



On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator

Milan Korda¹ · Igor Mezić¹

Received: 10 May 2017 / Accepted: 30 October 2017 / Published online: 13 November 2017
© Springer Science+Business Media, LLC 2017

Abstract Extended dynamic mode decomposition (EDMD) (Williams et al. in J Nonlinear Sci 25(6):1307–1346, 2015) is an algorithm that approximates the action of the Koopman operator on an N -dimensional subspace of the space of observables by sampling at M points in the state space. Assuming that the samples are drawn either independently or ergodically from some measure μ , it was shown in Klus et al. (J Comput Dyn 3(1):51–79, 2016) that, in the limit as $M \rightarrow \infty$, the EDMD operator $\mathcal{K}_{N,M}$ converges to \mathcal{K}_N , where \mathcal{K}_N is the $L_2(\mu)$ -orthogonal projection of the action of the Koopman operator on the finite-dimensional subspace of observables. We show that, as $N \rightarrow \infty$, the operator \mathcal{K}_N converges in the strong operator topology to the Koopman operator. This in particular implies convergence of the predictions of future values of a given observable over any finite time horizon, a fact important for practical applications such as forecasting, estimation and control. In addition, we show that accumulation points of the spectra of \mathcal{K}_N correspond to the eigenvalues of the Koopman operator with the associated eigenfunctions converging weakly to an eigenfunction of the Koopman operator, provided that the weak limit of the eigenfunctions is nonzero. As a by-product, we propose an analytic version of the EDMD algorithm which, under some assumptions, allows one to construct \mathcal{K}_N directly, without the use of sampling. Finally, under additional assumptions, we analyze convergence of $\mathcal{K}_{N,N}$ (i.e., $M = N$), proving convergence, along a subsequence, to weak eigenfunctions (or eigendistributions) related to the eigenmeasures of the Perron–Frobenius operator. No

Communicated by Bruno Eckhardt.

✉ Milan Korda
milan.korda@engineering.ucsb.edu

Igor Mezić
mezic@engineering.ucsb.edu

¹ University of California, Santa Barbara, Santa Barbara, CA, USA

assumptions on the observables belonging to a finite-dimensional invariant subspace of the Koopman operator are required throughout.

Keywords Koopman operator · Dynamic mode decomposition · Convergence · Spectrum

Mathematics Subject Classification 37-XX · 37A30 · 47Nxx

1 Introduction

Recently, there has been an expanding interest in utilizing the *spectral expansion*-based methodology that enabled progress in data-driven analysis of high-dimensional nonlinear systems. The research was initiated in Mezić (2005) and Mezić and Banaszuk (2004), using composition (Koopman) operator representation originally defined in Koopman (1931). The framework is being used for model reduction, identification, prediction, data assimilation and control of deterministic (e.g., Budisić et al. 2012; Mauroy and Goncalves 2017; Rowley et al. 2009; Williams et al. 2015; Brunton et al. 2016b; Giannakis et al. 2015; Korda and Mezić 2016) as well as stochastic dynamical systems (e.g., Takeishi et al. 2017; Riseth and Taylor-King 2017; Wu and Noé 2017). This has propelled the theory to wide use on a diverse set of applications such as fluid dynamics (Sharma et al. 2016; Glaz et al. 2016), power grid dynamics (Raak et al. 2016), neurodynamics (Brunton et al. 2016a), energy efficiency (Georgescu and Mezić 2015), molecular dynamics (Wu et al. 2017) and data fusion (Williams et al. 2015).

Numerical methods for approximation of the spectral properties of the Koopman operator have been considered since the inception of the data-driven analysis of dissipative dynamical systems (Mezić and Banaszuk 2004; Mezić 2005). These belong to the class of generalized Laplace analysis (GLA) methods (Mezić 2013). An alternative line of algorithms, called the dynamic mode decomposition (DMD) algorithm (Schmid 2010; Rowley et al. 2009), has also been advanced, enabling concurrent data-driven determination of approximate eigenvalues and eigenvectors of the underlying DMD operator. The examples of DMD-type algorithms are (1) the companion matrix method proposed by Rowley et al. (2009), (2) the SVD-based DMD developed by Schmid (2010), (3) the Exact DMD method introduced by Tu et al. (2014), (4) the Extended DMD (Williams et al. 2015). The relationship between these methods and the spectral operator properties of the Koopman operator was first noticed in Rowley et al. (2009), based on the spectral expansion developed in Mezić (2005). However, rigorous results in this direction are sparse. Williams et al. (2015) provided a result, the corollary of which is that the spectrum of the EDMD approximation is contained in the spectrum of the Koopman operator provided the observables belong to a finite-dimensional invariant subspace of the Koopman operator and the data matrix is of the same rank. The work of Arbabi and Mezić (2016) suggested that an alternative assumption to the finite rank is that the number of sampling points M goes to infinity even though the results of Arbabi and Mezić (2016) still implicitly rely on a finite-dimensional assumption.

The work (Klus et al. 2016) showed that, assuming either independent identically distributed (iid) or ergodic sampling from a measure μ , the EDMD operator on N observables constructed using M samples, $\mathcal{K}_{N,M}$, converges as $M \rightarrow \infty$ to \mathcal{K}_N , where \mathcal{K}_N is the $L_2(\mu)$ -orthogonal projection of the action of the Koopman operator on the finite-dimensional subspace of observables. In this work, we show that \mathcal{K}_N converges to \mathcal{K} in the strong operator topology. As a result, predictions of a given observable obtained using \mathcal{K}_N or $\mathcal{K}_{N,M}$ over any finite prediction horizon converge in the $L_2(\mu)$ norm (as N or N and M tend to infinity) to its true values. In addition, we show that, as $N \rightarrow \infty$, accumulation points of the spectra of \mathcal{K}_N correspond to eigenvalues of the Koopman operator and the associated eigenfunctions converge weakly to an eigenfunction of the Koopman operator, provided that the weak limit of the eigenfunctions is nonzero. The results hold in a very general setting with minimal underlying assumptions. In particular, we *do not* assume that the finite-dimensional subspace of observables is invariant under the action of the Koopman operator or that the dynamics is measure preserving.

As a by-product of our results, we propose an analytic version of the EDMD algorithm which allows one to construct \mathcal{K}_N directly, without the need for sampling, under the assumption that the transition mapping of the dynamical system is known analytically and the N -dimensional subspace of observables is such that the integrals of the products of the observables precomposed with the transition mapping can be evaluated in closed form. This method is not immediately useful for large-scale data-driven applications that the EDMD was originally designed for, but it may prove useful in control applications (e.g., Korda and Mezić 2016), where model is often known, or for numerical studies of Koopman operator approximations on classical examples, eliminating the sampling error in both cases.

Finally, we analyze convergence of $\mathcal{K}_{N,N}$, i.e., the situation where the number of samples M and the number of observables N are equal. Under the additional assumptions that the sample points lie on the same trajectory and the mapping T is a homeomorphism, we prove convergence along a subsequence to weak eigenfunctions (or eigendistributions in the sense of Gelfand and Shilov 1964) of the Koopman operator, which also turn out to be eigenmeasures of the Perron–Frobenius operator.

The paper is organized as follows: In Sect. 2, we introduce the setting of EDMD. In Sect. 3, we show that EDMD is an orthogonal projection of the action of the Koopman operator on a finite subspace of observables with respect to the empirical measure supported on sample points drawn from a measure μ . In Sect. 4, we show that this projection converges to the $L_2(\mu)$ -projection of the action of the Koopman operator. In Sect. 5, we analyze the convergence of the EDMD approximations as the dimension of the subspace N goes to infinity, showing convergence in strong operator topology and convergence of the eigenvalues along a subsequence plus weak convergence of the associated eigenfunctions. In Sect. 6, we show convergence of finite-horizon predictions. Section 7 describes the analytic construction of \mathcal{K}_N . Section 8 contains results for the case when $M = N$ and only convergence to weak eigenfunctions, along a subsequence, is proven. We conclude in Sect. 9.

Notation

The spaces of real and complex numbers are denoted by \mathbb{R} and \mathbb{C} with $\mathbb{R}^{n \times k}$ and $\mathbb{C}^{n \times k}$ denoting the corresponding real and complex $n \times k$ matrices; the real and complex column vectors are denoted by $\mathbb{R}^n := \mathbb{R}^{n \times 1}$ and $\mathbb{C}^n := \mathbb{C}^{n \times 1}$. The complex conjugate of $a \in \mathbb{C}$ is denoted by \bar{a} . Given a matrix $A \in \mathbb{C}^{n \times k}$, A^\top denotes its transpose and A^H denotes its Hermitian transpose (i.e., $A_{i,j}^\top = A_{j,i}$ and $A_{i,j}^H = \overline{A_{j,i}}$). The Moore–Penrose pseudoinverse of a matrix A is denoted by A^\dagger . The Frobenius norm of a matrix A is denoted by $\|A\|_F = \sqrt{\sum_{i,j} |A_{i,j}|^2}$. Given a vector $c \in \mathbb{C}^n$, the symbol $\|c\|_2 := \sqrt{\sum_i |c_i|^2}$ denotes its Euclidean norm.

2 Extended Dynamic Mode Decomposition

We consider a discrete time dynamical system

$$x^+ = T(x) \tag{1}$$

with $T: \mathcal{M} \rightarrow \mathcal{M}$, where \mathcal{M} is a topological space,¹ and we assume that we are given snapshots of data

$$X = [x_1, \dots, x_M], \quad Y = [y_1, \dots, y_M] \tag{2}$$

satisfying $y_i = T(x_i)$. We do not assume that the data points line on a single trajectory of (1).

Given a vector space of observables \mathcal{F} such that $\psi: \mathcal{M} \rightarrow \mathbb{C}$ and $\psi \circ T \in \mathcal{F}$ for every $\psi \in \mathcal{F}$, we define the Koopman $\mathcal{K}: \mathcal{F} \rightarrow \mathcal{F}$ by

$$\mathcal{K}\psi = \psi \circ T,$$

where \circ denotes the pointwise function composition. Given a set of linearly independent basis functions $\psi_i \in \mathcal{F}, i = 1, \dots, N$, and defining

$$\mathcal{F}_N := \text{span} \{ \psi_1, \dots, \psi_N \}, \tag{3}$$

the EDMD constructs a finite-dimensional approximation $\mathcal{K}_{N,M}: \mathcal{F}_N \rightarrow \mathcal{F}_N$ of the Koopman operator by solving the least-squares problem

$$\min_{A \in \mathbb{C}^{N \times N}} \|A\psi(X) - \psi(Y)\|_F^2 = \min_{A \in \mathbb{C}^{N \times N}} \sum_{i=1}^M \|A\psi(x_i) - \psi(y_i)\|_2^2, \tag{4}$$

¹ We choose to work in the general setting of dynamical systems on arbitrary topological spaces which encompass dynamical systems on finite-dimensional manifolds (in which case one can regard \mathcal{M} as a subset of \mathbb{R}^n), as well as infinite-dimensional dynamical systems, arising, for example, from the study of partial differential equations or dynamical systems with control inputs (Korda and Mezić 2016).

where

$$\boldsymbol{\psi}(X) = [\boldsymbol{\psi}(x_1), \dots, \boldsymbol{\psi}(x_M)], \quad \boldsymbol{\psi}(Y) = [\boldsymbol{\psi}(y_1), \dots, \boldsymbol{\psi}(y_M)]$$

and

$$\boldsymbol{\psi}(x) = [\psi_1(x), \dots, \psi_N(x)]^\top.$$

Denoting

$$A_{N,M} = \boldsymbol{\psi}(Y)\boldsymbol{\psi}(X)^\dagger, \tag{5}$$

a solution² to (4), the finite-dimensional approximation of the Koopman operator

$$\mathcal{K}_{N,M}: \mathcal{F}_N \rightarrow \mathcal{F}_N$$

is then defined by

$$\mathcal{K}_{N,M}\phi = c_\phi^H A_{N,M}\boldsymbol{\psi} \tag{6}$$

for any $\phi = c_\phi^H \boldsymbol{\psi}$, $c_\phi \in \mathbb{C}^N$ (i.e., for any $\phi \in \mathcal{F}_N$). The operator $\mathcal{K}_{N,M}$ will be referred to as the EDMD operator.

3 EDMD as L_2 Projection

To the best of our knowledge, the results of this section and Sect. 4 were first obtained in Klus et al. (2016, Section 3.4) and hinted at already in Williams et al. (2015). Here, we rephrase these results in a form more suitable for our purposes.

From here on, we assume³ that $\mathcal{F} = L_2(\mu)$, where μ is a given positive measure on \mathcal{M} . This assumption in particular implies that the basis functions ψ_i in (3) belong to $L_2(\mu)$ and hence \mathcal{F}_N is a closed (but not necessarily invariant) subspace of $L_2(\mu)$. Note that the measure μ is *not* required to be invariant for (1) and hence the Koopman operator \mathcal{K} is not necessarily unitary. In practical applications, the measure μ will typically be the uniform measure on \mathcal{M} or other measure from which samples can be drawn efficiently.

We recall that given an arbitrary positive measure ν on \mathcal{M} , the space $L_2(\nu)$ is the Hilbert space of all measurable functions $\phi: \mathcal{M} \rightarrow \mathbb{C}$ satisfying

$$\|\phi\|_{L_2(\nu)} := \sqrt{\int_{\mathcal{M}} |\phi(x)|^2 d\nu(x)} < \infty.$$

Assuming \mathcal{F}_N is a closed subspace of $L_2(\nu)$, the $L_2(\nu)$ -projection of a function $\phi \in L_2(\nu)$ onto $\mathcal{F}_N \subset L_2(\nu)$ is defined by

² In general, the solution to (4) may not be unique; however, the matrix $A_{N,M} = \boldsymbol{\psi}(Y)\boldsymbol{\psi}(X)^\dagger$, where \cdot^\dagger denotes the Moore–Penrose pseudoinverse, is always uniquely defined and $A_{N,M}$ is always a minimizer in (4).

³ Since $\mathcal{K}: \mathcal{F} \rightarrow \mathcal{F}$, the assumption of $\mathcal{F} = L_2(\mu)$ implies that the composition relation $\phi \circ T, \phi \in L_2(\mu)$, gives rise to a well-defined operator from $L_2(\mu)$ to $L_2(\mu)$. In particular, this implies that $\|\phi_1 \circ T - \phi_2 \circ T\|_{L_2(\mu)} = 0$ whenever $\|\phi_1 - \phi_2\|_{L_2(\mu)} = 0$ and that $\int_{\mathcal{M}} |\phi \circ T|^2 d\mu < \infty$ for all $\phi \in L_2(\mu)$.

$$P_N^v \phi = \arg \min_{f \in \mathcal{F}_N} \|f - \phi\|_{L_2(v)} = \arg \min_{f \in \mathcal{F}_N} \int_{\mathcal{M}} |f - \phi|^2 dv = \arg \min_{c \in \mathbb{C}^N} \int_{\mathcal{M}} |c^H \psi - \phi|^2 dv. \tag{7}$$

Now, given the data points x_1, \dots, x_M from (2), we define the empirical measure $\hat{\mu}_M$ by

$$\hat{\mu}_M = \frac{1}{M} \sum_{i=1}^M \delta_{x_i}, \tag{8}$$

where δ_{x_i} is the Dirac measure at x_i . In particular, the integral of a function ϕ with respect to $\hat{\mu}_M$ is given by

$$\int_{\mathcal{M}} \phi(x) d\hat{\mu}_M(x) = \frac{1}{M} \sum_{i=1}^M \phi(x_i).$$

We remark that the EDMD subspace \mathcal{F}_N defined in (3) is a closed subspace of both $L_2(\hat{\mu}_M)$ and $L_2(\mu)$ and hence the projections P_N^μ and $P_N^{\hat{\mu}_M}$ are well defined.

Now, we are ready to state the following characterization of $\mathcal{K}_{N,M}\phi$:

Theorem 1 *Let $\hat{\mu}_M$ denote the empirical measure (8) associated to the sample points x_1, \dots, x_M and assume that the $N \times N$ matrix*

$$M_{\hat{\mu}_M} = \frac{1}{M} \sum_{i=1}^M \psi(x_i) \psi(x_i)^H = \int_{\mathcal{M}} \psi \psi^H d\hat{\mu}_M \tag{9}$$

is invertible. Then, for any $\phi \in \mathcal{F}_N$

$$\mathcal{K}_{N,M}\phi = P_N^{\hat{\mu}_M} \mathcal{K}\phi = \arg \min_{f \in \mathcal{F}_N} \|f - \mathcal{K}\phi\|_{L_2(\hat{\mu}_M)}, \tag{10}$$

i.e.,

$$\mathcal{K}_{N,M} = P_N^{\hat{\mu}_M} \mathcal{K}|_{\mathcal{F}_N}, \tag{11}$$

where $\mathcal{K}|_{\mathcal{F}_N}: \mathcal{F}_N \rightarrow \mathcal{F}$ is the restriction of the Koopman operator to \mathcal{F}_N .

Proof Since the matrix $M_{\hat{\mu}_M}$ is invertible, the least-squares problem (4) has a unique solution given by

$$a_i = \left(\sum_{j=1}^M \psi(x_j) \psi(x_j)^H \right)^{-1} \sum_{j=1}^M \psi(x_j) \overline{\psi_i(y_j)},$$

where $a_i^H \in \mathbb{C}^{1 \times N}$ is the i th row of $A_{N,M}$. Therefore,

$$A_{N,M}^H = \left(\sum_{j=1}^M \psi(x_j) \psi(x_j)^H \right)^{-1} \sum_{j=1}^M \psi(x_j) \psi(y_j)^H.$$

On the other hand, analyzing the minimization problem on the right-hand side of (10), we get for any $\phi = c^H \psi$

$$\arg \min_{f \in \mathcal{F}_N} \|f - \mathcal{K}\phi\|_{L_2(\hat{\mu}_M)} = \arg \min_{c \in \mathbb{C}^N} \frac{1}{M} \sum_{i=1}^M \left(c^H \psi(x_i) - c_\phi^H \psi(y_i) \right)^2$$

with the unique minimizer (since the minimized functions is strictly convex in c)

$$c = \left(\sum_{j=1}^M \psi(x_j) \psi(x_j)^H \right)^{-1} \sum_{j=1}^M \psi(x_j) \psi(y_j)^H c_\phi = A_{N,M}^H c_\phi.$$

Hence, $\arg \min_{f \in \mathcal{F}_N} \|f - \mathcal{K}\phi\|_{L_2(\hat{\mu}_M)} = c^H \psi = c_\phi^H A_{N,M} \psi = \mathcal{K}_{N,M} \phi$ as desired, where we used (6) in the last equality. □

Theorem 1 says that for any function $\phi \in \mathcal{F}_N$, the EDMD operator $\mathcal{K}_{N,M}$ computes the $L_2(\hat{\mu}_M)$ -orthogonal projection of $\mathcal{K}\phi$ on the subspace spanned by ψ_1, \dots, ψ_N .

Remark 1 If the assumption that the matrix $M_{\hat{\mu}_M}$ is invertible does not hold, then the solution to the projection problem on the right-hand side of (10) may not be unique.⁴ The action of the EDMD operator $\mathcal{K}_{N,M}$ [which is defined uniquely by (5) and (6)] then selects one solution to the projection problem. In concrete terms, we have

$$\mathcal{K}_{N,M} \phi \in \text{Arg min}_{f \in \mathcal{F}_N} \|f - \mathcal{K}\phi\|_{L_2(\hat{\mu}_M)},$$

where $\text{Arg min}_{f \in \mathcal{F}_N} \|f - \mathcal{K}\phi\|_{L_2(\hat{\mu}_M)}$ denotes the set of all minimizers of $\|f - \mathcal{K}\phi\|_{L_2(\hat{\mu}_M)}$ among $f \in \mathcal{F}_N$.

4 Convergence of $\mathcal{K}_{N,M}$ as $M \rightarrow \infty$

The first step in understanding convergence of EDMD is to understand the convergence of $\mathcal{K}_{N,M}$ as the number of samples M tends to infinity. In this section, we prove that

$$\mathcal{K}_{N,M} \rightarrow \mathcal{K}_N,$$

where

$$\mathcal{K}_N = P_N^\mu \mathcal{K}_{|\mathcal{F}_N}, \tag{12}$$

provided that the samples x_1, \dots, x_M are drawn independently from a given probability distribution μ (e.g., uniform distribution for compact \mathcal{M} or Gaussian for $\mathcal{M} = \mathbb{R}^n$).

⁴ To be more specific, if the matrix $M_{\hat{\mu}_M}$ is not invertible, the solution to (10) may not be unique when viewed as a member of $L_2(\mu)$. When viewed as a member of $L_2(\hat{\mu}_M)$, the solution is unique (since it is a projection onto a closed subspace of a Hilbert space). This is because in this case two functions from \mathcal{F}_N belonging to distinct $L_2(\mu)$ equivalence classes may fall into the same $L_2(\hat{\mu}_M)$ equivalence class.

Assumption 1 (*μ independence*) The basis functions ψ_1, \dots, ψ_N are such that

$$\mu \left\{ x \in \mathcal{M} \mid c^H \boldsymbol{\psi}(x) = 0 \right\} = 0$$

for all nonzero $c \in \mathbb{C}^N$.

This is a natural assumption ensuring that the measure μ is not supported on a zero level set of a linear combination of the basis functions used. It is satisfied if μ is any measure with the support equal to \mathcal{M} in conjunction with the majority of most commonly used basis functions such as polynomials and radial basis functions with unbounded support (e.g., Gaussian, thin plate splines). This assumption in particular implies that the matrix $M_{\hat{\mu}_M}$ defined in (9) is invertible with probability one for $M \geq N$ if x_j 's are iid samples from μ .

Lemma 1 *If Assumption 1 holds, then for any $\phi \in \mathcal{F}$ we have with probability one*

$$\lim_{M \rightarrow \infty} \| P_N^{\hat{\mu}_M} \phi - P_N^\mu \phi \| = 0, \tag{13}$$

where $\| \cdot \|$ is any norm on \mathcal{F}_N (which are all equivalent since \mathcal{F}_N is finite dimensional).

Proof We have

$$\begin{aligned} P_N^\mu \phi &= \arg \min_{f \in \mathcal{F}_N} \int_{\mathcal{M}} |f - \phi|^2 d\mu = \boldsymbol{\psi}^H \arg \min_{c \in \mathbb{C}^N} \int_{\mathcal{M}} |c^H \boldsymbol{\psi} - \phi|^2 d\mu \\ &= \boldsymbol{\psi}^H \arg \min_{c \in \mathbb{C}^N} \left[c^H M_\mu c - 2 \operatorname{Re} \{ c^H b_{\mu, \phi} \} \right], \end{aligned}$$

where

$$M_\mu = \int_{\mathcal{M}} \boldsymbol{\psi} \boldsymbol{\psi}^H d\mu \in \mathbb{C}^{N \times N}, \quad b_{\mu, \phi} = \int_{\mathcal{M}} \boldsymbol{\psi} \bar{\phi} d\mu \in \mathbb{C}^N$$

and we dropped the constant term in the last equality which does not influence the minimizer. By Assumption 1, the matrix M_μ is invertible and hence Hermitian positive definite. Therefore, the unique minimizer is $c = M_\mu^{-1} b_{\mu, \phi}$. Hence,

$$P_N^\mu \phi = b_{\mu, \phi}^H M_\mu^{-1} \boldsymbol{\psi}.$$

On the other hand, the same computation shows that

$$P_N^{\hat{\mu}_M} \phi = b_{\hat{\mu}_M, \phi}^H M_{\hat{\mu}_M}^{-1} \boldsymbol{\psi}$$

with

$$b_{\hat{\mu}_M, \phi} = \int_{\mathcal{M}} \boldsymbol{\psi} \bar{\phi} d\hat{\mu}_M = \frac{1}{M} \sum_{i=1}^M \boldsymbol{\psi}(x_i) \overline{\phi(x_i)}$$

and with the matrix $M_{\hat{\mu}_M}$ defined in (9) guaranteed to be Hermitian positive definite by Assumption 1 with probability one for $M \geq N$. The result then follows by the strong law of large numbers which ensures that

$$\lim_{M \rightarrow \infty} \left(b_{\hat{\mu}_M, \phi}^H M_{\hat{\mu}_M}^{-1} \right) = b_{\mu, \phi}^H M_{\mu}^{-1}$$

with probability one since the matrix function $A \mapsto A^{-1}$ is continuous and the samples x_i are iid. □

Theorem 2 *If Assumption 1 holds, then we have with probability one for all $\phi \in \mathcal{F}_N$*

$$\lim_{M \rightarrow \infty} \|\mathcal{K}_{N,M}\phi - \mathcal{K}_N\phi\| = 0, \tag{14}$$

where $\|\cdot\|$ is any norm on \mathcal{F}_N . In particular

$$\lim_{M \rightarrow \infty} \|\mathcal{K}_{N,M} - \mathcal{K}_N\| = 0, \tag{15}$$

where $\|\cdot\|$ is any operator norm and

$$\lim_{M \rightarrow \infty} \text{dist}(\sigma(\mathcal{K}_{N,M}), \sigma(\mathcal{K}_N)) = 0, \tag{16}$$

where $\sigma(\cdot) \subset \mathbb{C}$ denotes the spectrum of an operator and $\text{dist}(\cdot, \cdot)$ the Hausdorff metric on subsets of \mathbb{C} .

Proof For any fixed $\phi \in \mathcal{F}_N$, we have by Theorem 1

$$\mathcal{K}_{N,M}\phi = P_N^{\hat{\mu}_M} \mathcal{K}\phi = P_N^{\hat{\mu}_M} (\phi \circ T).$$

By definition of $\mathcal{K}_N\phi$, we have $\mathcal{K}_N\phi = P_N^{\mu} (\phi \circ T)$ and therefore (14) holds from Lemma 1 with probability one. Since \mathcal{F}_N is finite dimensional, (14) holds with probability one for all basis functions of \mathcal{F}_N and hence by linearity for all $\phi \in \mathcal{F}_N$. Convergence in the operator norm (15) and spectral convergence (16) follows from (14) since the operators $\mathcal{K}_{N,M}$ and \mathcal{K}_N are finite dimensional. □

Theorem 2 tells us that in order to understand the convergence of $\mathcal{K}_{N,M}$ to \mathcal{K} it is sufficient to understand the convergence of \mathcal{K}_N to \mathcal{K} . This convergence is analyzed in Sect. 5.

4.1 Ergodic Sampling

The assumption that the samples x_1, \dots, x_M are drawn independently from the distribution μ can be replaced by the assumption that (T, \mathcal{M}, μ) is ergodic and the samples x_1, \dots, x_M are the iterates of the dynamical system starting from some initial condition $x \in \mathcal{M}$, i.e., $x_i = T^i(x)$. Provided that Assumption 1 holds, both Lemma 1

and Theorem 2 hold without change; the statement “with probability one” is now interpreted with respect to drawing the initial condition x from the distribution μ . The proofs follow exactly the same argument; only the strong law of large numbers is replaced by the Birkhoff’s ergodic theorem in Lemma 1.

5 Convergence of \mathcal{K}_N to \mathcal{K}

In this section, we investigate convergence of \mathcal{K}_N to \mathcal{K} in the limit as N goes to infinity. Since the operator \mathcal{K}_N is defined on \mathcal{F}_N rather than \mathcal{F} , we extend the operator to all of \mathcal{F} by precomposing with P_N^μ , i.e., we study the convergence of $\mathcal{K}_N P_N^\mu = P_N^\mu \mathcal{K} P_N^\mu: \mathcal{F} \rightarrow \mathcal{F}$ to $\mathcal{K}: \mathcal{F} \rightarrow \mathcal{F}$ as $N \rightarrow \infty$. Note that as far as spectrum is concerned, precomposing with P_N^μ just adds a zero to the spectrum.

To simplify notation, in what follows we denote the $L_2(\mu)$ norm of a function f by

$$\|f\| := \|f\|_{L_2(\mu)} = \sqrt{\int_{\mathcal{M}} |f|^2 d\mu}$$

and the usual $L_2(\mu)$ inner product by

$$\langle f, g \rangle := \int_{\mathcal{M}} f \bar{g} d\mu.$$

5.1 Preliminaries

Before stating our results, we recall several concepts from functional analysis and operator theory.

Definition 1 (*Bounded operator*) An operator $\mathcal{A}: \mathcal{F} \rightarrow \mathcal{F}$ defined on a Hilbert space \mathcal{F} is bounded if

$$\|\mathcal{A}\| := \sup_{f \in \mathcal{F}, \|f\|=1} \|\mathcal{A}f\| < \infty.$$

The quantity $\|\mathcal{A}\|$ is referred as the norm of \mathcal{A} .

Definition 2 (*Convergence in strong operator topology*) A sequence of bounded operators $\mathcal{A}_i: \mathcal{F} \rightarrow \mathcal{F}$ defined on a Hilbert space \mathcal{F} converges strongly (or in the strong operator topology) to an operator $\mathcal{A}: \mathcal{F} \rightarrow \mathcal{F}$ if

$$\lim_{i \rightarrow \infty} \|\mathcal{A}_i g - \mathcal{A}g\| \tag{17}$$

for all $g \in \mathcal{F}$.

Definition 3 (*Weak convergence*) A sequence of elements $f_i \in \mathcal{F}$ of a Hilbert space \mathcal{F} converges weakly to $f \in \mathcal{F}$, denoted $f_i \xrightarrow{w} f$, if

$$\lim_{i \rightarrow \infty} \langle f_i, g \rangle = \langle f, g \rangle \tag{18}$$

for all $g \in \mathcal{F}$.

We emphasize that Definition 2 pertains to convergence of operators defined on \mathcal{F} , whereas Definition 3 pertains to convergence of elements of \mathcal{F} . We also remark that strong convergence of $f_i \rightarrow f$ (i.e., $\|f_i - f\| \rightarrow 0$) implies weak convergence, but not vice versa. Similarly, convergence in the strong operator topology implies convergence in the weak operator topology (i.e., $\mathcal{A}_i g \xrightarrow{w} \mathcal{A}g$ for all g), but does not imply convergence in the operator norm (i.e., $\|\mathcal{A}_i - \mathcal{A}\| \rightarrow 0$).

In our setting of $\mathcal{F} = L_2(\mu)$, the statements (17) and (18) translate to the requirements that, respectively,

$$\lim_{i \rightarrow \infty} \sqrt{\int_{\mathcal{M}} |\mathcal{A}_i g - \mathcal{A}g|^2 d\mu} = 0 \quad \text{and} \quad \lim_{i \rightarrow \infty} \int_{\mathcal{M}} f_i \bar{g} d\mu = \int_{\mathcal{M}} f \bar{g} d\mu$$

for all $g \in L_2(\mu)$.

For the remainder of this work, we invoke the following assumption:

Assumption 2 The following conditions hold:

1. The Koopman operator $\mathcal{K}: \mathcal{F} \rightarrow \mathcal{F}$ is bounded.
2. The observables ψ_1, \dots, ψ_N defining \mathcal{F}_N are selected from a given orthonormal basis of \mathcal{F} , i.e., $(\psi_i)_{i=1}^\infty$ is an orthonormal basis of \mathcal{F} .

The first part of the assumption holds for instance when T is invertible, Lipschitz with Lipschitz inverse and μ is the Lebesgue measure on \mathcal{M} (or any measure absolutely continuous w.r.t. the Lebesgue measure with bounded and strictly positive density). The second part of the assumption is nonrestrictive as any countable dense subset of \mathcal{F} can be turned into an orthonormal basis using the Gram–Schmidt process.

5.2 Convergence in Strong Operator Topology

In this section, we prove convergence in the strong operator topology (Definition 2) of $\mathcal{K}_N P_N^\mu$ to \mathcal{K} . First, we need the following immediate lemma:

Lemma 2 *If $(\psi_i)_{i=1}^\infty$ form an orthonormal basis of $\mathcal{F} = L_2(\mu)$, then P_N^μ converge strongly to the identity operator I and in addition $\|I - P_N\| \leq 1$ for all N .*

Proof Let $\phi = \sum_{i=1}^\infty c_i \psi_i$ with $\|\phi\| = 1$. Then, by Parseval’s identity $\sum_{i=1}^\infty |c_i|^2 = 1$ and

$$\|P_N^\mu \phi - \phi\| = \left\| \sum_{i=N+1}^\infty c_i \psi_i \right\| = \sqrt{\sum_{i=N+1}^\infty |c_i|^2} \rightarrow 0$$

with $\sum_{i=N+1}^\infty |c_i|^2 \leq 1$ for all N . □

Now, we are ready to prove strong convergence of $P_N^\mu \mathcal{K} P_N^\mu$ to \mathcal{K} .

Theorem 3 *If Assumption 2 holds, then the sequence of operators $\mathcal{K}_N P_N^\mu = P_N^\mu \mathcal{K} P_N^\mu$ converges strongly to \mathcal{K} as $N \rightarrow \infty$, i.e.,*

$$\lim_{N \rightarrow \infty} \int_{\mathcal{M}} |\mathcal{K}_N P_N^\mu \phi - \mathcal{K} \phi|^2 d\mu = 0$$

for all $\phi \in \mathcal{F}$.

Proof Let $\phi \in \mathcal{F}$ be given. Then, writing $\phi = P_N^\mu \phi + (I - P_N^\mu) \phi$ we have

$$\begin{aligned} \|P_N^\mu \mathcal{K} P_N^\mu \phi - \mathcal{K} \phi\| &= \|(P_N^\mu - I) \mathcal{K} P_N^\mu \phi + \mathcal{K} (P_N^\mu - I) \phi\| \\ &\leq \|(P_N^\mu - I) \mathcal{K} P_N^\mu \phi\| + \|\mathcal{K}\| \|(I - P_N^\mu) \phi\| \\ &\leq \|(P_N^\mu - I) \mathcal{K} \phi\| + \|(P_N^\mu - I)\| \|\mathcal{K} P_N^\mu \phi - \mathcal{K} \phi\| \\ &\quad + \|\mathcal{K}\| \|(I - P_N^\mu) \phi\| \rightarrow 0 \end{aligned}$$

by Lemma 2 and by the fact that $\mathcal{K} P_N^\mu \phi \rightarrow \mathcal{K} \phi$ since \mathcal{K} is continuous by Assumption 2. □

5.3 Weak Spectral Convergence

Unfortunately, strong converge does not in general guarantee convergence of the spectra of the operators. This is guaranteed only if the operators converge in the operator norm.⁵ An important exception to this is the case of \mathcal{F}_N being an invariant subspace, i.e., $\mathcal{K} f \in \mathcal{F}_N$ for all $f \in \mathcal{F}_N$ in which case the spectra of \mathcal{K}_N and $\mathcal{K}|_{\mathcal{F}_N}$ coincide. Here, however, we *do not assume* that \mathcal{F}_N is invariant and prove certain spectral convergence results in a weak sense. In particular, we prove convergence of the eigenvalues of \mathcal{K}_N along a subsequence and weak convergence of the associated eigenfunctions (see Definition 3), provided that the weak limit of the eigenfunctions is nonzero.

Theorem 4 *If Assumption 2 holds and λ_N is a sequence of eigenvalues of \mathcal{K}_N with the associated normalized eigenfunctions $\phi_N \in \mathcal{F}_N$, $\|\phi_N\| = 1$, then there exists a subsequence $(\lambda_{N_i}, \phi_{N_i})$ such that*

$$\lim_{i \rightarrow \infty} \lambda_{N_i} = \lambda, \quad \phi_{N_i} \xrightarrow{w} \phi,$$

where $\lambda \in \mathbb{C}$ and $\phi \in \mathcal{F}$ are such that $\mathcal{K} \phi = \lambda \phi$. In particular, if $\|\phi\| \neq 0$, then λ is an eigenvalue of \mathcal{K} with eigenfunction ϕ .

Proof First, observe that since $\mathcal{K}_N \phi_N = \lambda_N \phi_N$ with $\phi_N \in \mathcal{F}_N$, we also have $P_N^\mu \mathcal{K} P_N^\mu \phi_N = \lambda_N \phi_N$. Hence, $|\lambda_N| \leq \|P_N^\mu \mathcal{K} P_N^\mu\| \leq \|\mathcal{K}\| < \infty$ by Assumption 2 and the fact that $\|P_N^\mu\| \leq 1$. Therefore, the sequence λ_N is bounded. Since ϕ_N is

⁵ A sequence of operators \mathcal{A}_i converges in the operator norm to an operator \mathcal{A} if $\lim_{i \rightarrow \infty} \|\mathcal{A}_i - \mathcal{A}\| = 0$.

normalized and hence bounded, by weak sequential compactness of the unit ball of a Hilbert space [which follows from the Banach–Alaoglu theorem (Rudin 1973, Theorems 3.15) and Eberlein–Šmulian theorem (Dunford and Schwartz 1971)], there exists a subsequence $(\lambda_{N_i}, \phi_{N_i})$ such that $\lambda_{N_i} \rightarrow \lambda$ and $\phi_{N_i} \xrightarrow{w} \phi$.

It remains to prove that (λ, ϕ) is an eigenvalue–eigenfunction pair of \mathcal{K} . For ease of notation, set $\lambda_i = \lambda_{N_i}$ and $\phi_i = \phi_{N_i}$. Denote $\hat{\mathcal{K}}_i = \mathcal{K}_{N_i} P_{N_i}^\mu = P_{N_i}^\mu \mathcal{K} P_{N_i}^\mu$ and observe that $\hat{\mathcal{K}}_i \phi_i = \lambda_i \phi_i$ for all i . Then, we have

$$\mathcal{K}\phi = \hat{\mathcal{K}}_i(\phi - \phi_i) + (\mathcal{K} - \hat{\mathcal{K}}_i)\phi + \hat{\mathcal{K}}_i\phi_i.$$

Taking the inner product with an arbitrary $f \in \mathcal{F}$ and using the fact that $\hat{\mathcal{K}}_i\phi_i = \lambda_i\phi_i$, we get

$$\langle \mathcal{K}\phi, f \rangle = \langle \hat{\mathcal{K}}_i(\phi - \phi_i), f \rangle + \langle (\mathcal{K} - \hat{\mathcal{K}}_i)\phi, f \rangle + \langle \lambda_i\phi_i, f \rangle.$$

Now, the second term on the right-hand side $\langle (\mathcal{K} - \hat{\mathcal{K}}_i)\phi, f \rangle \rightarrow 0$ since $\hat{\mathcal{K}}_i$ converges strongly to \mathcal{K} by Theorem 3. The last term $\langle \lambda_i\phi_i, f \rangle \rightarrow \langle \lambda\phi, f \rangle$ since $\lambda_i \rightarrow \lambda$ and $\phi_i \xrightarrow{w} \phi$. It remains to show that the first term converges to zero. We have

$$\langle \hat{\mathcal{K}}_i(\phi - \phi_i), f \rangle = \langle P_{N_i}^\mu \mathcal{K} P_{N_i}^\mu (\phi - \phi_i), f \rangle = \langle \mathcal{K}(P_{N_i}^\mu \phi - \phi_i), P_{N_i}^\mu f \rangle,$$

where we used the fact that $P_{N_i}^\mu$ is self-adjoint and $\phi_i \in \mathcal{F}_{N_i}$ and hence $P_{N_i}^\mu \phi_i = \phi_i$. Denote $h_i := \mathcal{K}(P_{N_i}^\mu \phi - \phi_i)$. We will show that $h_i \xrightarrow{w} 0$. Indeed, denoting \mathcal{K}^* the adjoint of \mathcal{K} , we have

$$\begin{aligned} \langle \mathcal{K}(P_{N_i}^\mu \phi - \phi_i), f \rangle &= \langle (P_{N_i}^\mu \phi - \phi + \phi - \phi_i), \mathcal{K}^* f \rangle = \langle P_{N_i}^\mu \phi - \phi, \mathcal{K}^* f \rangle \\ &\quad + \langle \phi - \phi_i, \mathcal{K}^* f \rangle \rightarrow 0, \end{aligned}$$

since $P_{N_i}^\mu$ converges strongly to the identity (Lemma 2) and $\phi_i \xrightarrow{w} \phi$. Finally, we show that $\langle h_i, P_{N_i}^\mu f \rangle \rightarrow 0$. We have

$$\langle h_i, P_{N_i}^\mu f \rangle = \langle h_i, P_{N_i}^\mu f - f \rangle + \langle h_i, f \rangle.$$

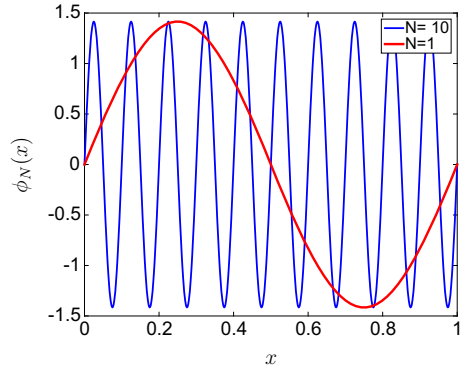
The second term goes to zero since $h_i \xrightarrow{w} 0$. For the first term, we have

$$\langle h_i, P_{N_i}^\mu f - f \rangle \leq \|h_i\| \|P_{N_i}^\mu f - f\| \rightarrow 0$$

since $P_{N_i}^\mu$ converges strongly to the identity operator (Lemma 2) and h_i is bounded since \mathcal{K} is bounded by Assumption 2, $\|P_{N_i}^\mu\| \leq 1$ and $\|\phi_i\| \leq 1$. Therefore, we conclude that

$$\langle \mathcal{K}\phi, f \rangle = \lim_{i \rightarrow \infty} \langle \lambda_i\phi_i, f \rangle = \langle \lambda\phi, f \rangle$$

Fig. 1 Graph of the functions $\phi_N(x) = \sqrt{2} \sin(2\pi Nx)$ for $N = 1$ and $N = 10$. These functions satisfy $\|\phi_N\|_{L_2} = 1$ and $\phi_N \xrightarrow{w} 0$ as $N \rightarrow \infty$. If such ϕ_N happen to be eigenfunctions of \mathcal{K}_N with eigenvalues λ_N , then the accumulation points of the sequence $(\lambda_N)_{N=1}^\infty$ need not be eigenvalues of the Koopman operator



for all $f \in \mathcal{F}$. Therefore, $\mathcal{K}\phi = \lambda\phi$. □

Example As an example demonstrating that the assumption that the weak limit of ϕ_N is nonzero is important, consider $\mathcal{M} = [0, 1]$, $T(x) = x$ and μ the Lebesgue measure on $[0, 1]$. In this setting, the Koopman operator $\mathcal{K}: L_2(\mu) \rightarrow L_2(\mu)$ is the identity operator with the spectrum being the singleton $\sigma(\mathcal{K}) = \{1\}$. However, given any $\lambda \in \mathbb{C}$ and the sequence of functions $\phi_N = \sqrt{2} \sin(2\pi Nx)$, we have

$$\mathcal{K}\phi_N - \lambda\phi_N = \phi_N - \lambda\phi_N = (1 - \lambda)\sqrt{2} \sin(2\pi Nx) \xrightarrow{w} 0$$

with $\|\phi_N\|^2 = \int_0^1 2 \sin^2(2\pi Nx) dx = 1$. Therefore, if ϕ_N were the eigenfunctions of \mathcal{K}_N with eigenvalues $\lambda_N \rightarrow \lambda \neq 1$, then the sequence λ_N would converge to a spurious eigenvalue λ . Fortunately, in this case, we have $\sigma(\mathcal{K}_N) = \{1\}$ and hence no spurious eigenvalues exist; however, in general, we cannot rule out this possibility, at least not as far as the statement of Theorem 4 goes. See Fig. 1 for illustration.

This example, with the highly oscillatory functions ϕ_N , may motivate practical considerations in detecting spurious eigenvalues, e.g., using Sobolev-type (pseudo) norms $\int_{\mathcal{M}} \|\nabla\phi_N\|^2 d\mu$ or other metrics of oscillatoriness. See e.g., Giannakis (2016) for the use of Sobolev norms in the context of Koopman data analysis and forecasting.

As an immediate corollary of Theorem 4, we get:

Corollary 1 *If Assumption 2 holds and $\lambda_{N,M}$ is a sequence of eigenvalues of $\mathcal{K}_{N,M}$ with the associated normalized eigenfunctions $\phi_{N,M} \in \mathcal{F}_N$, $\|\phi_{N,M}\| = 1$, then there exists a subsequence $(\lambda_{N_i,M_j}, \phi_{N_i,M_j})$ such that with probability one*

$$\lim_{i \rightarrow \infty} \lim_{j \rightarrow \infty} \lambda_{N_i,M_j} = \lambda, \quad \lim_{i \rightarrow \infty} \lim_{j \rightarrow \infty} \langle \phi_{N_i,M_j}, f \rangle = \langle \phi, f \rangle,$$

for all $f \in \mathcal{F}$, where $\lambda \in \mathbb{C}$ and $\phi \in \mathcal{F}$ are such that $\mathcal{K}\phi = \lambda\phi$. In particular, if $\|\phi\| \neq 0$, then λ is an eigenvalue of \mathcal{K} with eigenfunction ϕ .

Proof First notice that since $\mathcal{K}_{N,M} \rightarrow \mathcal{K}_N$ in the operator norm (Theorem 2) and $\|\mathcal{K}_N\| \leq \|\mathcal{K}\| < \infty$, the sequence $\lambda_{N,M}$ is bounded. Since $\phi_{N,M}$ are normalized and hence bounded, we can extract a subsequence $(\lambda_{N,M_j}, \phi_{N,M_j})$ such that

$\lim_{j \rightarrow \infty} \lambda_{N,M_j} = \lambda_N \in \mathbb{C}$ and $\lim_{j \rightarrow \infty} \phi_{N,M_j} = \phi_N \in \mathcal{F}_N$ (strong convergence as \mathcal{F}_N is finite dimensional). Then,

$$\mathcal{K}_N \phi_N = (\mathcal{K}_N - \mathcal{K}_{N,M_j})\phi_N + \mathcal{K}_{N,M_j}(\phi_N - \phi_{N,M_j}) + \mathcal{K}_{N,M_j}\phi_{N,M_j}.$$

Since \mathcal{K}_{N,M_j} converges strongly to \mathcal{K}_N with probability one (Theorem 2) and since ϕ_{N,M_j} converges strongly to ϕ_N , the first two terms go to zero with probability one as j tends to infinity. The last term is equal to $\lambda_{N,M_j}\phi_{N,M_j}$ and hence necessarily $\mathcal{K}\phi_N = \lambda_N\phi_N$, $\|\phi_N\| = 1$, with probability one. The result then follows from Theorem 4. \square

6 Implications for Finite-Horizon Predictions

One of the main roles of an approximation to the Koopman operator is to provide a *prediction* of the evolution of a given observable, whereas obtaining accurate predictions over an infinite-time horizon cannot be expected in general from the EDMD approximation of the Koopman operator; a prediction over any *finite horizon* is asymptotically, as $N \rightarrow \infty$, exact when the prediction error is measured in the $L_2(\mu)$ norm:

Theorem 5 *Let $f \in \mathcal{F}^n$ be a given (vector) observable⁶ and let Assumption 2 hold. Then, for any $\Omega \in \mathbb{N}$ we have*

$$\lim_{N \rightarrow \infty} \sup_{i \in \{1, \dots, \Omega\}} \|(\mathcal{K}_N)^i P_N^\mu f - \mathcal{K}^i f\| = 0. \tag{19}$$

In particular, if $f \in \mathcal{F}_{N_0}^n$ for some $N_0 \in \mathbb{N}$, then

$$\lim_{N \rightarrow \infty} \sup_{i \in \{1, \dots, \Omega\}} \|(\mathcal{K}_N)^i f - \mathcal{K}^i f\| = 0. \tag{20}$$

Proof We proceed by induction. Let $f \in \mathcal{F}$. For $\Omega = 1$, the result is exactly Theorem 3. Let the result hold for some $\Omega \in \mathbb{N}$. It is sufficient to prove that $\|(\mathcal{K}_N)^{\Omega+1} P_N^\mu f - \mathcal{K}^{\Omega+1} f\| \rightarrow 0$ as $N \rightarrow \infty$. We have

$$\begin{aligned} \|(\mathcal{K}_N)^{\Omega+1} P_N^\mu f - \mathcal{K}^{\Omega+1} f\| &= \|\mathcal{K}_N(\mathcal{K}_N)^\Omega P_N^\mu f - \mathcal{K}\mathcal{K}^\Omega f\| = \|\mathcal{K}_N g_N - \mathcal{K}g\| \\ &\leq \|\mathcal{K}_N g - \mathcal{K}g\| + \|\mathcal{K}_N(g_N - g)\| \\ &\leq \|\mathcal{K}_N g - \mathcal{K}g\| + \|\mathcal{K}\| \|g_N - g\|. \end{aligned}$$

where $g_N = (\mathcal{K}_N)^\Omega P_N^\mu f$ and $g = \mathcal{K}^\Omega f$; in the last inequality we used the fact that $\|\mathcal{K}_N\| \leq \|K\|$. The term $\|\mathcal{K}_N g - \mathcal{K}g\|$ tends to zero by Theorem 3, whereas the term $\|g_N - g\| \rightarrow 0$ by the induction hypothesis. This proves (19) for a scalar observable $f \in \mathcal{F}$. The general result with a vector valued observable $f \in \mathcal{F}^n$ follows

⁶ We choose to state the theorem for vector observables as this is the form of prediction typically encountered in practice. For a vector observable $f \in \mathcal{F}^n$, the norm $\|f\|$ is defined by $\sum_{i=1}^n \|f_i\|_{L_2(\mu)}$, where $f_i \in \mathcal{F}$ is the i^{th} component of f .

by applying the same reasoning to each component of f . The result (20) follows from (19) since if $f \in \mathcal{F}_{N_0}^n$ for some $N_0 \in \mathbb{N}$, then $P_N^\mu f = f$ for $N \geq N_0$. \square

To be more specific on practical use of \mathcal{K}_N for prediction, assume that $f \in \mathcal{F}_{N_0}^n$ for some $N_0 \in \mathbb{N}$. Then, for all $N \geq N_0$ there exists a matrix $C_N \in \mathbb{R}^{n \times N}$ such that $f = C_N \psi_N$, where $\psi_N = [\psi_1, \dots, \psi_N]^\top$ are the observables used in EDMD. Assume that an initial state x_0 is given and the values of the observables $\psi_N(x_0)$ are known and we wish to predict the value of the observable f at a state $x_i = T^i(x_0)$, i.e., i steps ahead in the future. Using \mathcal{K}_N , this prediction is given by $C_N A_N^i \psi_N(x_0)$, where

$$A_N = \lim_{M \rightarrow \infty} A_{N,M}$$

with $A_{N,M}$ defined⁷ in (5). Theorem 5 then says that

$$\lim_{N \rightarrow \infty} \int_{\mathcal{M}} \|C_N A_N^i \psi_N - f \circ T^i\|_2^2 d\mu = 0 \quad \forall i \in \mathbb{N}. \tag{21}$$

A typical application of Theorem 5 is the prediction of the future state x of the dynamical system (1) with a finite-dimensional state space $\mathcal{M} \subset \mathbb{R}^n$. In this case, one simply sets $f(x) = x$. A crucial feature of the predictor obtained in this way is its *linearity* in the “lifted state” $z = \psi_N(x)$, allowing linear tools to address a nonlinear problem. This concept was successfully applied to model predictive control in [Korda and Mezić \(2016\)](#) and to state estimation in [Surana and Banaszuk \(2016\)](#).

Remark 2 If $A_{N,M}$ is used instead of A_N in Theorem 5 and Eq. (21), then the same convergence results hold with a double limit, first taking the number of samples M to infinity and then the number of basis functions N . In particular, we get

$$\lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \int_{\mathcal{M}} \|C_N A_{N,M}^i \psi_N - f \circ T^i\|_2^2 d\mu = 0 \quad \forall i \in \mathbb{N}. \tag{22}$$

7 Analytic EDMD

The results of the previous sections suggest a variation of the EDMD algorithm provided that the mapping T is known in closed form and provided that the basis functions ψ_i are such that the integrals of $\int_{\mathcal{M}} \psi_i \overline{\psi_j} d\mu$ and $\int_{\mathcal{M}} (\psi_i \circ T) \overline{\psi_j} d\mu$ can be computed analytically. This is the case in particular if T and ψ_i are simple functions such as multivariate polynomials or trigonometric functions and μ is the uniform distribution over a simple domain \mathcal{M} such as a box or a ball, or, e.g., a Gaussian distribution over \mathbb{R}^n .

Provided that such analytical evaluation is possible, one can circumvent the sampling step of EDMD and construct directly \mathcal{K}_N rather than $\mathcal{K}_{N,M}$. Indeed, define

$$A_N = M_{T,\mu} M_\mu^{-1}, \tag{23}$$

⁷ In Sect. 7, we show how the matrix A_N can be constructed analytically.

where

$$M_\mu = \int_{\mathcal{M}} \psi \psi^H d\mu, \quad M_{T,\mu} = \int_{\mathcal{M}} (\psi \circ T) \psi^H d\mu.$$

Then, the operator from \mathcal{F}_N to \mathcal{F}_N defined by $c^H \psi \mapsto c^H A_N \psi$ is exactly $\mathcal{K}_N = P_N^\mu \mathcal{K}|_{\mathcal{F}_N}$.

Theorem 6 *If the matrix M_μ is invertible, then for any $\phi = c_\phi^H \psi \in \mathcal{F}_N$ we have*

$$c_\phi^H A_N \psi = \mathcal{K}_N \phi.$$

Proof Given $\phi = c_\phi^H \psi$, we get

$$\begin{aligned} \mathcal{K}_N \phi &= P_N^\mu \mathcal{K} \phi = \psi^H \arg \min_{c \in \mathbb{C}^N} \int_{\mathcal{M}} [c^H \psi - c_\phi^H (\psi \circ T)]^2 d\mu \\ &= \psi^H \arg \min_{c \in \mathbb{C}^N} [c^H M_\mu c - 2 \operatorname{Re} \{c^H M_{T,\mu}^H c_\phi\}] \end{aligned}$$

with the unique minimizer $c = M_\mu^{-1} M_{T,\mu}^H c_\phi$ (since M_μ is invertible, therefore Hermitian positive definite, and hence the minimized function is strictly convex). Therefore, as desired

$$\mathcal{K}_N \phi = c^H \psi = c_\phi^H M_{T,\mu} M_\mu^{-1} \psi = c_\phi^H A_N \psi.$$

□

Example In order to demonstrate the use of Analytic DMD, we compare the spectra of \mathcal{K}_N and $\mathcal{K}_{N,M}$ for various values of M . The system considered is the logistic map

$$x^+ = 2x^2 - 1, \quad x \in [-1, 1].$$

The measure μ is taken to be the uniform distribution on $[-1, 1]$, which is not invariant and hence the dynamics is not measure preserving. The finite-dimensional subspace \mathcal{F}_N is the space of all polynomials of degree no more than eight. For numerical stability, we chose a basis of this subspace to be the Laguerre polynomials scaled such that they are orthonormal with respect to the uniform measure on $[-1, 1]$. Spectra of \mathcal{K}_N and $\mathcal{K}_{N,M}$ for $M = 10^2$, $M = 10^3$ and $M = 10^5$ are depicted in Fig. 2. We observe that a relatively large number of samples M are required to obtain an accurate approximation of the spectrum of \mathcal{K}_N . This example demonstrates that a special care must be taken in practice when drawing conclusions about spectral quantities based on a computation with a small number of samples. On the other hand, Fig. 3 suggests that, at least on this example, predictions generated by $\mathcal{K}_{N,M}$ are less affected by sampling. Indeed, even for $M = 100$ the prediction accuracy of $\mathcal{K}_{N,M}$ is comparable to that of \mathcal{K}_N and for $M = 1000$ the two predictions almost coincide.

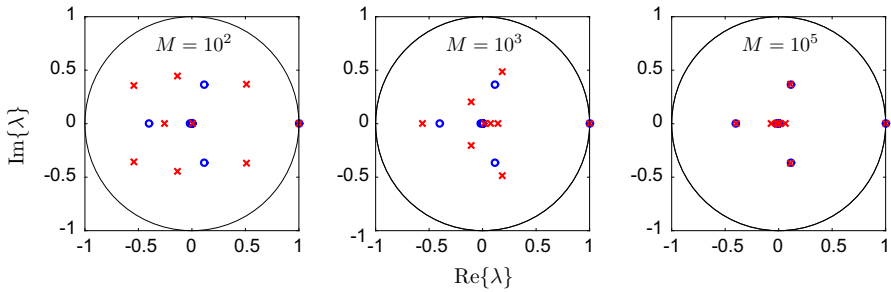
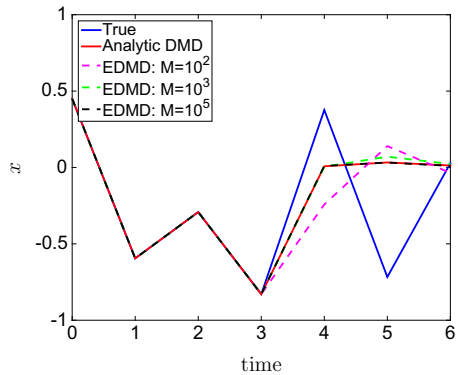


Fig. 2 Comparison of the spectra of \mathcal{K}_N (blue circles) computed using Analytic DMD and the spectra of $\mathcal{K}_{N,M}$ for different values of M (red crosses) (Color figure online)

Fig. 3 Comparison of predictions generated using \mathcal{K}_N and using $\mathcal{K}_{N,M}$ for different values of M



8 Convergence of $\mathcal{K}_{N,N}$

In this section, we investigate what happens when we simultaneously increase the number of basis functions N and the number of samples M . We treat the special case of $M = N$ for which interesting conclusions can be drawn. Set therefore $M = N$ and denote $\lambda_N = \lambda_{N,N}$ any eigenvalue of $\mathcal{K}_{N,N}$ and $\phi_N = \phi_{N,N} \in \mathcal{F}_N$, $\|\phi_N\|_{C(\mathcal{M})} = 1$, the associated eigenfunction, where $\|\phi\|_{C(\mathcal{M})} = \sup_{x \in \mathcal{M}} |\phi(x)|$; such normalization is possible if the basis functions ψ_i are continuous and \mathcal{M} compact, which we assume in this section. First notice that, assuming $M_{\hat{\mu}_N}$ invertible, for $N = M$ the system of equations

$$\psi(Y) = A\psi(X)$$

with the unknown $A \in \mathbb{R}^{N \times N}$ has a solution and hence the minimum in the least-squares problem (4) is zero. In other words, for any $f \in \mathcal{F}_N$, the EDMD operator $\mathcal{K}_{N,N}$ applied to f matches the value of the Koopman operator applied to f on the samples points x_1, \dots, x_N :

$$(\mathcal{K}f)(x_i) = (\mathcal{K}_{N,N}f)(x_i)$$

for all $f \in \mathcal{F}_N$. This relation is in particular satisfied for the eigenfunctions ϕ_N of $\mathcal{K}_{N,N}$, obtaining

$$(\phi_N \circ T)(x_i) = \lambda_N \phi_N(x_i).$$

Multiplying by an arbitrary $h \in \mathcal{F}$ and integrating with respect to the empirical measure supported on the sample points (8), we get

$$\int_{\mathcal{M}} h \cdot (\phi_N \circ T) \, d\hat{\mu}_N = \lambda_N \int_{\mathcal{M}} h \phi_N \, d\hat{\mu}_N. \tag{24}$$

Define the linear functional $L_N: C(\mathcal{M}) \rightarrow \mathbb{C}$ by

$$L_N(h) = \int_{\mathcal{M}} h \phi_N \, d\hat{\mu}_N,$$

and

$$(\mathcal{K}L_N)(h) = \int_{\mathcal{M}} h \cdot (\phi_N \circ T) \, d\hat{\mu}_N.$$

With this notation, the relationship (24) becomes

$$\mathcal{K}L_N = \lambda_N L_N.$$

Since $\|\phi_N\|_{C(\mathcal{M})} = 1$, we have $\|L_N\| = \sup_{h \in C(\mathcal{M})} \frac{|L_N(h)|}{\|h\|_{C(\mathcal{M})}} \leq 1$ and $\|\mathcal{K}L_N\| \leq 1$. Therefore, assuming separability⁸ of $C(\mathcal{M})$, by the Banach–Alaoglu theorem (e.g., Rudin 1973, Theorems 3.15, 3.17) there exists a subsequence, along which these functionals converge in the weak* topology⁹ to some functionals $L \in C(\mathcal{M})^*$ and $\mathcal{K}L \in C(\mathcal{M})^*$ satisfying

$$\mathcal{K}L = \lambda L,$$

where λ is an accumulation point of λ_N . Furthermore, by the Riesz representation theorem the bounded linear functionals L and $\mathcal{K}L$ can be represented by complex-valued measures ν and $\mathcal{K}\nu$ on \mathcal{M} satisfying

$$\mathcal{K}\nu = \lambda \nu.$$

We remark that $\mathcal{K}L$ and $\mathcal{K}\nu$ are here merely symbols for the weak* limit of $\mathcal{K}L_N$ and its representation as a measure; in particular, the functional $\mathcal{K}L$ is not necessarily of the form $(\mathcal{K}L)(h) = \int_{\mathcal{M}} h \cdot (\rho \circ T) \, d\mu$ for some function ρ .

⁸ A sufficient condition for $C(\mathcal{M})$ to be separable is \mathcal{M} compact and metrizable.

⁹ A sequence of functionals $L_i \in C(\mathcal{M})^*$ converges in the weak* topology if $\lim_{i \rightarrow \infty} L_i(f) = L(f)$ for all $f \in C(\mathcal{M})$.

In order to get more understanding of \mathcal{K}_v (and hence \mathcal{KL}), we need to impose additional assumptions on the structure of the problem. In particular, we assume that the mapping $T: \mathcal{M} \rightarrow \mathcal{M}$ is a homeomorphism and that the points x_1, \dots, x_N lie on a single trajectory, i.e., $x_{i+1} = T(x_i)$. With this assumption, Eq. (24) reads

$$\frac{1}{N} \sum_{i=1}^N h(x_i)\phi_N(x_{i+1}) = \lambda_N \frac{1}{N} \sum_{i=1}^N h(x_i)\phi_N(x_i), \tag{25}$$

where we set $x_{N+1} := T(x_N)$. The left-hand side of (25) is

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N h(x_i)\phi_N(x_{i+1}) &= \frac{1}{N} \sum_{i=1}^N h(T^{-1}x_i)\phi_N(x_i) + \frac{1}{N} [h(x_N)\phi_N(x_{N+1}) \\ &\quad - h(T^{-1}x_1)\phi_N(x_1)] \\ &= \int_{\mathcal{M}} h \circ T^{-1} \, d\nu_N + \frac{1}{N} [h(x_N)\phi_N(x_{N+1}) - h(T^{-1}x_1)\phi_N(x_1)], \end{aligned} \tag{26}$$

where ν_N is the measure $\phi_N d\hat{\mu}_N$. Setting $\xi_N := h(x_N)\phi_N(x_{N+1}) - h(T^{-1}x_1)\phi_N(x_1)$, the relation (25) becomes

$$\int_{\mathcal{M}} h \circ T^{-1} \, d\nu_N + \frac{1}{N} \xi_N = \lambda_N \int_{\mathcal{M}} h \, d\nu_N.$$

Since h is bounded on \mathcal{M} (h is continuous and \mathcal{M} compact) and $\|\phi_N\|_{C(\mathcal{M})} = 1$, the term ξ_N is bounded; in addition $h \circ T^{-1}$ is continuous since T is a homeomorphism by assumption. Therefore, taking a limit on both sides, along a subsequence such that $\nu_{N_i} \rightarrow \nu$ weakly,¹⁰ $\hat{\mu}_{N_i} \rightarrow \mu$ weakly and $\lambda_{N_i} \rightarrow \lambda$, we obtain

$$\int_{\mathcal{M}} h \circ T^{-1} \, d\nu = \lambda \int_{\mathcal{M}} h \, d\nu \tag{27}$$

for all $h \in C(\mathcal{M})$.

8.1 Weak Eigenfunctions/Eigendistributions

To understand relation (27), note that a completely analogous computation to (26) shows that the measure μ is invariant¹¹ and therefore the $L_2(\mu)$ -adjoint \mathcal{K}^* of the Koopman operator [viewed as an operator from $L_2(\mu)$ to $L_2(\mu)$] is given by

¹⁰ A sequence of Borel measures μ_i converges weakly to a measure μ if $\lim_{i \rightarrow \infty} \int f \, d\mu_i = \int f \, d\mu$ for all continuous bounded functions f . This convergence is also referred to as narrow convergence and it coincides with convergence in the weak* topology if the underlying space is compact (which is the case in our setting).

¹¹ A measure μ on \mathcal{M} is invariant if $\mu(T^{-1}(A)) = \mu(A)$ for all Borel sets $A \subset \mathcal{M}$ or equivalently if $\int_{\mathcal{M}} f \circ T \, d\mu = \int_{\mathcal{M}} f \, d\mu$ for all continuous bounded functions f .

$$\mathcal{K}^* f = f \circ T^{-1}.$$

To see this, write

$$\begin{aligned} \langle \mathcal{K}f, g \rangle &= \int_{\mathcal{M}} (f \circ T)\bar{g} \, d\mu = \int_{\mathcal{M}} (f \circ T)\overline{(g \circ T^{-1} \circ T)} \, d\mu \\ &= \int_{\mathcal{M}} f \cdot \overline{(g \circ T^{-1})} \, d\mu = \langle f, \mathcal{K}^*g \rangle, \end{aligned}$$

which means the operator $g \mapsto g \circ T^{-1}$ is indeed the $L_2(\mu)$ -adjoint of \mathcal{K} . The relation (27) then becomes

$$\int_{\mathcal{M}} \mathcal{K}^*h \, d\nu = \lambda \int_{\mathcal{M}} h \, d\nu \tag{28}$$

or

$$L(\mathcal{K}^*h) = \lambda L(h). \tag{29}$$

Functionals of the form (29) were called “generalized eigenfunctions” by Gelfand and Shilov (1964); here, we prefer to call them “weak eigenfunctions” or “eigendistributions” in order to avoid confusion with generalized eigenfunctions viewed as an extension of the notion of generalized eigenvectors from linear algebra. The measure ν in (28) is then called “eigenmeasure.” Here, again, we emphasize the requirement that the limiting functional L (or the measure ν) be nonzero in order for these objects to be called eigenfunctionals/eigenmeasures.

8.2 Eigenmeasures of Perron–Frobenius

We also observe an interesting connection to eigenmeasures of the Perron–Frobenius operator. To see this, set $h := g \circ T$ in (27) to obtain $\int_{\mathcal{M}} g \, d\nu = \lambda \int_{\mathcal{M}} g \circ T \, d\nu$ or, provided that $\lambda \neq 0$,

$$\int_{\mathcal{M}} g \circ T \, d\nu = \frac{1}{\lambda} \int_{\mathcal{M}} g \, d\nu. \tag{30}$$

In other words, if nonzero, the measure ν is the eigenmeasure of the Perron–Frobenius operator with eigenvalue $1/\lambda$. Here, the Perron–Frobenius operator $\mathcal{P}: M(\mathcal{M}) \rightarrow M(\mathcal{M})$, where $M(\mathcal{M})$ is the space of all complex-valued measures on \mathcal{M} , is defined for every $\eta \in M(\mathcal{M})$ and every Borel set A by

$$(\mathcal{P}\eta)(A) = \eta(T^{-1}(A)).$$

The results of Sect. 8 are summarized in the following theorem:

Theorem 7 *Suppose that \mathcal{M} is a compact metric space, T is a homeomorphism, $\mathcal{K}: L_2(\mu) \rightarrow L_2(\mu)$ is bounded, the observables ψ_1, \dots, ψ_N are continuous and the sample points x_1, \dots, x_N satisfy $x_{i+1} = T(x_i)$ for all $i \in \{1, \dots, N - 1\}$. Let λ_N be a bounded sequence of eigenvalues of $\mathcal{K}_{N,N}$, let $\phi_N, \|\phi_N\|_{C(\mathcal{M})} = 1$, be the*

associated normalized eigenfunctions and denote $v_N = \phi_N d\hat{\mu}_N$. Then, there exists a subsequence $(N_i)_{i=1}^\infty$ such that v_{N_i} and $\hat{\mu}_{N_i}$ converge weakly to complex-valued measures $v \in M(\mathcal{M})$, $\mu \in M(\mathcal{M})$ and $\lim_{i \rightarrow \infty} \lambda_{N_i} = \lambda \in \mathbb{C}$ such that

$$\int_{\mathcal{M}} h \circ T^{-1} dv = \lambda \int_{\mathcal{M}} h dv \quad \forall h \in C(\mathcal{M}).$$

In addition, the measure μ is invariant under the action of T and

$$\int_{\mathcal{M}} \mathcal{K}^* h dv = \lambda \int_{\mathcal{M}} h dv \quad \forall h \in C(\mathcal{M}),$$

where \mathcal{K}^* is the $L_2(\mu)$ adjoint of \mathcal{K} , i.e., if nonzero, v is a weak eigenfunction (or eigendistribution) of the Koopman operator. Furthermore, if $\lambda \neq 0$, then

$$\int_{\mathcal{M}} h \circ T dv = \frac{1}{\lambda} \int_{\mathcal{M}} h dv \quad \forall h \in C(\mathcal{M}),$$

i.e., if nonzero, v is an eigenmeasure of the Perron–Frobenius operator with eigenvalue $1/\lambda$.

9 Conclusions

This paper analyzes the convergence of the EDMD operator $\mathcal{K}_{N,M}$, where M is the number of samples and N the number of observables used in EDMD. It was proven in [Klus et al. \(2016\)](#) that as $M \rightarrow \infty$, the operator $\mathcal{K}_{N,M}$ converges to \mathcal{K}_N , the orthogonal projection of the action of the Koopman operator on the span of the observables used in EDMD. We analyzed the convergence of \mathcal{K}_N as $N \rightarrow \infty$, obtaining convergence in strong operator topology to the Koopman operator and weak convergence of the associated eigenfunctions along a subsequence together with the associated eigenvalues. In particular, any accumulation point of the spectra of \mathcal{K}_N corresponding to a nonzero weak accumulation point of the eigenfunctions lies in the point spectrum of the Koopman operator \mathcal{K} . In addition, we proved convergence of finite-horizon predictions obtained using \mathcal{K}_N in the L_2 norm, a result important for practical applications such as forecasting, estimation and control. Finally, we analyzed convergence of $\mathcal{K}_{N,N}$ (i.e., the situation where the number of samples and the number of basis functions is equal) under the assumptions that the sample points lie on the same trajectory. In this case, one obtains convergence, along a subsequence, to a weak eigenfunction (or eigendistribution) of the Koopman operator, provided that the weak limit is nonzero. This eigendistribution turns out to be also an eigenmeasure of the Perron–Frobenius operator. As a by-product of these results, we proposed an algorithm that, under some assumptions, allows one to construct \mathcal{K}_N directly, without the need for sampling, thereby eliminating the sampling error.

Future work should focus on non-asymptotic analysis, e.g., on selecting the subspace \mathcal{F}_N such that $\|\mathcal{K}_N - \mathcal{K}|_{\mathcal{F}_N}\|$ is minimized and at the same time such that \mathcal{F}_N is rich enough in the sense of containing observables of practical interest (e.g., the state

observable). This line of research was already investigated in the context of stochastic systems in [Wu and Noé \(2017\)](#), providing an interesting and actionable method for selecting \mathcal{F}_N .

Acknowledgements The first author would like to thank Mihai Putinar for discussions on a related topic that sparked the work on this paper. We also thank Clancy Rowley for helpful comments on an earlier version of this manuscript as well as to Péter Koltai for pointing out the reference ([Klus et al. 2016](#)). This research was supported in part by the ARO-MURI Grant W911NF-14-1-0359 and the DARPA Grant HR0011-16-C-0116. The research of M. Korda was supported by the Swiss National Science Foundation under Grant P2ELP2_165166.

References

- Arbabi, H., Mezić, I.: Ergodic Theory, Dynamic Mode Decomposition and Computation of Spectral Properties of the Koopman Operator (2016). arXiv preprint [arXiv:1611.06664](#)
- 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 (2016a)
- Brunton, S.L., Brunton, B.W., Proctor, J.L., Kutz, J.N.: Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS ONE* **11**(2), e0150171 (2016b)
- Budisić, M., Mohr, R., Mezić, I.: Applied koopmanism. *Chaos Interdiscip. J. Nonlinear Sci.* **22**(4), 047–510 (2012)
- Dunford, N., Schwartz, J.T.: *Linear Operators, Part I*. Wiley-interscience, New York (1971)
- Gelfand, I.M., Shilov, G.: *Generalized Functions: Properties and Operations*, vol. 1. Academic Press, New York (1964)
- Georgescu, M., Mezić, I.: Building energy modeling: a systematic approach to zoning and model reduction using Koopman mode analysis. *Energy Build.* **86**, 794–802 (2015)
- Giannakis, D.: Data-Driven Spectral Decomposition and Forecasting of Ergodic Dynamical Systems (2016). arXiv preprint [arXiv:1507.02338](#)
- Giannakis, D., Slawinska, J., Zhao, Z.: Spatiotemporal feature extraction with data-driven Koopman operators. In: *Proceedings of the 1st International Workshop on Feature Extraction: Modern Questions and Challenges, NIPS*, pp. 103–115 (2015)
- Glaz, B., Mezić, I., Fonoberova, M., Loire, S.: Quasi-Periodic Intermittency in Oscillating Cylinder Flow (2016). arXiv preprint [arXiv:1609.06267](#)
- Klus, S., Koltai, P., Schütte, C.: On the numerical approximation of the Perron–Frobenius and Koopman operator. *J. Comput. Dyn.* **3**(1), 51–79 (2016)
- Koopman, B.O.: Hamiltonian systems and transformation in Hilbert space. *Proc. Natl. Acad. Sci. USA.* **17**(5), 315 (1931)
- Korda, M., Mezić, I.: Linear Predictors for Nonlinear Dynamical Systems: Koopman Operator Meets Model Predictive Control (2016). arXiv preprint [arXiv:1611.03537](#)
- Mauroy, A., Goncalves, J.: Koopman-Based Lifting Techniques for Nonlinear Systems Identification (2017). arXiv preprint [arXiv:1709.02003](#)
- Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**(1–3), 309–325 (2005)
- Mezić, I.: Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45**, 357–378 (2013)
- Mezić, I., Banaszuk, A.: Comparison of systems with complex behavior. *Phys. D Nonlinear Phenom.* **197**(1), 101–133 (2004)
- Raak, F., Susuki, Y., Mezić, I., Hikiyara, T.: On Koopman and dynamic mode decompositions for application to dynamic data with low spatial dimension. In: *IEEE 55th Conference on Decision and Control (CDC)*, pp. 6485–6491 (2016)
- Riseth, A.N., Taylor-King, J.P.: Operator Fitting for Parameter Estimation of Stochastic Differential Equations (2017). arXiv preprint [arXiv:1709.05153](#)
- Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.: Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**(1), 115–127 (2009)

- Rudin, W.: *Functional Analysis*. McGraw-Hill, Inc., New York (1973)
- Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28 (2010)
- Sharma, A.S., Mezić, I., McKeon, B.J.: On the correspondence between Koopman mode decomposition, resolvent mode decomposition, and invariant solutions of the Navier–Stokes equations. *Phys. Rev. Fluids* **1**(3), 032402(R) (2016)
- Surana, A., Banaszuk, A.: Linear observer synthesis for nonlinear systems using Koopman operator framework. In: *IFAC Symposium on Nonlinear Control Systems (NOLCOS)* (2016)
- Takeishi, N., Kawahara, Y., Yairi, T.: Subspace Dynamic Mode Decomposition for Stochastic Koopman Analysis (2017). arXiv preprint [arXiv:1705.04908](https://arxiv.org/abs/1705.04908)
- 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**, 391–421 (2014)
- 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)
- Wu, H., Noé, F.: Variational Approach for Learning Markov Processes from Time Series Data (2017). arXiv preprint [arXiv:1707.04659](https://arxiv.org/abs/1707.04659)
- 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)