SIAM J. NUMER. ANAL.
Vol. 61, No. 3, pp. 1585–1608

THE mpEDMD ALGORITHM FOR DATA-DRIVEN COMPUTATIONS OF MEASURE-PRESERVING DYNAMICAL SYSTEMS*

MATTHEW J. COLBROOK†

Abstract. Koopman operators globally linearize nonlinear dynamical systems and their spectral information is a powerful tool for the analysis and decomposition of nonlinear dynamical systems. However, Koopman operators are infinite dimensional, and computing their spectral information is a considerable challenge. We introduce *measure-preserving extended dynamic mode decomposition* (`mpEDMD`), the first Galerkin method whose eigendecomposition converges to the spectral quantities of Koopman operators for general measure-preserving dynamical systems. `mpEDMD` is a data-driven algorithm based on an orthogonal Procrustes problem that enforces measure-preserving truncations of Koopman operators using a general dictionary of observables. It is flexible and easy to use with any preexisting dynamic mode decomposition (DMD)-type method, and with different types of data. We prove convergence of `mpEDMD` for projection-valued and scalar-valued spectral measures, spectra, and Koopman mode decompositions. For the case of delay embedding (Krylov subspaces), our results include the first convergence rates of the approximation of spectral measures as the size of the dictionary increases. We demonstrate `mpEDMD` on a range of challenging examples, its increased robustness to noise compared to other DMD-type methods, and its ability to capture the energy conservation and cascade of a turbulent boundary layer flow with Reynolds number $> 6 \times 10^4$ and state-space dimension $> 10^5$.

Key words. computational spectral problem, infinite dimensions, structure-preserving algorithms, dynamical systems, Koopman operator, dynamic mode decomposition

MSC codes. 65J10, 65P99, 47N40, 47A10, 47B33, 37M10, 37A05, 37N10

DOI. 10.1137/22M1521407

1. Introduction.

We consider dynamical systems whose state $\boldsymbol{x}$ evolves over a state space $\Omega \subseteq \mathbb{R}^d$ in discrete time steps according to a function $F : \Omega \to \Omega$, i.e.,

$$\boldsymbol{x}_{n+1} = F(\boldsymbol{x}_n), \qquad n \geq 0, \tag{1.1}$$

for an initial condition $\boldsymbol{x}_0 \in \Omega$. Throughout this paper, we assume that (1.1) is *measure-preserving* (also called volume-preserving) with respect to a positive measure ω on Ω. Measure preserving means that $\omega(E) = \omega(\{\boldsymbol{x} : F(\boldsymbol{x}) \in E\})$ for any measurable set $E \subset \Omega$. This assumption covers many systems of interest such as Hamiltonian flows [2], geodesic flows [31], Bernoulli schemes [83], physical systems in equilibrium [43], and ergodic systems [90]. Moreover, many dynamical systems admit invariant measures [56] or have measure-preserving posttransient behavior [66].

In many modern applications, the system's dynamics are too complicated to describe analytically, or we only have access to incomplete knowledge of its evolution. Therefore, we do not assume explicit knowledge of the function F in (1.1). Instead, we assume that we have access to discrete-time snapshots of this system, i.e., a dataset

$$\{\boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)}\}_{m=1}^{M} \quad \text{such that} \quad \boldsymbol{y}^{(m)} = F(\boldsymbol{x}^{(m)}), \quad m = 1, \dots, M. \tag{1.2}$$

*Received by the editors September 12, 2022; accepted for publication (in revised form) March 6, 2023; published electronically June 16, 2023.
https://doi.org/10.1137/22M1521407

Funding: This work was funded by a FSMP Fellowship at École Normale Supérieure.

†DAMTP, University of Cambridge, Cambridge, CB3 0WA, England (m.colbrook@damtp.cam.ac.uk).

Suitable data could be collected from one long trajectory, corresponding to $\boldsymbol{x}^{(m)} = F^{m-1}(\boldsymbol{x}_0)$ (i.e., $m-1$ applications of F), or from multiple shorter trajectories. The data could come from experimental observations or numerical simulations, both of which occur in this paper. We approximate quantities of (1.1) using the data (1.2). With the arrival of big data and machine learning, this data-driven viewpoint is currently undergoing a renaissance [48, 80, 13, 14, 16, 95, 37].

We introduce a new Galerkin discretization that maintains the measure-preserving nature of (1.1) and allows us to prove convergence to the spectral quantities of (1.1). Convergence of spectral quantities is crucial for recovering the correct dynamical behavior. Preserving the measure is crucial for stability, and improved qualitative and long-time behavior (see Figures 7 and 8). Moreover, the fact that our discretization preserves a measure is a key factor in proving convergence.

1.1. Koopman operators. Koopman operators are a widely used and powerful tool for the data-driven study of dynamical systems. First introduced by Koopman and von Neumann in the 1930's [52, 53], Koopman operators allow a global linearization of (1.1) using the space of scalar functions on Ω [70]. Their increasing popularity, known as "Koopmanism" [15], has led to thousands of articles over the last decade [12]. Popular applications include epidemiology [76], finance [63], fluid dynamics [79, 78, 67], neuroscience [10], molecular dynamics [50, 82], and robotics [6, 9].

Since (1.1) is measure preserving, its Koopman operator, $\mathcal{K}$, is defined by

$$[\mathcal{K}g](\boldsymbol{x}) = (g \circ F)(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega, \qquad g \in L^2(\Omega, \omega), \tag{1.3}$$

and is an isometry on the space $L^2(\Omega, \omega)$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$ [77]. The functions g are also called "observables" because they indirectly measure the state of the dynamical system. The Koopman operator transforms the nonlinear dynamics in the state variable $\boldsymbol{x}$ into equivalent linear dynamics in the observables g. Hence, the behavior of the dynamical system (1.1) is determined by the spectral information of $\mathcal{K}$. Obtaining linear representations has the potential to revolutionize our ability to predict and control nonlinear systems.

However, there is a price to pay for this linearization—$\mathcal{K}$ acts on an *infinite-dimensional* space. Therefore, its spectral information can be far more complicated and more difficult to compute than that of a finite matrix [91, 20, 5, 24, 22]. Common challenges include: computing spectral measures and continuous spectra [11, 61, 25]; spectral pollution [92], where discretizations cause spurious eigenvalues (and hence spurious coherent structures) to appear [58, 26]; and, in the context of this paper, preserving the isometric nature of $\mathcal{K}$. This last issue is often critical to ensuring that approximations retain the physical properties of the original system (e.g., energy conservation). Structure-preserving algorithms have a rich history in geometric integration [38] and have recently come to the fold in data-driven problems [42, 18, 49, 36, 41].

1.2. Existing work. Most existing approaches to approximate $\mathcal{K}$ and its spectral properties are based on dynamic mode decomposition (DMD) [79, 78, 88, 57] or its variants [19, 45, 75, 27, 23]. DMD approximates $\mathcal{K}$ via a best-fit linear model of (1.1) that advances spatial measurements from one time step to the next. However, DMD is based on linear observables, i.e., linear functions g in (1.3), which are not rich enough for many nonlinear systems. To overcome this, [92] introduced extended DMD (EDMD), a Galerkin approximation of $\mathcal{K}$ acting on a dictionary of nonlinear observables (see section 3). As the number of snapshots, M, increases, the eigenvalues computed by EDMD correspond to the so-called finite section method [8]. These

eigenvalues can suffer from spectral pollution [92]—persistent spurious modes that result from the approximation of infinite-dimensional dynamics in a chosen finite-dimensional computational basis. Moreover, as the dictionary becomes richer, the so-called spectral measures of EDMD, which generalize the notion of spectral projections, do not typically converge weakly to that of $\mathcal{K}$. Finally, although the Koopman mode decomposition (KMD) provided by EDMD converges in an appropriate sense (in contrast to DMD), it is not measure preserving, which can be a serious concern (see subsection 6.3 for a real-world example).

Recently, Baddoo et al. [4] introduced physics-informed DMD (piDMD), which enforces symmetry constraints on the DMD approximation. For conservative systems, piDMD forces the DMD matrix to be orthogonal. However, no convergence results are known for piDMD, and piDMD uses linear observables and implicitly assumes that these are orthonormal in $L^2(\Omega, \omega)$. This assumption typically does not hold and may not even be possible after re-weighting. We shall see that this issue can be overcome by working in a data-driven inner product space. The resulting Gram matrix of the observables must be included in a measure-preserving discretization, otherwise the wrong measure may be preserved.

There are several methods that are not based on the eigenvalues of a Galerkin approximation, and that approximate spectral measures. Similar to the Ulam approximation of the Perron–Frobenius operator, [34, 35] proved convergence of periodic approximations of $\mathcal{K}$ via a partitioning of the state space, and developed a numerical method for dealing with measure-preserving automorphisms on tori. In [55], the authors computed measures for ergodic systems by first computing moments and then making use of the Christoffel–Darboux kernel. Similar to DMD-type methods, this has the advantage over periodic approximations of being implementable in high dimensions. Residual DMD (ResDMD) [27, 23] provides a general computational framework to deal with continuous spectra and spectral measures through smoothed approximations of spectral measures, leading to explicit and rigorous high-order convergence. ResDMD deals jointly with discrete and continuous spectra, does not assume ergodicity, and can be applied to data collected from either short or long trajectories. Finally, the method in [29] constructs compact regularizations of the skew-adjoint generator of continuous-time, measure-preserving, ergodic systems. Though based on different techniques special to continuous-time systems, [29] is similar in spirit to the method we propose—it preserves the measure and has convergence of spectral measures.

1.3. Contributions. We introduce a new approximation of $\mathcal{K}$ that is measure preserving and that converges to the correct spectral information. Our method uses an orthogonal Procrustes problem using *general dictionaries* and *nonlinear measurements*, and we call our algorithm measure-preserving EDMD (`mpEDMD`). Table 1 compares DMD, EDMD, piDMD, and `mpEDMD`, and highlights some of the benefits of `mpEDMD`. The fact that `mpEDMD` produces a normal truncation of $\mathcal{K}$ is an important property used in the proofs of our convergence results. Our contributions include

- we introduce `mpEDMD` to deal with generic measure-preserving systems.[1] `mpEDMD` is simple and easy to use with any preexisting DMD-type method, it is measure preserving, and it can be used with a range of different data structures and acquisition methods (e.g., single trajectories or multiple trajectories);

[1]For example, we do not assume in this paper that the system is ergodic or invertible.

TABLE 1

Comparisons of Galerkin discretizations discussed in this paper. $X = [\boldsymbol{x}^{(1)} \cdots \boldsymbol{x}^{(M)}], Y = [\boldsymbol{y}^{(1)} \cdots \boldsymbol{y}^{(M)}] \in \mathbb{C}^{d\times M}$ *are matrices of the snapshots (linear dictionary) and it is common to combine DMD with a truncated SVD (see section 3).* $G = \Psi_X^* W \Psi_X$ *and* $A = \Psi_X^* W \Psi_Y$*, where* Ψ_X*,* Ψ_Y *are given in* (3.3) *and* $W = \mathrm{diag}(w_1, \dots, w_M)$ *is a diagonal matrix of quadrature weights.*

	DMD	**EDMD**	**piDMD**	**mpEDMD**
Aux. SVD matrices	n/a	n/a	$YX^* = V_1 S V_2^*$	$G^{-\frac{1}{2}} A^* G^{-\frac{1}{2}} = U_1 \Sigma U_2^*$
Koopman matrix	$(YX^\dagger)^\top$	$G^\dagger A$	$V_2 V_1^*$	$G^{-\frac{1}{2}} U_2 U_1^* G^{\frac{1}{2}}$
Nonlinear dictionary	✘	✓	✘	✓
Conv. spec. meas.	✘	✘	✘	✓
Conv. spectra	✘	✘	✘	✓
Conv. KMD	✘	✓	✘	✓
Measure preserving	✘	✘	✘/✓‡	✓

‡*Note: piDMD is measure-preserving only if* XX^* *and* W *are multiples of the identity,* I*.*

TABLE 2

Lookup table of the approximated spectral quantities using `mpEDMD` *(Algorithm* 4.1*) and the relevant convergence results of this paper. The vectors* $\{v_j\}_{j=1}^N$ *denote the eigenvectors of* $\mathbb{K}$ *with corresponding eigenvalues* $\{\lambda_j\}_{j=1}^N$*, and* δ_{λ_j} *denotes a Dirac delta distribution centered at* λ_j*.*

Spectral quantity	**Approximation**	**Convergence results**
Spec. measure $\mathcal{E}$	$\mathcal{E}_{N,M} = \sum_{j=1}^N v_j v_j^* G \delta_{\lambda_j}$	Theorems A.1 and 5.1
Spec. measures μ_g	$\mu_{\boldsymbol{g}}^{(N,M)} = \sum_{j=1}^N \lvert v_j^* G \boldsymbol{g}\rvert^2 \delta_{\lambda_j}$	Theorem 5.3, Corollaries 5.4 and 5.5
Approx. pt. spec. $\sigma_{\mathrm{ap}}(\mathcal{K})$	$\{\lambda_1, \dots, \lambda_N\}$	Theorem 5.6 and Equation (5.7)
Koop. mode decomp.	Eq. (5.2) for $g(\boldsymbol{x}_n)$.	Lemma A.2 and Remark 5.2

- we prove convergence of `mpEDMD` for various spectral quantities of interest, summarized in Table 2. Our results include weak convergence[2] of projection-valued and scalar-valued spectral measures, convergence of spectra (including spectral inclusion and ways to deal with spectral pollution), and convergence of KMDs in $L^2(\Omega, \omega)$. `mpEDMD` is the first truncation method whose eigendecomposition converges to these spectral quantities for general measure-preserving dynamical systems. Corollary 5.5 is the first result in the literature on convergence rates of the approximation of spectral measures as the size of the dictionary increases;
- we demonstrate our convergence results and the use of `mpEDMD` on several examples, including numerically simulated data and experimental data. These examples also demonstrate the increased robustness of `mpEDMD` to noise compared with other DMD-type methods, and the ability to deal with difficult problems such as capturing the energy conservation and statistics of a turbulent boundary layer flow.

1.4. Paper structure. In section 2 we introduce various concepts and notation, and motivate the computation of spectral properties of $\mathcal{K}$. Section 3 recalls the basics of EDMD. This section can be read independently and the interested reader may find the Galerkin interpretation helpful. In section 4 we introduce `mpEDMD` and state its convergence properties in section 5. Proofs can be found in Appendix A. A range of numerical examples are presented in section 6 and we conclude

[2]This means convergence after integrating against a Lipschitz continuous test function on the unit circle (where the measures are supported) [7, Chap. 1]. The computation of spectral measures poses a serious numerical challenge [25] and can only ever be done in this weak sense [21]. For example, the spectral type of $\mathcal{K}$ is well known to be sensitive to arbitrarily small perturbations.

in section 7. General purpose code for `mpEDMD` and the examples of this paper can be found at https://github.com/MColbrook/Measure-preserving-Extended-Dynamic-Mode-Decomposition.

2. Mathematical preliminaries. Here we provide the background material on spectra and spectral measures needed to understand later sections.

2.1. Spectral measures and unitary extensions of $\mathcal{K}$. Spectral measures provide a way of diagonalizing normal operators, i.e., those that commute with their adjoint. However, a Koopman operator that is an isometry does not necessarily commute with its adjoint because it may not be invertible. A famous simple example is the Koopman operator of the tent map

$$F(\boldsymbol{x}) = 2\min\{\boldsymbol{x}, 1-\boldsymbol{x}\}, \quad \Omega = [0,1].$$

If ω is the usual measure, the system is measure preserving. However, the corresponding Koopman operator, $\mathcal{K}$, cannot be unitary since any function $\mathcal{K}g$ is symmetric about $\boldsymbol{x} = 1/2$ and hence $\mathcal{K}$ is not onto.

Despite this example, a Koopman operator $\mathcal{K} : L^2(\Omega,\omega) \to L^2(\Omega,\omega)$ of a measure-preserving dynamical system has a unitary extension $\mathcal{K}'$ defined on an extended Hilbert space $\mathcal{H}'$ with $L^2(\Omega,\omega) \subset \mathcal{H}'$ [73, Chapter I]. Such an extension is not unique, but it still allows us to understand the spectral information of $\mathcal{K}$ by considering $\mathcal{K}'$, which is a normal operator. We shall see that after projecting back onto $L^2(\Omega,\omega)$, the measure is independent of the extension. If F is invertible and measure preserving, $\mathcal{K}$ is unitary and we can simply take $\mathcal{K}' = \mathcal{K}$ and $\mathcal{H}' = L^2(\Omega,\omega)$.

The spectral theorem for a normal matrix $B \in \mathbb{C}^{n\times n}$, i.e., $B^*B = BB^*$, states that there exists an orthonormal basis of eigenvectors $v_1, \dots, v_n$ for $\mathbb{C}^n$ such that

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \quad v \in \mathbb{C}^n \qquad \text{and} \qquad Bv = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \quad v \in \mathbb{C}^n, \tag{2.1}$$

where $\lambda_1, \dots, \lambda_n$ are eigenvalues of B, i.e., $Bv_k = \lambda_k v_k$ for $1 \le k \le n$. In other words, the projections $v_k v_k^*$ simultaneously decompose the space $\mathbb{C}^n$ and diagonalize the operator B. This intuition carries over to the infinite-dimensional setting of this paper, by replacing $v \in \mathbb{C}^n$ by $f \in \mathcal{H}'$, and B by a normal operator $\mathcal{K}'$. However, if $\mathcal{K}'$ has a nonempty continuous spectrum, then the eigenvectors of $\mathcal{K}'$ do not form a basis for $\mathcal{H}'$ or diagonalize $\mathcal{K}'$. Instead, the spectral theorem for normal operators states that the projections $v_k v_k^*$ in (2.1) can be replaced by a projection-valued measure $\mathcal{E}'$ supported on the spectrum of $\mathcal{K}'$ [28, Thm. X.4.11]. In our setting, $\mathcal{K}'$ is unitary and hence its spectrum is contained inside the unit circle $\mathbb{T}$. The measure $\mathcal{E}'$ assigns an orthogonal projector to each Borel measurable subset of $\mathbb{T}$ such that

$$f = \left(\int_{\mathbb{T}} d\mathcal{E}'(\lambda)\right) f \qquad \text{and} \qquad \mathcal{K}'f = \left(\int_{\mathbb{T}} \lambda\, d\mathcal{E}'(\lambda)\right) f, \qquad f \in \mathcal{H}'.$$

Analogously to (2.1), $\mathcal{E}'$ decomposes $\mathcal{H}'$ and diagonalizes the operator $\mathcal{K}'$. For example, if $U \subset \mathbb{T}$ contains only discrete eigenvalues of $\mathcal{K}'$ and no other types of spectra, then $\mathcal{E}'(U)$ is simply the spectral projector onto the invariant subspace spanned by the corresponding eigenfunctions. More generally, $\mathcal{E}'$ decomposes elements of $\mathcal{H}'$ along the discrete and continuous spectrum of $\mathcal{K}'$. An excellent and readable introduction to the spectral theorem can be found in Halmos' article [39].

PROPOSITION 2.1. *Let $\mathcal{P}$ denote the orthogonal projection from $\mathcal{H}'$ to $L^2(\Omega,\omega)$ and define $\mathcal{E} = \mathcal{P}\mathcal{E}'\mathcal{P}^*$. Then $\mathcal{E}$ is independent of the choice of unitary extension.*

Proof. For any $g, h \in L^2(\Omega, \omega)$ and Borel measurable set $U \subset \mathbb{T}$, $\langle \mathcal{E}(U)g, h\rangle = \langle \mathcal{E}'(U)g, h\rangle_{\mathcal{H}'}$. Hence, it is enough to show that the scalar-valued measures $\mu_{g,h}$, $\mu_{g,h}(U) := \langle \mathcal{E}'(U)g, h\rangle_{\mathcal{H}'}$, are independent of the choice of $\mathcal{K}'$. For $n \in \mathbb{Z}$,

$$\int_{\mathbb{T}} \lambda^n \, d\mu_{g,h}(\lambda) = \langle (\mathcal{K}')^n g, h\rangle_{\mathcal{H}'} = \begin{cases} \langle \mathcal{K}^n g, h\rangle & \text{if } n \geq 0, \\ \langle g, \mathcal{K}^{-n} h\rangle & \text{otherwise.} \end{cases}$$

Since $\mu_{g,h}$ is determined by these moments, the result follows. ∎

Proposition 2.1 shows that the choice of unitary extension is immaterial. Henceforth, we dispense with the extension $\mathcal{K}'$, and call $\mathcal{E}$ the spectral measure of $\mathcal{K}$. The approximation of $\mathcal{E}$ plays a critical role in many applications. For example, in model reduction, the approximate spectral projections provide a low-order model [71, 66]. A related example is the KMD in Remark 5.2. Furthermore, the decomposition of $\mathcal{E}$ into atomic and continuous parts often characterizes a dynamical system. For example, suppose F is measure preserving and bijective, and ω is a probability measure. Then, the dynamical system is (1) ergodic if and only if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$, (2) weakly mixing if and only if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$ and there are no other eigenvalues, and (3) mixing if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$ and $\mathcal{K}$ has an absolutely continuous spectrum on $\text{span}\{1\}^{\perp}$ [40]. Different spectral types also have interpretations in various applications such as fluid mechanics [67], anomalous transport [94], and the analysis of invariants/exponents of trajectories [48].

Given an observable $g \in L^2(\Omega, \omega)$ of interest that is normalized to have $\|g\| = 1$, the spectral measure of $\mathcal{K}$ with respect to g is a probability measure defined as $\mu_g(U) := \langle \mathcal{E}(U)g, g\rangle$, where $U \subset \mathbb{T}$ is a Borel measurable set. The proof of Proposition 2.1 shows that the moments of the measure μ_g are the correlations $\langle \mathcal{K}^n g, g\rangle$ and $\langle g, \mathcal{K}^n g\rangle$ for $n \in \mathbb{Z}_{\geq 0}$. For example, if our system corresponds to the dynamics on an attractor, these statistical properties allow comparison of complex dynamics [71]. More generally, the spectral measure of $\mathcal{K}$ with respect to almost every $g \in L^2(\Omega, \omega)$ is a signature for the forward-time dynamics of (1.1). This is because μ_g completely determines $\mathcal{K}$ when g is cyclic, i.e., when the closure of $\text{span}\{g, \mathcal{K}g, \mathcal{K}^2 g, \ldots\}$ is $L^2(\Omega, \omega)$. If g is not cyclic, then μ_g only determines the action of $\mathcal{K}$ on the closure of $\text{span}\{g, \mathcal{K}g, \mathcal{K}^2 g, \ldots\}$, which can still be useful if one is interested in particular observables. The choice of g is up to the practitioner and application.

2.2. Approximate point spectra. The spectrum of $\mathcal{K}$ is defined as

$$\sigma(\mathcal{K}) := \left\{ \lambda \in \mathbb{C} : (\mathcal{K} - \lambda)^{-1} \text{ does not exist} \right\}.$$

The spectrum includes the set of eigenvalues, but can also include points that are not eigenvalues. Since $\mathcal{K}$ is an isometry, any eigenvalue of $\mathcal{K}$ must lie in $\mathbb{T}$. The approximate point spectrum generalizes the notion of eigenvalues:

$$\sigma_{\text{ap}}(\mathcal{K}) := \left\{ \lambda \in \mathbb{C} : \exists \{g_n\}_{n \in \mathbb{N}} \subset L^2(\Omega, \omega) \text{ such that } \|g_n\| = 1, \lim_{n \to \infty} \|(\mathcal{K} - \lambda) g_n\| = 0 \right\}.$$

Any observable g with $\|g\| = 1$ and $\lambda \in \mathbb{C}$ such that $\|(\mathcal{K} - \lambda)g\| \leq \epsilon$ is known as an (ϵ-)approximate eigenfunction. Such observables are important for the dynamical system (1.1) since $\mathcal{K}^n g = \lambda^n g + \mathcal{O}(n\epsilon)$. In other words, λ describes the coherent oscillation and decay/growth of the observable g with time. The approximate eigenfunctions and $\sigma_{\text{ap}}(\mathcal{K})$ encode information about the underlying dynamical system (1.1) [70]. For example, the level sets of certain eigenfunctions determine the invariant manifolds [68]

(e.g., Figure 4) and isostables [65], and the global stability of equilibria and ergodic partitions can be characterized by approximate eigenfunctions and $\sigma_{\text{ap}}(\mathcal{K})$ [64, 15].

We can approximate $\sigma_{\text{ap}}(\mathcal{K})$ using the eigenvalues computed by Algorithm 4.1 and verify the computations using residuals (see subsection 5.4). If $\mathcal{K}$ is unitary, then $\sigma_{\text{ap}}(\mathcal{K}) = \sigma(\mathcal{K}) \subset \mathbb{T}$. Otherwise, $\sigma_{\text{ap}}(\mathcal{K}) = \mathbb{T}$ and $\sigma(\mathcal{K})$ is the closed unit disc in $\mathbb{C}$.

3. EDMD. Given a dictionary of functions $\{\psi_1, \dots, \psi_N\} \subset L^2(\Omega, \omega)$, EDMD [92] constructs a matrix $\mathbb{K}_{\text{EDMD}} \in \mathbb{C}^{N\times N}$ from the snapshot data (1.2) that approximate the action of $\mathcal{K}$ on the finite-dimensional subspace $V_N = \text{span}\{\psi_1, \dots, \psi_N\}$. The choice of the dictionary is up to the user, with some common handcrafted choices given in [92, Table 1]. When the state-space dimension d is large, it is beneficial to use a data-driven choice of dictionary [57, 93], which can be verified a posteriori to capture the relevant dynamics via residual techniques [27]. We define the vector-valued function or "quasimatrix" Ψ via

$$\Psi(\boldsymbol{x}) = \begin{bmatrix} \psi_1(\boldsymbol{x}) & \cdots & \psi_N(\boldsymbol{x}) \end{bmatrix} \in \mathbb{C}^{1\times N}.$$

Any function $g \in V_N$ can then be written as $g(\boldsymbol{x}) = \sum_{j=1}^N \psi_j(\boldsymbol{x}) g_j = \Psi(\boldsymbol{x})\,\boldsymbol{g}$ for some vector of constant coefficients $\boldsymbol{g} \in \mathbb{C}^N$. It follows from (1.3) that

$$[\mathcal{K}g](\boldsymbol{x}) = \Psi(\boldsymbol{x})(\mathbb{K}_{\text{EDMD}}\,\boldsymbol{g}) + R(\boldsymbol{g}, \boldsymbol{x}), \quad R(\boldsymbol{g}, \boldsymbol{x}) := \Psi(F(\boldsymbol{x}))\,\boldsymbol{g} - \Psi(\boldsymbol{x})(\mathbb{K}_{\text{EDMD}}\,\boldsymbol{g}).$$

Typically, the subspace V_N generated by the dictionary is not an invariant subspace of $\mathcal{K}$. Hence there is no choice of $\mathbb{K}_{\text{EDMD}}$ that makes the error $R(\boldsymbol{g}, \boldsymbol{x})$ zero for all choices of $g \in V_N$ and $\boldsymbol{x} \in \Omega$. Instead, it is natural to select $\mathbb{K}_{\text{EDMD}}$ as a solution of

$$\underset{B\in\mathbb{C}^{N\times N}}{\text{argmin}} \left\{ \int_\Omega \max_{\|\boldsymbol{g}\|_2=1} |R(\boldsymbol{g}, \boldsymbol{x})|^2 \, d\omega(\boldsymbol{x}) = \int_\Omega \|\Psi(F(\boldsymbol{x})) - \Psi(\boldsymbol{x})B\|_2^2 \, d\omega(\boldsymbol{x}) \right\}. \tag{3.1}$$

Here, $\|\cdot\|_2$ denotes the standard Euclidean norm of a vector. Given a finite amount of snapshot data, we cannot directly evaluate the integral in (3.1). Instead, we approximate it via a quadrature rule by treating the data points $\{\boldsymbol{x}^{(m)}\}_{m=1}^M$ as quadrature nodes with weights $\{w_m\}_{m=1}^M$. Note that in the original definition of EDMD, ω is a probability measure and the quadrature weights are $w_m = 1/M$. General weights are an important consideration when we sample according to a measure different from ω or if we are free to chose $\{\boldsymbol{x}^{(m)}\}_{m=1}^M$ according to a high-order quadrature rule. The discretized version of (3.1) is

$$\mathbb{K}_{\text{EDMD}} \in \underset{B\in\mathbb{C}^{N\times N}}{\text{argmin}} \sum_{m=1}^M w_m \left\| \Psi(\boldsymbol{y}^{(m)}) - \Psi(\boldsymbol{x}^{(m)})B \right\|_2^2. \tag{3.2}$$

For notational convenience, we define the following two matrices,

$$\Psi_X = \begin{pmatrix} \Psi(\boldsymbol{x}^{(1)}) \\ \vdots \\ \Psi(\boldsymbol{x}^{(M)}) \end{pmatrix} \in \mathbb{C}^{M\times N}, \quad \Psi_Y = \begin{pmatrix} \Psi(\boldsymbol{y}^{(1)}) \\ \vdots \\ \Psi(\boldsymbol{y}^{(M)}) \end{pmatrix} \in \mathbb{C}^{M\times N}, \tag{3.3}$$

and let $W = \text{diag}(w_1, \dots, w_M)$ be the diagonal weight matrix of the quadrature rule. We define the Gram matrix $G = \Psi_X^* W \Psi_X$ and the matrix $A = \Psi_X^* W \Psi_Y$. Letting "$\dagger$" denote the pseudoinverse, a solution to (3.2) is

$$\mathbb{K}_{\text{EDMD}} = G^\dagger A = (\Psi_X^* W \Psi_X)^\dagger (\Psi_X^* W \Psi_Y) = (\sqrt{W}\Psi_X)^\dagger \sqrt{W}\Psi_Y.$$

In some applications, the matrix G may be ill-conditioned and it is common to consider truncated singular value decompositions or other forms of regularization [57]. For simplicity, we assume throughout the paper that G is invertible. The convergence analysis of our algorithm under regularization is left for future work.

If the quadrature approximation converges, then

$$\lim_{M\to\infty} G_{jk} = \langle \psi_k, \psi_j \rangle \quad \text{and} \quad \lim_{M\to\infty} A_{jk} = \langle \mathcal{K}\psi_k, \psi_j \rangle. \tag{3.4}$$

Let $\mathcal{P}_{V_N}$ denote the orthogonal projection onto V_N. As $M \to \infty$, the convergence in (3.4) means that $\mathbb{K}_{\text{EDMD}}$ approaches a matrix representation of $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}^*_{V_N}$. Thus, EDMD is a Galerkin method in the large data limit $M \to \infty$. As a special case, if $\psi_j(\boldsymbol{x}) = e_j^*\boldsymbol{x}$ for $j = 1, \dots, d = N$ and $w_m = 1/M$, then $\mathbb{K}_{\text{EDMD}} = (\sqrt{W}\Psi_X)^\dagger \sqrt{W}\Psi_Y$. In this case, $\mathbb{K}_{\text{EDMD}}$ is the transpose of the usual DMD matrix,

$$\mathbb{K}_{\text{DMD}} = \Psi_Y^\top \Psi_X^{\top\dagger} = \Psi_Y^\top \sqrt{W}(\Psi_X^\top \sqrt{W})^\dagger = ((\sqrt{W}\Psi_X)^\dagger \sqrt{W}\Psi_Y)^\top = \mathbb{K}_{\text{EDMD}}^\top.$$

Thus, DMD can be interpreted as producing a Galerkin approximation of the Koopman operator using the set of linear monomials as basis functions. When d is large, it is common to form a low-rank approximation of $\sqrt{W}\Psi_X$ via a truncated SVD [57].

There are typically three scenarios for which the convergence in (3.4) holds:

(i) *Random sampling:* In the initial definition of EDMD, ω is a probability measure and $\{\boldsymbol{x}^{(m)}\}_{m=1}^M$ are drawn independently according to ω with the quadrature weights $w_m = 1/M$. The strong law of large numbers shows that (3.4) holds with probability one [51, section 3.4], provided that ω is not supported on a zero level set that is a linear combination of the dictionary [54, section 4]. Convergence is typically at a Monte Carlo rate of $\mathcal{O}(M^{-1/2})$ [17].

(ii) *High-order quadrature:* If the dictionary and F are sufficiently regular and we are free to choose the $\{\boldsymbol{x}^{(m)}\}_{m=1}^M$, then it is beneficial to select $\{\boldsymbol{x}^{(m)}\}_{m=1}^M$ as an M-point quadrature rule with weights $\{w_m\}_{m=1}^M$. This can lead to much faster convergence rates in (3.4) [27], but can be difficult if d is large.

(iii) *Ergodic sampling:* For a single fixed initial condition $\boldsymbol{x}_0$ and $\boldsymbol{x}^{(m)} = F^{m-1}(\boldsymbol{x}_0)$ (i.e., data collected along one trajectory), if the dynamical system is ergodic, then one can use Birkhoff's ergodic theorem to show (3.4) [54]. One chooses $w_m = 1/M$ but the convergence rate is problem dependent [46].

If one is entirely free to select the initial conditions of the trajectory data, and d is not too large, then we recommend picking them based on a high-order quadrature rule. Random and ergodic sampling have the advantage of being practical even when d is large. Ergodic sampling is particularly useful when we have access to only one trajectory of the dynamical system. Ergodic sampling does not require knowledge of ω (e.g., if one wishes to study the dynamics near attractors).

4. mpEDMD. We now seek a matrix $\mathbb{K} \in \mathbb{C}^{N\times N}$ that approximates the action of $\mathcal{K}$ on the finite-dimensional subspace V_N and, in addition, corresponds to a unitary operator on V_N. Given the Gram matrix $G = \Psi_X^* W \Psi_X$, we can approximate the inner product $\langle \cdot, \cdot \rangle$ via the inner product induced by G:

$$\boldsymbol{h}^* G \boldsymbol{g} = \sum_{j,k=1}^N \overline{h_j} g_k G_{j,k} \approx \sum_{j,k=1}^N \overline{h_j} g_k \langle \psi_k, \psi_j \rangle = \langle \Psi\boldsymbol{g}, \Psi\boldsymbol{h} \rangle. \tag{4.1}$$

If (3.4) holds, then this approximation converges to the inner product on $L^2(\Omega, \omega)$ as $M \to \infty$. Hence we have $\|\Psi\boldsymbol{g}\|^2 \approx \boldsymbol{g}^* G \boldsymbol{g}$ and $\|\Psi\mathbb{K}\boldsymbol{g}\|^2 \approx \boldsymbol{g}^*\mathbb{K}^* G \mathbb{K}\boldsymbol{g}$. Since $\mathcal{K}$ is an isometry, we must have that $\|\Psi\mathbb{K}\boldsymbol{g}\|^2 = \|\Psi\boldsymbol{g}\|^2$. It is therefore natural to enforce

$$\boldsymbol{g}^* G \boldsymbol{g} = \boldsymbol{g}^* \mathbb{K}^* G \mathbb{K} \boldsymbol{g} \quad \forall \boldsymbol{g} \in \mathbb{C}^N.$$

This condition holds if and only if $\mathbb{K}^* G \mathbb{K} = G$. Hence, we replace (3.1) by

(4.2)
$$\operatorname*{argmin}_{\substack{B \in \mathbb{C}^{N\times N} \\ B^* G B = G}} \left\{ \int_\Omega \max_{\|G^{\frac{1}{2}}\boldsymbol{g}\|_2 = 1} |R(\boldsymbol{g}, \boldsymbol{x})|^2 \, d\omega(\boldsymbol{x}) = \int_\Omega \left\| \Psi(F(\boldsymbol{x})) G^{-\frac{1}{2}} - \Psi(\boldsymbol{x}) B G^{-\frac{1}{2}} \right\|_2^2 d\omega(\boldsymbol{x}) \right\}.$$

We therefore enforce that our Galerkin approximation is an isometry with respect to the learned inner product induced by G. This is a key difference to piDMD [4] which forces the DMD approximation matrix to be orthogonal with respect to the Euclidean inner product in which the observables are written down, i.e., $B^* B = I$ (note also that the "coordinate" inner product used by piDMD is not canonical or basis independent).

After applying the quadrature rule, the discretized version of (4.2) is

$$\operatorname*{argmin}_{\substack{B \in \mathbb{C}^{N\times N} \\ B^* G B = G}} \sum_{m=1}^{M} w_m \left\| \Psi(\boldsymbol{y}^{(m)}) G^{-\frac{1}{2}} - \Psi(\boldsymbol{x}^{(m)}) B G^{-\frac{1}{2}} \right\|_2^2. \tag{4.3}$$

Letting $B = G^{-1/2} C G^{1/2}$ for some matrix C, the problem in (4.3) is equivalent to

$$\operatorname*{argmin}_{\substack{C \in \mathbb{C}^{N\times N} \\ C^* C = I}} \left\| W^{\frac{1}{2}} \Psi_X G^{-\frac{1}{2}} C - W^{\frac{1}{2}} \Psi_Y G^{-\frac{1}{2}} \right\|_F^2, \tag{4.4}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The problem (4.4) is known as the *orthogonal Procrustes problem* [81, 3]. The predominant method for computing a solution is via the SVD. First, we compute an SVD of

$$G^{-\frac{1}{2}} \Psi_Y^* W \Psi_X G^{-\frac{1}{2}} = G^{-\frac{1}{2}} A^* G^{-\frac{1}{2}} = U_1 \Sigma U_2^*.$$

A solution of (4.4) is then $C = U_2 U_1^*$ and we take $\mathbb{K} = G^{-1/2} U_2 U_1^* G^{1/2}$.

Since $\mathbb{K}$ is similar to a unitary matrix, its eigenvalues lie along the unit circle. For stability purposes, the best way to compute the eigendecomposition of $\mathbb{K}$ is to do so for the unitary matrix $U_2 U_1^*$. To numerically ensure an orthonormal basis of eigenvectors, we use the MATLAB `schur` command in the examples of section 6. The computation of $\mathbb{K}$ and its eigendecomposition are summarized in Algorithm 4.1. Note that once the matrices G and A are given, the cost of Algorithm 4.1 is $\mathcal{O}(N^3)$ and hence comparable to other methods such as EDMD or piDMD.

The following proposition lists some useful properties of Algorithm 4.1. In part (iii), we use the notion of pseudospectra [86] of an operator or matrix $\mathcal{A}$:

$$\sigma_\epsilon(\mathcal{A}) = \{\lambda \in \mathbb{C} : \|(\mathcal{A} - \lambda)^{-1}\| \geq 1/\epsilon\} = \cup_{\|\mathcal{B}\| \leq \epsilon} \sigma(\mathcal{A} + \mathcal{B}), \quad \epsilon > 0.$$

In this paper, the Koopman operator is an isometry and hence $\sigma_\epsilon(\mathcal{K}) = \{\lambda \in \mathbb{C} : \operatorname{dist}(\lambda, \sigma(\mathcal{K})) \leq \epsilon\}$ so that the spectrum is stable to perturbations.

PROPOSITION 4.1. *The output of Algorithm* 4.1 *has the following properties.*

(i) *If* (3.4) *holds, then any limit point of the matrices* $\mathbb{K}$ *as* $M \to \infty$ *corresponds to an operator that is the unitary part of a polar decomposition of* $\mathcal{P}_{V_N} \mathcal{K} \mathcal{P}_{V_N}^*$.
(ii) *If* (3.4) *holds and* $g = \Psi \boldsymbol{g}$ *is such that* $\mathcal{K} g \in V_N$, *then* $\lim_{M\to\infty} \mathbb{K}\boldsymbol{g}$ *exists and* $\lim_{M\to\infty} \Psi \mathbb{K} \boldsymbol{g} = \mathcal{K} g$.
(iii) *For any* $\epsilon \geq 0$, $\sigma_\epsilon(\mathbb{K}) \subset \sigma_{\epsilon\kappa(G^{1/2})}(U_2 U_1^*) \subset \{z : ||z| - 1| \leq \epsilon\kappa(G^{1/2})\}$.
(iv) $\kappa(V) \leq \kappa(G^{1/2})$.

Algorithm 4.1: `mpEDMD` for approximating spectral properties of $\mathcal{K}$.

Input: Snapshot data $\{\boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)} = F(\boldsymbol{x}^{(m)})\}_{m=1}^{M}$, quadrature weights $\{w_m\}_{m=1}^{M}$, and a dictionary of functions $\{\psi_j\}_{j=1}^{N}$.

1: Compute $G = \Psi_X^* W \Psi_X$ and $A = \Psi_X^* W \Psi_Y$, where Ψ_X, Ψ_Y are given in (3.3).
2: Compute an SVD of $G^{-1/2} A^* G^{-1/2} = U_1 \Sigma U_2^*$.
3: Compute the eigendecomposition $U_2 U_1^* = \hat{V} \Lambda \hat{V}^*$.
4: Compute $\mathbb{K} = G^{-1/2} U_2 U_1^* G^{1/2}$ and $V = G^{-1/2} \hat{V}$.

Output: Koopman matrix $\mathbb{K}$ with eigenvectors V and eigenvalues Λ.

Proof. Suppose that $B \in \mathbb{C}^{N \times N}$ is a limit point of the matrix $\mathbb{K}$ as $M \to \infty$. By taking subsequences if necessary (all matrices are bounded), we may assume that $B = \hat{G}^{-\frac{1}{2}} \hat{U}_2 \hat{U}_1^* \hat{G}^{\frac{1}{2}}$, $\hat{U}_j = \lim_{M\to\infty} U_j$, $\hat{G} = \lim_{M\to\infty} G$, and $\hat{\Sigma} = \lim_{M\to\infty} \Sigma$. In the large data limit, the problem (4.3) is independent of the choice of basis for V_N, and property (i) is also basis independent. Hence, we may assume without loss of generality that $\hat{G}$ is the identity matrix corresponding to an orthonormal basis. It follows that $\mathbb{K}_{\text{EDMD}} = \hat{U}_2 \hat{\Sigma} \hat{U}_1^*$. Part (i) now follows.

For part (ii), since $\mathcal{K} g \in V_N$, we have $\|g\| = \|\mathcal{K} g\| = \lim_{M\to\infty} \|U_2 \Sigma U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2 = \lim_{M\to\infty} \|\Sigma U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2$, where the last equality holds because U_2 is unitary. Similarly, $\|g\| = \lim_{M\to\infty} \|G^{\frac{1}{2}} \boldsymbol{g}\|_2 = \lim_{M\to\infty} \|U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2$. Hence $\lim_{M\to\infty} \|U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2 = \lim_{M\to\infty} \|\Sigma U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2$. Σ is a diagonal matrix and all of its entries are in $[0, 1]$. We claim that $\lim_{M\to\infty} [U_1^* G^{\frac{1}{2}} - \Sigma U_1^* G^{\frac{1}{2}}] \boldsymbol{g} = 0$. If not, then by taking a subsequence if necessary, we may assume that $\lim_{M\to\infty} \Sigma$ and $\lim_{M\to\infty} U_j$ exist with $\lim_{M\to\infty} [U_1^* G^{\frac{1}{2}} - \Sigma U_1^* G^{\frac{1}{2}}] \boldsymbol{g} \neq 0$. But this contradicts $\lim_{M\to\infty} \|U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2 = \lim_{M\to\infty} \|\Sigma U_1^* G^{\frac{1}{2}} \boldsymbol{g}\|_2$. Since $\lim_{M\to\infty} G^{-\frac{1}{2}} U_2 \Sigma U_1^* G^{\frac{1}{2}} \boldsymbol{g} = \lim_{M\to\infty} \mathbb{K}_{\text{EDMD}} \boldsymbol{g}$ exists, $\lim_{M\to\infty} \mathbb{K} \boldsymbol{g}$ exists and $\lim_{M\to\infty} \Psi \mathbb{K} \boldsymbol{g} = \lim_{M\to\infty} \Psi \mathbb{K}_{\text{EDMD}} \boldsymbol{g} = g$.

For part (iii), for any $z \notin \sigma(\mathbb{K})$ we have $\|(\mathbb{K} - z)^{-1}\| = \|G^{-\frac{1}{2}} (U_2 U_1^* - z)^{-1} G^{\frac{1}{2}}\| \leq \kappa(G^{\frac{1}{2}}) \|(U_2 U_1^* - z)^{-1}\|$. Hence, $\sigma_\epsilon(\mathbb{K}) \subset \sigma_{\epsilon\kappa(G^{1/2})}(U_2 U_1^*)$. $U_2 U_1^*$ is unitary, and hence $\sigma_{\epsilon\kappa(G^{\frac{1}{2}})}(U_2 U_1^*) \subset \{z : ||z| - 1| \leq \epsilon\kappa(G^{\frac{1}{2}})\}$. Finally, $V = G^{-\frac{1}{2}} \hat{V}$ for unitary $\hat{V}$ so (iv) holds. ∎

Part (i) of Proposition 4.1 provides a geometric interpretation of Algorithm 4.1, that we use to prove convergence of spectral measures in section 5. Part (ii) shows that Algorithm 4.1 respects the invariance properties of $\mathcal{K}$. This is particularly useful for delay embedding (see Corollary 5.5). Parts (iii) and (iv) provide conditioning bounds on the eigendecomposition of $\mathbb{K}$. This is useful since we can only ever approximate the eigendecomposition using finite M. In contrast, conditioning bounds for $\mathbb{K}_{\text{EDMD}}$ cannot hold in general. In fact, $\mathbb{K}_{\text{EDMD}}$ need not even be diagonalizable (see subsection 5.2). Further stability properties of `mpEDMD` are investigated in subsection 6.2.

Note that `mpEDMD` can be used with generic choices of dictionary which generate the matrices G and A. Data-driven choices of dictionary include diffusion kernels [32] and trained neural networks [59, 72]. For example, the kernel trick has been used for implementing EDMD when the state-space dimension is large [93]. It is straightforward to adapt Algorithm 4.1 along the same lines.

5. Convergence theory.

We now show convergence of Algorithm 4.1 to the spectral information of $\mathcal{K}$. Throughout, $\{v_j\}_{j=1}^{N}$ denotes the eigenvectors of $\mathbb{K}$ with corresponding eigenvalues $\{\lambda_j\}_{j=1}^{N}$, where $\mathbb{K}$ is the matrix output of Algorithm 4.1. For ease of reading, the proofs of the results in this section are collected in Appendix A.

5.1. Approximation of projection-valued spectral measures. To approximate the spectral measure $\mathcal{E}$, we consider the spectral measure, $\mathcal{E}_{N,M}$, of the matrix $\mathbb{K}$ on the Hilbert space $\mathbb{C}^N$ with the inner product in (4.1) induced by G:

$$d\mathcal{E}_{N,M}(\lambda) := \sum_{j=1}^{N} v_j v_j^* G \delta(\lambda - \lambda_j)\, d\lambda.$$

We consider a sequence of vectors spaces $\{V_N\}_{N=1}^{\infty}$ and the large data limit $M \to \infty$. The following theorem shows weak convergence[3] of $\Psi\mathcal{E}_{N,M}$ if $\mathcal{K}$ is unitary. For example, if F is invertible and measure-preserving, $\mathcal{K}$ is unitary. A key part of the proof is that $\mathbb{K}$ represents a normal truncation of $\mathcal{K}$.

THEOREM 5.1. *Suppose that* $\lim_{N\to\infty} \operatorname{dist}(h, V_N) = 0$ *for all* $h \in L^2(\Omega,\omega)$, (3.4) *holds,* $\mathcal{K}$ *is unitary, and that* $\phi : \mathbb{T} \to \mathbb{R}$ *is Lipschitz continuous. Then for any* $g \in L^2(\Omega,\omega)$ *and* $\boldsymbol{g}_N \in \mathbb{C}^N$ *with* $\lim_{N\to\infty} \|g - \Psi \boldsymbol{g}_N\| = 0$,

$$\lim_{N\to\infty} \limsup_{M\to\infty} \left\| \int_{\mathbb{T}} \phi(\lambda)\, d\mathcal{E}(\lambda) g - \Psi \int_{\mathbb{T}} \phi(\lambda)\, d\mathcal{E}_{N,M}(\lambda) \boldsymbol{g}_N \right\| = 0. \tag{5.1}$$

The proof of Theorem 5.1 shows that the rate of convergence in (5.1) as $N \to \infty$ depends on the regularity of ϕ (how fast its Laurent series converges) as well as how well the powers of $\mathcal{K}$ are captured by the powers of $\mathbb{K}$.

Remark 5.2 (computing suitable $\boldsymbol{g}_N$ and the Koopman mode decomposition). Given $g \in L^2(\Omega,\omega)$, we can compute a suitable $\boldsymbol{g}_N$ in Theorem 5.1 via

$$\boldsymbol{g}_N = G^{-1} \Psi_X^* W \begin{pmatrix} g(\boldsymbol{x}^{(1)}) & \cdots & g(\boldsymbol{x}^{(M)}) \end{pmatrix}^\top \in \mathbb{C}^N.$$

If the quadrature rule converges, $\Psi \boldsymbol{g}_N$ converges to $\mathcal{P}_{V_N} g$ in the large data limit and $\lim_{N\to\infty} \|g - \mathcal{P}_{V_N} g\| = 0$ under the first condition of the theorem. We obtain

$$[\mathcal{P}_{V_N} g](\boldsymbol{x}) \approx \Psi(\boldsymbol{x}) V \left[V^{-1} (\sqrt{W}\Psi_X)^\dagger \sqrt{W} \begin{pmatrix} g(\boldsymbol{x}^{(1)}) & \cdots & g(\boldsymbol{x}^{(M)}) \end{pmatrix}^\top \right].$$

Hence, we have the approximate factorization

$$\begin{aligned} g(\boldsymbol{x}_n) &\approx \Psi(\boldsymbol{x}_0) \mathbb{K}^n V \left[V^{-1} (\sqrt{W}\Psi_X)^\dagger \sqrt{W} \begin{pmatrix} g(\boldsymbol{x}^{(1)}) & \cdots & g(\boldsymbol{x}^{(M)}) \end{pmatrix}^\top \right] \\ &= [\Psi(\boldsymbol{x}_0) V]\, \Lambda^n \left[V^{-1} (\sqrt{W}\Psi_X)^\dagger \sqrt{W} \begin{pmatrix} g(\boldsymbol{x}^{(1)}) & \cdots & g(\boldsymbol{x}^{(M)}) \end{pmatrix}^\top \right]. \end{aligned} \tag{5.2}$$

The factor ΨV is a quasimatrix of approximate Koopman eigenfunctions. The columns of the final factor in square brackets are known as Koopman modes [66]. The first part of Lemma A.2 shows the convergence of this approximation.

5.2. Warning example. We cannot drop the condition that $\mathcal{K}$ is unitary from Theorem 5.1. In general, since $\mathcal{K}$ is an isometry, it is unitarily equivalent to a direct sum of unilateral shifts and a unitary operator [73, Chapter I]. The general case can be understood by supposing that $\mathcal{K}$ is unitarily equivalent to a single unilateral shift. Let $V_N = \operatorname{span}\{e_1, \ldots, e_N\}$, where the $\{e_j\}$ is a basis so that $\mathcal{K} e_j = e_{j+1}$ for $j \in \mathbb{N}$. We have

$$\lim_{M\to\infty} \mathbb{K}_{\mathrm{EDMD}} = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}, \quad \lim_{M\to\infty} \mathbb{K} = \begin{pmatrix} 0 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

[3] This is not to be confused with weak operator convergence of the operator-valued measures.

Note that for this example, $\mathbb{K}_{\text{EDMD}}$ is not even diagonalizable. Let $\phi(\lambda) = 1/\lambda$, then

$$\int_{\mathbb{T}} \phi(\lambda)\, d\mathcal{E}(\lambda) e_1 = \mathcal{K}^* e_1 = 0, \qquad \lim_{M\to\infty} \Psi \int_{\mathbb{T}} \phi(\lambda)\, d\mathcal{E}_{N,M}(\lambda) e_1 = e_N,$$

where we also use e_1 to denote the first canonical basis vector of $\mathbb{C}^N$. Clearly e_N does not converge in the norm to e_1. However, e_N does converge weakly to 0 in $L^2(\Omega,\omega)$. Motivated by this, we remove the need for $\mathcal{K}$ to be unitary when considering scalar-valued spectral measures in the next subsection.

5.3. Approximation of scalar-valued spectral measures. Let $g \in L^2(\Omega,\omega)$ with $\|g\| = 1$. We approximate the spectral measure μ_g by $\mu_{\boldsymbol{g}}^{(N,M)}$, where

$$\mu_{\boldsymbol{g}}^{(N,M)}(U) := \boldsymbol{g}^* G \mathcal{E}_{N,M}(U) \boldsymbol{g} = \sum_{\lambda_j \in U} |v_j^* G \boldsymbol{g}|^2 \tag{5.3}$$

and $\boldsymbol{g}$ is normalized so that $\boldsymbol{g}^* G \boldsymbol{g} = 1$. Since $\{G^{1/2} v_j\}_{j=1}^N$ is an orthonormal basis for $\mathbb{C}^N$, $\mu_{\boldsymbol{g}}^{(N,M)}$ is a probability measure. To measure the distance between probability measures, we use the 1-Wasserstein distance, W_1. For two Borel probability measures μ and ν on $\mathbb{T}$, the W_1 distance is defined as

$$W_1(\mu,\nu) := \sup\left\{ \int_{\mathbb{T}} \phi(\lambda)\, d(\mu-\nu)(\lambda) : \phi : \mathbb{T} \to \mathbb{R} \text{ Lip. cts., Lip. constant} \le 1 \right\}.$$

Convergence in this metric is equivalent to the usual weak convergence of measures. The following theorem explicitly bounds $W_1(\mu_g, \mu_{\boldsymbol{g}}^{(N,M)})$, before taking any limits.

THEOREM 5.3. *For any $L \in \mathbb{N}$, $g \in L^2(\Omega,\omega)$, and $\boldsymbol{g} \in \mathbb{C}^N$ with $\boldsymbol{g}^* G \boldsymbol{g} = 1$,*

$$W_1\left(\mu_g, \mu_{\boldsymbol{g}}^{(N,M)}\right) \le C \left(\frac{\log(L)}{L} + \sum_{1 \le l \le L} \frac{|\langle \mathcal{K}^{|l|} g, g\rangle - \boldsymbol{g}^* G \mathbb{K}^{|l|} \boldsymbol{g}|}{l} \right)$$

for a universal constant C that can be made explicit.

Theorem 5.3 and the first part of Lemma A.2, show the following corollary.

COROLLARY 5.4. *Suppose that $\lim_{N\to\infty} \operatorname{dist}(h, V_N) = 0$ for all $h \in L^2(\Omega,\omega)$, and* (3.4) *holds. Then for any $g \in L^2(\Omega,\omega)$ and $\boldsymbol{g}_N \in \mathbb{C}^N$ with $\lim_{N\to\infty} \|g - \Psi \boldsymbol{g}_N\| = 0$,*

$$\lim_{N\to\infty} \limsup_{M\to\infty} W_1\left(\mu_g, \mu_{\boldsymbol{g}}^{(N,M)}\right) = 0. \tag{5.4}$$

A popular choice of dictionary is a Krylov subspace, i.e., $\operatorname{span}\{g, \mathcal{K}g, \ldots, \mathcal{K}^{L-1} g\}$. This corresponds to time-delay embedding, which is a popular method for DMD-type algorithms [1, 47, 74]. More generally, we can consider the case $\{g, \mathcal{K}g, \ldots, \mathcal{K}^L g\} \subset V_N$. Part (ii) of Proposition 4.1 shows that if $g, \mathcal{K}g, \ldots, \mathcal{K}^l g \in V_N$ and $g = \Psi \boldsymbol{g}$, then $\lim_{M\to\infty} |\langle \mathcal{K}^{|l|} g, g\rangle - \boldsymbol{g}^* G \mathbb{K}^{|l|} \boldsymbol{g}| = 0$. Combining this with Theorem 5.3 shows the following corollary, which provides an explicit rate of convergence.

COROLLARY 5.5. *If $\{g, \mathcal{K}g, \ldots, \mathcal{K}^L g\} \subset V_N$, $g = \Psi \boldsymbol{g}$, and* (3.4) *holds, then*

$$\limsup_{M\to\infty} W_1\left(\mu_g, \mu_{\boldsymbol{g}}^{(N,M)}\right) \le C \log(L)/L \tag{5.5}$$

for a universal constant C that can be made explicit.

5.4. Approximation of spectra. We end this section with the convergence to $\sigma_{\mathrm{ap}}(\mathcal{K})$. The following theorem shows that the eigenvalues computed by Algorithm 4.1 approximate the whole of $\sigma_{\mathrm{ap}}(\mathcal{K})$ as $N \to \infty$ and the subspace V_N becomes richer.

THEOREM 5.6. *If* $\lim_{N\to\infty} \operatorname{dist}(h, V_N) = 0 \ \forall h \in L^2(\Omega, \omega)$ *and* (3.4) *holds, then*

$$\lim_{N\to\infty} \limsup_{M\to\infty} \sup_{\lambda\in\sigma_{\mathrm{ap}}(\mathcal{K})} \operatorname{dist}(\lambda, \{\lambda_1, \ldots, \lambda_N\}) = 0. \tag{5.6}$$

Despite this result, $\sigma(\mathbb{K}) = \{\lambda_1, \ldots, \lambda_N\}$ can suffer from spectral pollution. That is, eigenvalues of $\mathbb{K}$ may approximate points that are not in $\sigma(\mathcal{K})$. We can avoid spectral pollution by computing residuals and discarding eigenpairs with a large residual. Suppose that (3.4) holds, $v \in \mathbb{C}^N$, and $\lambda \in \mathbb{C}$. Since $\mathcal{K}^*\mathcal{K}$ is the identity,

$$\|(\mathcal{K}-\lambda)\Psi v\| = \sqrt{\langle(\mathcal{K}-\lambda)\Psi v, (\mathcal{K}-\lambda)\Psi v\rangle} = \lim_{M\to\infty} \sqrt{v^*[(1+|\lambda|^2)G - \overline{\lambda}A - \lambda A^*]v}. \tag{5.7}$$

Since $\mathcal{K}$ is an isometry, this residual provides a good error estimate. In particular, if v is normalized so that $\lim_{M\to\infty} \|G^{1/2}v\|_2 = 1$, then $\|(\mathcal{K} - \lambda)\Psi v\| \geq \operatorname{dist}(\lambda, \sigma(\mathcal{K}))$.

6. Numerical examples. We consider three numerical examples, two with data from numerical simulations, and one with experimentally collected data. Each example demonstrates different aspects and advantages of `mpEDMD`.

6.1. Lorenz system and convergence of spectral measures. The Lorenz system [60] is the following system of three coupled ordinary differential equations:

$$\dot{X} = 10\,(Y - X)\,, \quad \dot{Y} = X\,(28 - Z) - Y, \quad \dot{Z} = XY - 8Z/3.$$

We consider the dynamics of $\boldsymbol{x} = (X, Y, Z)$ on the Lorenz attractor. The system is chaotic and strongly mixing [62] (and hence ergodic). Hence the only eigenvalue (including multiplicities) of $\mathcal{K}$ is the trivial eigenvalue $\lambda = 1$ corresponding to a constant eigenfunction. We consider the corresponding discrete-time dynamical system by sampling with a time step $\Delta_t = 0.1$. We use the `ode45` command in MATLAB to collect data along a single trajectory with M snapshots, from an initial point on the attractor. The quadrature rule in section 3 corresponds to ergodic sampling.[4]

We first consider the scalar-valued spectral measures $\mu_{\boldsymbol{g}_j}^{(N,M)}$, where $g_j(\boldsymbol{x}) = c_j[\boldsymbol{x}]_j$ is the jth coordinate normalized to have norm 1 with respect to the ergodic measure ω. For each j, we use $\{g_j, \mathcal{K}g_j, \ldots, \mathcal{K}^{N-1}g_j\}$ as the dictionary. This choice corresponds to time-delay embedding. Figure 1 (left) shows the convergence as $M \to \infty$ (large data limit) for a fixed $N = 50$. The convergence is at a Monte Carlo rate of $\mathcal{O}(M^{-1/2})$. Figure 1 (middle) shows the convergence as $N \to \infty$, where $M = 10^6$ is selected large enough to have negligible effect on the shown errors. The W_1 distance to μ_{g_j} is computed by comparing it to an approximation with larger N selected large enough to have negligible effect on the shown errors. The plot demonstrates the rate $\mathcal{O}(N^{-1})$ from Corollary 5.5. For either convergence in M or N, we do not observe monotonic decrease of errors, which is to be expected. Figure 1 (right) plots the cumulative distribution functions (cdfs) of $\mu_{\boldsymbol{g}_j}^{(N,M)}$ for $N = 10^3$ and $M = 10^6$. For this example, the cdf of μ_{g_j} is continuous away from $\lambda = 1$ and hence the cdf of $\mu_{\boldsymbol{g}_3}^{(N,M)}$ converges pointwise on $\mathbb{T}\backslash\{1\}$. The cdf for $\mu_{\boldsymbol{g}_3}^{(N,M)}$ suggests an atom at $\lambda = 1$ with a small absolutely continuous spectrum in the vicinity of $\lambda = 1$. In contrast, $\mu_{\boldsymbol{g}_1}^{(N,M)}$ and $\mu_{\boldsymbol{g}_2}^{(N,M)}$ are more uniform.

[4]Though we cannot accurately numerically integrate for long time periods since the system is chaotic, this does not affect the convergence of the quadrature rule. This effect is known as shadowing.

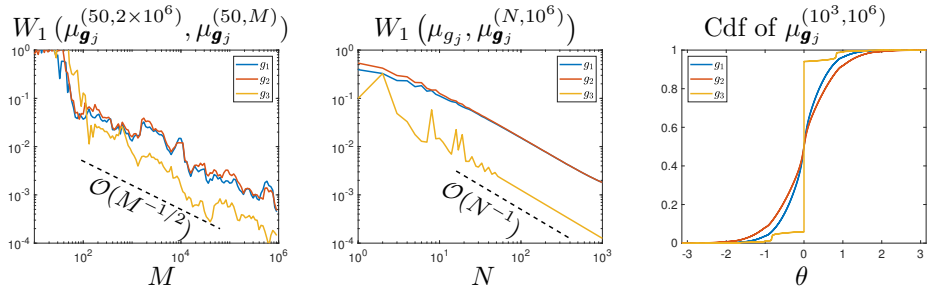

FIG. 1. *Left: Convergence of* $\mu_{\boldsymbol{g}_j}^{(50,M)}$ *as* $M\to\infty$. *Middle: Convergence to the scalar-valued measure as* $N\to\infty$. *Right: Cdf of* $\mu_{\boldsymbol{g}_j}^{(10^3,10^6)}$ *plotted against the phase,* θ, *of the spectral parameter* $\lambda=e^{i\theta}$. *In all cases, the* W_1 *distance is computed using the* L^1 *distance between the cdfs.*

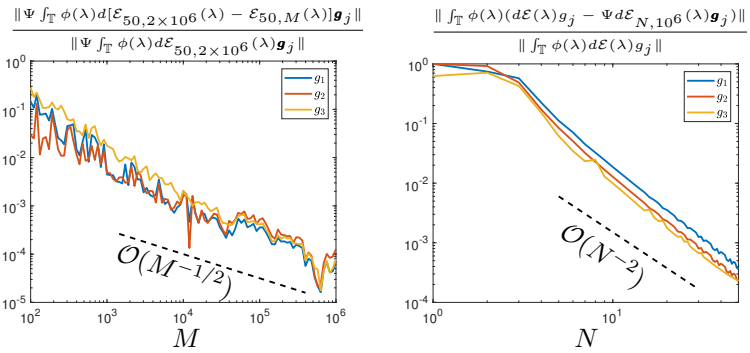

FIG. 2. *Convergence of integrals as* $M\to\infty$ *(left) and* $N\to\infty$ *(right).*

Next, we approximate the projection-valued spectral measures and demonstrate Theorem 5.1. We use the same dictionaries as before, take $\phi(\lambda)=(1-\lambda)^2\log(1-\lambda)$, and compute $\int_{\mathbb{T}}\phi(\lambda)\,d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g}_j$. Figure 2 (left) shows the convergence as $M\to\infty$ for a fixed $N=50$. Again, we see the Monte Carlo rate of convergence $\mathcal{O}(M^{-1/2})$. Figure 2 (right) shows the convergence as $N\to\infty$. Theorem 5.1 does not provide a rate of convergence. However, we may use the proof of Theorem A.1. The function ϕ was chosen so that its second derivative has a logarithmic blowup at $\lambda=1$. Figure 1 suggests a convergence rate of approximately $\mathcal{O}(N^{-2})$. In general, the rate depends on how fast the truncated Laurent series of ϕ converges and how fast powers of $\mathbb{K}$ converge strongly to powers of $\mathcal{K}$. Figure 3 shows the outputs as functions on the Lorenz attractor using $N=50$ basis functions and $M=10^6$.

6.2. Nonlinear pendulum, approximate eigenfunctions, and robustness to noise. We now consider the dynamical system of the nonlinear pendulum. Let the state variables $\boldsymbol{x}=(x_1,x_2)$ be governed by the following equations of motion,

$$\dot{x_1}=x_2,\quad \dot{x_2}=-\sin(x_1)\quad \text{with}\quad \Omega=[-\pi,\pi]_{\text{per}}\times\mathbb{R},$$

where ω is the standard Lebesgue measure on Ω. We consider the corresponding discrete-time dynamical system by sampling with a time step $\Delta_t=0.5$. The system is nonchaotic and Hamiltonian, with challenging Koopman operator theory [61].

We use the dictionary $\{g,\mathcal{K}g,\dots,\mathcal{K}^{99}g\}$ with $g(x_1,x_2)=\exp(ix_1)x_2\exp(-x_2^2/2)$. We collect data points on an equispaced tensor product grid corresponding to the

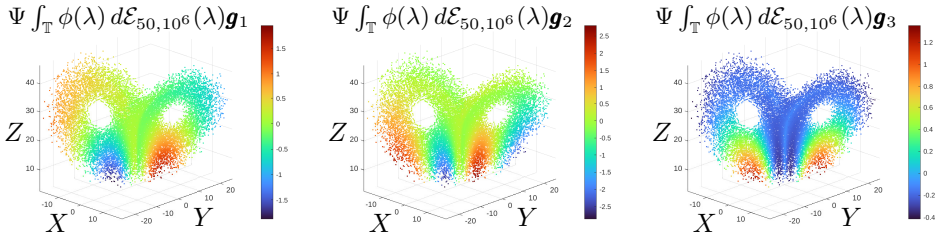

FIG. 3. *Real part of* $\Psi \int_{\mathbb{T}} \phi(\lambda)\, d\mathcal{E}_{50,10^6}(\lambda)\boldsymbol{g}_j$ *plotted at* 2×10^4 *points on the attractor.*

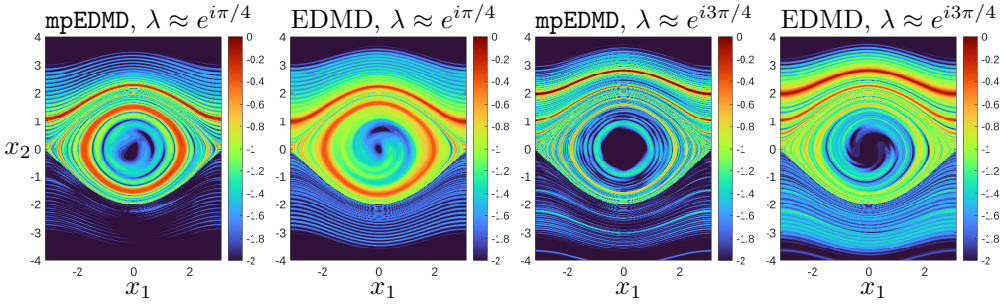

FIG. 4. *Eigenfunctions* $\log_{10}(|v|)$*, where* $v = \Psi \boldsymbol{v}$ *is normalized and* $\boldsymbol{v}$ *is the eigenvector of* $\mathbb{K}$ *(*`mpEDMD`*) or* $\mathbb{K}_{\mathrm{EDMD}}$ *(EDMD). In each case we plot the eigenfunction with eigenvalue nearest to the shown value of* λ*. Taking* $\lambda \to \overline{\lambda}$ *yields the corresponding eigenfunctions reflected in* $x_2 = 0$*.*

periodic trapezoidal quadrature rule with M_1 points in the x_1 direction and a truncated trapezoidal quadrature rule with $M_2 = M_1$ points in the x_2 direction. For our problem, these quadrature rules have exponential [87] and $\mathcal{O}(\exp(-CM_2^{2/3}))$ [85] convergence, respectively. To simulate the collection of trajectory data, we compute trajectories starting at each initial condition using the `ode45` command in MATLAB.

Figure 4 shows approximate eigenfunctions on a log-scale, computed using $M_1 = 200$. The Koopman operator $\mathcal{K}$ has no normalizable eigenfunctions, but has generalized eigenfunctions supported along unions of contour lines of the action variable [69]. The eigenfunctions produced by `mpEDMD` are much more localized along these contour lines and better approximate the generalized eigenfunctions than EDMD, whose approximate eigenfunctions are blurred. Figure 5 (left) shows the eigenvalues of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{EDMD}}$. The eigenvalues of $\mathbb{K}_{\mathrm{EDMD}}$ lie strictly inside the unit disc, corresponding to spectral pollution. Note that this spectral pollution has nothing to do with any stability issues, but instead is due to the discretization of the infinite-dimensional operator $\mathcal{K}$ by a finite matrix. In contrast, `mpEDMD` does not suffer from spectral pollution.

Noise is a substantial problem for most DMD methods, and a common remedy is to consider a total least squares (TLS) problem [30]. The solution to the orthogonal Procrustes problem (4.4) is also the solution to the corresponding constrained TLS problem [3]. Hence, `mpEDMD` is optimally robust when noise is present in both data matrices in (4.4) [89]. We test the robustness to noise by adding τ Gaussian random noise to the measurement matrices Ψ_X and Ψ_Y in (3.3). Figure 5 (right) shows the effect of noise on the eigenvalues of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{EDMD}}$ for $\tau = 0.1$ (10% noise). The deterioration of the spectrum of $\mathbb{K}_{\mathrm{EDMD}}$ is clear. To further investigate robustness, we compute the (relative) residual of eigenpairs using (5.7) with noise-free matrices G, A computed using large M_1. Figure 6 plots the mean residual over all $N = 200$

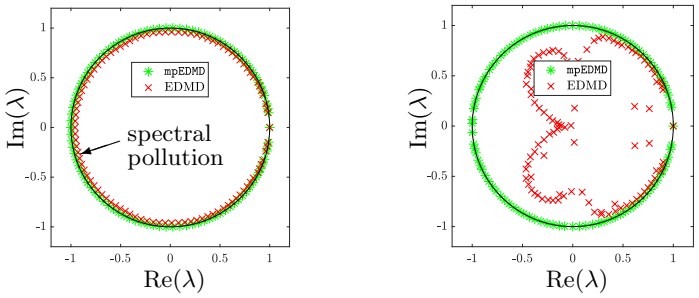

FIG. 5. *Eigenvalues of* $\mathbb{K}_{\text{EDMD}}$ *(EDMD) and* $\mathbb{K}$ *(*`mpEDMD`*). Left: Noise-free case. Right:* 10% *Gaussian random noise added to* Ψ_X *and* Ψ_Y.

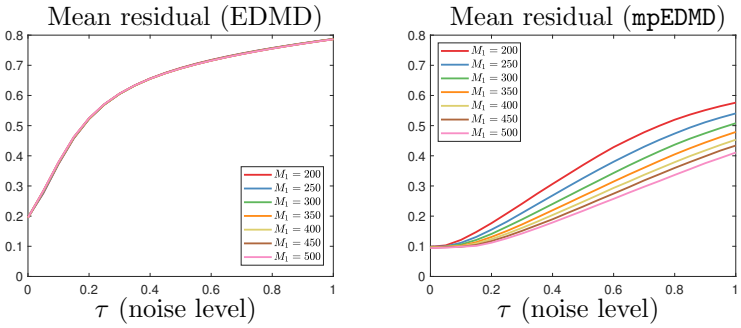

FIG. 6. *Mean residual over all* N *eigenpairs and* 10 *independent realizations per noise level* τ. *Residuals are computed using* (5.7) *with matrices* G *and* A *computed using a larger* M_1 *and* $\tau = 0$.

eigenpairs and 10 independent noise realizations against the noise level τ. We see that `mpEDMD` is much more robust to noise than EDMD. Moreover, for a given noise level τ, the accuracy of `mpEDMD` increases as M_1 increases. A full statistical analysis of this phenomenon is beyond the scope of this paper, but we note that this type of behavior, known as strongly consistent estimation, is typical of TLS [89, Chapter 8]. This phenomenon does not happen with EDMD in Figure 6.

6.3. Conservation of energy and statistics for turbulent boundary layer flow. We now consider the boundary layer generated by a thin jet of height 12.7 mm injecting air onto a smooth flat wall. Experiments are performed at the wind tunnel of Virginia Tech [84]. A two-component time-resolved particle image velocimetry system is used to capture 1000 snapshots of the two-dimensional velocity field of the wall-jet flow over a spatial grid and a time period of 1 s. The streamwise origin of the field-of-view is 1282.7 mm downstream of the wall-jet nozzle. We use a jet velocity of $U_{\text{jet}} = 50$ m/s, corresponding to a jet Reynolds number of 6.4×10^4. The length and height of the field-of-view is approximately 75 mm $\times$ 40 mm, and the spatial resolution of the measurements is $\Delta x = \Delta y \approx 0.24$ mm. This corresponds to dimension $d = 102300$ in (1.1). We use a full SVD of the data matrix to form a dictionary, as outlined in section 3. The flow consists of two main regions. Within the region bounded by the wall and the peak in the velocity profile at $y \approx 15.5$ mm, the flow exhibits the properties of a zero pressure gradient turbulent boundary layer. Above this fluid portion, the flow is dominated by a two-dimensional shear layer consisting of large, energetic flow structures. This example is a considerable challenge for regular DMD approaches due to multiple turbulent scales expected within the boundary layer.

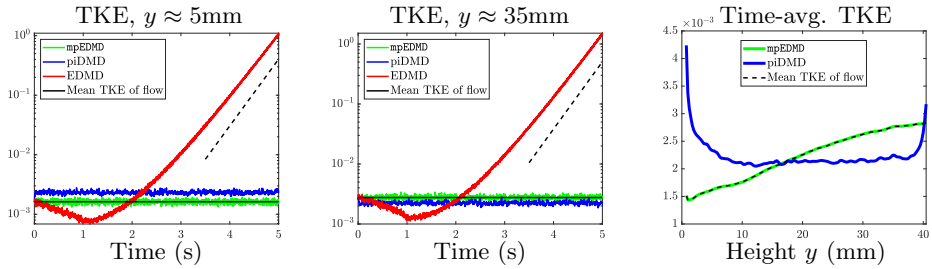

FIG. 7. *Left and middle: Turbulent kinetic energy (TKE) as a function of time, averaged over the homogeneous horizontal direction. The dashed line shows the expected growth rate of EDMD from the eigenvalues of* $\mathbb{K}_{\text{EDMD}}$. *Right: TKE as a function of vertical height, averaged over time, and the homogeneous horizontal direction. The relative error of* `mpEDMD` *is bounded by* 0.001.

We investigate the conservation of energy and statistics of the flow when using the KMD in Remark 5.2, in particular (5.2), to make future state predictions. We consider the velocity profiles predicted by `mpEDMD`, EDMD, and piDMD over a time period of 5 s (five times the window of observations) from an initialization $\boldsymbol{x}_0$ selected at random from the trajectory data. The results are averaged over 100 such random initializations.

Figure 7 (left, middle) shows the TKE of the predictions, averaged in the (homogenous) horizontal direction and normalized by U_{jet}^2. We show the TKE at vertical heights in the boundary layer (left panel) and in the shear layer (middle panel). The instability of the KMD for EDMD is clear. Whilst piDMD is stable, it does not preserve the correct values of TKE. This is because piDMD is preserving the standard Euclidean inner product, as opposed to the inner product in (4.1) induced by the matrix G. In contrast, `mpEDMD` preserves the correct inner product and conserves the correct TKE. Figure 7 (right) highlights this by showing the TKE prediction of `mpEDMD` and piDMD as a function of the vertical height, and averaged over the whole time period of 5 s. The maximum relative error of `mpEDMD` is less than 0.001. These results underline the importance, even for dictionaries of linear functions, of the nontrivial matrix G in Algorithm 4.1.

Figure 8 (top row) shows characteristic predictions of the horizontal component of the velocity field at prediction time 4 s. `mpEDMD` captures the larger-scale structures above the boundary layer, whereas piDMD does not, and EDMD overpredicts the velocity magnitude. Note that the system is chaotic so we can only expect to predict the qualitative behavior. To investigate the statistics of the predictions, Figure 8 (bottom row) shows the wavenumber spectrum, computed by applying the Fourier transform to spatial autocorrelations of the predictions in the horizontal direction [33, Chapter 8]. The wavenumber spectrum provides a measure of energy content of various turbulent structures as a function of their size, and provides an efficient measure of how well a flow-reconstruction method performs for various spatial scales. The wavenumber spectrum of `mpEDMD` shows excellent agreement with the wavenumber spectrum of the flow. In contrast, EDMD and piDMD do not capture the correct turbulent statistics. While we can only ever capture the statistics to the resolution of the collected data, this example provides very promising results for the use of `mpEDMD` in real-world applications and methods such as model order reduction.

7. Conclusion. We formulated a structure-preserving data-driven approximation of Koopman operators for measure-preserving dynamical systems, `mpEDMD`, summarized in Algorithm 4.1. We proved the convergence of `mpEDMD` to various

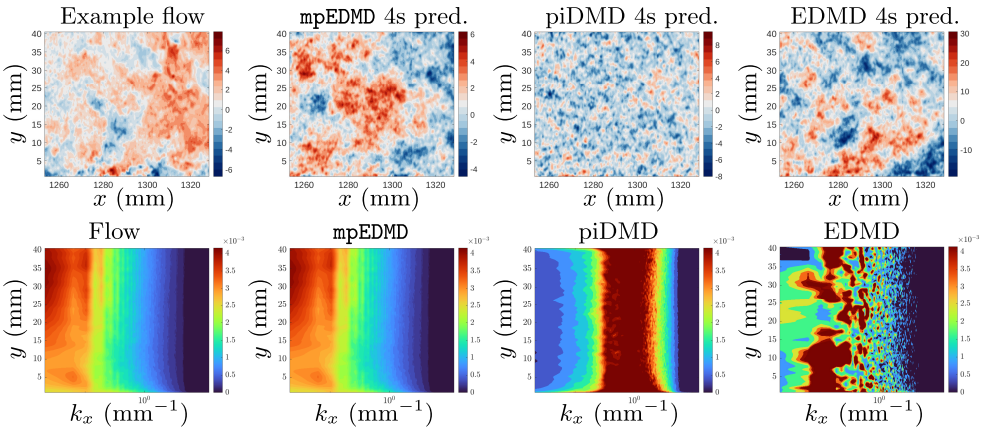

FIG. 8. *Top row: Horizontal velocity profiles predicted at* 4*s. Bottom row: Wavenumber spectra, which measure the energy content of various turbulent structures at different scales, thus providing an efficient measure of a flow reconstruction method's performance over various spatial scales.*

infinite-dimensional spectral quantities of interest, summarized in Table 2. In particular, `mpEDMD` is the first truncation method whose eigendecomposition converges to these spectral quantities for general measure-preserving dynamical systems. We also proved the first results on convergence rates of the approximation in the size of the dictionary. As well as the convergence theory, our numerical examples show the increased robustness of `mpEDMD` to noise compared with other DMD-type methods, and the ability to capture energy conservation and statistics of a real-world turbulent boundary layer flow. These results open the door to future extensions to more general structure-preserving methods for Koopman operators and data-driven dynamical systems.

Appendix A. Proofs of results in section 5. To prove Theorem 5.1, we begin with the following bound.

THEOREM A.1. *Suppose that* $\phi : \mathbb{T} \to \mathbb{R}$ *is Lipschitz continuous with Lipschitz constant bounded by* 1*. Then for any* $L \in \mathbb{N}$*,* $g \in L^2(\Omega, \omega)$*, and* $\boldsymbol{g} \in \mathbb{C}^N$*,*

$$\left\| \int_{\mathbb{T}} \phi(\lambda)[d\mathcal{E}(\lambda)g - \Psi d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g}] \right\| \leq C\left(\frac{\log(L)}{L}\left[\|g\| + \|\Psi G^{-\frac{1}{2}}\| \|G^{\frac{1}{2}}\boldsymbol{g}\|_2 \right] + \|g - \Psi\boldsymbol{g}\| \|\phi\|_\infty \right.$$
$$\left. + \sum_{1 \leq l \leq L} \frac{\left[\|\mathcal{K}^l g - \Psi \mathbb{K}^l \boldsymbol{g}\| + \|(\mathcal{K}^*)^l g - \Psi \mathbb{K}^{-l}\boldsymbol{g}\| \right]}{l} \right),$$

where C *is a universal constant.*

Proof. Consider the Laurent series of ϕ, $\phi(\lambda) = \sum_{l=-\infty}^{\infty} c_l \lambda^l$, and let

$$S_L\phi(\lambda) = \sum_{|l| \leq L} c_l \lambda^l, \quad \text{where } c_l = \frac{1}{2\pi i} \int_{\mathbb{T}} \lambda^{-(l+1)} \phi(\lambda)\, d\lambda.$$

For $|l| \geq 1$, since ϕ is Lipschitz continuous with Lipschitz constant bounded by 1, $|c_l| \lesssim 1/|l|$. Arguing as in the proof of Proposition 2.1,

$$\int_{\mathbb{T}} \lambda^l\, d\mathcal{E}(\lambda)g = \begin{cases} \mathcal{K}^l g & \text{if } l \geq 0, \\ (\mathcal{K}^*)^{-l} g & \text{otherwise.} \end{cases}$$

Arguing directly, we see that $\Psi \int_{\mathbb{T}} \lambda^l \, d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g} = \Psi \mathbb{K}^l \boldsymbol{g}$. It follows that

$$\begin{aligned} &\left\| \int_{\mathbb{T}} S_L\phi(\lambda)\, d\mathcal{E}(\lambda) g - \Psi \int_{\mathbb{T}} S_L\phi(\lambda)\, d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g} \right\| \\ &\qquad \lesssim \|g - \Psi\boldsymbol{g}\| \|\phi\|_\infty + \sum_{1\le l\le L} \frac{1}{l} \left[\|\mathcal{K}^l g - \Psi\mathbb{K}^l\boldsymbol{g}\| + \|(\mathcal{K}^*)^l g - \Psi\mathbb{K}^{-l}\boldsymbol{g}\| \right]. \end{aligned} \tag{A.1}$$

Let $\hat{\phi} = \phi - \phi(0)$, then since the Lipschitz constant of ϕ is bounded by 1, $\|\hat{\phi}\|_\infty \le \pi$. Since $\|\phi - S_L\phi\|_\infty = \|\hat{\phi} - S_L\hat{\phi}\|_\infty \lesssim \log(L)/L$ [44, Chapter I.3], it follows that

$$\left\| \int_{\mathbb{T}} (\phi - S_L\phi)(\lambda)\, d\mathcal{E}(\lambda) g \right\| \lesssim \frac{\log(L)}{L} \|g\|. \tag{A.2}$$

For functions f defined on $\mathbb{T}$,

$$\Psi \int_{\mathbb{T}} f(\lambda)\, d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g} = \Psi \sum_{j=1}^{N} f(\lambda_j) v_j v_j^* G\boldsymbol{g} = \Psi G^{-1/2} f(U_2U_1^*) G^{1/2}\boldsymbol{g},$$

where we have used the fact that the eigenvectors of $U_2U_1^*$ are $G^{1/2}v_j$. It follows that

$$\Psi \int_{\mathbb{T}} (\phi - S_L\phi)(\lambda)\, d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g} = \Psi G^{-1/2} (\phi - S_L\phi)(U_2U_1^*) G^{1/2}\boldsymbol{g}.$$

Since $U_2U_1^*$ is unitary, $\|(\phi - S_L\phi)(U_2U_1^*)\| \le \|\phi - S_L\phi\|_\infty$. It follows that

$$\left\| \Psi \int_{\mathbb{T}} (\phi - S_L\phi)(\lambda)\, d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g} \right\| \lesssim \frac{\log(L)}{L} \|\Psi G^{-1/2}\| \|G^{1/2}\boldsymbol{g}\|_2. \tag{A.3}$$

Theorem A.1 follows by combining (A.1), (A.2), and (A.3). ∎

The following lemma shows that the first summation term in Theorem A.1 converges to zero as $N \to \infty$ if the sequence of vector spaces is dense, and that the second summation term also converges to zero if, in addition, $\mathcal{K}$ is unitary. This result shows strong operator convergence of $\mathbb{K}^l$.

LEMMA A.2. *Suppose that* $\lim_{N\to\infty} \operatorname{dist}(h, V_N) = 0$ *for all* $h \in L^2(\Omega,\omega)$ *and* (3.4) *holds. Then for any* $g \in L^2(\Omega,\omega)$ *and* $\boldsymbol{g}_N \in \mathbb{C}^N$ *with* $\lim_{N\to\infty} \|g - \Psi\boldsymbol{g}_N\| = 0$,

$$\lim_{N\to\infty} \limsup_{M\to\infty} \|\mathcal{K}^l g - \Psi\mathbb{K}^l\boldsymbol{g}_N\| = 0 \quad \forall l \in \mathbb{N}. \tag{A.4}$$

If, in addition, $\mathcal{K}$ *is unitary, then*

$$\lim_{N\to\infty} \limsup_{M\to\infty} \|(\mathcal{K}^*)^l g - \Psi\mathbb{K}^{-l}\boldsymbol{g}_N\| = 0 \quad \forall l \in \mathbb{N}. \tag{A.5}$$

Proof. Recall that $\mathcal{P}_N$ is the orthogonal projection onto V_N so that $\mathcal{P}_N\mathcal{P}_N^*$ is the identity on V_N. For notational convenience, let $\mathcal{Q}_N = \mathcal{P}_N^*\mathcal{P}_N$. The assumption that $\lim_{N\to\infty} \operatorname{dist}(h, V_N) = 0$ for all $h \in L^2(\Omega,\omega)$ implies that $\mathcal{Q}_N$ converges strongly to the identity on $L^2(\Omega,\omega)$, denoted by I. It follows that $\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N$ converges strongly to $\mathcal{K}$ and that $(\mathcal{Q}_N\mathcal{K}^*\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N)^{1/2}$ converges strongly to $(\mathcal{K}^*\mathcal{K})^{1/2} = I$.

To prove (A.4), we may assume without loss of generality, by taking subsequences if necessary, that the large data limit $\lim_{M\to\infty} \mathbb{K}$ exists for each fixed N. Let $\mathcal{K}_N$ denote the operator on V_N represented by $\lim_{M\to\infty} \mathbb{K}$. Proposition 4.1(i) shows that

$$\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N(\mathcal{Q}_N\mathcal{K}^*\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N)^{1/2} = \mathcal{Q}_N\mathcal{K}\mathcal{Q}_N.$$

Let $h \in L^2(\Omega,\omega)$ and $\epsilon > 0$. Choose $N_0 \in \mathbb{N}$ so that if $N \geq N_0$, then $\|\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N h - \mathcal{K}h\| \leq \epsilon$ and $\|(\mathcal{Q}_N\mathcal{K}^*\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N)^{1/2}h - h\| \leq \epsilon$. It follows that if $N \geq N_0$, then

$$\|\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N h - \mathcal{K}h\| \leq \|\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N h - \mathcal{K}h\| + \|\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N[(\mathcal{Q}_N\mathcal{K}^*\mathcal{Q}_N\mathcal{K}\mathcal{Q}_N)^{1/2}h - h]\| \leq 2\epsilon,$$

where we have used the fact that $\|\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N\| \leq 1$. Since $h \in L^2(\Omega,\omega)$ and $\epsilon > 0$ were arbitrary, it follows that $\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N$ converges strongly to $\mathcal{K}$ as $N \to \infty$ and hence $[\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N]^l$ converges strongly to $\mathcal{K}^l$ for any $l \in \mathbb{N}$. Let $g_N = \Psi\boldsymbol{g}_N$, then

$$\lim_{M\to\infty} \Psi\mathbb{K}^l\boldsymbol{g}_N = \mathcal{P}_N^*\mathcal{K}_N^l\mathcal{P}_N g_N = [\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N]^l g_N \quad \forall l \in \mathbb{N}.$$

Since g_N converges to g, $[\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N]^l$ converges strongly to $\mathcal{K}^l$, and all relevant operators are uniformly bounded, the limit in (A.4) holds.

Now suppose that $\mathcal{K}$ is unitary so that $\mathcal{K}\mathcal{K}^*$ is the identity. Again, we may assume without loss of generality that $\lim_{M\to\infty}\mathbb{K}$ exists for each fixed N. Let $\mathcal{K}_N$ denote the operator on V_N represented by $\lim_{M\to\infty}\mathbb{K}$. Since $\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N$ converges strongly to $\mathcal{K}$ as $N \to \infty$ for all $h \in L^2(\Omega,\omega)$, we must have

$$\limsup_{N\to\infty} \|[\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N]^*\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N\mathcal{K}^*h - [\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N]^*h\| \leq \lim_{N\to\infty} \|\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N\mathcal{K}^*h - h\| = 0.$$

Since $\mathcal{K}_N^*\mathcal{K}_N = \mathcal{P}_N\mathcal{P}_N^*$ are the identity on V_N, $[\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N]^*\mathcal{P}_N^*\mathcal{K}_N\mathcal{P}_N\mathcal{K}^*h = \mathcal{Q}_N\mathcal{K}^*h$ converges to $\mathcal{K}^*h$. It follows that $\mathcal{P}_N^*\mathcal{K}_N^*\mathcal{P}_N$ converges strongly to $\mathcal{K}^*$ as $N \to \infty$. Since $\lim_{M\to\infty}\Psi\mathbb{K}^{-l}\boldsymbol{g}_N = \mathcal{P}_N^*(\mathcal{K}_N^*)^l\mathcal{P}_N g_N$, we argue as before to show (A.5). ∎

Using Lemma A.2, we now prove Theorem 5.1.

Proof of Theorem 5.1. By rescaling, we may assume without loss of generality that the Lipschitz constant of ϕ is bounded by 1. We use the bound in Theorem A.1, replacing $\boldsymbol{g}$ by $\boldsymbol{g}_N$. We have $\lim_{M\to\infty}\|\Psi G^{-1/2}\| = 1$ and $\lim_{M\to\infty}\|G^{1/2}\boldsymbol{g}_N\|_2 = \|\Psi\boldsymbol{g}_N\|$. Since $\lim_{N\to\infty}\|g - \Psi\boldsymbol{g}_N\| = 0$, it follows that for any $L \in \mathbb{N}$,

$$\begin{aligned}\limsup_{N\to\infty}\limsup_{M\to\infty}&\left\|\int_{\mathbb{T}}\phi(\lambda)\,d\mathcal{E}(\lambda)g - \Psi\int_{\mathbb{T}}\phi(\lambda)\,d\mathcal{E}_{N,M}(\lambda)\boldsymbol{g}_N\right\| \\ &\lesssim \frac{\log(L)}{L}\|g\| + \limsup_{N\to\infty}\limsup_{M\to\infty}\sum_{1\leq l\leq L}\frac{1}{l}\left[\|\mathcal{K}^l g - \Psi\mathbb{K}^l\boldsymbol{g}_N\| + \|(\mathcal{K}^*)^l g - \Psi\mathbb{K}^{-l}\boldsymbol{g}_N\|\right].\end{aligned}$$

Since $L \in \mathbb{N}$ is arbitrary, to prove the theorem it is enough to show that

$$\limsup_{N\to\infty}\limsup_{M\to\infty}\|\mathcal{K}^l g - \Psi\mathbb{K}^l\boldsymbol{g}_N\| + \|(\mathcal{K}^*)^l g - \Psi\mathbb{K}^{-l}\boldsymbol{g}_N\| = 0 \quad \forall l \in \mathbb{N}.$$

This follows from Lemma A.2. ∎

Proof of Theorem 5.3. The proof is almost identical to that of Theorem A.1. Let $\phi : \mathbb{T} \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constant bounded by 1. Since μ_g and $\mu_{\boldsymbol{g}}^{(N,M)}$ are probability measures, we may assume that $\phi(0) = 0$. Moreover,

$$\int_{\mathbb{T}}\lambda^l\,d\mu_g(\lambda) = \begin{cases}\langle\mathcal{K}^l g, g\rangle & \text{if } l \geq 0,\\ \langle(\mathcal{K}^{-l})^* g, g\rangle = \langle g, \mathcal{K}^{-l}g\rangle & \text{otherwise}\end{cases}$$

and $\boldsymbol{g}^*G\int_{\mathbb{T}}\lambda^l\,d\mu_{\boldsymbol{g}}^{(N,M)}(\lambda)\boldsymbol{g} = \boldsymbol{g}^*G\mathbb{K}^l\boldsymbol{g} = \boldsymbol{g}^*G^{1/2}[U_2U_1^*]^lG^{1/2}\boldsymbol{g}$. In particular, if $l < 0$, then $\boldsymbol{g}^*G\int_{\mathbb{T}}\lambda^l\,d\mu_{\boldsymbol{g}}^{(N,M)}(\lambda)\boldsymbol{g} = \overline{\boldsymbol{g}^*G\mathbb{K}^{|l|}\boldsymbol{g}}$. It follows that

$$\left|\int_{\mathbb{T}}\lambda^l\,d(\mu_g - \mu_{\boldsymbol{g}}^{(N,M)})(\lambda)\right| = \left|\langle\mathcal{K}^{|l|}g, g\rangle - \boldsymbol{g}^*G\mathbb{K}^{|l|}\boldsymbol{g}\right| \quad \forall l \in \mathbb{Z}.$$

Arguing as in the proof of Theorem A.1, it follows that

$$\left|\int_{\mathbb{T}} S_L\phi(\lambda)\, d(\mu_g - \mu_{\boldsymbol{g}}^{(N,M)})(\lambda)\right| \lesssim \sum_{1\le l\le L} \frac{1}{l}\left|\langle \mathcal{K}^{|l|}g, g\rangle - \boldsymbol{g}^* G\mathbb{K}^{|l|}\boldsymbol{g}\right|. \tag{A.6}$$

Since μ_g and $\mu_{\boldsymbol{g}}^{(N,M)}$ are probability measures and $\|\phi - S_L\phi\|_\infty \lesssim \log(L)/L$,

$$\left|\int_{\mathbb{T}} (\phi - S_L\phi)(\lambda)\, d(\mu_g - \mu_{\boldsymbol{g}}^{(N,M)})(\lambda)\right| \lesssim \frac{\log(L)}{L}. \tag{A.7}$$

The result follows by combining (A.6) and (A.7) and taking suprema over such ϕ. ∎

Proof of Theorem 5.6. To prove (5.6), we may assume without loss of generality, by taking subsequences if necessary, that the large data limit $\lim_{M\to\infty}\mathbb{K}$ exists for each fixed N. Let $\mathcal{K}_N$ denote the operator on V_N represented by $\lim_{M\to\infty}\mathbb{K}$.

Let $\delta > 0$ and let $\{z_1,\dots,z_k\} \subset \sigma_{\mathrm{ap}}(\mathcal{K})$ be such that $\operatorname{dist}(\lambda, \{z_1,\dots,z_k\}) \le \delta$ for any $\lambda \in \sigma_{\mathrm{ap}}(\mathcal{K})$. Such a δ-net exists since $\sigma_{\mathrm{ap}}(\mathcal{K})$ is compact. For $j = 1,\dots,k$ there exists $g_j \in L^2(\Omega,\omega)$ of norm 1 such that $\|(\mathcal{K} - z_j)g_j\| \le \delta$. Since $\lim_{N\to\infty}\operatorname{dist}(h, V_N) = 0$ for any $h \in L^2(\Omega,\omega)$, we may choose $g_{j,N} = \Psi\boldsymbol{g}_{j,N} \in V_N$, each of norm 1, such that $\lim_{N\to\infty}\|g_j - g_{j,N}\| = 0$ for $j = 1,\dots,k$. Using the first part of Lemma A.2,

$$\limsup_{N\to\infty}\|(\mathcal{K}_N - z_j)g_{j,N}\| = \limsup_{N\to\infty}\lim_{M\to\infty}\|\Psi(\mathbb{K} - z_j)\boldsymbol{g}_{j,N}\| = \|(\mathcal{K} - z_j)g_j\| \le \delta.$$

Since $\mathcal{K}_N$ is unitary, $\limsup_{N\to\infty}\operatorname{dist}(z_j, \sigma(\mathcal{K}_N)) \le \delta$ and hence

$$\limsup_{N\to\infty}\limsup_{M\to\infty}\operatorname{dist}(z_j, \sigma(\mathbb{K})) = \limsup_{N\to\infty}\operatorname{dist}(z_j, \sigma(\mathcal{K}_N)) \le \delta.$$

Since $\sup_{\lambda\in\sigma_{\mathrm{ap}}(\mathcal{K})}\operatorname{dist}(\lambda, \sigma(\mathbb{K})) \le \sup_{j=1,\dots,k}\operatorname{dist}(z_j, \sigma(\mathbb{K})) + \delta$, we have

$$\limsup_{N\to\infty}\limsup_{M\to\infty}\sup_{\lambda\in\sigma_{\mathrm{ap}}(\mathcal{K})}\operatorname{dist}(\lambda, \sigma(\mathbb{K})) \le 2\delta.$$

Since $\delta > 0$ was arbitrary, the theorem follows. ∎

Acknowledgments. I would like to thank Steve Brunton, Andrew Stuart and Alex Townsend for helpful feedback on this work, and Máté Szőke for providing the experimental PIV data for the last example.

REFERENCES

[1] H. Arbabi and I. Mezić, *Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator*, SIAM J. Appl. Dyn. Syst., 16 (2017), pp. 2096–2126.

[2] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer, New York, 1989.

[3] K. S. Arun, *A unitarily constrained total least squares problem in signal processing*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 729–745.

[4] P. J. Baddoo, B. Herrmann, B. J. McKeon, J. N. Kutz, and S. L. Brunton, *Physics-informed dynamic mode decomposition,* Proc. A, 479 (2023), 20220576.

[5] J. Ben-Artzi, M. J. Colbrook, A. C. Hansen, O. Nevanlinna, and M. Seidel, *Computing Spectra - On the Solvability Complexity Index Hierarchy and Towers of Algorithms*, preprint, arXiv:1508.03280, 2020.

[6] E. Berger, M. Sastuba, D. Vogt, B. Jung, and H. Ben Amor, *Estimation of perturbations in robotic behavior using dynamic mode decomposition*, Adv. Robot., 29 (2015), pp. 331–343.

[7] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley, New York, 1999.

[8] A. Böttcher and B. Silbermann, *The finite section method for Toeplitz operators on the quarter-plane with piecewise continuous symbols*, Math. Nachr., 110 (1983), pp. 279–291.
[9] D. Bruder, B. Gillespie, C. D. Remy, and R. Vasudevan, *Modeling and Control of Soft Robots Using the Koopman Operator and Model Predictive Control*, preprint, arXiv:1902.02827, 2019.
[10] B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz, *Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition*, J. Neurosci. Methods, 258 (2016), pp. 1–15.
[11] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, *Chaos as an intermittently forced linear system*, Nature Commun., 8 (2017), pp. 1–9.
[12] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, *Modern Koopman theory for dynamical systems*, SIAM Rev., 64 (2022), pp. 229–340.
[13] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press, Cambridge, 2019.
[14] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Natl. Acad. Sci. USA, 113 (2016), pp. 3932–3937.
[15] M. Budišić, R. Mohr, and I. Mezić, *Applied Koopmanism*, Chaos, 22 (2012), 047510.
[16] D. Burov, D. Giannakis, K. Manohar, and A. Stuart, *Kernel analog forecasting: Multiscale test problems*, Multiscale Model. Simul., 19 (2021), pp. 1011–1040.
[17] R. E. Caflisch, *Monte Carlo and quasi-Monte Carlo methods*, Acta Numer., 7 (1998), pp. 1–49.
[18] E. Celledoni, M. J. Ehrhardt, C. Etmann, R. I. McLachlan, B. Owren, C.-B. Schönlieb, and F. Sherry, *Structure-preserving deep learning*, European J. Appl. Math., 32 (2021), pp. 888–936.
[19] K. K. Chen, J. H. Tu, and C. W. Rowley, *Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses*, J. Nonlinear Sci., 22 (2012), pp. 887–915.
[20] M. J. Colbrook, *The Foundations of Infinite-Dimensional Spectral Computations*, Ph.D. thesis, University of Cambridge, Cambridge, 2020.
[21] M. J. Colbrook, *Computing spectral measures and spectral types*, Comm. Math. Phys., 384 (2021), pp. 433–501.
[22] M. J. Colbrook, *On the computation of geometric features of spectra of linear operators on Hilbert spaces*, Found. Comput. Math., (2022), pp. 1–82.
[23] M. J. Colbrook, L. J. Ayton, and M. Szőke, *Residual dynamic mode decomposition: Robust and verified Koopmanism*, J. Fluid Mech., 955 (2023), A21.
[24] M. J. Colbrook and A. C. Hansen, *The foundations of spectral computations via the solvability complexity index hierarchy*, J. Eur. Math. Soc. (JEMS), to appear.
[25] M. Colbrook, A. Horning, and A. Townsend, *Computing spectral measures of self-adjoint operators*, SIAM Rev., 63 (2021), pp. 489–524.
[26] M. J. Colbrook, B. Roman, and A. C. Hansen, *How to compute spectra with error control*, Phys. Rev. Lett., 122 (2019), 250201.
[27] M. J. Colbrook and A. Townsend, *Rigorous Data-Driven Computation of Spectral Properties of Koopman Operators for Dynamical Systems*, preprint, arXiv:2111.14889, 2021.
[28] J. B. Conway, *A Course in Functional Analysis*, Grad. Texts in Math. 96, 2nd ed., Springer, New York, 1990.
[29] S. Das, D. Giannakis, and J. Slawinska, *Reproducing kernel Hilbert space compactification of unitary evolution groups*, Appl. Comput. Harmon. Anal., 54 (2021), pp. 75–136.
[30] S. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley, *Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition*, Exp. Fluids, 57 (2016), pp. 1–19.
[31] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry — Methods and Applications Part* I, Grad. Texts in Math. 104, Springer, New York, 1984.
[32] D. Giannakis, A. Kolchinskaya, D. Krasnov, and J. Schumacher, *Koopman analysis of the long-term evolution in a turbulent convection cell*, J. Fluid Mech., 847 (2018), pp. 735–767.
[33] S. Glegg and W. Devenport, *Aeroacoustics of Low Mach Number Flows: Fundamentals, Analysis, and Measurement*, Academic Press, London, 2017.
[34] N. Govindarajan, R. Mohr, S. Chandrasekaran, and I. Mezic, *On the approximation of Koopman spectra for measure preserving transformations*, SIAM J. Appl. Dyn. Syst., 18 (2019), pp. 1454–1497.

[35] N. Govindarajan, R. Mohr, S. Chandrasekaran, and I. Mezic, *On the approximation of Koopman spectra of measure-preserving flows*, SIAM J. Appl. Dyn. Syst., 20 (2021), pp. 232–261.
[36] S. Greydanus, M. Dzamba, and J. Yosinski, *Hamiltonian neural networks*, in Advances in Neural Information Processing Systems, Vol. 32, Curran Associates, Red Hook, NY, (2019), 1378.
[37] M. Guo and J. S. Hesthaven, *Data-driven reduced order modeling for time-dependent problems*, Comput. Methods Appl. Mech. Engrg., 345 (2019), pp. 75–99.
[38] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration*, 2nd ed., Springer Ser. Comput. Math. 31, Springer, Berlin, 2006.
[39] P. R. Halmos, *What does the spectral theorem say?*, Amer. Math. Monthly, 70 (1963), pp. 241–247.
[40] P. R. Halmos, *Lectures on Ergodic Theory*, Dover Publications, Mineola, NY, 2017.
[41] Q. Hernández, A. Badías, D. González, F. Chinesta, and E. Cueto, *Structure-preserving neural networks*, J. Comput. Phys., 426 (2021), 109950.
[42] J. S. Hesthaven, C. Pagliantini, and G. Rozza, *Reduced basis methods for time-dependent problems*, Acta Numer., 31 (2022), pp. 265–345.
[43] T. L. Hill, *An Introduction to Statistical Thermodynamics*, Dover, New York, 1986.
[44] D. Jackson, *The Theory of Approximation*, Amer. Math. Soc. Colloq. Publ. 11, American Mathematical Society, New York, 1930.
[45] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, *Sparsity-promoting dynamic mode decomposition*, Phys. Fluids, 26 (2014), 024103.
[46] A. G. Kachurovskii, *The rate of convergence in ergodic theorems*, Russian Math. Surveys, 51 (1996), pp. 653–703.
[47] M. Kamb, E. Kaiser, S. L. Brunton, and J. N. Kutz, *Time-delay observables for Koopman: Theory and applications*, SIAM J. Appl. Dyn. Syst., 19 (2020), pp. 886–917.
[48] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Nonlinear Sci. Ser. 7, Cambridge University Press, Cambridge, 2004.
[49] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, *Physics-informed machine learning*, Nature Rev. Phys., 3 (2021), pp. 422–440.
[50] S. Klus, et al., *Data-driven model reduction and transfer operator approximation*, J. Nonlinear Sci., 28 (2018), pp. 985–1010.
[51] S. Klus, P. Koltai, and C. Schütte, *On the numerical approximation of the Perron-Frobenius and Koopman operator*, J. Comput. Dyn., 3 (2016), pp. 51–79.
[52] B. O. Koopman, *Hamiltonian systems and transformation in Hilbert space*, Proc. Natl. Acad. Sci. USA, 17 (1931), pp. 315–318.
[53] B. O. Koopman and J. von Neumann, *Dynamical systems of continuous spectra*, Proc. Natl. Acad. Sci. USA, 18 (1932), pp. 255–263.
[54] M. Korda and I. Mezić, *On convergence of extended dynamic mode decomposition to the Koopman operator*, J. Nonlinear Sci., 28 (2018), pp. 687–710.
[55] M. Korda, M. Putinar, and I. Mezić, *Data-driven spectral analysis of the Koopman operator*, Appl. Comput. Harmon. Anal., 48 (2020), pp. 599–629.
[56] N. Kryloff and N. Bogoliouboff, *La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire*, Ann. of Math. (2), (1937), pp. 65–113.
[57] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, Other Titles Appl. Math. 149, SIAM, Philadelphia, 2016.
[58] M. Lewin and É. Séré, *Spectral pollution and how to avoid it*, Proc. Lond. Math. Soc. (3), 100 (2010), pp. 864–900.
[59] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis, *Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator*, Chaos, 27 (2017), 103111.
[60] E. N. Lorenz, *Deterministic nonperiodic flow*, J. Atmosph. Sci., 20 (1963), pp. 130–141.
[61] B. Lusch, J. N. Kutz, and S. L. Brunton, *Deep learning for universal linear embeddings of nonlinear dynamics*, Nature Commun., 9 (2018), pp. 1–10.
[62] S. Luzzatto, I. Melbourne, and F. Paccaut, *The Lorenz attractor is mixing*, Comm. Math. Phys., 260 (2005), pp. 393–401.
[63] J. Mann and J. N. Kutz, *Dynamic mode decomposition for financial trading strategies*, Quant. Finance, 16 (2016), pp. 1643–1655.
[64] A. Mauroy and I. Mezić, *Global stability analysis using the eigenfunctions of the Koopman operator*, IEEE Trans. Automat. Control, 61 (2016), pp. 3356–3369.

[65] A. Mauroy, I. Mezić, and J. Moehlis, *Isostables, isochrons, and Koopman spectrum for the action-angle representation of stable fixed point dynamics*, Phys. D, 261 (2013), pp. 19–30.
[66] I. Mezić, *Spectral properties of dynamical systems, model reduction and decompositions*, Nonlinear Dynam., 41 (2005), pp. 309–325.
[67] I. Mezić, *Analysis of fluid flows via spectral properties of the Koopman operator*, Annu. Rev. Fluid Mech., 45 (2013), pp. 357–378.
[68] I. Mezić, *On applications of the spectral theory of the Koopman operator in dynamical systems and control theory*, in 2015 54th IEEE CDC, IEEE, Piscataway, NJ, 2015, pp. 7034–7041.
[69] I. Mezić, *Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry*, J. Nonlinear Sci., 30 (2020), pp. 2091–2145.
[70] I. Mezić, *Koopman operator, geometry, and learning of dynamical systems*, Notices Amer. Math. Soc., 68 (2021), pp. 1087–1105.
[71] I. Mezić and A. Banaszuk, *Comparison of systems with complex behavior*, Phys. D, 197 (2004), pp. 101–133.
[72] T. Murata, K. Fukami, and K. Fukagata, *Nonlinear mode decomposition with convolutional neural networks for fluid dynamics*, J. Fluid Mech., 882 (2020), A13.
[73] B. S. Nagy, C. Foias, H. Bercovici, and L. Kérchy, *Harmonic Analysis of Operators on Hilbert Space*, Springer, New York, 2010.
[74] S. Pan and K. Duraisamy, *On the structure of time-delay embedding in linear models of non-linear dynamical systems*, Chaos, 30 (2020), 073135.
[75] J. L. Proctor, S. L. Brunton, and J. N. Kutz, *Dynamic mode decomposition with control*, SIAM J. Appl. Dyn. Syst., 15 (2016), pp. 142–161.
[76] J. L. Proctor and P. A. Eckhoff, *Discovering dynamic patterns from infectious disease data using dynamic mode decomposition*, Internat. Health, 7 (2015), pp. 139–145.
[77] M. Reed and B. Simon, *Methods of Modern Mathematical Physics.* I, 2nd ed., Academic Press, New York, 1980.
[78] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, *Spectral analysis of nonlinear flows*, J. Fluid Mech., 641 (2009), pp. 115–127.
[79] P. J. Schmid, *Dynamic mode decomposition of numerical and experimental data*, J. Fluid Mech., 656 (2010), pp. 5–28.
[80] M. Schmidt and H. Lipson, *Distilling free-form natural laws from experimental data*, Science, 324 (2009), pp. 81–85.
[81] P. H. Schönemann, *A generalized solution of the orthogonal Procrustes problem*, Psychometrika, 31 (1966), pp. 1–10.
[82] C. R. Schwantes and V. S. Pande, *Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9*, J. Chem. Theory Comput., 9 (2013), pp. 2000–2009.
[83] P. C. Shields, *The Theory of Bernoulli Shifts*, University of Chicago Press, Chicago, 1973.
[84] M. Szőke, N. Nurani Hari, W. J. Devenport, S. A. Glegg, and T.-R. Teschner, *Flow field analysis around pressure shielding structures*, in AIAA Aviation 2021 (2021), 2293.
[85] L. N. Trefethen, *Exactness of quadrature formulas*, SIAM Rev., 64 (2022), pp. 132–150.
[86] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.
[87] L. N. Trefethen and J. A. C. Weideman, *The exponentially convergent trapezoidal rule*, SIAM Rev., 56 (2014), pp. 385–458.
[88] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, *On dynamic mode decomposition: Theory and applications*, J. Comput. Dyn., 1 (2014), pp. 391–421.
[89] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers Appl. Math. 9, SIAM, Philadelphia, 1991.
[90] P. Walters, *An Introduction to Ergodic Theory*, Springer, New York, 2000.
[91] M. Webb and S. Olver, *Spectra of Jacobi operators via connection coefficient matrices*, Comm. Math. Phys., 382 (2021), pp. 657–707.
[92] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, *A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition*, J. Nonlinear Sci., 25 (2015), pp. 1307–1346.
[93] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, *A kernel-based method for data-driven Koopman spectral analysis*, J. Comput. Dyn., 2 (2015), pp. 247–265.
[94] G. M. Zaslavsky, *Chaos, fractional kinetics, and anomalous transport*, Phys. Rep., 371 (2002), pp. 461–580.
[95] Z. Zhao and D. Giannakis, *Analog forecasting with dynamics-adapted kernels*, Nonlinearity, 29 (2016), 2888.