenvalue problem (16), the relation between the eigenvectors is given by

$$a^\ell = C_0^{-1/2} \, \tilde{a}^\ell.$$

$\square$

5 Conclusion

In this review paper, we established connections between different data-driven model reduction and transfer operator approximation methods developed independently by the dynamical systems, fluid dynamics, machine learning, and molecular dynamics communities. Although the derivations of these methods differ, we have shown that the resulting algorithms share many similarities.

DMD, TICA and MSMs are popular methods to approximate the dynamics of high-dimensional systems. Due to their simple basis functions, they conduct relatively rough approximations, but when only a few spectral components are required, the approximation error can be controlled by choosing sufficiently large lag times τ (Sarich et al. 2010). The more general methods VAC and EDMD are better suited to obtain accurate approximations of eigenfunctions. However, to ensure such an accurate approximation, one would have to deploy multiple basis functions in all coordinates and their combinations, which is unfeasible for high-dimensional systems, and would also lead to overfitting when estimating the eigenfunctions of the Koopman operator from a finite data set (McGibbon and Pande 2015).

A natural approach to mitigate these problems is to construct an iterative or "deep" approach in which the dynamical systems subspace in which a high resolution of basis functions is required is found by multiple successive analysis steps. A common approach is to first reduce the dimension by an inexpensive method such as TICA,

in order to have a relatively low-dimensional space in which the eigenfunctions are approximated with a higher-resolution method. Another possibility is to exploit low-rank tensor approximations of transfer operators and their eigenfunctions. Tensor- and sparse-grid-based reformulations of some of the methods described in this paper can be found in Nüske et al. (2016), Klus and Schütte (2016), Klus et al. (2016), and in Junge and Koltai (2009), respectively. The efficiency of these tensor decomposition approaches depends strongly on the coupling structure; strong coupling between different variables typically leads to high ranks. Furthermore, some tensor formats also depend on the ordering of variables and a permutation of the variable's indices would lead to different tensor decompositions. Yet another approach might be to exploit sparsity-promoting methods using L_1-regularization techniques. Basis functions that are not required to represent the eigenfunctions of an operator can thus be eliminated and refined adaptively. Moreover, dictionary-learning methods could be applied to *learn* a basis set and to adapt the dictionary to specific data (Mairal et al. 2009). Future work includes evaluating and combining different dimensionality reduction, tensor decomposition, and sparsification methods to mitigate the curse of dimensionality.

References

Bandle, C.: Isoperimetric inequalities and applications. Pitman, Monographs and studies in mathematics (1980)

Baxter, J.R., Rosenthal, J.S.: Rates of convergence for everywhere-positive Markov chains. Stat. Probab. Lett. **22**(4), 333–338 (1995)

Brunton, S.L., Proctor, J.L., Tu, J.H., Kutz, J.N.: Compressed sensing and dynamic mode decomposition. J. Comput. Dyn. **2**(2), 165–191 (2015)

Brunton, B.W., Johnson, L.A., Ojemann, J.G., Kutz, J.N.: Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. J. Neurosci. Methods **258**, 1–15 (2016)

Budišić, M., Mohr, R., Mezić, I.: Applied koopmanism. Appl. Chaos An Interdiscip. J. Nonlinear Sci. **22**(4), 047510 (2012)

Chen, K.K., Tu, J.H., Rowley, C.W.: Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses. J. Nonlinear Sci. **22**(6), 887–915 (2012)

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

Dellnitz, M., Junge, O.: On the approximation of complicated dynamical behavior. SIAM J. Numer. Anal. **36**(2), 491–515 (1999)

Deuflhard, P., Weber, M.: Robust Perron cluster analysis in conformation dynamics. Linear Algebra Appl. **398**, 161–184 (2005)

Djurdjevac, N., Sarich, M., Schütte, C.: Estimating the eigenvalue error of Markov state models. Multiscale Model. Simul. **10**(1), 61–81 (2012)

Ferguson, A.L., Panagiotopoulos, A.Z., Kevrekidis, I.G., Debenedetti, P.G.: Nonlinear dimensionality reduction in molecular simulation: the diffusion map approach. Chem. Phys. Lett. **509**(1), 1–11 (2011)

Froyland, G., Padberg, K.: Almost-invariant sets and invariant manifolds-connecting probabilistic and geometric descriptions of coherent structures in flows. Physica D **238**, 1507–1523 (2009)

Froyland, G., Padberg-Gehle, K.: Almost-invariant and finite-time coherent sets: directionality, duration, and diffusion. In: Bahsoun, W., Bose, C.H., Froyland, G. (eds.) Ergodic Theory. Open Dynamics, and Coherent Structures, pp. 171–216. Springer, New York (2014)

Froyland, G., Junge, O., Koltai, P.: Estimating long term behavior of flows without trajectory integration: the infinitesimal generator approach. SIAM J. Numer. Anal. **51**(1), 223–247 (2013)

Froyland, G., Gottwald, G., Hammerlindl, A.: A computational method to extract macroscopic variables and their dynamics in multiscale systems. SIAM J. Appl. Dyn. Syst. **13**(4), 1816–1846 (2014)

Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. rXiv e-prints (2015)

Hopf, E.: The general temporally discrete Markoff process. J. Ration. Mech. Anal. **3**(1), 13–45 (1954)

Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, Hoboken (2001)

Jovanović, M.R., Schmid, P.J., Nichols, J.W.: Sparsity-promoting dynamic mode decomposition. Phys. Fluids **26**(2), 024103 (2014)

Junge, O., Koltai, P.: Discretization of the Frobenius–Perron operator using a sparse Haar tensor basis: the Sparse Ulam method. SIAM J. Numer. Anal. **47**, 3464–3485 (2009)

Klus, S., Gelß, P., Peitz, S. and Schütte, Ch.: Tensor-based dynamic mode decomposition. ArXiv e-prints (2016)

Klus, S. and Schütte, CH.: Towards tensor-based methods for the numerical approximation of the Perron–Frobenius and Koopman operator. J. Comput. Dyn. 3(2) (2016)

Klus, S., Koltai, P., Schütte, Ch.: On the numerical approximation of the Perron–Frobenius and Koopman operator. J. Comput. Dyn. **3**(1), 51–79 (2016)

Koopman, B.: Hamiltonian systems and transformation in Hilbert space. Proc. Ntl. Acad. Sci. USA **17**(5), 315 (1931)

Korda, M. and Mezić, I.: On convergence of Extended Dynamic Mode Decomposition to the Koopman operator. arXiv preprint rXiv:1703.04680 (2017)

Krengel, U.: Ergodic Theorems, Volume 6 of de Gruyter Studies in Mathematics, vol. 6. Walter de Gruyter & Co., Berlin (1985)

Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: Dynamic mode decomposition: data-driven modeling of complex systems. SIAM (2016)

Lasota, A.: Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics volume 97 of Applied Mathematical Sciences, 2nd edn. Springer, Berlin (1994)

Leimkuhler, B., Chipot, Ch., Elber, R., Laaksonen, A., Mark, A., Schlick, T., Schütte, Ch., Skeel, R.: New Algorithms for Macromolecular Simulation (Lecture Notes in Computational Science and Engineering). Springer, New York (2006)

Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09, pp. 689–696. ACM (2009)

McGibbon, R.T., Pande, V.S.: Variational cross-validation of slow dynamical modes in molecular kinetics. J. Chem. Phys. **142**(12), 03B621 (2015)

Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn. **41**(1), 309–325 (2005)

Mezić, I.: Analysis of fluid flows via spectral properties of the Koopman operator. Ann. Rev. Fluid Mech. **45**(1), 357–378 (2013)

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

Nadler, B., Lafon, S., Coifman, R.R., Kevrekidis, I.G.: Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Appl. Comput. Harmonic Anal. **21**(1), 113–127 (2006)

Noé, F., Clementi, C.: Kinetic distance and kinetic maps from molecular dynamics simulation. J. Chem. Theory Comput. **11**(10), 5002–5011 (2015)

Noé, F., Nüske, F.: A variational approach to modeling slow processes in stochastic dynamical systems. Multiscale Model. Simul. **11**(2), 635–655 (2013)

Noé, F., Wu, H., Prinz, J.-H., Plattner, N.: Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. J. Chem. Phys. **139**, 184114 (2013)

Nüske, F., Keller, B.G., Pérez-Hernández, G., Mey, A.S.J.S., Noé, F.: Variational approach to molecular kinetics. J. Chem. Theory Comput. **10**(4), 1739–1752 (2014)

Nüske, F., Keller, B.G., Pérez-Hernández, G., Mey, A.S.J.S., Noé, F.: Variational approach to molecular kinetics. J. Chem. Theory Comput. **10**, 1739–1752 (2014)

Nüske, F., Schneider, R., Vitalini, F., Noé, F.: Variational tensor approach for approximating the rare-event kinetics of macromolecular systems. J. Chem. Phys. **144**(5), 054105 (2016)

Øksendal, B.: Stochastic Differential Equations, 6th edn. Springer, Berlin (2003)

Pavliotis, G.A.: Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations volume 60 of Texts in Applied Mathematics. Springer, Berlin (2014)

Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Springer, Berlin (1983)

Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., Noé, F.: Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. **139**(1), 07B604 (2013)

Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J.D., Schütte, C., Noé, F.: Markov models of molecular kinetics: generation and validation. J. Chem. Phys. **134**, 174105 (2011)

Röblitz, S., Weber, M.: Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. Adv. Data Anal. Classif. **7**(2), 147–179 (2013)

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

Sarich, M., Noé, F., Schütte, C.: On the approximation quality of Markov state models. Multiscale Model. Simul. **8**, 1154–1177 (2010)

Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5–28 (2010)

Schütte, Ch. and Sarich, M.: *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*. Number 24 in Courant Lecture Notes. American Mathematical Society, (2013)

Schütte, Ch., Fischer, A., Huisinga, W., Deuflhard, P.: A direct approach to conformational dynamics based on hybrid Monte Carlo. J. Comput. Phys. **151**(1), 146–168 (1999)

Schwantes, C.R., Pande, V.S.: Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. J. Chem. Theory. Comput. **9**, 2000–2009 (2013)

Shadden, S.C., Lekien, F., Marsden, J.E.: Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. Phys. D Nonlinear Phenom. **212**(3), 271–304 (2005)

Tong, L., Soon, V.C., Huang, Y.F., Liu, R.: AMUSE: a new blind identification algorithm. In: IEEE International Symposium on Circuits and Systems, pp. 1784–1787 (1990)

Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L. ,Kutz, J.N.: On dynamic mode decomposition: theory and applications. J. Comput. Dyn. 1(2), (2014)

Ulam, S.M.: A Collection of Mathematical Problems. Interscience Publisher, Hoboken (1960)

Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. **25**(6), 1307–1346 (2015)

Williams, M.O., Rowley, C.W., Kevrekidis, I.G.: A kernel-based method for data-driven Koopman spectral analysis. J. Comput. Dyn. **2**(2), 247–265 (2015)

Wu, H., Nüske, F., Paul, F., Klus, S., Koltai, P., Noé, F.: Variational Koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations. J. Chem. Phys. **146**(15), 154104 (2017)

Ziehe, A. and Müller, K.-R.: TDSEP—an efficient algorithm for blind separation using time structure. In: CANN 98, pp. 675–680. Springer Science and Business Media (1998)