Chapter 4

The multiverse of dynamic mode decomposition algorithms

Matthew J. Colbrook
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom
e-mail address: m.colbrook@damtp.cam.ac.uk

Contents

1 Introduction 129
2 The basics of DMD 134
2.1 The underlying theory: Koopman operators and spectra 134
2.1.1 What is a Koopman operator? 134
2.1.2 Crash course on spectral properties of Koopman operators 136
2.2 The fundamental DMD algorithm 141
2.2.1 The linear regression interpretation 141
2.2.2 The Galerkin interpretation: connecting to Koopman operators 143
2.2.3 The Koopman mode decomposition 145
2.3 Three canonical examples 145
2.3.1 Flow past a cylinder wake 146
2.3.2 Lorenz system 147
2.3.3 Duffing oscillator 150
2.4 The goals and challenges of DMD 152
3 Variants from the regression perspective 153
3.1 Increasing robustness to noise 154
3.1.1 The problem of noise 154
3.1.2 Forward-backward dynamic mode decomposition (fbDMD) 155
3.1.3 Total least-squares dynamic mode decomposition (tlsDMD) 156
3.1.4 Optimized dynamic mode decomposition (optDMD) 157
3.1.5 Examples 158
3.2 Compression and randomized linear algebra 160
3.2.1 Compressed sensing meets DMD (cDMD and csDMD) 161
3.2.2 Randomized dynamic mode decomposition (rDMD) 163
3.2.3 Examples 165
3.3 Multiresolution dynamic mode decomposition (mrDMD) 166
3.3.1 The algorithm 167
3.3.2 Example 169
3.3.3 Nonautonomous systems 170
3.4 Control 170
3.4.1 Dynamic mode decomposition with control (DMDc) 171
3.4.2 Example 172

Handbook of Numerical Analysis, Volume 25, ISSN 1570-8659, https://doi.org/10.1016/bs.hna.2024.05.004
Copyright © 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

3.4.3 Extensions and connection with Koopman operators 173
4 Variants from the Galerkin perspective 174
4.1 Nonlinear observables: extended dynamic mode decomposition (EDMD) 175
4.1.1 The algorithm 176
4.1.2 Choices of dictionary 179
4.1.3 Convergence theory 181
4.1.4 Infinitesimal generators 183
4.1.5 Example 184
4.2 Time-delay embedding 185
4.2.1 Hankel dynamic mode decomposition (Hankel-DMD) 186
4.2.2 Hankel alternative view of Koopman (HAVOK) 187
4.3 Controlling projection errors: residual dynamic mode decomposition (ResDMD) 189
4.3.1 The algorithm 189
4.3.2 Convergence theory 192
4.3.3 Examples 194
5 Variants that preserve structure 195
5.1 Physics-informed dynamic mode decomposition (piDMD) 196
5.1.1 The framework 196
5.1.2 Examples 197
5.1.3 Future work 197
5.2 Measure-preserving extended dynamic mode decomposition (mpEDMD) 199
5.2.1 The algorithm 199
5.2.2 Convergence theory 201
5.2.3 Examples 203
5.3 Compactification methods for continuous-time systems 205
5.3.1 Example 207
5.4 Further methods 209
6 Further topics and open problems 210
6.1 Transfer operators 210
6.2 Continuous spectra and spectral measures 211
6.3 Stochastic dynamical systems 212
6.4 Some open problems 213
Acknowledgments 215
References 215

Abstract

Dynamic Mode Decomposition (DMD) is a popular data-driven analysis technique used to decompose complex, nonlinear systems into a set of modes, revealing underlying patterns and dynamics through spectral analysis. This review presents a comprehensive and pedagogical examination of DMD, emphasizing the role of Koopman operators in transforming complex nonlinear dynamics into a linear framework. A distinctive feature of this review is its focus on the relationship between DMD and the spectral properties of Koopman operators, with particular emphasis on the theory and practice of DMD algorithms for spectral computations. We explore the diverse "multiverse" of DMD methods, categorized into three main areas: linear regression-based methods, Galerkin approximations, and structure-preserving techniques. Each category is studied for its unique contributions and challenges, providing a detailed overview of significant algorithms and their applications as outlined in Table 1. We include a MATLAB® package with examples and applications to enhance the practical understanding of these methods. This review serves as both a practical guide and a theoretical reference for various DMD methods, accessible to both experts and newcomers, and enabling readers to explore their areas of interest in the expansive field of DMD.

Keywords

Dynamical systems, Koopman operator, Data-driven discovery, Dynamic mode decomposition, Spectral theory, Spectral computations

MSC Codes
37A30, 37Mxx, 37Nxx, 47A10, 47B33, 65J10, 65P99

1 Introduction

Dynamical systems provide a powerful framework for modeling the evolution of various scientific and engineering systems over time. They are crucial for understanding complex phenomena ranging from weather patterns and population growth to stock market fluctuations. We consider discrete-time dynamical systems represented as:

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n), \qquad n = 0, 1, 2, \ldots, \tag{1.1}$$

where $\mathbf{x} \in \Omega$ denotes the state of the system, and $\Omega \subseteq \mathbb{R}^d$ is the state space. The function $\mathbf{F} : \Omega \to \Omega$ governs the system's evolution. The classical approach to analyzing such systems, tracing back over a century to the seminal work of Poincaré (1899), is geometric. It involves local analysis of fixed points, periodic orbits, and stable or unstable manifolds. While Poincaré's framework has significantly advanced our understanding of dynamical systems, it faces two main challenges in modern applications:

- **Global understanding of nonlinear dynamics:** Unlike linear systems, there is no comprehensive mathematical framework for nonlinear systems. The principle of linear superposition is not applicable in this context. Local models can predict long-term dynamics near fixed points and attracting manifolds but have limited predictive power for other initial conditions. Consequently, the global understanding of nonlinear dynamics in state space is predominantly qualitative.
- **Incomplete knowledge of evolution:** Many systems cannot be analytically described due to their complexity or our incomplete understanding. Typically, our knowledge is limited to discrete-time snapshots of the system, i.e., a finite dataset

$$\left\{\mathbf{x}^{(m)}, \mathbf{y}^{(m)}\right\}_{m=1}^{M} \quad \text{such that} \quad \mathbf{y}^{(m)} = \mathbf{F}(\mathbf{x}^{(m)}), \quad m = 1, \ldots, M.$$

  We concisely write this data in the form of snapshot matrices

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \cdots & \mathbf{x}^{(M)} \end{pmatrix} \in \mathbb{R}^{d \times M},$$
$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}^{(1)} & \mathbf{y}^{(2)} & \cdots & \mathbf{y}^{(M)} \end{pmatrix} \in \mathbb{R}^{d \times M}. \tag{1.2}$$

  Advances in measurement technologies have significantly enhanced our ability to collect detailed multimodal and multi-fidelity snapshot data. Data could be collected from one long trajectory or multiple shorter trajectories. It can come from experimental observations or numerical simulations. The question

becomes how to use this data to meaningfully study the dynamical system in (1.1).

The advent of big data (Hey et al., 2009), coupled with strides in modern statistical learning (Hastie et al., 2009) and machine learning (Mohri et al., 2018), has heralded a new era of data-driven algorithms to address these issues. This review will focus on one of the most prominent of these algorithms, *Dynamic Mode Decomposition* (DMD), closely connected with *Koopman operators*.

Koopman Operators — In 1931, Koopman introduced his operator-theoretic approach to dynamical systems, initially to describe Hamiltonian systems (Koopman, 1931). This theory was further expanded by Koopman and von Neumann (1932) to include systems with continuous spectra. Koopman operators offer a powerful alternative to the classical geometric view of dynamical systems by addressing the fundamental issue of *nonlinearity*. We lift a nonlinear system (1.1) into an infinite-dimensional space of observable functions $g : \Omega \to \mathbb{C}$ using a Koopman operator $\mathcal{K}$:

$$[\mathcal{K}g](\mathbf{x}) = g(\mathbf{F}(\mathbf{x})), \quad \text{so that} \quad [\mathcal{K}g](\mathbf{x}_n) = g(\mathbf{x}_{n+1}).$$

Through this approach, the evolution dynamics become linear, enabling the use of generic solution techniques based on spectral decompositions. Initially, the primary application of Koopman operators was in ergodic theory (Eisner et al., 2015), notably playing a pivotal role in proving the ergodic theorem by von Neumann (Neumann, 1932) and Birkhoff (Birkhoff, 1931; Birkhoff and Koopman, 1932). More recently, they have been extensively used in data-driven methods for studying dynamical systems.

Dynamic Mode Decomposition — A significant objective of modern Koopman operator theory is to identify a coordinate transformation under which even strongly nonlinear dynamics may be approximated by a linear system. This coordinate system is related to the spectrum of the Koopman operator. DMD was initially developed by Schmid (2009, 2010) (see also Schmid and Sesterhenn, 2008) in the context of fluid dynamics. Mezić (2005) introduced the Koopman mode decomposition, providing a theoretical basis for Rowley et al. (2009) to connect DMD with Koopman operators. This connection validated DMD's application in nonlinear systems and offered a powerful yet straightforward, data-driven approach for approximating Koopman operators. The fusion of contemporary Koopman theory with an efficient numerical algorithm has led to significant advancements and a surge in research. DMD is now the central algorithm for computational approximations of Koopman operators with applications in various fields beyond fluid mechanics, such as neuroscience, disease modeling, robotics, video processing, power grids, financial markets, and plasma physics. The simplicity and effectiveness of DMD have led to numerous innovations, giving rise to a diverse array of DMD methods, playfully described here as a "multiverse", aimed at addressing specific challenges.

This Review — We provide a comprehensive tour of this "multiverse" of DMD methods. Our primary focus is on the interplay between DMD, the spec-

tral properties of Koopman operators, and their numerical computations. At the time of writing, these methods can be broadly categorized into three main areas:

- DMD methods based on linear regression;
- DMD methods utilizing Galerkin approximations;
- DMD methods aimed at preserving structures or symmetries of (1.1).

These distinctions are not rigid, and some methods encompass multiple flavors. This review navigates these key areas and variants, summarized in Table 1, where we also highlight the unique challenges each algorithm addresses (see also Section 2.4). We provide detailed summaries and examples of these algorithms in action. Accompanying this review is a MATLAB package:

https://github.com/MColbrook/DMD-Multiverse

featuring user-friendly implementations and examples from the paper, most of which are new. We aim for readers to utilize this paper as a practical manual for various DMD methods. Although extensive, the review is structured modularly, enabling readers to selectively engage with DMD versions and topics that interest them most.

Differentiating itself from prior reviews, this review specifically focuses on the theory and practice of DMD algorithms for computing spectral properties of Koopman operators, complementing other reviews on the subject. Mezić (2013) and the more recent review of Schmid (2022) (see also Taira et al., 2017, 2020) focus on developments associated with applications in fluid dynamics. While the initial applications of Koopman and DMD techniques were in fluid problems, their utility has been demonstrated in a broader range of fields. We also briefly explore the applications of DMD in control theory, and readers seeking further exploration in this area are encouraged to read (Otto and Rowley, 2021). An excellent early review of "Applied Koopmanism" is presented by Budišić et al. (2012). More recently, Brunton et al. (2022) have provided a broad overview of Koopman operators and their applications, with connections to other fields.

This review is organized into several sections. In Section 2, we introduce the basic DMD algorithm, offering a concise introduction to Koopman operators and their spectral properties, before presenting the fundamental DMD algorithm and its two key interpretations: regression and projection. We discuss three canonical examples, followed by an examination of the goals and challenges of DMD. Section 3 focuses on variants from the regression viewpoint, including noise reduction, compression, randomized linear algebra, multiscale dynamics, and control. The connection with Koopman operators is further explored in Section 4, where we discuss nonlinear observables, time-delay embedding methods, and methods for controlling the infinite-dimensional projection error of DMD (e.g., to ensure convergence). In Section 5, we review recent methods that preserve the structures of dynamical systems. These methods often lead to greater noise resistance, improved generalization, and reduced data demands for training. We conclude in Section 6 by discussing further connections and open

TABLE 1 Executive summary of the DMD methods discussed in detail in this review. Numerous others are also discussed. The bold horizontal lines separate the different flavors of regression (top), Galerkin (middle), and structure-preserving (bottom). The fundamental DMD algorithm, exact DMD, is given in Algorithm 1. **NB:** For measure-preserving systems, discretizations that preserve the measure are crucial for convergence, recovering the correct dynamical behavior, stability, robustness to noise, and improved qualitative and long-time behavior.

DMD Method	Challenges Overcome	Key Insight/Development	Key Reference(s)
Forward-Backward DMD	Sensor noise bias.	Take geometric mean of forward and backward propagators for the data.	Dawson et al. (2016)
Total Least-Squares DMD	Sensor noise bias.	Replace least-squares problem with total least-squares problem.	Hemati et al. (2017) Dawson et al. (2016)
Optimized DMD Bagging Optimized DMD	Sensor noise bias. Optimal collective processing of snapshots.	Exponential fitting problem, solve using variable projection method. Statistical bagging sampling strategy.	Chen et al. (2012) Askham and Kutz (2018) Sashidhar and Kutz (2022)
Compressed Sensing	Computational efficiency. Temporal or spatial undersampling.	Unitary invariance of DMD extended to settings of compressed sensing (e.g., RIP, sparsity-promoting regularizers).	Tu et al. (2014a) S.L. Brunton et al. (2016b) Erichson et al. (2019a)
Randomized DMD	Computational efficiency. Memory usage.	Sketch data matrix for computations in reduced-dimensional space.	Erichson et al. (2019b)
Multiresolution DMD	Multiscale dynamics.	Filtered decomposition across scales.	Kutz et al. (2016b)
DMD with Control	Separation of unforced dynamics and actuation.	Generalized regression for globally linear control framework.	Proctor et al. (2016)

continued on next page

TABLE 1 (*continued*)

DMD Method	Challenges Overcome	Key Insight/Development	Key Reference(s)
Extended DMD	Nonlinear observables.	Arbitrary (nonlinear) dictionaries, recasting of DMD as a Galerkin method.	Williams et al. (2015a)
Hankel DMD	Delay-embedding for ergodic systems. Convergence under invariant subspace assumption.	Connection with Krylov subspace methods and Birkhoff's ergodic theorem.	Arbabi and Mezić (2017a)
HAVOK	Lack of closed linear models for chaotic systems.	Delay-embedding with chaos as forcing.	Brunton et al. (2017)
Residual DMD	Infinite-dimensional projection errors, verification (general systems). Computation of Koopman spectra (general systems). Spectral measures (measure-preserving systems).	Append EDMD with additional matrix (available from the snapshot data) to compute infinite-dimensional residuals and overcome the nonconvergence of EDMD.	Colbrook and Townsend (2023) Colbrook et al. (2023a)
Physics-Informed DMD	Preserving structure of dynamical systems. Numerous instances given in general framework.	Restrict the least-squares optimization to lie on a matrix manifold.	Baddoo et al. (2023)
Measure-Preserving Extended DMD	Measure-preserving discretizations of system. Convergence to Koopman spectral properties (including continuous spectra/spectral measures).	Alter EDMD to be measure-preserving with respect to a learned inner product. (via a polar decomposition of EDMD)	Colbrook (2023)
Compactification Methods	Continuous-time generator of measure-preserving system. Convergence to Koopman spectral properties (including continuous spectra/spectral measures). Conditioning of dictionary.	Compactification of generator or its resolvent using kernel integral operators, dictionary of kernel eigenvectors.	Das et al. (2021) Valva and Giannakis (2023)

problems. I hope the reader enjoys this tour of the DMD multiverse as much as I have enjoyed writing it!

Due to the sheer breadth and thousands of papers written on DMD, it is impossible for this review to cover every version of DMD in great detail. Significant DMD algorithms that are not discussed in their own sections are still discussed in some detail. If the reader searches this paper, they will find dozens of DMD algorithms. I have included all significant references I am aware of, but many others may not have been included. I apologize for that in advance and encourage all readers to inform the author about results that deserve more discussion.

2 The basics of DMD

To understand the DMD "multiverse", we must first study the basic DMD algorithm. We begin with Koopman operator theory, the theoretical underpinning of DMD, before moving on to the fundamental DMD algorithm and two important viewpoints. We then provide three canonical examples and discuss the goals and challenges of DMD.

2.1 The underlying theory: Koopman operators and spectra

In this section, we recall the definition of Koopman operators and equip the reader with a crash course on their relevant spectral properties. At its core, DMD is an algorithm that uses the snapshot data in (1.2) to approximate the spectral properties of Koopman operators.

2.1.1 What is a Koopman operator?

To define a Koopman operator, we begin with a space $\mathcal{F}$ of functions $g : \Omega \to \mathbb{C}$, where Ω is the state space of our dynamical system. The functions g, referred to as *observables*, serve as tools for indirectly measuring the state of the system described in (1.1). Specifically, $g(\mathbf{x}_n)$ indirectly measures the state $\mathbf{x}_n$. Koopman operators enable us to capture the time evolution of these observables through a linear operator framework. For a suitable domain $\mathcal{D}(\mathcal{K}) \subset \mathcal{F}$, we define the Koopman operator via the composition formula:

$$[\mathcal{K}g](\mathbf{x}) = [g \circ \mathbf{F}](\mathbf{x}) = g(\mathbf{F}(\mathbf{x})), \qquad g \in \mathcal{D}(\mathcal{K}). \tag{2.1}$$

In this context, $[\mathcal{K}g](\mathbf{x}_n) = g(\mathbf{F}(\mathbf{x}_n)) = g(\mathbf{x}_{n+1})$ represents the measurement of the state one time step ahead of $g(\mathbf{x}_n)$. This process effectively captures the dynamic progression of the system. The overarching concept is summarized in Fig. 1.

The key property of the Koopman operator $\mathcal{K}$ is its *linearity*. This linearity holds irrespective of whether the system's dynamics, as represented in (1.1), are linear or nonlinear. Consequently, the spectral properties of $\mathcal{K}$ become a powerful tool in analyzing the dynamical system's behavior. To study spectra, we

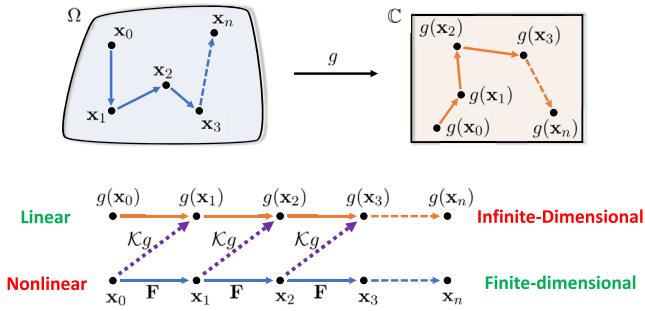

FIGURE 1 Summary of the idea of Koopman operators. By lifting to a space of observables, we trade a nonlinear finite-dimensional system for a linear infinite-dimensional system.

assume that $\mathcal{F}$ is a Banach space.[1] For the spectrum of $\mathcal{K}$ to be meaningful and nontrivial, we assume that its domain, $\mathcal{D}(\mathcal{K})$, is dense in $\mathcal{F}$ and that $\mathcal{K}$ itself is a closed operator.[2] If these conditions are not met, the spectrum would encompass the entirety of $\mathbb{C}$. It is crucial to recognize that the Koopman operator is not uniquely defined by the dynamical system in (1.1); rather, it is fundamentally dependent on the choice of the space of observables $\mathcal{F}$. In this review, we focus on cases where $\mathcal{F}$ is defined as the following Hilbert space:

$$\mathcal{F} = L^2(\Omega, \omega) \quad \text{with inner product} \quad \langle g_1, g_2 \rangle = \int_\Omega g_1(\mathbf{x})\overline{g_2(\mathbf{x})}\, d\omega(\mathbf{x}) \quad \text{and}$$
$$\text{norm } \|g\| = \sqrt{\langle g, g \rangle},$$

for some positive measure ω.[3] In going from a pointwise definition in (2.1) to the space $L^2(\Omega, \omega)$, a little care is needed since $L^2(\Omega, \omega)$ consists of equivalence classes of functions. We assume that the map $\mathbf{F}$ is nonsingular with respect to ω, meaning that

$$\omega(E) = 0 \quad \text{implies that} \quad \omega(\{\mathbf{x} : \mathbf{F}(\mathbf{x}) \in E\}) = 0.$$

This ensures that the Koopman operator is well-defined since $g_1(\mathbf{x}) = g_2(\mathbf{x})$ for ω-almost every $\mathbf{x}$ implies that $g_1(\mathbf{F}(\mathbf{x})) = g_2(\mathbf{F}(\mathbf{x}))$ for ω-almost every $\mathbf{x}$. The above Hilbert space setting is standard in most of the Koopman literature for two

[1] A Banach space is a normed vector space that is complete, i.e., every Cauchy sequence converges. Thus, a Banach space has no 'holes' in it. We have deliberately kept the background functional analysis to a minimum in this review.

[2] An operator being 'closed' means that its graph $\{(g, \mathcal{K}g) : g \in \mathcal{D}(\mathcal{K})\}$ is a closed subset within the product space $\mathcal{F} \times \mathcal{F}$.

[3] We do not assume that this measure is invariant. For Hamiltonian systems, a common choice of ω is the standard Lebesgue measure, for which the Koopman operator is unitary on $L^2(\Omega, \omega)$. For other systems, we can select ω according to the region where we wish to study the dynamics, such as a Gaussian measure. In many applications, ω corresponds to an unknown ergodic physical measure on an attractor.

reasons. First, it is a reasonable assumption for many dynamical systems, particularly if we study the dynamics on an attractor. Second, working with operators in a Hilbert space is much easier computationally than in a Banach space. For Koopman operators on more general spaces, see Mezić (2020). Practical algorithms for Koopman operators on more general Banach spaces remain a largely open problem (see Section 6.4).

Since $\mathcal{K}$ acts on an *infinite-dimensional* function space, we have exchanged the nonlinearity in (1.1) for an infinite-dimensional linear system. This means that the spectral properties of $\mathcal{K}$ can be significantly more complex than those of a finite matrix, making them more challenging to compute. While this might seem disheartening, as we will explore in Section 4, methodologies exist that enable the analysis of infinite-dimensional spectral properties through a series of finite-dimensional approximations.

2.1.2 Crash course on spectral properties of Koopman operators

We will now review the relevant spectral properties of $\mathcal{K}$. Readers primarily interested in applying DMD algorithms will still find the dynamical interpretations of these properties insightful. The sole assumption made throughout this paper is that $\mathcal{K}$ is a closed and densely defined operator. Specifically, unless stated otherwise, we do not presuppose that $\mathcal{K}$ possesses a nontrivial[4] finite-dimensional invariant subspace, nor do we assume it has an eigenvector basis. These two assumptions are often implicitly (and sometimes wrongly) assumed in DMD papers and can lead to confusion if care is not taken.

Koopman spectra

If $g \in L^2(\Omega, \omega)$ is an *eigenfunction* of $\mathcal{K}$ with *eigenvalue* λ, then g exhibits perfect coherence[5] with

$$g(\mathbf{x}_n) = [\mathcal{K}^n g](\mathbf{x}_0) = \lambda^n g(\mathbf{x}_0) \quad \forall n \in \mathbb{N}. \tag{2.2}$$

The oscillation and decay/growth of the observable g are dictated by the complex argument and absolute value of the eigenvalue λ, respectively. In infinite dimensions, the appropriate generalization of the set of eigenvalues of $\mathcal{K}$ is the *spectrum*, denoted by $\mathrm{Sp}(\mathcal{K})$, and defined as

$$\mathrm{Sp}(\mathcal{K}) = \left\{ z \in \mathbb{C} : (\mathcal{K} - zI)^{-1} \text{ does not exist as a bounded operator} \right\} \subset \mathbb{C}.$$

Here, I denotes the identity operator. The spectrum $\mathrm{Sp}(\mathcal{K})$ includes the set of eigenvalues of $\mathcal{K}$, but in general, $\mathrm{Sp}(\mathcal{K})$ contains points that are not eigenvalues.

[4] If the measure ω is finite, then the constant function $g(\mathbf{x}) = 1$ is a trivial eigenfunction with eigenvalue 1. It is deemed trivial because the dynamics of a constant observable lack useful information.

[5] In the setting of dynamical systems, coherent sets or structures are subsets of the phase space where elements (e.g., particles, agents, etc.) exhibit similar behavior over some time interval. This behavior remains relatively consistent despite potential perturbations or the chaotic nature of the system.

This is because there are more ways for $(\mathcal{K}-zI)^{-1}$ to not exist in infinite dimensions than in finite dimensions. For example, we may have continuous spectra. The standard Lorenz system on the Lorenz attractor gives rise to a Koopman operator that has no nontrivial eigenvalues, yet the spectrum is the whole unit circle!

In general, we cannot numerically approximate an eigenfunction perfectly. Moreover, the operator $\mathcal{K}$ may not have any nontrivial eigenfunctions, for instance, if the system is weakly mixing. Instead, the so-called *approximate point spectrum* is the following subset of $\mathrm{Sp}(\mathcal{K})$:

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

An observable g with $\|g\| = 1$ and $\|(\mathcal{K}-\lambda I)g\| \leq \epsilon$ for $\lambda \in \mathbb{C}$ is known as ϵ-pseudoeigenfunction. Such observables are important for the dynamical system (1.1) since

$$\|\mathcal{K}^n g - \lambda^n g\| = \mathcal{O}(n\epsilon) \quad \forall n \in \mathbb{N}.$$

In other words, λ describes an approximate coherent oscillation and decay/growth of the observable g with time. The pseudoeigenfunctions and $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$ encode information about the underlying dynamical system (Mezić, 2021). For example, the level sets of certain eigenfunctions determine the invariant manifolds (Mezić, 2015) and isostables (Mauroy et al., 2013), and the global stability of equilibria (Mauroy and Mezić, 2016) and ergodic partitions (Budišić et al., 2012; Mezić and Wiggins, 1999) can be characterized by pseudoeigenfunctions and $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$.

Koopman pseudospectra

Approximate point spectra and pseudoeigenfunctions are related to the notion of *pseudospectra* (Trefethen and Embree, 2005). For a finite matrix $A \in \mathbb{C}^{n\times n}$ and $\epsilon > 0$, the ϵ-pseudospectrum of A is the set[6]

$$\mathrm{Sp}_\epsilon(A) = \Big\{\lambda \in \mathbb{C} : \|(A-\lambda I)^{-1}\| \geq 1/\epsilon\Big\} = \bigcup_{B\in\mathbb{C}^{n\times n}, \|B\|\leq\epsilon} \mathrm{Sp}(A+B).$$

The ϵ-pseudospectra of A are regions in the complex plane enclosing the eigenvalues of A. These regions tell us how far an ϵ-sized perturbation can perturb an eigenvalue. Pseudospectra of Koopman operators must be defined with some care because $\mathcal{K}$ may be an unbounded operator and hence the resolvent norm $\|(\mathcal{K}-\lambda I)^{-1}\|$ can be constant on open subsets of $\mathbb{C}\backslash\mathrm{Sp}(\mathcal{K})$ (Shargorodsky,

[6] Some authors use a strict inequality in the definition of ϵ-pseudospectra. We prefer the given definition since then the pseudospectrum is a *closed* subset of $\mathbb{C}$.

2008). We define the ϵ-pseudospectrum of $\mathcal{K}$ as (Roch and Silbermann, 1996, Prop. 4.15)[7]:

$$\mathrm{Sp}_\epsilon(\mathcal{K}) = \mathrm{Cl}\left(\{\lambda \in \mathbb{C} : \|(\mathcal{K} - \lambda I)^{-1}\| > 1/\epsilon\}\right) = \mathrm{Cl}\left(\bigcup_{\|\mathcal{B}\|<\epsilon} \mathrm{Sp}(\mathcal{K} + \mathcal{B})\right), \tag{2.3}$$

where Cl denotes the closure of a set. To see the connection with $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$, note that if $\|(\mathcal{K} - \lambda I)g\| \leq \epsilon$ for an observable g with $\|g\| = 1$, then $\|(\mathcal{K} - \lambda I)^{-1}\| \geq 1/\epsilon$. We care about pseudospectra for several reasons, but two stand out as the most important:

- Pseudospectra allow us to determine which regions of computed spectra are accurate and trustworthy. This could be in terms of the numerical stability, but also pseudospectra aid in detecting so-called *spectral pollution* (see Figs. 7 and 9). These are spurious eigenvalues arising from discretization that are unrelated to the underlying Koopman operator. The term spectral pollution refers to the accumulation of these spurious eigenvalues at points outside the spectrum of $\mathcal{K}$ as the discretization size increases (Lewin and Séré, 2010). This occurs even when $\mathcal{K}$ is a normal operator (see Fig. 8). It is essential to realize that spectral pollution leads to spurious modes that are not linked to stability issues but are a consequence of discretizing the infinite-dimensional operator $\mathcal{K}$ to a finite matrix.
- If the Koopman operator is nonnormal, the system's transient behavior can differ significantly from the asymptotic behavior captured by $\mathrm{Sp}(\mathcal{K})$. Pseudospectra can be employed to detect and quantify transients not represented by the spectrum (Trefethen et al., 1993) (Trefethen and Embree, 2005, Section IV).

Pseudospectra also provide a means of computing spectra since $\lim_{\epsilon \downarrow 0} \mathrm{Sp}_\epsilon(\mathcal{K}) = \mathrm{Sp}(\mathcal{K})$. This convergence occurs in the so-called Attouch–Wets metric space (Beer, 1993), which roughly means that we obtain uniform convergence on any compact region of $\mathbb{C}$. This observation goes beyond Koopman operators and has been behind some recent breakthroughs in the computation of spectra in infinite dimensions (Ben-Artzi et al., 2020; Colbrook, 2020, 2022; Colbrook and Hansen, 2022; Colbrook et al., 2019).

Koopman mode decompositions and spectral theorems beyond eigenvalues

One of the most useful features of Koopman operators is the *Koopman Mode Decomposition* (KMD) (Mezić, 2005). The KMD expresses the state $\mathbf{x}$ or an observable $g(\mathbf{x})$ as a linear combination of dominant coherent structures. It can be considered a diagonalization of the Koopman operator. As a result, the KMD

[7] While Roch and Silbermann (1996, Prop. 4.15) consider bounded operators, it can be adjusted to cover unbounded operators (Trefethen and Embree, 2005, Thm. 4.3).

is invaluable for tasks such as dimensionality and model reduction. It generalizes the space-time separation of variables typically achieved through the Fourier transform or singular value decomposition (SVD). It is crucial to realize that an exact KMD is rigorously justified only if $\mathcal{K}$ possesses some form of spectral theorem, which extends the concept of diagonalization to infinite dimensions. Nevertheless, obtaining an *approximate* KMD is still possible even without a spectral theorem.

For example, suppose that the system (1.1) is *measure-preserving* with respect to the positive measure ω. This means that $\omega(E) = \omega(\{\mathbf{x} : \mathbf{F}(\mathbf{x}) \in E\})$ for any measurable set $E \subset \Omega$. In other words, the dynamical system preserves a volume. Measure-preserving systems encompass many systems of interest such as Hamiltonian flows (Arnold, 1989), geodesic flows (Dubrovin et al., 1984), Bernoulli schemes (Shields, 1973), physical systems in equilibrium (Hill, 1986), and ergodic systems (Walters, 2000). Furthermore, many dynamical systems either admit invariant measures (Kryloff and Bogoliouboff, 1937) or exhibit measure-preserving post-transient behavior (Mezić, 2005). In fact, if Ω is a compact metric space and $\mathbf{F}$ is continuous, then there is an invariant measure Mañé (1987, Prop. 8.1).[8] For a measure-preserving system, the Koopman operator $\mathcal{K}$ is an isometry, i.e., $\|\mathcal{K}g\| = \|g\|$ for all observables $g \in \mathcal{D}(\mathcal{K}) = L^2(\Omega, \omega)$. For simplicity, we further assume that $\mathcal{K}$ is unitary, implying that it is normal (it commutes with its adjoint).[9]

Under these conditions, the spectral theorem (Conway, 2007, Thm. X.4.11) allows us to diagonalize the Koopman operator $\mathcal{K}$. There is a *projection-valued measure* $\mathcal{E}$ supported on $\mathrm{Sp}(\mathcal{K})$. For readers unfamiliar with the spectral theorem, Halmos (1963) provides an excellent and readable introduction. In our example, $\mathcal{K}$ is unitary, which implies that $\mathrm{Sp}(\mathcal{K})$ lies within the unit circle $\mathbb{T}$. The measure $\mathcal{E}$ associates an orthogonal projector with each Borel measurable subset of $\mathbb{T}$. For such a subset $S \subset \mathbb{T}$, $\mathcal{E}(S)$ is a projection onto the spectral elements of $\mathcal{K}$ inside S. For any observable $g \in L^2(\Omega, \omega)$,

$$g = \left(\int_{\mathbb{T}} \mathrm{d}\mathcal{E}(\lambda) \right) g \qquad \text{and} \qquad \mathcal{K}g = \left(\int_{\mathbb{T}} \lambda \, \mathrm{d}\mathcal{E}(\lambda) \right) g.$$

The essence of this formula is the decomposition of g according to the spectral content of $\mathcal{K}$. The projection-valued measure $\mathcal{E}$ simultaneously decomposes the space $L^2(\Omega, \omega)$ and diagonalizes the Koopman operator. For example, we have

$$g(\mathbf{x}_n) = [\mathcal{K}^n g](\mathbf{x}_0) = \left[\left(\int_{\mathbb{T}} \lambda^n \, \mathrm{d}\mathcal{E}(\lambda) \right) g \right](\mathbf{x}_0). \tag{2.4}$$

[8] Of course, whether or not this is useful or whether our chosen ω is invariant is another matter.

[9] A Koopman operator that is an isometry need not be unitary, e.g., the Koopman operator associated with the tent map. However, an isometry can always be extended to a unitary operator, and the spectral measures associated with forward-time dynamics are independent of the chosen extension (Colbrook, 2023).

This directly extends (2.2). The spectral theorem can be perceived as offering a custom Fourier-type transform specifically for the operator $\mathcal{K}$ that extracts coherent features. Of particular interest are *scalar-valued* spectral measures. Given a normalized observable $g \in L^2(\Omega, \omega)$ with $\|g\| = 1$, the scalar-valued spectral measure of $\mathcal{K}$ with respect to g is a probability measure defined as

$$\mu_g(S) = \langle \mathcal{E}(S)g, g \rangle.$$

These measures can be further refined using Lebesgue's decomposition into a pure point part, supported on the eigenvalues of $\mathcal{K}$, and a continuous part. The continuous part can further be decomposed into an absolutely continuous part with a density function and a singular continuous part. The moments of the measure μ_g are the correlations

$$\langle \mathcal{K}^n g, g \rangle = \int_{\mathbb{T}} \lambda^n \, \mathrm{d}\mu_g(\lambda), \quad n \in \mathbb{Z}.$$

For example, if our system corresponds to the dynamics on an attractor, these statistical properties allow comparison of complex dynamics (Mezić and Banaszuk, 2004). More generally, the spectral measure of $\mathcal{K}$ with respect to $g \in L^2(\Omega, \omega)$ is a signature for the forward-time dynamics of (1.1).

Going one step further, $\mathcal{E}$ leads to a decomposition of $L^2(\Omega, \omega)$ into parts associated with quasiperiodic evolution and weak-mixing dynamics. Namely, we have the following orthogonal decomposition into two $\mathcal{K}$-invariant subspaces (Halmos, 2017)

$$L^2(\Omega, \omega) = \mathcal{H}_{\mathrm{pp}} \oplus \mathcal{H}_{\mathrm{c}}.$$

Here, the subspace $\mathcal{H}_{\mathrm{pp}}$ consists of the closure of the linear span of eigenvectors and admits an orthonormal basis of eigenvectors $\{\phi_j\}$ of $\mathcal{K}$ with eigenvalues $\{\lambda_j\}$. This means that we can write

$$\mathcal{K}^n g = \sum_j \lambda_j^n \langle g, \phi_j \rangle \phi_j \quad \forall g \in \mathcal{H}_{\mathrm{pp}}, n \in \mathbb{N}. \tag{2.5}$$

The spectrum of $\mathcal{K} \restriction_{\mathcal{H}_{\mathrm{pp}}}$ need not be a discrete subset of $\mathbb{T}$. For example, an ergodic rotation on the circle has eigenvalues that densely fill $\mathbb{T}$. In contrast to (2.5), observables in the continuous part $\mathcal{H}_{\mathrm{c}}$ exhibit a decay of correlations that is typical of chaotic systems. Namely, for any $\epsilon > 0$ (Katznelson, 2004, p.45),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \left| \langle \mathcal{K}^j g, f \rangle \right|^{\epsilon} = 0 \quad \forall g \in \mathcal{H}_{\mathrm{c}}, f \in L^2(\Omega, \omega).$$

This result says that $|\langle \mathcal{K}^j g, f \rangle|$ converges to zero in density, that is, for any $\delta > 0$, the proportion in all sufficiently large intervals of integers j such that $\left|\langle \mathcal{K}^j g, f \rangle\right| > \delta$ is arbitrarily small.

The above dichotomy is an example of how the decomposition of $\mathcal{E}$ into atomic and continuous parts often characterizes a dynamical system. For example, suppose that $\mathbf{F}$ is measure-preserving and bijective, and ω is a probability measure. Then, the dynamical system is (Halmos, 2017)

- **Ergodic** if and only if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$,
- **Weakly mixing** if and only if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$ and there are no other eigenvalues,
- **Mixing** if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$, and $\mathcal{K}$ has absolutely continuous spectrum on $\text{span}\{1\}^{\perp}$.

Different spectral types find interpretations across various applications, including fluid mechanics (Mezić, 2013), anomalous transport (Zaslavsky, 2002), and the analysis of invariants/exponents related to trajectories (Kantz and Schreiber, 2006). The approximation of $\mathcal{E}$ is critical in many applications. For example, the approximate spectral projections provide reduced-order models (Mezić and Banaszuk, 2004; Mezić, 2005). The computation of spectral measures is discussed in Section 6.2.

2.2 The fundamental DMD algorithm

With the definition of a Koopman operator in hand, we can now present the fundamental DMD algorithm and two interpretations. The first interpretation of DMD is as a linear regression. The second is as a projection method. Both interpretations are instrumental, and understanding their interplay is often key to unlocking the power of DMD.

2.2.1 The linear regression interpretation

The simplest and historically first interpretation of DMD is as a linear regression. Given the snapshot matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{d\times M}$ in (1.2), we seek a matrix $\mathbf{K}_{\text{DMD}}$ such that $\mathbf{Y} \approx \mathbf{K}_{\text{DMD}}\mathbf{X}$. We can think of this as constructing a *linear and approximate* dynamical system. To find a suitable matrix $\mathbf{K}_{\text{DMD}}$, we consider the minimization problem

$$\min_{\mathbf{K}_{\text{DMD}}\in\mathbb{C}^{d\times d}} \|\mathbf{Y} - \mathbf{K}_{\text{DMD}}\mathbf{X}\|_{\text{F}}, \tag{2.6}$$

where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm. Similar optimization problems will be at the heart of the various DMD-type algorithms we consider in this review. A solution to the problem in (2.6) is

$$\mathbf{K}_{\text{DMD}} = \mathbf{Y}\mathbf{X}^{\dagger} \in \mathbb{C}^{d\times d},$$

where $\dagger$ denotes the Moore–Penrose pseudoinverse. Often, the matrices $\mathbf{X}$ and $\mathbf{Y}$ are tall and skinny, meaning that $d \gg M$. In this scenario, we typically first project onto a low-dimensional subspace to reconstruct the leading nonzero eigenvalues and eigenvectors of the matrix $\mathbf{K}_{\text{DMD}}$ without explicitly computing it. The standard DMD algorithm does this using an SVD and is summarized

Algorithm 1 The exact DMD algorithm (Tu et al., 2014b), which has become the workhorse DMD algorithm.

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, rank $r \in \mathbb{N}$.

1: Compute a truncated SVD of the data matrix $\mathbf{X} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, $\mathbf{U} \in \mathbb{C}^{d\times r}$, $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$, $\mathbf{V} \in \mathbb{C}^{M\times r}$. The columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal and $\mathbf{\Sigma}$ is diagonal.
2: Compute the compression $\tilde{\mathbf{K}}_{\text{DMD}} = \mathbf{U}^*\mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1} \in \mathbb{C}^{r\times r}$.
3: Compute the eigendecomposition $\tilde{\mathbf{K}}_{\text{DMD}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$.
The columns of $\mathbf{W}$ are eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues.
4: Compute the modes $\mathbf{\Phi} = \mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}$.

Output: The eigenvalues $\mathbf{\Lambda}$ and modes $\mathbf{\Phi} \in \mathbb{C}^{d\times r}$.

in Algorithm 1, where we have assumed that the projected DMD matrix is diagonalizable.[10] Algorithm 1 is known as *exact DMD* (Tu et al., 2014b) and often the modes are further scaled by Λ^{-1}. There are several remarks about this algorithm that are worth mentioning:

- The rank r is usually chosen based on the decay of singular values of $\mathbf{X}$. If low-dimensional structure is present in the data (Udell and Townsend, 2019), the singular values decrease rapidly, and small r captures the dominant modes. Moreover, the lowest energy modes may be corrupted by noise, and low-dimensional projection is a form of spectral filtering which has the positive effect of dampening the influence of noise (Hansen et al., 2006).[11] The question of how best to truncate is difficult to answer and is often performed heuristically. If the measurement error is additive white noise, there are algorithmic choices (Gavish and Donoho, 2014). In the context of Koopman operators, r is equivalent to the size of the space spanned by basis functions, and a good choice depends on the chosen observables. For example, we shall see below that Algorithm 1 corresponds to a linear set of basis functions, which may not capture the relevant nonlinear dynamics. Hence, a larger r may be suitable for other basis choices. Often, the choice of r is modest, meaning that randomized methods (Halko et al., 2011) for computing the SVD can significantly reduce the computational cost. We will explore this and other compression methods in Section 3.2.

[10] We make this assumption about various matrices throughout. Mathematically, a Jordan decomposition may be substituted for an eigendecomposition, and the modes corresponding to a single Jordan block can be considered as interacting modes. However, computing a Jordan block should be avoided. A stable alternative is a Schur decomposition that provides an orthogonal set of interacting modes (in sharp contrast with what is typically considered a DMD mode) or a block-diagonal Schur decomposition with nearly confluent eigenvalues grouped together.

[11] Low-energy modes can be important though, for example, in optimized control (Rowley, 2005; Rowley et al., 2006).

- We can interpret the algorithm as constructing a linear model of the dynamical system on projected coordinates $\tilde{\mathbf{x}} = \mathbf{U}^*\mathbf{x}$. Namely, $\tilde{\mathbf{x}}_{n+1} \approx \tilde{\mathbf{K}}_{\text{DMD}}\tilde{\mathbf{x}}_n$. The left singular vectors $\mathbf{U}$ are known as proper orthogonal decomposition (POD) modes (Berkooz et al., 1993).
- If the SVD is exact, so that $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, then

$$\mathbf{K}_{\text{DMD}} = \mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^*.$$

  Using this relation, we have

$$\mathbf{K}_{\text{DMD}}[\mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}] = \mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\underbrace{\mathbf{U}^*\mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}}_{\tilde{\mathbf{K}}_{\text{DMD}}}\mathbf{W} = [\mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}]\mathbf{\Lambda},$$

  and hence Algorithm 1 computes exact eigenvalues and eigenvectors of $\mathbf{K}_{\text{DMD}}$. Moreover, one can show that this process identifies all of the nonzero eigenvalues of $\mathbf{K}_{\text{DMD}}$ (Tu et al., 2014b, Thm. 1). It is common to call $\mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}$ *exact modes* and $\mathbf{U}\mathbf{W}$ *projected modes*.
- Originally, DMD was developed in connection with Krylov subspaces and the Arnoldi algorithm. In this version, it is assumed that data is gathered along a single trajectory. The SVD version, on the other hand, is capable of handling more general trajectory data. Strategies for using this flexibility to reduce computational cost and average snapshot data noise are given in Tu et al. (2014b). The SVD version is also more numerically stable. Drmač has carefully analyzed the stability of DMD (Drmač et al., 2018, 2019; Drmač, 2020; Drmač et al., 2020).
- Centering the data before applying DMD can be helpful if the mean-subtracted data have linearly dependent columns, especially if the dynamics are perturbations about an equilibrium (Hirsh et al., 2020). This is equivalent to including an affine term in the linear regression. However, computing the DMD of centered data can have undesirable consequences (Chen et al., 2012).

The above interpretation of DMD is simple and intuitive. However, using DMD to analyze nonlinear dynamics globally seems dubious, as there is an underlying assumption of approximately linear dynamics in (2.6). Nevertheless, we shall now demonstrate that DMD can be interpreted as an approximation to Koopman spectral analysis. This provides a solid theoretical foundation for applying DMD in analyzing nonlinear dynamics, which will be further elaborated upon in Section 4.

2.2.2 *The Galerkin interpretation: connecting to Koopman operators*

The connection between Algorithm 1 and Koopman operators is revealed once we interpret DMD as a Galerkin method. Consider the two correlation matrices

$$\mathbf{G} = \frac{1}{M}\overline{\mathbf{U}^*\mathbf{X}}(\mathbf{U}^*\mathbf{X})^\top, \quad \mathbf{A} = \frac{1}{M}\overline{\mathbf{U}^*\mathbf{X}}(\mathbf{U}^*\mathbf{Y})^\top.$$

We can think of the jth row of the POD matrix $\mathbf{U}^*\mathbf{X}$ as an affine function u_j on the state space Ω evaluated at the snapshot data:

$$u_j(\mathbf{x}) = [\mathbf{U}_{:,j}]^*\mathbf{x}, \quad u_j(\mathbf{x}^{(m)}) = [\mathbf{U}^*\mathbf{X}]_{jm}.$$

It follows that $\mathbf{G}$ can be interpreted as a Gram matrix with respect to the positive semidefinite Hermitian form induced by the probability measure with equal point masses at the $\{\mathbf{x}^{(m)}\}$. Namely,

$$\mathbf{G}_{jk} = \frac{1}{M}\sum_{m=1}^{M} \overline{u_j(\mathbf{x}^{(m)})} u_k(\mathbf{x}^{(m)}) = \int_\Omega \overline{u_j(\mathbf{x})} u_k(\mathbf{x}) \, \mathrm{d}\omega_M(\mathbf{x}),$$
$$\omega_M = \frac{1}{M}\sum_{m=1}^{M} \delta_{\mathbf{x}^{(m)}}.$$

Writing $\langle \cdot, \cdot \rangle_M$ for the form induced by ω_M, we can argue similarly for $\mathbf{A}$ and succinctly write

$$\mathbf{G}_{jk} = \langle u_k, u_j \rangle_M,$$
$$\mathbf{A}_{jk} = \frac{1}{M}\sum_{m=1}^{M} \overline{u_j(\mathbf{x}^{(m)})} u_k(\mathbf{y}^{(m)}) = \int_\Omega \overline{u_j(\mathbf{x})} u_k(\mathbf{F}(\mathbf{x})) \, \mathrm{d}\omega_M(\mathbf{x}) = \langle \mathcal{K} u_k, u_j \rangle_M.$$

Assuming that the matrix $\mathbf{U}^*\mathbf{X}$ is of rank r, and using $\mathbf{U}^*\mathbf{X} = \mathbf{\Sigma}\mathbf{V}^*$, we can write

$$\tilde{\mathbf{K}}_{\mathrm{DMD}}^\top = \mathbf{\Sigma}^{-1}\mathbf{V}^\top\mathbf{Y}^\top\overline{\mathbf{U}} = (\mathbf{X}^\top\overline{\mathbf{U}})^\dagger\mathbf{Y}^\top\overline{\mathbf{U}} = \mathbf{G}^{-1}\mathbf{A}.$$

The matrix $\mathbf{G}^{-1}\mathbf{A}$ is an approximation of the action of $\mathcal{K}$ on the subspace spanned by the functions $\{u_j\}_{j=1}^r$. Namely, if g is an observable that can be expressed as the linear combination

$$g(\mathbf{x}) = \sum_{j=1}^{r} u_j(\mathbf{x})\mathbf{g}_j, \quad \text{for some} \quad \mathbf{g} \in \mathbb{C}^r,$$

then

$$[\mathcal{K}g](\mathbf{x}) \approx \sum_{j=1}^{r} u_j(\mathbf{x})(\mathbf{G}^{-1}\mathbf{A}\mathbf{g})_j = \sum_{j=1}^{r} u_j(\mathbf{x})(\tilde{\mathbf{K}}_{\mathrm{DMD}}^\top\mathbf{g})_j.$$

In other words, $\tilde{\mathbf{K}}_{\mathrm{DMD}}^\top$ is a matrix that approximates the action of the Koopman operator on expansion coefficients. More precisely, it is a Galerkin method corresponding to $\mathcal{K}$ and the form $\langle \cdot, \cdot \rangle_M$. This connection is explored more deeply in Section 4.1.

If $\langle \cdot, \cdot \rangle_M$ converges to $\langle \cdot, \cdot \rangle$ in the large data limit $M \to \infty$, then DMD can be considered to be a numerical approximation to Koopman spectral analysis. The

terms *DMD mode* and *Koopman mode* are often used interchangeably in the literature. It is important to note that the Koopman modes and eigenfunctions are distinct mathematical objects, requiring different approaches for approximation. The right eigenvectors of $\tilde{\mathbf{K}}_{\text{DMD}}$ give rise to time-invariant directions in the state space $\mathbf{x}$, whereas the left-eigenvectors give rise to Koopman eigenfunctions, which are similarly time-invariant directions in the space of observables.

2.2.3 The Koopman mode decomposition

We can now connect DMD with the spectral expansions discussed in Section 2.1.2. First, we approximate an initial condition $\mathbf{x}_0$ in the eigenvector coordinates via

$$\mathbf{x}_0 \approx \mathbf{\Phi}\mathbf{b}, \quad \mathbf{b} = \mathbf{\Phi}^{\dagger}\mathbf{x}_0.$$

This is not the only choice, but it is the simplest. The KMD then provides an approximation of the dynamics by

$$\mathbf{x}_n \approx \mathbf{K}_{\text{DMD}}^n \mathbf{x}_0 \approx \mathbf{K}_{\text{DMD}}^n \mathbf{\Phi}\mathbf{b} = \mathbf{\Phi}\mathbf{\Lambda}^n\mathbf{b}, \tag{2.7}$$

which echoes (2.4). Since zero eigenvalues do not contribute to the dynamics, this decomposition further justifies the compression in Algorithm 1. The KMD has also been related to other decompositions in various situations, particularly those that have arisen in the fluid dynamics community (Taira et al., 2017, 2020). These include POD (Towne et al., 2018), optimal mode decomposition (Wynn et al., 2013), and resolvent analysis (Sharma et al., 2016; Herrmann et al., 2021). Under suitable conditions, the KMD converges as we increase the dimension of the projected Koopman operator (see Section 4.1.3 and Section 5.2.2).

2.3 Three canonical examples

Having grasped the notion of Koopman operators and the basic DMD algorithm, it is time for some examples. As a warm-up for the reader, we consider three canonical well-studied examples of Algorithm 1, each with a unitary Koopman operator:

- The flow past a cylinder wake at $Re = 100$ with a state space dimension $d = 160{,}000$ that corresponds to the number of spatial measurement points in the flow. The associated Koopman operator has a pure point spectrum consisting of powers of a fundamental eigenvalue.
- The Lorenz system on the Lorenz attractor with a state space dimension $d = 3$. The associated Koopman operator possesses no eigenvalues, except for the simple eigenvalue $\lambda = 1$ whose eigenfunction is the constant function. The rest of the spectrum is continuous.
- The Duffing oscillator with a state space dimension $d = 2$. The associated Koopman operator possesses no eigenvalues, except for $\lambda = 1$, whose eigenspace now corresponds to the conserved Hamiltonian energy of the system and indicator functions associated with invariant sets of positive area.

Despite its large ambient space dimension, the first example is the easiest to address using Algorithm 1. This is because the cylinder wake exhibits an attracting limit cycle, and the Koopman operator has a basic spectrum. The other two examples demonstrate three difficulties of DMD: noise, projection error (which can lead to spurious modes and missing parts of the spectrum), and continuous spectra. These and further challenges are discussed in Section 2.4.

2.3.1 Flow past a cylinder wake

We first consider the classic DMD example of low Reynolds number flow past a circular cylinder. Due to its simplicity and relevance in engineering, this is one of the most studied examples in modal-analysis techniques (Rowley and Dawson, 2017, Table 3) (Chen et al., 2012; Taira et al., 2020). $Re = 100$ is chosen so that it is larger than the critical Reynolds number at which the flow undergoes a supercritical Hopf bifurcation, resulting in laminar vortex shedding (Jackson, 1987; Zebib, 1987). This limit cycle is stable and is representative of the three-dimensional flow (Noack and Eckelmann, 1994; Noack et al., 2003). The Koopman operator of the post-transient flow has a pure point spectrum with a lattice structure on the unit circle (Bagheri, 2013).

To collect snapshot data, we numerically compute the velocity field of a flow around a circular cylinder of diameter $D = 1$ using an incompressible, two-dimensional lattice-Boltzmann solver (Józsa et al., 2016; Szőke et al., 2017). The temporal resolution of the flow is chosen so that approximately 24 snapshots of the flow field correspond to the period of vortex shedding. The computational domain size is $18D$ in length and $5D$ in height, with a 800×200 grid resolution. The cylinder is positioned $2D$ downstream of the inlet at the mid-height of the domain. The cylinder side walls are defined as bounce-back and no-slip walls, and a parabolic velocity profile is given at the inlet of the domain. The outlet is defined as a nonreflecting outflow. After simulations converge to steady-state vortex shedding, we collect $M = 120$ snapshots for the DMD algorithm and a further 880 snapshots to test the prediction of the KMD. One should think of this as training data and test data, respectively. Letting $\mathbf{V}_x(t)$ denote the vectorized horizontal velocity field at time t, our snapshot matrices have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{V}_x(0) & \mathbf{V}_x(\Delta t) & \cdots & \mathbf{V}_x(119\Delta t) \end{pmatrix},$$
$$\mathbf{Y} = \begin{pmatrix} \mathbf{V}_x(\Delta t) & \mathbf{V}_x(2\Delta t) & \cdots & \mathbf{V}_x(120\Delta t) \end{pmatrix}.$$

We use a rank of $r = 47$ to recover the trivial mode corresponding to $\lambda = 1$ and 24 conjugate pairs of modes up to the timescale of vortex shedding. The eigenvalues come in conjugate pairs due to processing real-valued data $\mathbf{X}$ and $\mathbf{Y}$.

Fig. 2 shows the output of Algorithm 1. In the left panel, we see the lattice structure of the DMD modes correctly identified by Algorithm 1. In the middle plot, we show the predictive error of (2.7). The relative error is computed by taking the 2-norm of the error in the velocity field $\mathbf{V}_x$ and normalizing it by the 2-norm of the mean-subtracted flow at each time step. Due to the periodic

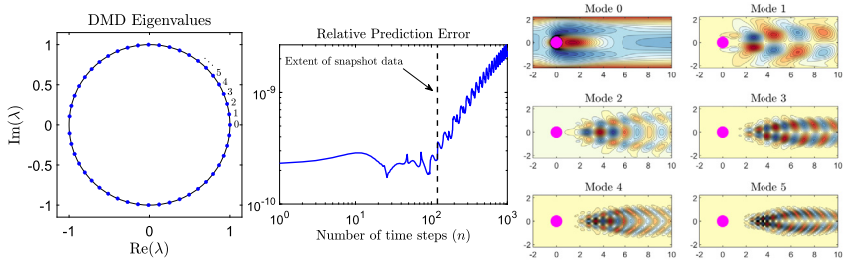

FIGURE 2 Output of Algorithm 1 for the flow past a cylinder wake. Left: The DMD eigenvalues. Middle: The total relative prediction error of (2.7). Right: Example modes of the horizontal component of the velocity field. The magenta disc corresponds to the cylinder. The zeroth mode corresponds to the time-averaged flow.

nature of the flow, there is excellent agreement between the KMD and flow, with slow algebraic growth of the error beyond the snapshot data time window. The right panel of Fig. 2 shows the real part of some of the Koopman modes for the horizontal velocity field.

2.3.2 *Lorenz system*

The Lorenz (63) system (Lorenz, 1963) is the following 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 $\mathbf{x} = (x, y, z)$ on the Lorenz attractor. The system is chaotic and strongly mixing (Luzzatto et al., 2005). It follows that $\lambda = 1$ is the only eigenvalue of $\mathcal{K}$, corresponding to a constant eigenfunction, and that this eigenvalue is simple. We consider a discrete-time dynamical system by sampling with a time-step $\Delta t = 0.001$. We use time-delay embedding, which is a popular method for DMD-type algorithms (Susuki and Mezić, 2015; Arbabi and Mezić, 2017a; Brunton et al., 2017; Das and Giannakis, 2019; Kamb et al., 2020; Pan and Duraisamy, 2020a) and corresponds to building a Krylov subspace. This technique is justified through Takens' embedding theorem (Takens, 2006), which says that under certain technical conditions, delay embedding a signal coordinate of the system can reconstruct the attractor of the original system up to a diffeomorphism. In this example, we augment $\mathbf{x}$ by $N - 1$ further time-delays of length $\Delta t' = 0.2$ and consider $M = 5 \times 10^5$ snapshots along a single trajectory. Specifically, our snapshot matrices have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}(0) & \mathbf{x}(\Delta t) & \cdots & \mathbf{x}((M-1)\Delta t) \\ \mathbf{x}(\Delta t') & \mathbf{x}(\Delta t'+\Delta t) & \cdots & \mathbf{x}(\Delta t'+(M-1)\Delta t) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}((N-1)\Delta t') & \mathbf{x}((N-1)\Delta t'+\Delta t) & \cdots & \mathbf{x}((N-1)\Delta t'+(M-1)\Delta t) \end{pmatrix} \in \mathbb{R}^{3N\times M},$$

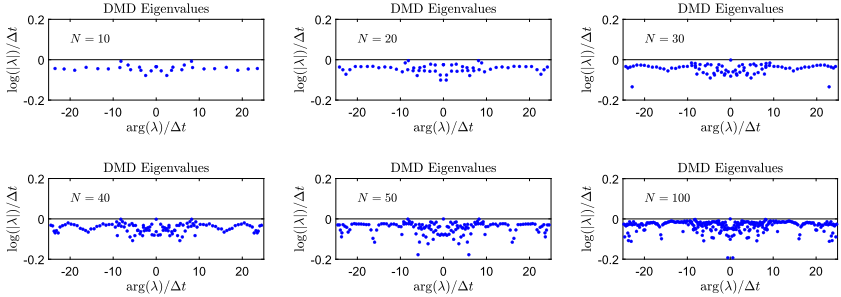

FIGURE 3 DMD eigenvalues for the Lorenz system and different choices of N (the number of eigenvalues is $3N$). The logarithm of the eigenvalues is plotted to align with the continuous-time system. As N increases, the eigenvalues cluster to approximate the continuous spectrum.

$$\mathbf{Y} = \begin{pmatrix} \mathbf{x}(\Delta t) & \mathbf{x}(2\Delta t) & \cdots & \mathbf{x}(M\Delta t) \\ \mathbf{x}(\Delta t'+\Delta t) & \mathbf{x}(\Delta t'+2\Delta t) & \cdots & \mathbf{x}(\Delta t'+M\Delta t) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}((N-1)\Delta t'+\Delta t) & \mathbf{x}((N-1)\Delta t'+2\Delta t) & \cdots & \mathbf{x}((N-1)\Delta t'+M\Delta t) \end{pmatrix} \in \mathbb{R}^{3N\times M}.$$

We use the `ode45` command in MATLAB to collect the data after an initial burn-in time to ensure that the initial point $\mathbf{x}(0)$ is (approximately) on the Lorenz attractor. The system is chaotic, so we cannot hope to integrate for long periods accurately numerically. However, convergence is still obtained in the large data limit $M \to \infty$ due to an effect known as shadowing.

In addition to a discrete time-step Δt, we can consider Koopman operators associated with continuous-time dynamical systems. The continuous-time infinitesimal generator is defined by

$$\mathcal{L}g = \lim_{\Delta t \downarrow 0} \frac{\mathcal{K}_{\Delta t} g - g}{\Delta t}, \tag{2.8}$$

where $\mathcal{K}_{\Delta t}$ is the Koopman operator corresponding to a time-step Δt. The generator satisfies

$$\mathcal{K}_{\Delta t} = \exp(\Delta t \mathcal{L}),$$

which can be made precise through the theory of semigroups (Pazy, 2010). Hence, in this example, we consider the following time-scaled logarithms of the eigenvalues:

$$\log(\lambda)/\Delta t = \log(|\lambda|)/\Delta t + i\arg(\lambda)/\Delta t.$$

Fig. 3 shows the DMD eigenvalues for various choices of N. The horizontal line corresponds to a portion of the spectrum of the Koopman operator. The DMD

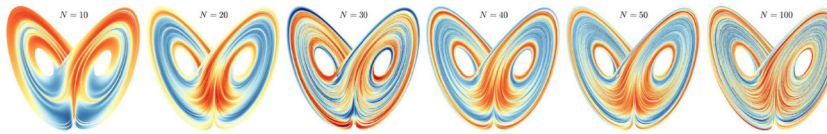

FIGURE 4 DMD eigenfunctions corresponding to $\log(\lambda)/(i\Delta t) \approx \pm 8.2$ rad/s. These are similar to the singularity in spectral measures detected in Korda et al. (2020).

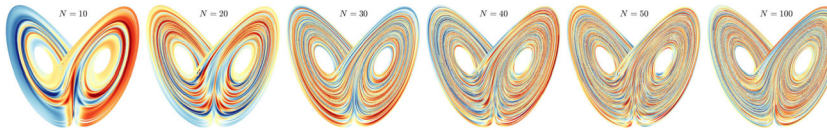

FIGURE 5 DMD eigenfunctions corresponding to $\log(\lambda)/(i\Delta t) \approx 6$ rad/s. We see increasing oscillations as N gets larger and the eigenfunctions resemble unstable periodic orbits, see also Colbrook and Townsend (2023).

eigenvalues fall below this line, corresponding to a dampening effect in the dynamics encapsulated by DMD. For these choices of parameters, this error is largely due to the finite amount of trajectory data and noise in the data matrices from the numerical solver. In Section 3.1.5, we shall see that this effect can be reduced using DMD methods designed to be robust to noise and by increasing M. Another reason for the dampening of the eigenvalues is the projection error (from projecting onto a finite matrix). Another interesting feature of Fig. 3 is the clustering of the DMD eigenvalues with increasing N as they attempt to approximate the continuous spectrum. Recall from above that the Koopman operator for this example has no eigenvalues except the trivial eigenvalue $\lambda = 1$. In Section 5.2, we shall see that *Measure-Preserving Extended DMD* (Colbrook, 2023) can deal with continuous spectra (see also the discussion in Section 6.2 for further methods).

We next show DMD eigenfunctions, but associated with the matrix $\mathbf{X}^{\dagger}\mathbf{Y}$. The discussion in Section 2.2.2 shows that these correspond to pseudoeigenfunctions of $\mathcal{K}$. Note that these pseudoeigenfunctions do not approximate true eigenfunctions - since eigenfunctions do not exist for this system! For visualization over the attractor, we plot function values along the trajectory of snapshot data. In Fig. 3 there are DMD eigenvalues close to the horizontal line with $\log(\lambda)/(i\Delta t) \approx \pm 8.2$ rad/s. These correspond to an apparent singularity in the spectral measure detected in Korda et al. (2020). Fig. 4 shows the corresponding pseudoeigenfunctions. These bear a striking resemblance to the local spectral projections in Korda et al. (2020, Figure 13), which the authors attributed to an almost-periodic motion of the z component during the time that the state resides in either of the two lobes of the Lorenz attractor. In Fig. 5, we plot the pseudoeigenfunctions corresponding to the DMD eigenvalue with $\log(\lambda)/(i\Delta t)$ closest to 6 rad/s. In this case, we see increasing oscillations as N gets larger, and the pseudoeigenfunctions resemble unstable periodic orbits, which in a sense, form a backbone of the attractor (Eckmann and Ruelle, 1985;

Tufillaro et al., 1993). For further examples of these kinds of pseudoeigenfunctions, see Colbrook and Townsend (2023).

2.3.3 *Duffing oscillator*

We now consider the Hamiltonian system

$$\dot{x} = y, \quad \dot{y} = x - x^3,$$

known as the (undamped nonlinear) Duffing oscillator with state $\mathbf{x} = (x, y) \in \Omega = \mathbb{R}^2$. This dynamical system has three fixed points at $\mathbf{x} = (0, 0)$ (a saddle), and $\mathbf{x} = (\pm 1, 0)$ (centers). The Hamiltonian for this system is $H = y^2 - x^2/2 + x^4/2$. We consider the corresponding discrete-time dynamical system by sampling with a time-step $\Delta t = 0.25$. Instead of using the state $\mathbf{x}$ or time-delay embedding to form our snapshot matrices, we consider an example of *Extended DMD* (Williams et al., 2015a), discussed in more detail in Section 4.1. Specifically, we consider 10^3 random points sampled uniformly in $[-2, 2]^2$, and then the trajectory of these points for 50 times steps. This leads to $M = 5 \times 10^4$ snapshots $\{\mathbf{x}^{(m)}, \mathbf{y}^{(m)}\}_{m=1}^{M}$. We then partition these into N clusters using k-means, and use these as centers $\mathbf{c}_j$ for N radial basis functions of the form

$$\psi_j(\mathbf{x}) = \exp(-\gamma \|\mathbf{x} - \mathbf{c}_j\|),$$

where γ is the squared reciprocal of the average ℓ^2-norm of the snapshot data after it is shifted to mean zero. Our snapshot matrices are then given by

$$\mathbf{X} = \begin{pmatrix} \psi_1(\mathbf{x}^{(1)}) & \psi_1(\mathbf{x}^{(2)}) & \cdots & \psi_1(\mathbf{x}^{(M)}) \\ \psi_2(\mathbf{x}^{(1)}) & \psi_2(\mathbf{x}^{(2)}) & \cdots & \psi_2(\mathbf{x}^{(M)}) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_N(\mathbf{x}^{(1)}) & \psi_N(\mathbf{x}^{(2)}) & \cdots & \psi_N(\mathbf{x}^{(M)}) \end{pmatrix},$$

$$\mathbf{Y} = \begin{pmatrix} \psi_1(\mathbf{y}^{(1)}) & \psi_1(\mathbf{y}^{(2)}) & \cdots & \psi_1(\mathbf{y}^{(M)}) \\ \psi_2(\mathbf{y}^{(1)}) & \psi_2(\mathbf{y}^{(2)}) & \cdots & \psi_2(\mathbf{y}^{(M)}) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_N(\mathbf{y}^{(1)}) & \psi_N(\mathbf{y}^{(2)}) & \cdots & \psi_N(\mathbf{y}^{(M)}) \end{pmatrix}.$$

Fig. 6 shows some of the Koopman eigenfunctions corresponding to $\lambda = 1$ and computed using $N = 1000$. For visualization, the values of the functions are plotted at the data points. We see that the level sets of the eigenfunctions correspond to trajectories, as expected. However, moving away from $\lambda = 1$ in the spectral plane becomes more challenging. Fig. 7 shows the DMD eigenvalues for various choices of N, along with the unit circle, which is the spectrum of the Koopman operator. Most of the DMD eigenvalues are spurious and correspond

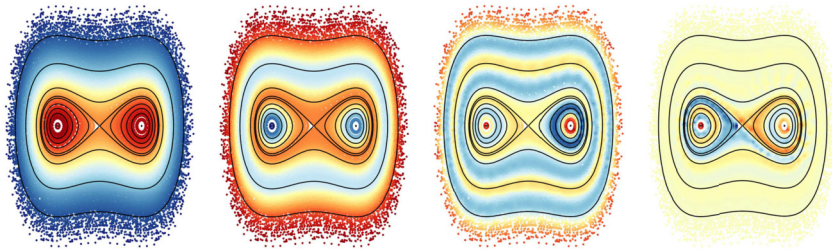

FIGURE 6 Some examples of computed eigenfunctions of the Duffing oscillator corresponding to $\lambda = 1$. The black lines show trajectory orbits and correspond to level sets of the eigenfunctions, which are invariant in time.

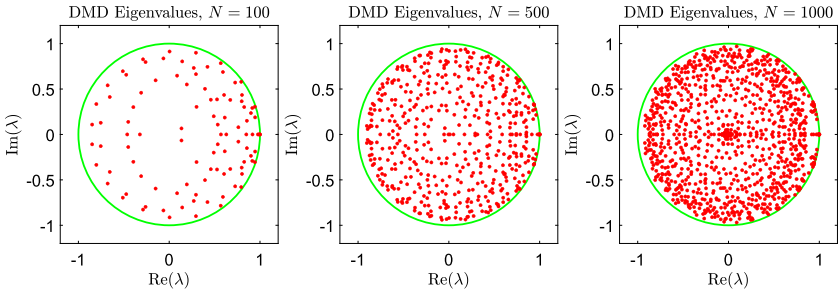

FIGURE 7 DMD eigenvalues (red dots) computed for various choices of N. The spectrum is the unit circle (green); hence, most eigenvalues are spurious. This occurs because a projection error occurs when the Koopman operator $\mathcal{K}$ is approximated by a finite DMD matrix.

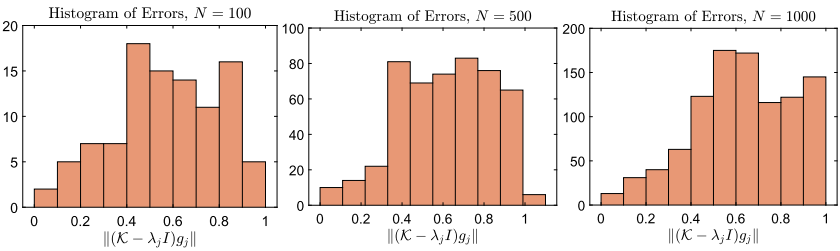

FIGURE 8 Histograms of the errors of the DMD eigenpairs. The errors are computed using ResDMD in Algorithm 11 and show the persistence of heavy spectral pollution of DMD as N increases.

to spectral pollution. This occurs because we have approximated the infinite-dimensional Koopman operator $\mathcal{K}$, by a finite matrix. These errors persist, even as we increase N.

To measure the errors, we can use *Residual DMD* (ResDMD) (Colbrook and Townsend, 2023; Colbrook et al., 2023a) to compute the error $\|(\mathcal{K} - \lambda_j I)g_j\|$ associated with a DMD eigenfunction g_j and eigenvalue λ_j. In other words, we can compute the projection error of DMD. This is detailed in Section 4.3. Fig. 8 shows the histograms of these projection errors. Only a small proportion of reliable DMD eigenvalues persist as N increases. Finally, we can also use ResDMD

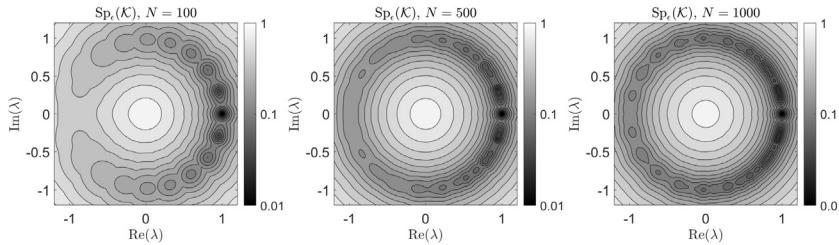

FIGURE 9 Pseudospectra (see (2.3)) computed using ResDMD (see Algorithm 12) and visualized by plotting several contour plots of ϵ on a logarithmic scale. The pseudospectra demonstrate the heavy spectral pollution present in Fig. 7. Note also that these pseudospectra are computed using the same snapshot data and dictionary used for Figs. 6 to 8. As $N \to \infty$, the algorithm converges to the pseudospectra.

to compute pseudospectra. Algorithm 12 converges to the pseudospectrum as $N \to \infty$. The output is shown in Fig. 9, where we visualize pseudospectra by plotting several contour plots of ϵ on a logarithmic scale. For this example, the pseudospectra are the annuli $\mathrm{Sp}_\epsilon(\mathcal{K}) = \{\lambda \in \mathbb{C} : ||\lambda| - 1| \leq \epsilon\}$. The projection errors are computed using the same snapshot data and dictionary used for Figs. 6 to 8. ResDMD allows us to compute and minimize projection errors directly in infinite dimensions to avoid spectral pollution and spurious modes.

In summary, DMD can suffer from closure issues (projection errors) associated with approximating the infinite-dimensional Koopman operator by a finite-dimensional matrix. This phenomenon is well-known (S.L. Brunton et al., 2016a; Kaiser et al., 2021). Nevertheless, by computing projection errors, we can avoid difficulties (such as spurious modes) associated with the infinite-dimensional nature of the Koopman operator.

2.4 The goals and challenges of DMD

The core goal of DMD is to apply linear algebra and spectral techniques to the analysis, prediction, and control of nonlinear dynamical systems. However, DMD often faces several challenges (Kutz et al., 2016a), many of which are discussed in Table 1. These challenges have been a driving force for the many versions of the DMD algorithm that have appeared.

For example, the KMD in (2.7) highlights the potential usefulness of DMD in forecasting. In instances where DMD is applied to noise-free data, such as in generating reduced-order models from high-fidelity numerical simulations (Kutz et al., 2016b; Alla and Kutz, 2017; Lu and Tartakovsky, 2020b), DMD proves effective for both reconstruction and accurate forecasting of the solutions. However, practitioners familiar with DMD's performance in noisy conditions recognize its shortcomings; the algorithm often fails to forecast and reconstruct even the time series it was trained on. In particular, the prediction error in Fig. 2 is somewhat misleading of what a user might expect in the general case. Even after over a decade, the application of DMD for forecasting or recon-

structing time-series data remains limited, typically restricted to high-quality, low-noise scenarios. In Section 3.1, we will focus on methods that mitigate the effect of noise in snapshot data. Many of the structure-preserving methods we discuss in Section 5 have an in-built robustness to noise.

Generally speaking, the error of DMD and its approximate KMD can be split into three types:

- The *projection error* is due to projecting/truncating the Koopman operator onto a finite-dimensional space of observables. This is linked to the issue of closure and lack of (or lack of knowledge of) nontrivial finite-dimensional Koopman invariant subspaces.
- The *estimation error* is due to estimating the matrices that represent the projected Koopman operator from a finite set of potentially noisy trajectory data.
- *Numerical errors* (e.g., roundoff, stability, further compression, etc.) incurred when processing the finite DMD matrix.

In particular, Wu et al. (2021) highlight the issues of robustness to noise and closure/projection errors as the two fundamental challenges for DMD methods. In Section 4 we will consider methods that directly connect DMD with Koopman operators through the Galerkin perspective. Ways of controlling and measuring the projection error are discussed in Section 4.3.

DMD's primary value has been as a diagnostic tool, and the interpretability of DMD modes and frequencies is crucial to this role. Most DMD papers focus on analyzing DMD modes and eigenvalues. This emphasis shapes much of this review. The KMD approximated by DMD modes and eigenvalues facilitates dimensionality reduction and model simplification, analogous to classical methods like the Fourier transform or SVD (Brunton and Kutz, 2022). There are numerous software packages for DMD methods, including https://github.com/dynamicslab/pykoopman, https://github.com/mathLab/PyDMD, and https://github.com/decargroup/pykoop. Moreover, there are numerous repositories connected to the papers cited below. Drmač has implemented the DMD algorithm and extensions in LAPACK (Drmač, 2022a,b).

3 Variants from the regression perspective

This section gives the reader a flavor of DMD variants from the regression perspective.[12] We focus on four key aspects that have proved influential over the last decade or so:

- Noise reduction;
- Compression and randomized linear algebra;
- Multiscale dynamics; and
- Control.

[12] As we shall see, there is less convergence theory (e.g., in the large data limit $M \to \infty$ or as the number of observables increases) for DMD methods based on this viewpoint than for those based on the Galerkin viewpoint in Section 4. This is due to a looser connection with Koopman operators.

The methods we discuss are only some of the variants - it is impossible to do justice to the breadth of techniques! Notable omissions include the following. *Bayesian DMD* (Takeishi et al., 2017a) transfers the Bayesian formulation into DMD. *Higher order DMD* (Le Clainche and Vega, 2017) applies time-delay embedding to build a larger state space after projecting onto POD modes. *Parametric DMD* (Huhn et al., 2023) performs DMD independently per parameter realization and interpolates the resulting Koopman operators. See also Andreuzzi et al. (2023). *Refined Rayleigh–Ritz data driven modal decomposition* (Drmač et al., 2018) produces refined Ritz pairs of the finite DMD matrix $\mathbf{K}_{\mathrm{DMD}}$. *Spatio-temporal Koopman decomposition* (Vega and Le Clainche, 2021) approximates spatio-temporal data as a linear combination of (possibly growing or decaying exponentially) standing or traveling waves. Klus et al. (2018b); Klus and Schütte (2016) develop tensor-based DMD methods for computing eigenfunctions of the Koopman operator. For example, *tensor-based DMD* exploits low-rank tensor decompositions of the data matrices to improve efficiency and memory use. There are extensions of this approach based on reproducing kernel Hilbert spaces (RKHSs) (Fujii and Kawahara, 2019) and EDMD (Nüske et al., 2021). Recent work has also explored connections between DMD and tensor factorizations (Redman, 2021).

3.1 Increasing robustness to noise

A challenge of DMD is that the computed eigenvalues are biased in the presence of sensor noise. Noise typically dampens the eigenvalues, meaning that for discrete-time systems, the absolute values of the eigenvalues are decreased, and the Koopman modes become distorted. For studies of this effect in various physical systems, see Duke et al. (2012); Bagheri (2014); Pan et al. (2015). Dawson et al. (2016) provide an exceptionally clear discussion of this topic. The bias occurs because standard algorithms treat the data "snapshot to snapshot" rather than as a whole and favor one direction (forward in time). Several variants of DMD aim to address this bias. In addition to the methods presented below, other techniques include utilizing Kalman filters (Nonomura et al., 2018, 2019; Jiang and Liu, 2022), adapting DMD to online data (Hemati et al., 2014, 2016), robust principal component analysis (Scherl et al., 2020), and using a second set of noisy observables that meet some independence requirements (Wanner and Mezić, 2022). Moreover, the structure-preserving methods we discuss in Section 5 often have an inbuilt robustness to noise. Finally, in Section 6.3, we discuss the stochastic Koopman operator, which can handle both system and sensor noise.

3.1.1 The problem of noise

We can understand the bias often encountered in DMD as follows. Assume that the snapshots come with additive sensor noise that affects only our measurements of a given system and does not interact with the true dynamics. This

means that we have access to noisy data matrices

$$\mathbf{X}_s = \mathbf{X} + \mathbf{N}_X, \quad \mathbf{Y}_s = \mathbf{Y} + \mathbf{N}_Y,$$

where $\mathbf{N}_X$ and $\mathbf{N}_Y$ are random matrices representing sensor noise, and $\mathbf{X}$ and $\mathbf{Y}$ are the noise-free snapshots. We then represent the data in a truncated POD mode basis so that

$$\tilde{\mathbf{X}}_s = \tilde{\mathbf{X}} + \tilde{\mathbf{N}}_X, \quad \tilde{\mathbf{Y}}_s = \tilde{\mathbf{Y}} + \tilde{\mathbf{N}}_Y,$$

and assume that a subset of POD modes has been selected so that $\tilde{\mathbf{X}}_s\tilde{\mathbf{X}}_s^*$ is invertible. Assuming the noise is sufficiently small, the DMD matrix can be expanded as

$$\begin{aligned}
\tilde{\mathbf{K}}_{\text{DMD}} &= \tilde{\mathbf{Y}}_s\tilde{\mathbf{X}}_s^\dagger = \tilde{\mathbf{Y}}_s\tilde{\mathbf{X}}_s^*(\tilde{\mathbf{X}}_s\tilde{\mathbf{X}}_s^*)^{-1} \\
&= (\tilde{\mathbf{Y}} + \tilde{\mathbf{N}}_Y)(\tilde{\mathbf{X}} + \tilde{\mathbf{N}}_X)^* \left[(\tilde{\mathbf{X}} + \tilde{\mathbf{N}}_X)(\tilde{\mathbf{X}} + \tilde{\mathbf{N}}_X)^*\right]^{-1} \\
&= (\tilde{\mathbf{Y}} + \tilde{\mathbf{N}}_Y)(\tilde{\mathbf{X}} + \tilde{\mathbf{N}}_X)^*(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^*)^{-1} \\
&\quad \times \left[\mathbf{I} - (\tilde{\mathbf{N}}_X\tilde{\mathbf{X}}^* + \tilde{\mathbf{X}}\tilde{\mathbf{N}}_X^* + \tilde{\mathbf{N}}_X\tilde{\mathbf{N}}_X^*)(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^*)^{-1} + \cdots\right].
\end{aligned}$$

Dawson et al. (2016) discard high-order terms in the expectation of this expansion to arrive at

$$\mathbb{E}(\tilde{\mathbf{K}}_{\text{DMD}}) \approx \tilde{\mathbf{Y}}\tilde{\mathbf{X}}^{-1}(\mathbf{I} - \mathbb{E}(\tilde{\mathbf{N}}_X\tilde{\mathbf{N}}_X^*)(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^*)^{-1}). \tag{3.1}$$

This indicates that DMD has an inherent bias due to sensor noise, causing a dampening effect. Interestingly, this bias depends only on $\tilde{\mathbf{N}}_X$ and not $\tilde{\mathbf{N}}_Y$. The reason is that the least squares problem in (2.6) is optimal only when assuming that all of the noise is in $\tilde{\mathbf{Y}}$, but not in $\tilde{\mathbf{X}}$. Another way of seeing this is that the expression $\tilde{\mathbf{Y}}\tilde{\mathbf{X}}^{-1}$ is linear in $\tilde{\mathbf{Y}}$, but not in $\tilde{\mathbf{X}}$, which is why perturbations to $\tilde{\mathbf{X}}$ do not have to propagate through the equation in an unbiased manner.

If the noise structure is known, DMD can be adjusted using a method called *noise-corrected DMD* (ncDMD) (Dawson et al., 2016). However, it is preferable to have methods that correct for noise without requiring explicit knowledge of its structure. We will now outline three popular DMD variants that address this bias without specific assumptions about the noise. The first two can be executed directly using SVDs. The final method requires an iterative method for solving an optimization problem and is more expensive yet more robust.

3.1.2 Forward-backward dynamic mode decomposition (fbDMD)

Forward-Backward DMD (fbDMD) can be considered a correction to the unidirectional bias of Algorithm 1 (Dawson et al., 2016). Let $\mathbf{X} = \mathbf{U}_X\mathbf{\Sigma}_X\mathbf{V}_X^*$ and $\mathbf{Y} = \mathbf{U}_Y\mathbf{\Sigma}_Y\mathbf{V}_Y^*$ be truncated SVDs of the matrices $\mathbf{X}$ and $\mathbf{Y}$, respectively. We define

$$\tilde{\mathbf{K}}_f = \mathbf{U}_X^*\mathbf{Y}\mathbf{V}_X\mathbf{\Sigma}_X^{-1}, \quad \tilde{\mathbf{K}}_b = \mathbf{U}_Y^*\mathbf{X}\mathbf{V}_Y\mathbf{\Sigma}_Y^{-1},$$

Algorithm 2 Forward-backward DMD (Dawson et al., 2016).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, rank $r \in \mathbb{N}$.

1: Compute a truncated SVD $\mathbf{X} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, $\mathbf{U} \in \mathbb{C}^{d\times r}$, $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$, $\mathbf{V} \in \mathbb{C}^{M\times r}$.
2: Compute the projected data matrices $\tilde{\mathbf{X}} = \mathbf{U}^*\mathbf{X}$, $\tilde{\mathbf{Y}} = \mathbf{U}^*\mathbf{Y}$ and their economized SVDs $\tilde{\mathbf{X}} = \mathbf{U}_X\mathbf{\Sigma}_X\mathbf{V}_X^*$, $\tilde{\mathbf{Y}} = \mathbf{U}_Y\mathbf{\Sigma}_Y\mathbf{V}_Y^*$.
3: Compute the forward and backward matrices $\tilde{\mathbf{K}}_f = \mathbf{U}_X^*\tilde{\mathbf{Y}}\mathbf{V}_X\mathbf{\Sigma}_X^{-1}$, $\tilde{\mathbf{K}}_b = \mathbf{U}_Y^*\tilde{\mathbf{X}}\mathbf{V}_Y\mathbf{\Sigma}_Y^{-1}$.
4: Compute the matrices $\mathbf{S}_f = \tilde{\mathbf{Y}}\mathbf{V}_X\mathbf{\Sigma}_X^{-1}$, $\mathbf{S}_b = \tilde{\mathbf{X}}\mathbf{V}_Y\mathbf{\Sigma}_Y^{-1}$, and $\mathbf{K}_f = \mathbf{S}_f\tilde{\mathbf{K}}_f\mathbf{S}_f^\dagger$, $\mathbf{K}_b = \mathbf{S}_b\tilde{\mathbf{K}}_b\mathbf{S}_b^\dagger$.
5: Compute the DMD matrix $\tilde{\mathbf{K}} = \left(\mathbf{K}_f\mathbf{K}_b^{-1}\right)^{1/2}$ and its eigendecomposition $\tilde{\mathbf{K}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$.
6: Compute the modes $\mathbf{\Phi} = \mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}$.

Output: The eigenvalues $\mathbf{\Lambda}$ and modes $\mathbf{\Phi} \in \mathbb{C}^{d\times r}$.

which represent forward and backward propagators for the data, analogous to Algorithm 1. Assuming the system's dynamics are invertible and $\tilde{\mathbf{K}}_b$ is also invertible, the matrix

$$\tilde{\mathbf{K}} = \left(\tilde{\mathbf{K}}_f\tilde{\mathbf{K}}_b^{-1}\right)^{1/2}$$

provides a debiased estimate of the forward propagator. The method is presented in Algorithm 2. Nonetheless, caution is required due to the nonuniqueness of the matrix square root (Higham, 2008). Dawson et al. (2016) suggest selecting the square root that is closest to $\tilde{\mathbf{K}}_f$ in norm, although this can be computationally costly. A more economical alternative involves measuring closeness in the computed eigencoordinates. Sometimes, the nonuniqueness can be avoided. For instance, if the samples are snapshots from a continuous system whose signal has a bandwidth of λ_B and the time-step satisfies $\Delta t < \pi/(2\lambda_B)$, then the discrete eigenvalues expected to be recovered will have a positive real part, which resolves the ambiguity mentioned previously. The square root issue is further analyzed in Drmač et al. (2018, Section 5.4). Finally, Askham and Kutz (2018) recommend first projecting onto r POD modes before applying fbDMD, an alteration that has demonstrated superior performance in practice. For a variational problem involving forward and backward dynamics, see *Consistent DMD* (Azencot et al., 2019).

3.1.3 Total least-squares dynamic mode decomposition (tlsDMD)

Total Least-Squares DMD (tlsDMD) addresses the asymmetric treatment of noise in $\mathbf{X}$ and $\mathbf{Y}$ by Algorithm 1. The least-squares problem in (2.6) can be formulated as

$$\min_{\mathbf{K}} \|\mathbf{E}_Y\|_{\mathrm{F}} \quad \text{such that} \quad \mathbf{Y} + \mathbf{E}_Y = \mathbf{K}\mathbf{X}.$$

Algorithm 3 Total least-squares DMD (Dawson et al., 2016; Hemati et al., 2017).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, rank $r \in \mathbb{N}$.

1: Compute a truncated SVD $\mathbf{X} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, $\mathbf{U} \in \mathbb{C}^{d\times r}$, $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$, $\mathbf{V} \in \mathbb{C}^{M\times r}$.
2: Compute the projected data matrices $\tilde{\mathbf{X}} = \mathbf{U}^*\mathbf{X}$, $\tilde{\mathbf{Y}} = \mathbf{U}^*\mathbf{Y}$.
3: Form the matrix $\mathbf{Z} = \begin{pmatrix}\tilde{\mathbf{X}}^\top & \tilde{\mathbf{Y}}^\top\end{pmatrix}^\top$ and compute its reduced SVD $\mathbf{Z} = \mathbf{U}_Z\mathbf{\Sigma}_Z\mathbf{V}_Z^*$.
4: Set $\mathbf{U}_1 = \mathbf{U}_Z(1:r, 1:r)$, $\mathbf{U}_2 = \mathbf{U}_Z(r+1:2r, 1:r)$.
5: Compute the DMD matrix $\tilde{\mathbf{K}} = \mathbf{U}_2\mathbf{U}_1^{-1}$ and its eigendecomposition $\tilde{\mathbf{K}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$.
6: Compute the modes $\mathbf{\Phi} = \mathbf{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}$.

Output: The eigenvalues $\mathbf{\Lambda}$ and modes $\mathbf{\Phi} \in \mathbb{C}^{d\times r}$.

Considering the reverse time direction, as in fbDMD, leads to the problem

$$\min_{\mathbf{K}} \|\mathbf{E}_X\|_{\mathrm{F}} \quad \text{such that} \quad \mathbf{Y} = \mathbf{K}(\mathbf{X} + \mathbf{E}_X).$$

While fbDMD accounts for both directions of error, a more direct approach utilizes the total least-squares problem (Van Huffel and Vandewalle, 1991):

$$\min_{\mathbf{K}} \left\| \begin{pmatrix}\mathbf{E}_X \\ \mathbf{E}_Y\end{pmatrix} \right\|_{\mathrm{F}} \quad \text{such that} \quad \mathbf{Y} + \mathbf{E}_Y = \mathbf{K}(\mathbf{X} + \mathbf{E}_X).$$

This problem can be solved via an SVD, and we follow the version presented by Dawson et al. (2016), which is similar in spirit to that of Hemati et al. (2017). First, we project $\mathbf{X}$ and $\mathbf{Y}$ onto $r < M/2$ POD modes to obtain $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$. We then define

$$\mathbf{Z} = \begin{pmatrix}\tilde{\mathbf{X}} \\ \tilde{\mathbf{Y}}\end{pmatrix}$$

and compute its reduced SVD $\mathbf{Z} = \mathbf{U}_Z\mathbf{\Sigma}_Z\mathbf{V}_Z^*$. The matrix $\tilde{\mathbf{K}} = \mathbf{U}_Z(r+1:2r, 1:r)\mathbf{U}_Z(1:r, 1:r)^{-1}$ then provides a debiased estimate of the forward propagator. The method is summarized in Algorithm 3.

3.1.4 Optimized dynamic mode decomposition (optDMD)

Optimized DMD (optDMD) is a variation of DMD that processes all data snapshots collectively (Chen et al., 2012). This approach reduces much of the bias associated with exact DMD. Nonetheless, it necessitates solving a nonlinear optimization problem, initially thought to hinder its practical application. However, Askham and Kutz (2018) demonstrated that an approximate solution to the optimization problem can be efficiently computed using the variable projection

method (Golub and Pereyra, 1973). In this framework, DMD is reformulated as an exponential data fitting problem (Pereyra and Scherer, 2010), which brings an additional advantage: the data snapshots do not have to be equidistant in time. For further DMD methodologies tailored for data with irregular time intervals, see Tu et al. (2014a); Guéniat et al. (2015); Leroux and Cordier (2016).

Initially, we project onto r POD modes to construct the data matrix $\tilde{\mathbf{X}} = [\mathbf{z}_0\ \mathbf{z}_1 \cdots \mathbf{z}_M]$, which corresponds to the projected data at times $t_0, t_1, \ldots, t_M$. Depending on the data structure, the projected matrix $\mathbf{Y}$ may also be incorporated into this matrix. We posit that the data represents the solution to a linear system of differential equations, expressed as

$$\mathbf{z}(t) \approx \mathbf{S} e^{\mathbf{\Lambda} t} \mathbf{S}^{\dagger} \mathbf{z}_0 \,,$$

where $\mathbf{S} \in \mathbb{C}^{r \times r}$ and $\mathbf{\Lambda} \in \mathbb{C}^{r \times r}$. This representation can be reformulated to

$$\tilde{\mathbf{X}}^{\top} \approx \mathbf{\Phi}(\boldsymbol{\alpha})\mathbf{B}, \quad \mathbf{B}_{i,j} = \mathbf{S}_{j,i} \left(\mathbf{S}^{\dagger}\mathbf{z}_0\right)_i \,,$$

where $\mathbf{\Phi}(\boldsymbol{\alpha}) \in \mathbb{C}^{(m+1)\times r}$, whose elements are $\Phi(\boldsymbol{\alpha})_{i,j} = \exp(\boldsymbol{\alpha}_j t_i)$. From this, we arrive at an exponential fitting problem:

$$\min_{\boldsymbol{\alpha} \in \mathbb{C}^r, \mathbf{B} \in \mathbb{C}^{r \times r}} \left\| \tilde{\mathbf{X}}^{\top} - \mathbf{\Phi}(\boldsymbol{\alpha})\mathbf{B} \right\|_{\mathrm{F}} .$$

The optimized DMD eigenvalues are determined by $\lambda_j = \boldsymbol{\alpha}_j$. This optimization problem is solved using the variable projection method, which exploits the specific structure of the exponential data fitting problem to eliminate many of the variables from the optimization process. A summary is provided in Algorithm 4, with practical details given in Askham and Kutz (2018), including strategies for selecting the initial guess (e.g., employing an alternate DMD algorithm).

While this nonlinear, nonconvex optimization problem is not guaranteed to be solved globally, and the method may be computationally intensive due to its iterative nature, optDMD often yields significant enhancements over traditional DMD approaches. Moreover, optDMD's efficacy can be further heightened by employing Breiman's statistical bagging sampling strategy (Breiman et al., 2017), which assembles a collection of models to reduce model variance, mitigate overfitting, and facilitate uncertainty quantification. This augmented method is referred to as *bagging optimized DMD* (bopDMD) (Sashidhar and Kutz, 2022).

3.1.5 Examples

For simplicity, we focus on the error associated with the approximated eigenvalues. Other error metrics related to the modes or the accuracy of the decomposition in fitting the data or forecasting are also frequently considered in the literature, often yielding similar results.

Algorithm 4 Optimized DMD, algorithmic details are given in Askham and Kutz (2018).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times(M+1)}$, rank $r \in \mathbb{N}$, initial guess for $\boldsymbol{\alpha}$.

1: Compute a truncated SVD $\mathbf{X} \approx \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$, $\mathbf{U} \in \mathbb{C}^{d\times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$, $\mathbf{V} \in \mathbb{C}^{(M+1)\times r}$.

2: Compute the projected data matrix $\tilde{\mathbf{X}} = \mathbf{U}^*\mathbf{X}$.

3: Solve the problem

$$\min_{\boldsymbol{\alpha}\in\mathbb{C}^r, \mathbf{B}\in\mathbb{C}^{r\times r}} \left\| \tilde{\mathbf{X}}^\top - \boldsymbol{\Phi}(\boldsymbol{\alpha})\mathbf{B} \right\|_{\mathrm{F}}$$

using a variable projection algorithm.

4: Set $\lambda_j = \boldsymbol{\alpha}_j$ and $\boldsymbol{\Phi}(:, i) = (\|\mathbf{U}\mathbf{B}^\top(:, i)\|_{\ell^2})^{-1}\mathbf{U}\mathbf{B}^\top(:, i)$.

Output: The eigenvalues $\boldsymbol{\Lambda}$ and modes $\boldsymbol{\Phi} \in \mathbb{C}^{d\times r}$.

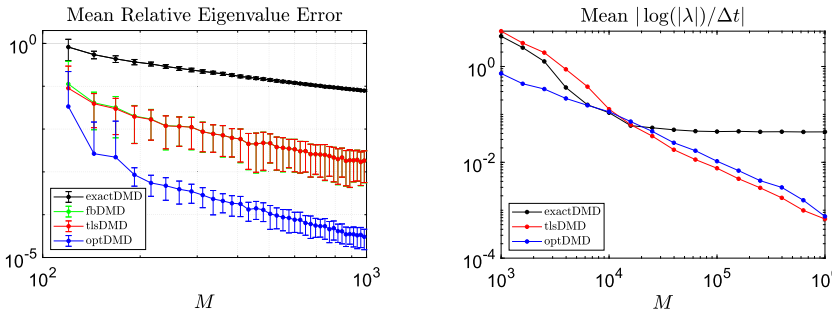

FIGURE 10 Left: Mean error (error bars correspond to the standard deviation across noise realizations) in the first 11 eigenvalues of the cylinder example. Right: Mean value of $|\log(|\lambda|)/\Delta t|$ for the DMD eigenvalues for the Lorenz system. We have not shown the results for fbDMD since they are almost identical to tlsDMD.

Noisy cylinder wake

We revisit the example of flow past a cylinder from Section 2.3.1. We center and normalize the data grid-wise before adding 40% Gaussian random noise to the measurements. Fig. 10 (left) shows the mean relative ℓ^2 error of the first 11 eigenvalues (see Fig. 2), averaged over 100 realizations of random noise. The errors are calculated by comparison with eigenvalues computed from noise-free snapshots that have converged in terms of both the size of the truncated SVD and the number of snapshots. The error bars represent one standard deviation from the mean. All methods exhibit a decreasing error as M increases, which is largely attributable to the truncation in the SVD used in DMD. As often noted in the literature, the fbDMD and tlsDMD methods perform comparably. However, optDMD demonstrates a significantly smaller error.

Lorenz system

We revisit the Lorenz system example from Section 2.3.2. Since the spectrum is continuous (apart from the trivial eigenvalue $\lambda = 1$), measuring the error of individual DMD eigenvalues is meaningless unless methods such as the residual in Section 4.3 are used. However, $|\log(|\lambda|)/\Delta t|$ vanishes on the spectrum of the Koopman operator. Therefore, we select $N = 10$ and compute the mean value of $|\log(|\lambda|)/\Delta t|$ over the DMD eigenvalues. Fig. 10 (right) shows the results, averaged over 50 randomly selected initial conditions on the attractor for the initial value $\mathbf{x}(0)$. For exact DMD, this error metric plateaus as M increases. Generally, the eigenvalues computed using DMD with delay embedding are damped and lie strictly within the unit disk (Korda et al., 2020, Corollary 2). Consequently, their logarithms are in the left-half plane, corresponding to positions below the horizontal line in Fig. 3. Conversely, the eigenvalues computed by tlsDMD and optDMD approach the unit disk with increasing M and exhibit greater robustness to noise in the measurements.

3.2 Compression and randomized linear algebra

With ever-increasing volumes of measurement data from simulations and experiments, modal extraction algorithms such as DMD can become prohibitively expensive, particularly for online or real-time analysis. Dynamics often evolve on low-dimensional attractors, indicating sparsity in a suitable coordinate system or an intrinsic low rankness. However, the SVD used in Algorithm 1 scales with the dimension of the measurements, not with the intrinsic dimension of the data. This section explores two principles aimed at mitigating this computational cost:

- **Compressed sensing** (Donoho, 2006; Candes et al., 2006) facilitates the reconstruction of sparse signals from a limited number of measurements, allowing for undersampling below traditional Shannon–Nyquist limits (Nyquist, 1928; Shannon, 1948). Applying compressed sensing to DMD can substantially improve computational efficiency, particularly during the SVD step of the algorithm. Acquiring high-resolution, time-resolved measurements can be challenging. Nevertheless, temporally and spatially sparse signals may be sampled less frequently than traditionally expected, which is crucial if data acquisition is costly.
- **Randomized numerical linear algebra** (Martinsson and Tropp, 2020) offers a way to solve certain linear algebra problems much faster than classical methods. The randomized SVD is a fast and straightforward technique for computing an approximate low-rank SVD (Halko et al., 2011). It is robust and amenable to parallelization and can benefit from GPU architectures. When coupled with randomized SVD, DMD scales with the intrinsic rank of the data matrices rather than the measurement dimension. The approximation error is manageable through oversampling and power iterations, providing a balance between computational speed and accuracy. Moreover, it can accom-

modate large datasets that exceed the capacity of fast memory by using a blocked matrix approach.

Beyond the methods detailed here, several DMD variants are based on related principles. For instance, due to the nonorthogonality of DMD modes, choosing an appropriate low-rank representation can be difficult (Kou and Zhang, 2017). *Sparsity-Promoting DMD* (Jovanović et al., 2014) aims to strike a balance between accuracy and the number of modes by identifying a sparse subset of modes. Other techniques for selecting dominant modes include ranking each DMD mode's importance by time integration (Kou and Zhang, 2017) or by assessing the time-averaged modal energy contribution (Tissot et al., 2014). Furthermore, one can apply DMD recursively to achieve orthogonality (Noack et al., 2016), a method termed *Recursive DMD*, which blends the principles of POD and DMD. Additionally, regularization terms can be imposed to encourage sparsity in the Koopman matrix (Sinha et al., 2019).

Finally, it is crucial to recognize that the usefulness of the methods in this section presumes the dynamics are evolving on a low-dimensional subspace characterized by a quickly decaying singular value spectrum. While not a fundamental limitation of DMD, this is a common underlying assumption which may not hold for all dynamical systems. Erichson et al. (2019b) provide a turbulent flow example that demonstrates the limits of the approaches in this section when this assumption does not hold.

3.2.1 Compressed sensing meets DMD (cDMD and csDMD)

A full description of the extensive field of compressed sensing is beyond the scope of this review. We outline the key points to understand its interplay with DMD. The reader is encouraged to consult the excellent textbooks (Foucart and Rauhut, 2013; Adcock and Hansen, 2021) for a comprehensive understanding or (Candes and Wakin, 2008) for a concise introductory tutorial. Compressed sensing is founded on two central principles: *sparsity*, which pertains to the signals of interest, and *incoherence*, which relates to the sensing methodology.

Consider a signal $\mathbf{x} \in \mathbb{C}^d$ that is approximately sparse in some basis $\mathbf{B} \in \mathbb{C}^{d \times d}$, meaning that $\mathbf{x} = \mathbf{B}\mathbf{z}$, where the vector $\mathbf{z}$ can be well approximated by a sparse vector. Many natural signals, such as images and audio, are approximately sparse in specific bases like the Fourier or wavelet bases. When we transform an image using Fourier or wavelet transformations, most coefficients are small and can be disregarded while still retaining the quality of the image. We assume that we have access to measurements:

$$\mathbf{x}_c = \mathbf{C}\mathbf{x} = \mathbf{C}\mathbf{B}\mathbf{z},$$

where $\mathbf{C} \in \mathbb{C}^{p \times d}$ is a measurement matrix with $p < d$. Compressed sensing theory implies that, under suitable conditions, we can recover an accurate approximation of $\mathbf{z}$ (and hence $\mathbf{x}$) from the subsampled measurements $\mathbf{x}_c$. For

Algorithm 5 Compressed DMD (S.L. Brunton et al., 2016b), suitable when given access to the full snapshots.

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, rank $r \in \mathbb{N}$, and measurement matrix $\mathbf{C} \in \mathbb{C}^{p\times d}$.

1: Compress $\mathbf{X}$ and $\mathbf{Y}$ to $\mathbf{X}_c = \mathbf{CX}$ and $\mathbf{Y}_c = \mathbf{CY}$.
2: Apply Algorithm 1 with input $\mathbf{X}_c$ and $\mathbf{Y}_c$ and outputs $\mathbf{\Lambda}_c$, $\mathbf{W}_c$, $\mathbf{V}_c$ and $\mathbf{\Sigma}_c$.
3: Reconstruct full-state modes via $\mathbf{\Phi} = \mathbf{Y}\mathbf{V}_c\mathbf{\Sigma}_c^{-1}\mathbf{W}_c$.

Output: The eigenvalues $\mathbf{\Lambda}_c$ and DMD modes $\mathbf{\Phi} \in \mathbb{C}^{d\times r}$.

example, consider the ℓ^1-minimization problem

$$\min \|\mathbf{z}\|_{\ell^1} \quad \text{subject to} \quad \mathbf{x}_c = \mathbf{CBz}. \tag{3.2}$$

Specifically, the measurement matrix $\mathbf{C}$ must be incoherent with respect to the sparse basis $\mathbf{B}$, meaning that the rows of $\mathbf{C}$ are uncorrelated with the columns of $\mathbf{B}$. If the matrix $\mathbf{CB}$ satisfies a restricted isometry property (RIP)[13]:

$$(1-\delta_k)\|\mathbf{z}\|_{\ell^2}^2 \le \|\mathbf{CBz}\|_{\ell^2}^2 \le (1+\delta_k)\|\mathbf{z}\|_{\ell^2}^2 \quad \text{for } k\text{-sparse vectors } \mathbf{z},$$

then we can prove results about how close solutions of (3.2) are to the true $\mathbf{z}$, how issues such as only approximately numerically solving (3.2) affect the solution, robustness to noise, and so forth. Beyond the above ℓ^1-minimization problem, many successful optimization problems and algorithms approximate their solutions in compressed sensing.

Tu et al. (2014a) combine *temporal* compressed sensing with ideas from DMD to recover POD modes. For the remainder of this section, we focus instead on *spatial* compressed sensing, following the methods of S.L. Brunton et al. (2016b). In essence, S.L. Brunton et al. (2016b) demonstrated that the unitary invariance of the DMD algorithm can be extended to approximate invariance under transformations satisfying a RIP, provided that the data is sparse in a basis that is incoherent with respect to the measurements. For compressed data matrices

$$\mathbf{X}_c = \mathbf{CX}, \quad \mathbf{Y}_c = \mathbf{CY},$$

there are essentially two approaches, depending on whether one has access to the matrix $\mathbf{Y}$ or not. Algorithm 5 illustrates *compressed DMD* (cDMD) (see also Erichson et al., 2019a), where one performs the standard DMD algorithm on the compressed data matrices, and then reconstructs the full-state modes using $\mathbf{Y}$.

[13] There are no known large matrices with bounded restricted isometry constants since computing these constants is NP-hard and hard to approximate. Typically, one builds random matrices so that the RIP holds with overwhelming probability. For example, Bernoulli and Gaussian random measurement matrices satisfy the RIP for a generic basis $\mathbf{B}$ with high probability (Candes and Tao, 2006).

Algorithm 6 Compressed sensing DMD (S.L. Brunton et al., 2016b), suitable when given access to only compressed data. The ℓ^1-minimization can be replaced with a plethora of similar minimization problems from the compressed sensing literature.

Input: Compressed snapshot data $\mathbf{X}_c \in \mathbb{C}^{p\times M}$ and $\mathbf{Y}_c \in \mathbb{C}^{p\times M}$, measurement matrix $\mathbf{C} \in \mathbb{C}^{p\times d}$, and basis $\mathbf{B} \in \mathbb{C}^{d\times d}$.

1: Apply Algorithm 1 with input $\mathbf{X}_c$ and $\mathbf{Y}_c$ and outputs $\mathbf{\Lambda}_c$ and $\mathbf{\Phi}_c$.
2: Apply ℓ^1-minimization (3.2) columnwise to reconstruct modes $\mathbf{\Phi}_s \in \mathbb{C}^{d\times r}$.
3: Recover full-state modes via $\mathbf{\Phi} = \mathbf{B}\mathbf{\Phi}_s$.

Output: The eigenvalues $\mathbf{\Lambda}_c$ and DMD modes $\mathbf{\Phi} \in \mathbb{C}^{d\times r}$.

If access to $\mathbf{Y}$ is not available, we can use an optimization problem such as (3.2) to recover the modes in the sparse basis $\mathbf{\Phi}_s$, and then reconstruct the full-state modes. This approach, known as *compressed sensing DMD* (csDMD), is outlined in Algorithm 6.

3.2.2 Randomized dynamic mode decomposition (rDMD)

Early uses of DMD with randomized SVD include Erichson and Donovan (2016), who utilized it to expedite DMD applications in video background subtraction, and Bistrian and Navon (2017), who applied it as a component of a reduced-order model for two-dimensional fluid flows. Although this method is reliable and robust to noise, it only accelerates the computation of the SVD, with subsequent computational steps in the DMD algorithm remaining costly. Instead, Erichson et al. (2019b) developed a *randomized DMD* (rDMD) algorithm. This algorithm relies on sketching the range of $\mathbf{X}$ and executing the entire DMD process in a reduced-dimensional space, ultimately recovering the DMD of the original system at the end.

The idea is to use randomness as a computational strategy to find a smaller representation, known as a *sketch*. This smaller matrix sketch can be used to compute an approximate low-rank factorization for the high-dimensional data matrix. rDMD utilizes the off-the-shelf probabilistic framework proposed in the seminal work of Halko et al. (2011). Given a target rank r, the aim is to compute a near-optimal basis $\mathbf{Q} \in \mathbb{C}^{d\times r}$ for the input matrix $\mathbf{X}$ such that $\mathbf{X} \approx \mathbf{Q}\mathbf{Q}^*\mathbf{X}$. A test matrix $\mathbf{\Omega} \in \mathbb{R}^{M\times r}$ is drawn from a normal Gaussian distribution to sample the range of $\mathbf{X}$ via

$$\mathbf{Z} = \mathbf{X}\mathbf{\Omega}.$$

To mitigate the $\mathcal{O}(dMr)$ cost of dense matrix multiplication, more sophisticated random test matrices, such as the subsampled randomized Hadamard transform, can also be used, leading to a complexity of $\mathcal{O}(dM\log(r))$. The orthonormal basis $\mathbf{Q}$ is then obtained via QR decomposition of $\mathbf{Z}$. In practice, we slightly oversample the desired rank r by a constant factor (typically, 10 suffices). A

Algorithm 7 Randomized range finder. Other choices of random test matrices may be employed for computational efficiency. The QR algorithm is written using MATLAB notation.

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$, target rank $r \in \mathbb{N}$, oversampling factor $p \in \mathbb{N}$, and power iteration factor $q \in \mathbb{N} \cup \{0\}$.

1: Generate a random Gaussian matrix $\boldsymbol{\Omega} \in \mathbb{R}^{M\times(r+p)}$ and form the matrix $\mathbf{Z} = \mathbf{X}\boldsymbol{\Omega}$.
2: **for** $j = 1, \ldots, q$ **do**
3: $\quad[\mathbf{Q}, \sim] = \texttt{qr}(\mathbf{Z}, \text{'econ'})$
4: $\quad[\mathbf{C}, \sim] = \texttt{qr}(\mathbf{X}^*\mathbf{Q}, \text{'econ'})$
5: $\quad\mathbf{Z} = \mathbf{X}\mathbf{C}$
6: **end for**
7: $[\mathbf{Q}, \sim] = \texttt{qr}(\mathbf{Z}, \text{'econ'})$.

Output: Range matrix $\mathbf{Q} \in \mathbb{C}^{d\times(r+p)}$.

Algorithm 8 Randomized DMD (Erichson et al., 2019b).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, target rank $r \in \mathbb{N}$, oversampling factor $p \in \mathbb{N}$, and power iteration factor $q \in \mathbb{N} \cup \{0\}$.

1: Run Algorithm 7 to generate the matrix $\mathbf{Q}$.
2: Compress $\mathbf{X}$ and $\mathbf{Y}$ to $\mathbf{X}_c = \mathbf{Q}^*\mathbf{X}$ and $\mathbf{Y}_c = \mathbf{Q}^*\mathbf{Y}$.
3: Apply Algorithm 1 with input $\mathbf{X}_c$ and $\mathbf{Y}_c$ and outputs $\boldsymbol{\Lambda}_c$ and $\boldsymbol{\Phi}_c$.
4: Reconstruct full-state modes via $\boldsymbol{\Phi} = \mathbf{Q}\boldsymbol{\Phi}_c$.

Output: The eigenvalues $\boldsymbol{\Lambda}_c$ and DMD modes $\boldsymbol{\Phi} \in \mathbb{C}^{d\times r}$.

second strategy to improve performance involves power iterations (Rokhlin et al., 2010; Gu, 2015). Particularly, a slowly decaying singular value spectrum of the input matrix can significantly affect the quality of the approximated basis matrix $\mathbf{Q}$. Power iterations are employed to preprocess the input matrix to promote a more rapidly decaying spectrum. The sampling matrix obtained is

$$\mathbf{Z} = (\mathbf{X}\mathbf{X}^*)^q\mathbf{X}\boldsymbol{\Omega},$$

and as few as $q = 2$ power iterations can considerably improve the approximation quality, even when the singular values of the input matrix decay slowly. This procedure is outlined in Algorithm 7, and we direct the reader to Martinsson and Tropp (2020, Section 11) for probabilistic error bounds. With the matrix $\mathbf{Q}$ in hand, we can perform DMD on the lower-dimensional space, as summarized in Algorithm 8. One can also simultaneously sketch the range and corange of $\mathbf{X}$. This method, known as *sketchy DMD*, was proposed by Ahmed et al. (2022).

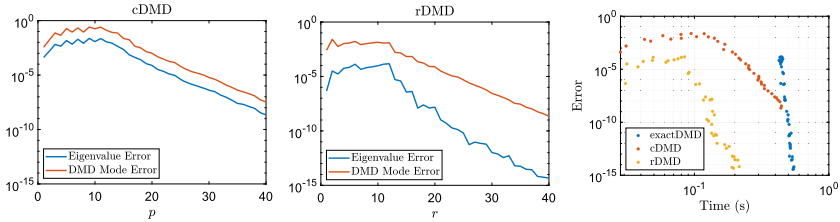

FIGURE 11 Left: Errors of cDMD (Algorithm 5) for the first 6 eigenvalues and DMD modes. Middle: Errors of rDMD (Algorithm 8) for the first six eigenvalues and DMD modes. Right: Computational times vs eigenvalue error.

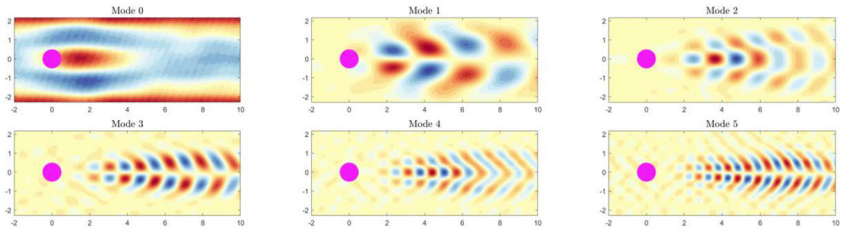

FIGURE 12 DMD modes constructed using csDMD (Algorithm 6) with a 99.5% compression in the dimension of the measurements.

3.2.3 *Examples*

Cylinder wake

We return to the cylinder wake discussed in Section 2.3.1 as an illustrative example. For cDMD and csDMD, we use the (inverse) two-dimensional discrete Fourier transform as our basis $\mathbf{B}$ and Gaussian random measurements $\mathbf{C}$. Fig. 11 shows the results for cDMD and csDMD when recovering the first six modes plotted in Fig. 2. In the left and middle panels, we have shown the mean eigenvalue error and the DMD mode error (computed as a subspace angle) averaged over 20 random realizations. The errors quickly become negligible for $p, r = \mathcal{O}(10)$. In the right panel, we have displayed the mean execution times on a standard laptop (without GPU) against the eigenvalue error. Even for this simple example with rapidly decreasing singular values, cDMD performs better than exact DMD, while rDMD is the clear winner.

Fig. 12 shows the modes computed by csDMD with $p = 800$, which corresponds to a 99.5% compression in the dimension of the snapshot matrices. We employ the CoSaMP algorithm (Needell and Tropp, 2009) to perform the ℓ^1-minimization step. When only compressed measurements are available, it is still possible to reconstruct full-state modes using compressed sensing. However, this typically requires more measurements and computational resources than cDMD or rDMD.

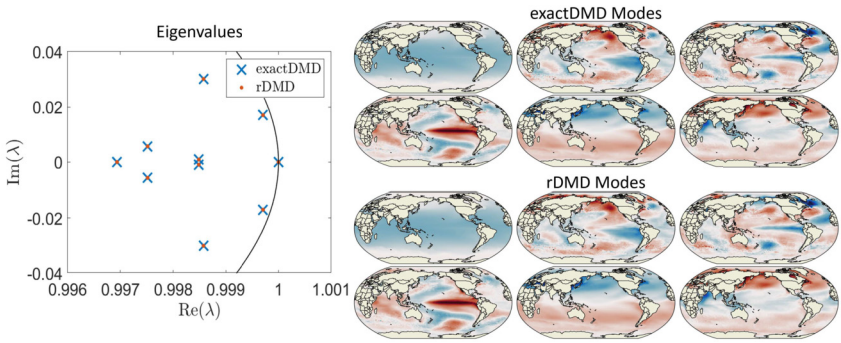

FIGURE 13 Eigenvalues and dynamic modes of the SST data set computed using exact DMD and rDMD. The eigenvalue plot shows the unit circle as a black line.

Sea surface data

We now consider high-resolution sea surface temperature (SST) data. SST data are widely studied in climate science for climate monitoring and prediction (Reynolds et al., 2007, 2002; Smith and Reynolds, 2005), and measurements are constructed by combining infrared satellite data with observations provided by ships and buoys. The data are available from the National Oceanic and Atmospheric Administration at https://www.esrl.noaa.gov/psd/ for the years 1981 to 2023, with a grid resolution of 0.25°. Omitting data over land results in $d = 691{,}150$ spatial grid points. The following experiments, similar in spirit to Erichson et al. (2019b), were performed using a system with Intel(R) Xeon(R) Gold 6126 CPU at 2.60GHz (48 cores) and 767GiB system memory.

We first consider a temporal resolution of one day and a data matrix $\mathbf{X} \in \mathbb{R}^{691{,}150 \times 15{,}097}$. Fig. 13 shows the eigenvalues and dynamic modes computed using exact DMD and rDMD (with $r = 10$), demonstrating the accuracy of rDMD. The bottom left mode is reminiscent of an El Niño mode generated from the El Niño-Southern Oscillation (ENSO). El Niño is the warm phase of the ENSO cycle. It is associated with a band of warm ocean water that develops in the central and east-central equatorial Pacific, including off the Pacific coast of South America (see also Fig. 15).

Next, we compare the accuracy and computational times for a temporal resolution of one week and a data matrix $\mathbf{X} \in \mathbb{R}^{691{,}150 \times 2156}$ of weekly averages. Fig. 14 shows the relative error in the Frobenius norm of the reconstructed data matrix and time taken for exact DMD, rDMD, and blocked rDMD (using four blocks). We observe substantial gains in computational time when using rDMD while maintaining an accuracy similar to the full deterministic exact DMD.

3.3 Multiresolution dynamic mode decomposition (mrDMD)

Multiscale systems are widespread across various scientific disciplines. Modeling the interactions between microscale and macroscale phenomena, which may

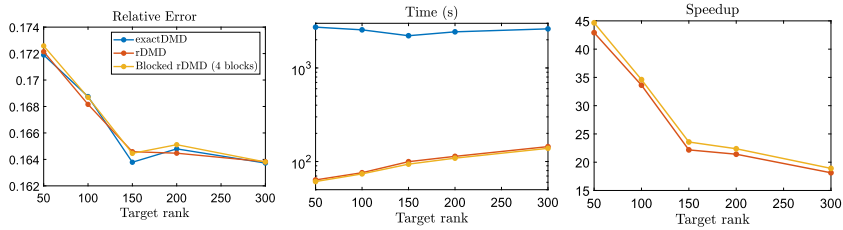

FIGURE 14 Left: Accuracy of exact DMD, rDMD, and batched rDMD for the SST data set. Middle: The computational times of each method. Right: The speedup of rDMD and batched rDMD compared to exact DMD. All plots show the mean of 10 independent runs.

differ by orders of magnitude either spatially or temporally, poses a considerable challenge. Wavelet-based methods and windowed Fourier transforms are well-suited for multiresolution analysis (MRA), as they systematically isolate temporal or spatial features through recursive refinement when sampling from the targeted data (Daubechies, 1992). Typically, MRA is employed separately in either space or time, but it is seldom applied to both simultaneously.

Multiresolution DMD (mrDMD) (Kutz et al., 2016b) integrates DMD with core principles from wavelet theory and MRA. It adjusts the sampling window of the data collection process in line with wavelet theory, filtering information across various scales. The process is iteratively refined through progressively shorter snapshot sampling windows, leading to the recursive extraction of DMD modes from slow to rapidly changing timescales. The benefits of this approach include enhanced prediction of the near-future state of the system, which is vital for control; effective management of transient phenomena; and improved handling of moving (translating/rotating) structures within the data. The latter two points underscore significant challenges inherent in standard DMD methods. The mrDMD algorithm has led to practical applications such as determining optimal sensor placement (Manohar et al., 2019).

3.3.1 The algorithm

When using mrDMD, it is typical to work with an M such that a full-rank approximation with $r = M$ in Algorithm 1 is feasible and such that high- and low-frequency content is present. We assume that data is collected along a single time trajectory with time-step Δt and express the eigenvalues in terms of their time-scaled logarithms $\eta = \log(\lambda)/\Delta t$. In the first pass, mrDMD separates the approximation in (2.7) into slow modes and fast modes:

$$\mathbf{x}(t) \approx \underbrace{\sum_{|\eta_k| \le \tau} \boldsymbol{\phi}_k^{(1)} \exp(\eta_k t)\mathbf{b}(k)}_{\text{slow modes}} + \underbrace{\sum_{|\eta_k| > \tau} \boldsymbol{\phi}_k^{(1)} \exp(\eta_k t)\mathbf{b}(k)}_{\text{fast modes}}, \tag{3.3}$$

where $\boldsymbol{\phi}_k^{(1)} = \boldsymbol{\Phi}(:, k)$, and the superscript (1) indicates the level. The first sum in the expression (3.3) represents the slow-mode dynamics, whereas the second

sum is everything else. How to choose the slow modes is important in the practical implementation. In the original mrDMD paper, it is suggested to set the threshold τ to select eigenvalues whose temporal behavior allows for at most one wavelength to fit within the sampling window.

The fast modes in (3.3) can be collected into a data matrix $\mathbf{X}_{M/2}$, where we let m_1 denote the number of slow modes in (3.3). The matrix $\mathbf{X}_{M/2}$ is now split into two matrices, where the first matrix contains the first $M/2$ snapshots, and the second matrix contains the remaining $M/2$ snapshots. The process is now repeated, where m_2 slow modes are collected at the second level and computed separately in the first and second intervals of snapshots. This process is repeated to obtain the decomposition

$$\begin{aligned}\mathbf{x}(t) \approx &\sum_{k=1}^{m_1} b_k^{(1)} \boldsymbol{\phi}_k^{(1)} \exp(\eta_k^{(1)} t) + \sum_{k=1}^{m_2} b_k^{(2)} \boldsymbol{\phi}_k^{(2)} \exp(\eta_k^{(2)} t) \\ &+ \sum_{k=1}^{m_3} b_k^{(3)} \boldsymbol{\phi}_k^{(3)} \exp(\eta_k^{(3)} t) + \cdots,\end{aligned}$$

where the $\boldsymbol{\phi}_k^{(\ell)}$ and $\eta_k^{(\ell)}$ are the DMD modes and eigenvalues at the ℓth level of the decomposition, the $b_k^{(\ell)}$ are the initial projections of the data onto the time interval of interest, and the m_ℓ are the number of slow modes retained at each level. The idea is that different spatiotemporal DMD modes are used to represent key multiresolution features. Thus, no single set of modes dominates the SVD and potentially marginalizes features at other time scales.

We can make the mrDMD more precise, letting L denote the number of levels of the decomposition. The solution is a sum with ℓ indexing the level, $j = 1, \ldots, 2^{(\ell-1)}$ indexing the time bins $[t_j^{(\ell)}, t_{j+1}^{(\ell)}]$ in each level and $k = 1, \ldots, m_\ell$ indexing the modes extracted at each level. To simplify the sum, define the following indicator function

$$f_{\ell,j}(t) = \begin{cases} 1, & \text{if } t \in [t_j^{(\ell)}, t_{j+1}^{(\ell)}] \\ 0, & \text{otherwise} \end{cases},$$

$$\text{so that} \quad \mathbf{x}_{\text{mrDMD}}(t) = \sum_{\ell=1}^{L} \sum_{j=1}^{2^{\ell-1}} \sum_{k=1}^{m_\ell} f_{\ell,j}(t) b_k^{(\ell,j)} \boldsymbol{\phi}_k^{(\ell,j)} \exp(\eta_k^{(\ell,j)} t).$$

In particular, each mode is represented in its respective time bin and level. Alternatively, this solution can be interpreted as yielding the least-squares fit to the dynamics within a given time bin at each level of the decomposition.

Numerous innovations enhance the practical implementation of mrDMD. Since only slow modes matter within a window, we can limit sampling to a fixed number of points per window, reducing the data matrix size for more manageable SVD computations. The sampling window locations are flexible, and

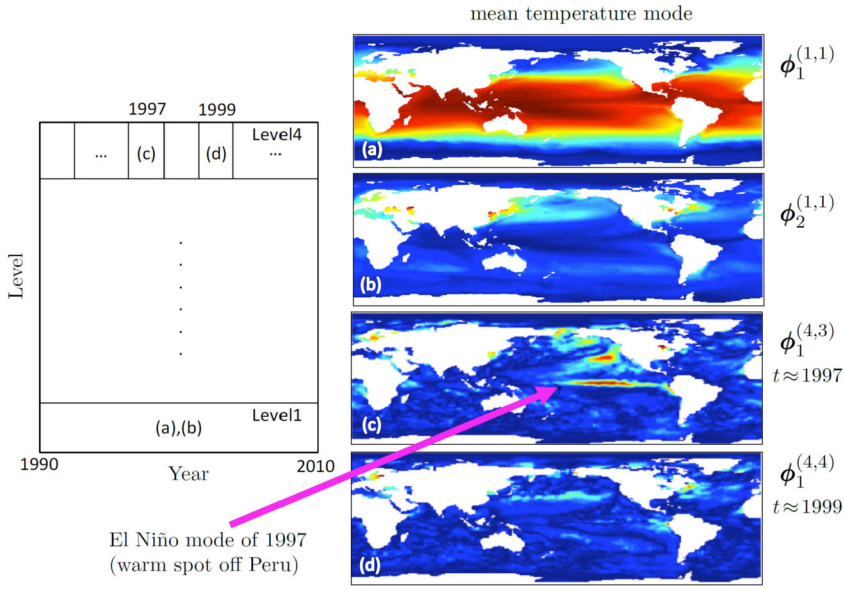

FIGURE 15 Application of mrDMD on sea surface temperature data from 1990 to 2010. The left panel illustrates the process for a 4-level decomposition. At each level, the slowest modes are extracted. Mode (c) clearly shows the El Niño mode of interest that develops in the central and east-central equatorial Pacific. The El Niño mode was absent in 1999, as is clear from mode (d). Reproduced with permission from Kutz et al. (2016b), copyright © 2016 Society for Industrial and Applied Mathematics, all rights reserved.

smoothing their edges can prevent the Gibbs phenomenon due to abrupt data cutoffs. One can also employ wavelet functions, like Haar, Daubechies, or Mexican Hat, for the sifting function $f_{\ell,j}(t)$. Finally, overlapping windows prevent data loss during sampling. A sliding window approach can robustly track data features, enabling pattern correlation across windows akin to a Gabor transform-generated spectrogram.

3.3.2 Example

The following example is from Kutz et al. (2016b). We consider the global sea surface temperature (see Section 3.2.3) with data that spans 20 years from 1990 to 2010. Fig. 15 illustrates the outcomes of employing a 4-level mrDMD decomposition. At the first level, mrDMD identifies two modes: the mean ocean temperature, denoted as $\boldsymbol{\phi}_1^{(1,1)}$, and an annual cycle, represented by $\boldsymbol{\phi}_2^{(1,1)}$. Intriguingly, at the fourth level, the approximate zero mode of the sampling window uncovers noteworthy phenomena; specifically, it isolates the 1997 El Niño event. In contrast, when the same sampling window is applied to the year 1999, the El Niño mode is absent, aligning with the recognized oceanic patterns of that year. These insights would not have been obtainable using traditional DMD

without preselecting the correct sampling windows. Moreover, even if such a step were taken, the slow modes identified at the first level, as shown in Fig. 15 (a) and (b), would pollute the data at the level of investigation. For additional techniques identifying approximate eigenfunctions that provide a rectified representation of the ENSO and function as (approximate) semiconjugacies or factor maps with circle rotations, see Froyland et al. (2021).

3.3.3 Nonautonomous systems

The use of mrDMD to detect transient behavior is tantalizing! Most DMD methods are designed for autonomous dynamical systems, where the function **F** on the right-hand side of (1.1) has no time dependence. However, there has been some recent initial work on nonautonomous systems, and we expect this area to grow significantly over the next few years. Mezić and Surana (2016) were the first to extend the Koopman operator framework to nonautonomous dynamical systems, applying the methodology to linear-periodic and quasiperiodic nonautonomous systems. Giannakis (2019) developed a strategy inspired by time-changed dynamical systems that involves rescaling the generator; this can be applied to a class of time-changed mixing systems. Further development of this approach, using delay-coordinate maps for recovering the dynamical system on tori with multiple time scales, is presented in Das and Giannakis (2019). The extraction of spatiotemporal patterns using an extension of approximation techniques developed in Giannakis (2019) on the space-time manifold defined as a skew-product structure is considered in Giannakis and Das (2020); Giannakis et al. (2019). For the online computation of windowed DMD using rank-one updates, see *Online DMD* (Zhang et al., 2019). Maćešić et al. (2018) provides an error analysis for DMD with moving stencils. For extensions to actuated systems, see Williams et al. (2016); Bai et al. (2020). Redman et al. (2023) develop an episodic memory approach that saves spectral objects associated with temporally local approximations of the Koopman operator, and utilizes this information to make new predictions. Nonautonomous systems have also been studied using transfer operators, which are the dual of Koopman operators (see the discussion in Section 6.1), and space-time manifolds (Froyland and Koltai, 2023).

3.4 Control

One of the most successful applications of the Koopman operator framework lies in control (Mauroy et al., 2020; Otto and Rowley, 2021), with demonstrated successes in various challenging applications. These include fluid dynamics (Arbabi et al., 2018; Peitz and Klus, 2020), robotics (Abraham et al., 2017; Bruder et al., 2019; Mamakoukas et al., 2019; Haggerty et al., 2023), power grids (Korda et al., 2018; Netto and Mili, 2018), biology (Hasnain et al., 2020), and chemical processes (Narasingam and Kwon, 2019). The key point is that Koopman operators represent nonlinear dynamics within a globally linear

framework. This approach leads to tractable convex optimization problems and circumvents theoretical and computational limitations associated with nonlinearity. Moreover, it is amenable to data-driven, model-free approaches (Proctor et al., 2016; Williams et al., 2016; Korda and Mezić, 2018a; Proctor et al., 2018; Surana, 2016; Kaiser et al., 2021, 2018a; Peitz and Klus, 2019; Abraham and Murphey, 2019). The resulting models reveal insights into global stability properties (Sootla and Mauroy, 2016; Mauroy and Mezić, 2016), observability/controllability (Vaidya, 2007; Goswami and Paley, 2017; Yeung et al., 2018), and sensor/actuator placement (Sinha et al., 2016; Sharma et al., 2019) for the underlying nonlinear systems.

Koopman operator theory was first extended to actuated systems by Mezić and Banaszuk (2004), with stochastic forcing interpreted as actuation. Proctor et al. (2016) developed the first control schemes based on DMD. A significant strength of DMD is the ability to describe complex and high-dimensional dynamical systems with a few dominant modes. Reducing the system's dimensionality enables faster and lower-latency prediction and estimation, leading to high-performance, robust controllers.

3.4.1 Dynamic mode decomposition with control (DMDc)

We will focus on the *DMD with control* (DMDc) algorithm (Proctor et al., 2016). DMDc extends DMD to disambiguate between unforced dynamics and the effect of actuation. The DMD regression of Section 2.2.1 is generalized to

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n, \mathbf{u}_n) \approx \mathbf{A}\mathbf{x}_n + \mathbf{B}\mathbf{u}_n,$$

where $\mathbf{u}_n \in \mathbb{C}^q$ is a vector of control inputs for each time-step. Here $\mathbf{A} \in \mathbb{C}^{d\times d}$ and $\mathbf{B} \in \mathbb{C}^{d\times q}$ are unknown matrices. Snapshot triplets of the form $\{\mathbf{x}^{(m)}, \mathbf{y}^{(m)}, \mathbf{u}^{(m)}\}_{m=1}^{M}$ are collected, where we assume that

$$\mathbf{y}^{(m)} \approx \mathbf{F}(\mathbf{x}^{(m)}, \mathbf{u}^{(m)}), \quad m = 1, \dots, M.$$

The control portion of the snapshots is arranged into the matrix $\boldsymbol{\Upsilon} = \begin{pmatrix}\mathbf{u}^{(1)}\,\mathbf{u}^{(2)}\,\cdots\,\mathbf{u}^{(M)}\end{pmatrix} \in \mathbb{C}^{q\times M}$. The optimization problem in (2.6) is replaced by

$$\min_{(\mathbf{A}\ \mathbf{B})} \|\mathbf{Y} - (\mathbf{A}\ \mathbf{B})\boldsymbol{\Omega}\|_{\mathrm{F}}^2, \quad \text{where} \quad \boldsymbol{\Omega} = \begin{pmatrix}\mathbf{X}\\ \boldsymbol{\Upsilon}\end{pmatrix} \in \mathbb{C}^{(d+q)\times M}.$$

A solution is given as $(\mathbf{A}\ \mathbf{B}) = \mathbf{Y}\boldsymbol{\Omega}^{\dagger}$. In practice, we seek a reduced-order model by performing a truncated SVD on both the input and output space. The full algorithm is summarized in Algorithm 9 and is an extension of Algorithm 1. DMDc has been used with Model-Predictive Control (MPC) for enhanced control of nonlinear systems in Korda and Mezić (2018a); Kaiser et al. (2018b), with the DMDc method performing surprisingly well, even for strongly nonlinear systems. Extensions are discussed in Section 3.4.3.

Algorithm 9 DMD with control (Proctor et al., 2016).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$, $\mathbf{Y} \in \mathbb{C}^{d\times M}$ and $\boldsymbol{\Upsilon} \in \mathbb{C}^{q\times M}$, ranks $r, p \in \mathbb{N}$.

1: Compute a truncated SVD of the input matrix $\begin{pmatrix}\mathbf{X}\\ \boldsymbol{\Upsilon}\end{pmatrix} \approx \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^*$, $\tilde{\mathbf{U}} \in \mathbb{C}^{(d+q)\times p}$, $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{p\times p}$, $\tilde{\mathbf{V}} \in \mathbb{C}^{M\times p}$. Break up the matrix $\tilde{\mathbf{U}}$ into $\tilde{\mathbf{U}}^* = [\tilde{\mathbf{U}}_1^*\ \tilde{\mathbf{U}}_2^*]$ where $\tilde{\mathbf{U}}_1 \in \mathbb{C}^{d\times p}$ and $\tilde{\mathbf{U}}_2 \in \mathbb{C}^{q\times p}$.

2: Compute a truncated SVD of $\mathbf{Y} \approx \widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^*$, $\widehat{\mathbf{U}} \in \mathbb{C}^{d\times r}$, $\widehat{\boldsymbol{\Sigma}} \in \mathbb{R}^{r\times r}$, $\widehat{\mathbf{V}} \in \mathbb{C}^{M\times r}$.

3: Compute the compressions $\tilde{\mathbf{A}} = \widehat{\mathbf{U}}^*\mathbf{Y}\tilde{\mathbf{V}}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{U}}_1^*\widehat{\mathbf{U}} \in \mathbb{C}^{r\times r}$ and $\tilde{\mathbf{B}} = \widehat{\mathbf{U}}^*\mathbf{Y}\tilde{\mathbf{V}}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{U}}_2^* \in \mathbb{C}^{r\times q}$.

4: Compute the eigendecomposition $\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\boldsymbol{\Lambda}$. The columns of $\mathbf{W}$ are eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues.

5: Compute the modes $\boldsymbol{\Phi} = \mathbf{Y}\tilde{\mathbf{V}}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{U}}_1^*\widehat{\mathbf{U}}\mathbf{W}$.

Output: The eigenvalues $\boldsymbol{\Lambda}$ and modes $\boldsymbol{\Phi} \in \mathbb{C}^{d\times r}$.

3.4.2 Example

We illustrate DMDc for system identification on a high-dimensional, linear system with spectral sparsity following Proctor et al. (2016, Section 4.3). We consider a two-dimensional torus discretized by a 128×128 equispaced grid such that $\mathbf{x} \in \mathbb{R}^{128\times 128} \cong \mathbb{R}^{16,384}$. Taking the two-dimensional discrete Fourier transform of $\mathbf{x}$, we obtain $\hat{\mathbf{x}}$. The system evolves according to

$$\hat{\mathbf{x}}_{n+1} = \hat{\mathbf{A}}\hat{\mathbf{x}}_n + \hat{\mathbf{B}}\hat{\mathbf{u}}_n.$$

Here, $\hat{\mathbf{A}}$ is a diagonal matrix with five nonzero entries representing the modes, each with a randomly chosen frequency and small damping. The random input signal, $\hat{\mathbf{u}}$, is one-dimensional and directly influences the sparse modes, resulting in a localized negative control input when transformed back to the spatial domain. This back-transformation yields our dynamical system in physical space. The system is constructed by sampling a continuous-time system at time steps of $\Delta t = 0.01$. We collect $M = 400$ snapshots of the data for our analysis. Further details of this system can be found in S.L. Brunton et al. (2016b).

Fig. 16 displays the true eigenvalues alongside those computed by DMDc and exact DMD. Ten eigenvalues are present in conjugate pairs due to processing real-valued data. DMDc demonstrates greater accuracy than exact DMD, which inaccurately estimates some eigenvalues and generates unstable modes. The true DMD modes for this system appear at the top Fig. 17. The DMDc modes in the middle row align almost perfectly with the true modes. The subspace angle between the true modes and the DMDc-computed modes is on the order of machine precision. In contrast, the modes produced by exact DMD show significant distortion.

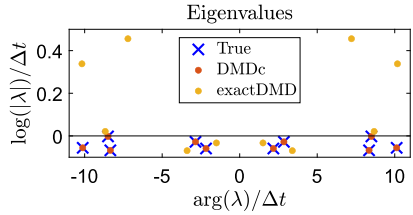

FIGURE 16 True eigenvalues of the torus example and those computed by DMDc and exact DMD. The logarithm of the eigenvalues are plotted to align with the continuous-time system.

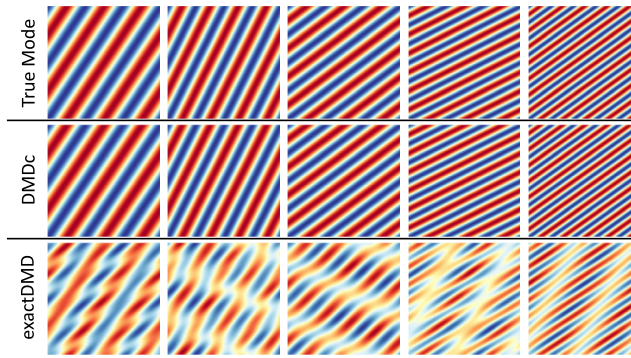

FIGURE 17 The true DMD modes for the torus example, alongside those computed by DMDc and exact DMD. The modes obtained from DMDc are accurate to machine precision, whereas those computed using exact DMD are significantly distorted.

3.4.3 *Extensions and connection with Koopman operators*

Koopman theory has been used in combination with the Linear Quadratic Regulator (LQR) (S.L. Brunton et al., 2016a; Mamakoukas et al., 2019, 2021), state-dependent LQR (Kaiser et al., 2021), and MPC (Korda and Mezić, 2018a; Kaiser et al., 2018b). Other noteworthy directions include optimal control for switching control problems (Peitz and Klus, 2019, 2020), Lyapunov-based stabilization (Huang et al., 2018, 2020), eigenstructure assignment (Hemati and Yao, 2017), and active learning (Abraham and Murphey, 2019). Additionally, deep learning architectures have been employed to represent the nonlinear observables in combination with MPC (Li et al., 2019), see also (Liu et al., 2018; Han et al., 2020), and (Peitz and Klus, 2019; Peitz et al., 2020; Klus et al., 2020b) for parametrized models.

Koopman theory is closely related to Carleman linearization (Carleman, 1932), which embeds finite-dimensional dynamics into infinite-dimensional linear systems using a polynomial basis. Carleman linearization has been used for decades to obtain truncated linear (and bilinear) state estimators (Krener, 1974; Brockett, 1976) and to examine stability, observability, and controllability of the underlying nonlinear system (Loparo and Blankenship, 1978).

The DMDc framework may be extended to nonlinear observables using EDMD (see Section 4), an approach called eDMDc (Williams et al., 2016). Korda and Mezić (2018a) integrated eDMDc into MPC. Here, the Koopman operator is characterized as an autonomous operator on the extended state vector $(\mathbf{x}^\top, \mathbf{u}^\top)^\top$, with observables that may be nonlinear functions of both the state and the input. In practical applications, simplifications are employed to ensure the control problem remains convex (Korda and Mezić, 2018a; Proctor et al., 2018). This method has been applied for control in the coordinates of Koopman eigenfunctions (Kaiser et al., 2021, 2018a; Folkestad et al., 2020) and in interpolated Koopman models (Peitz, 2018; Peitz et al., 2020). Convergence can be established under the assumption of an infinite amount of data and an infinite number of basis functions. Koopman Lyapunov-based MPC guarantees closed-loop stability and controller feasibility (Narasingam and Kwon, 2019; Son et al., 2020). However, general guarantees regarding the optimality, stability, and robustness of the controlled dynamical system are still limited.

The Koopman operator's eigenfunctions (or approximate eigenfunctions) are a natural choice of observables due to their simple temporal behavior. It is crucial to validate computed eigenfunctions to ensure that their evolution is consistent with the predictions of their associated eigenvalues, particularly for prediction tasks. They have been used for observer design within the Koopman canonical transform (Surana, 2016; Surana and Banaszuk, 2016) and within the Koopman reduced-order nonlinear identification and control framework (Kaiser et al., 2021), which both typically yield a global bilinear representation of the underlying system. Subsequent research has focused on directly identifying Koopman eigenfunctions (Korda and Mezić, 2020; Pan et al., 2021) and approximate invariant subspaces (Haseli and Cortés, 2023).

The efficacy of Koopman-based MPC is currently at odds with the difficulties of approximating the Koopman operator and its spectra. Only a limited number of systems with a known Koopman-invariant subspace and verifiable eigenfunctions exist for model analysis and evaluation. Furthermore, the linearity of Koopman eigenfunctions is seldom validated. Nevertheless, Koopman-based MPC demonstrates remarkable resilience with models of marginal predictive ability. Despite notable successes, understanding how well the Koopman operator is actually approximated and producing error bounds remains largely incomplete.

4 Variants from the Galerkin perspective

We now explore variants of DMD from the Galerkin (or projection) perspective, building on the connection established in Section 2.2.2. This approach particularly focuses on addressing the infinite-dimensional nature of Koopman operators. Given that a Koopman operator transforms a finite-dimensional nonlinear system into an infinite-dimensional linear one, a significant part of this section will address nonlinear observables. We will concentrate on three methods designed to tackle these challenges:

- **Extended DMD:** This represents a fundamental extension of DMD that treats it as a Galerkin method.[14] In particular, it introduces nonlinear observables to form a dictionary, which generates a subspace within $L^2(\Omega, \omega)$. Adopting the Galerkin perspective enables the application of numerical tools for addressing infinite-dimensional spectral problems. However, the well-studied challenges of infinite-dimensional spectral computations are significant. Generally, EDMD will not converge to the spectral properties of the Koopman operator, either theoretically or practically (see Section 4.1.3 and common pitfalls in Section 4.3).
- **Time-delay Embedding:** This technique is commonly used to construct a dictionary of observables for EDMD and generates a Krylov subspace. Our focus will be on two methods: Hankel-DMD, which is a widely used technique suitable for ergodic systems that have a low-dimensional attractor, and HAVOK (Hankel Alternative View Of Koopman) analysis, which produces a linear model using the leading delay coordinates and includes forcing terms represented by low-energy delay coordinates.
- **Residual DMD:** This algorithm computes verified spectral properties of Koopman operators via an infinite-dimensional residual corresponding to the projection error of (E)DMD. This residual is computed from the snapshot data by augmenting EDMD with an additional matrix. This leads to the computation of spectra and pseudospectra without spectral pollution (general systems) and can be used to compute spectral measures (measure-preserving systems). Since the algorithms have error control, ResDMD allows a posteriori verification of spectral quantities, Koopman mode decompositions, and learned dictionaries.[15]

4.1 Nonlinear observables: extended dynamic mode decomposition (EDMD)

The standard DMD algorithm can accurately characterize periodic and quasi-periodic behaviors in nonlinear systems. However, DMD models based on linear observables generally fail to capture truly nonlinear phenomena. To address this limitation, Williams et al. (2015a) introduced *Extended DMD* (EDMD), which also elucidated the interpretation of DMD as a Galerkin method. Specifically, they demonstrated that EDMD converges to the numerical approximation obtained by a Galerkin method in the limit of large data sets. Prior research in a similar vein includes (Tu et al., 2014b). Moreover, the connection between EDMD and the earlier variational approach of conformation dynamics (Noé and Nüske, 2013; Nüske et al., 2014) from molecular dynamics is explored in Wu et al. (2017); Klus et al. (2018c).

[14] Though once nonlinear observables have been chosen, one can also apply the regression interpretation of Section 3.

[15] One can often show that a priori error control is impossible (Colbrook et al., 2019).

4.1.1 The algorithm

Following the discussion of Koopman operators in Section 2.1, the objective of EDMD is to approximate the Koopman operator with a matrix. For the sake of simplicity, the initial formulation of EDMD assumes that the columns of the snapshot matrix $\mathbf{X}$ are independently sampled from the distribution ω. In our discussion, we extend EDMD to accommodate any given snapshot matrices

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \dots & \mathbf{x}^{(M)} \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} \mathbf{y}^{(1)} & \mathbf{y}^{(2)} & \dots & \mathbf{y}^{(M)} \end{pmatrix},$$

and consider the $\mathbf{x}^{(m)}$ as *quadrature nodes* used for integration with respect to ω. This adaptability permits the application of various quadrature weights tailored to the specific scenario. It will be shown that EDMD generalizes the setup of Section 2.2.2.

One first chooses a dictionary $\{\psi_1, \dots, \psi_N\}$, i.e., a list of observables, in the space $L^2(\Omega, \omega)$. These observables form a finite-dimensional subspace $V_N = \operatorname{span}\{\psi_1, \dots, \psi_N\}$. EDMD selects a matrix $\mathbf{K} \in \mathbb{C}^{N \times N}$ that approximates the action of $\mathcal{K}$ confined to this subspace. We desire that

$$[\mathcal{K}\psi_j](\mathbf{x}) = \psi_j(\mathbf{F}(\mathbf{x})) \approx \sum_{i=1}^{N} \mathbf{K}_{ij} \psi_i(\mathbf{x}), \quad 1 \le j \le N.$$

Define the vector-valued feature map

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

Any $g \in V_N$ can be written as $g(\mathbf{x}) = \sum_{j=1}^{N} \psi_j(\mathbf{x}) \mathbf{g}_j = \mathbf{\Psi}(\mathbf{x})\,\mathbf{g}$ for some vector $\mathbf{g} \in \mathbb{C}^N$. Hence

$$[\mathcal{K}g](\mathbf{x}) = \mathbf{\Psi}(\mathbf{F}(\mathbf{x}))\,\mathbf{g} = \mathbf{\Psi}(\mathbf{x})(\mathbf{K}\,\mathbf{g}) + \underbrace{\left(\sum_{j=1}^{N} \psi_j(\mathbf{F}(\mathbf{x})) \mathbf{g}_j - \mathbf{\Psi}(\mathbf{x})(\mathbf{K}\,\mathbf{g}) \right)}_{=: R(\mathbf{g}, \mathbf{x})}.$$

Typically, V_N is not an invariant subspace of $\mathcal{K}$. Hence, there is no choice of $\mathbf{K}$ that makes $R(\mathbf{g}, \mathbf{x})$ zero for all $g \in V_N$ and ω-almost every $\mathbf{x} \in \Omega$. Instead, it is natural to select $\mathbf{K}$ as a solution of

$$\min_{\mathbf{K} \in \mathbb{C}^{N \times N}} \left\{ \int_{\Omega} \max_{\mathbf{g} \in \mathbb{C}^N, \|\mathbf{C}\mathbf{g}\|_{\ell^2} = 1} |R(\mathbf{g}, \mathbf{x})|^2 \, \mathrm{d}\omega(\mathbf{x}) \right.$$
$$\left. = \int_{\Omega} \left\| \mathbf{\Psi}(\mathbf{F}(\mathbf{x}))\mathbf{C}^{-1} - \mathbf{\Psi}(\mathbf{x})\mathbf{K}\mathbf{C}^{-1} \right\|_{\ell^2}^2 \, \mathrm{d}\omega(\mathbf{x}) \right\}. \tag{4.1}$$

Here, $\|\cdot\|_{\ell^2}$ denotes the standard Euclidean norm of a vector, and $\mathbf{C}$ is a positive self-adjoint matrix that controls the size of $g = \mathbf{\Psi}\mathbf{g}$. One should think of this $\mathbf{C}$ as choosing an appropriate norm. This is important since not all norms on an infinite-dimensional vector space are equivalent (for an example in DMD analysis of fluid flow, see Colbrook, 2023, Figure 7).

In practical, data-driven contexts, it is not possible to directly evaluate the integral in (4.1). Instead, we approximate it via a quadrature rule with nodes $\{\mathbf{x}^{(m)}\}_{m=1}^{M}$ and weights $\{w_m\}_{m=1}^{M}$. For notational convenience, let $\mathbf{D} = \mathrm{diag}(w_1, \dots, w_M)$ and

$$\mathbf{\Psi}_X = \begin{pmatrix} \mathbf{\Psi}(\mathbf{x}^{(1)}) \\ \vdots \\ \mathbf{\Psi}(\mathbf{x}^{(M)}) \end{pmatrix} \in \mathbb{C}^{M\times N}, \quad \mathbf{\Psi}_Y = \begin{pmatrix} \mathbf{\Psi}(\mathbf{y}^{(1)}) \\ \vdots \\ \mathbf{\Psi}(\mathbf{y}^{(M)}) \end{pmatrix} \in \mathbb{C}^{M\times N}. \tag{4.2}$$

The discretized version of (4.1) is the following weighted least-squares problem:

$$\min_{\mathbf{K}\in\mathbb{C}^{N\times N}} \left\{ \sum_{m=1}^{M} w_m \left\| \mathbf{\Psi}(\mathbf{y}^{(m)})\mathbf{C}^{-1} - \mathbf{\Psi}(\mathbf{x}^{(m)})\mathbf{K}\mathbf{C}^{-1} \right\|_{\ell^2}^2 = \left\| \mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{C}^{-1} - \mathbf{D}^{1/2}\mathbf{\Psi}_X\mathbf{K}\mathbf{C}^{-1} \right\|_{\mathrm{F}}^2 \right\}, \tag{4.3}$$

where we remind the reader that $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. By reducing the size of the dictionary if necessary, we may assume without loss of generality that $\mathbf{D}^{1/2}\mathbf{\Psi}_X$ has rank N. For example, we can do this in DMD by projecting onto POD modes. Regularization through a truncated singular value decomposition may also be considered. A solution to (4.3) is

$$\mathbf{K} = (\mathbf{D}^{1/2}\mathbf{\Psi}_X)^{\dagger}\mathbf{D}^{1/2}\mathbf{\Psi}_Y = (\mathbf{\Psi}_X^{*}\mathbf{D}\mathbf{\Psi}_X)^{\dagger}\mathbf{\Psi}_X^{*}\mathbf{D}\mathbf{\Psi}_Y,$$

where '†' denotes the pseudoinverse. Note that this solution is independent of the matrix $\mathbf{C}$. However, a suitable choice of $\mathbf{C}$ is vital once we add constraints to the optimization problem in (4.1), see Section 5.2.1. As observed in Section 2.2.2, if the quadrature weights are equal and $\mathbf{\Psi} = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}$ constitutes an appropriate linear dictionary, then $\mathbf{K}$ is the transpose of the DMD matrix. Conceptually, DMD can be regarded as a particular instance of EDMD employing a set of linear basis functions.

We now generalize Section 2.2.2 by defining the two correlation matrices

$$\mathbf{G} = \mathbf{\Psi}_X^{*}\mathbf{D}\mathbf{\Psi}_X = \sum_{m=1}^{M} w_m \mathbf{\Psi}(\mathbf{x}^{(m)})^{*}\mathbf{\Psi}(\mathbf{x}^{(m)}),$$

Algorithm 10 The EDMD algorithm (Williams et al., 2015a).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, quadrature weights $\{w_m\}_{m=1}^M$, and a dictionary of functions $\{\psi_j\}_{j=1}^N$.

1: Compute the matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ defined in (4.2) and $\mathbf{D} = \operatorname{diag}(w_1, \ldots, w_M)$.
2: Compute the EDMD matrix $\mathbf{K} = (\mathbf{D}^{1/2}\mathbf{\Psi}_X)^\dagger \mathbf{D}^{1/2}\mathbf{\Psi}_Y \in \mathbb{C}^{N\times N}$.
3: Compute the eigendecomposition $\mathbf{KV} = \mathbf{V\Lambda}$.
The columns of $\mathbf{V}$ are eigenvector coefficients and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues.

Output: The eigenvalues $\mathbf{\Lambda}$ and eigenvector coefficients $\mathbf{V} \in \mathbb{C}^{N\times N}$.

$$\mathbf{A} = \mathbf{\Psi}_X^* \mathbf{D} \mathbf{\Psi}_Y = \sum_{m=1}^{M} w_m \mathbf{\Psi}(\mathbf{x}^{(m)})^* \mathbf{\Psi}(\mathbf{y}^{(m)}). \tag{4.4}$$

If we consider the discrete measure $\omega_M = \sum_{m=1}^M w_m \delta_{\mathbf{x}^{(m)}}$, then

$$\mathbf{G}_{jk} = \int_\Omega \overline{\psi_j(\mathbf{x})}\psi_k(\mathbf{x})\, \mathrm{d}\omega_M(\mathbf{x}), \quad \mathbf{A}_{jk} = \int_\Omega \overline{\psi_j(\mathbf{x})}\psi_k(\mathbf{F}(\mathbf{x}))\, \mathrm{d}\omega_M(\mathbf{x}).$$

If the quadrature converges, then

$$\lim_{M\to\infty} \mathbf{G}_{jk} = \langle \psi_k, \psi_j \rangle \quad \text{and} \quad \lim_{M\to\infty} \mathbf{A}_{jk} = \langle \mathcal{K}\psi_k, \psi_j \rangle, \tag{4.5}$$

where $\langle \cdot, \cdot \rangle$ is the inner product associated with $L^2(\Omega, \omega)$. Hence, in the large data limit, $\mathbf{K} = \mathbf{G}^\dagger \mathbf{A}$ approaches a matrix representation of $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}_{V_N}^*$, where $\mathcal{P}_{V_N}$ denotes the orthogonal projection onto V_N. In essence, EDMD is a Galerkin method. The EDMD eigenvalues thus approach the spectrum of $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}_{V_N}^*$, and EDMD is an example of the so-called *finite section method* (Böttcher and Silbermann, 1983) (Mezić, 2022, Section 4). Since the finite section method can suffer from spectral pollution (spurious modes), spectral pollution is a concern for EDMD (Williams et al., 2015a). We saw an explicit example in Section 2.3.3. See also Mezić (2022, Example 2) for the worked example $\mathbf{F}(\mathbf{x}) = \mathbf{x}^2$ on the unit circle.

Algorithm 10 summarizes the procedure for computing eigenvalues and eigenvectors. We can also use EDMD to compute Koopman modes. Given an observable $g = \mathbf{\Psi g} \in V_N$, we may expand g in terms of the eigenvectors of $\mathbf{K}$ as

$$g = \mathbf{\Psi g} = \mathbf{\Psi V}\left[\mathbf{V}^{-1}\mathbf{g}\right], \tag{4.6}$$

where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{K}$ with eigenvalues $\{\lambda_j\}_{j=1}^N$. Similarly, for general $g \in L^2(\Omega, \omega) \backslash V_N$, we obtain an approximate expansion

$$g \approx \mathbf{\Psi V} \left[\mathbf{V}^{-1} (\mathbf{D}^{1/2} \mathbf{\Psi}_X)^\dagger \mathbf{D}^{1/2} \left(g(\mathbf{x}^{(1)}), \ldots, g(\mathbf{x}^{(M)}) \right)^\top \right]. \tag{4.7}$$

This expansion is called the KMD of g.[16] With an abuse of notation, if $g \notin V_N$, we set

$$\mathbf{g} = (\mathbf{D}^{1/2} \mathbf{\Psi}_X)^\dagger \mathbf{D}^{1/2} \left(g(\mathbf{x}^{(1)}), \ldots, g(\mathbf{x}^{(M)}) \right)^\top.$$

As $M \to \infty$, assuming that the quadrature rule underlying EDMD converges, the approximation $g \approx \mathbf{\Psi g}$ converges to the projected observable $\mathcal{P}_{V_N} g$. As a particular case, we can vectorize and obtain

$$\mathbf{x} \approx \mathbf{\Psi}(\mathbf{x}) \mathbf{V} \left[\mathbf{V}^{-1} (\mathbf{D}^{1/2} \mathbf{\Psi}_X)^\dagger \mathbf{D}^{1/2} \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)} \right)^\top \right].$$

The jth row of the matrix in square brackets is known as the jth Koopman mode, which we denote as $\boldsymbol{\xi}_j \in \mathbb{C}^{1 \times d}$. Generalizing (2.7), the KMD provides an approximation of the dynamics by

$$\mathbf{x}_n \approx \mathbf{\Psi}(\mathbf{x}_0) \mathbf{K}^n \mathbf{V} \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_N \end{pmatrix} = \mathbf{\Psi}(\mathbf{x}_0) \mathbf{V} \mathbf{\Lambda}^n \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_N \end{pmatrix} = \sum_{j=1}^N \mathbf{\Psi}(\mathbf{x}_0) \mathbf{V}(:, j) \lambda_j^n \boldsymbol{\xi}_j.$$

Similarly, for general g, we obtain

$$g(\mathbf{x}_n) \approx \mathbf{\Psi}(\mathbf{x}_0) \mathbf{K}^n \mathbf{V} \mathbf{V}^{-1} \mathbf{g} = \mathbf{\Psi}(\mathbf{x}_0) \mathbf{V} \lambda^n \mathbf{V}^{-1} \mathbf{g} = \sum_{j=1}^N \mathbf{\Psi}(\mathbf{x}_0) \mathbf{V}(:, j) \lambda_j^n [\mathbf{V}^{-1} \mathbf{g}]_j,$$

which includes the triple of Koopman eigenvectors, eigenvalues, and modes.

4.1.2 Choices of dictionary

We have already met two examples of EDMD in this review: the Lorenz system discussed in Section 2.3.2, where a dictionary was constructed from delay embedding, and the Duffing oscillator discussed in Section 2.3.3, where we utilized a dictionary of radial basis functions. The selection of the dictionary significantly affects the efficacy of EDMD. In their original formulation, Williams et al. (2015a) suggest various dictionary choices, such as polynomials, Fourier

[16] Unfortunately, there are numerous meanings of the term KMD in the literature. There is the KMD of Mezić (2005), which we discussed in Section 2.1.2 and is based on the spectral theorem for unitary Koopman operators. There is also the (typically approximate) KMD produced by DMD and EDMD.

modes, spectral elements, and radial basis functions. Subsequent extensions have primarily focused on addressing the challenges of large state-space dimensions and mitigating the curse of dimensionality.

Kernelized EDMD (Williams et al., 2015b) (developed in parallel in Kawahara, 2016) uses the kernel trick (Scholkopf, 2001) to perform EDMD with a choice of dictionary determined implicitly by a choice of a kernel function. This approach can help circumvent the curse of dimensionality and can be very effective when the state-space dimension d is large. Numerous papers have been written on the approximation of Koopman operators in a RKHS (Klus et al., 2018a; Fujii and Kawahara, 2019; DeGennaro and Urban, 2019; Alexander and Giannakis, 2020; Das and Giannakis, 2020; Klus et al., 2020c,a; Mezić, 2020; Burov et al., 2021; Baddoo et al., 2022; Kostic et al., 2022; Khosravi, 2023; Philipp et al., 2023). This also includes methods for continuous-time dynamical systems (Das et al., 2021; Rosenfeld et al., 2022). A challenge associated with RKHS techniques is that a general RKHS does not exhibit invariance under the action of the Koopman operator. This situation renders the selection of a reproducing kernel a delicate task. Ideally, one should choose the kernel so that the Koopman operator on the RKHS is not only densely defined but also closable. Finding such a kernel is generally nontrivial, as indicated in Ikeda et al. (2022).

Kernel analog forecasting (KAF) (Zhao and Giannakis, 2016) is a kernel method used for nonparametric statistical forecasting of dynamically generated time series data. Under measure-preserving and ergodic dynamics, KAF consistently approximates the conditional expectation of observables that are acted upon by the Koopman operator of the dynamical system and are conditioned on the observed data at forecast initialization (Alexander and Giannakis, 2020). KAF yields optimal predictions in the sense of minimal root mean square error with respect to the invariant measure in the asymptotic limit of large data. This connection facilitates the analysis of generalization error and uncertainty quantification. KAF has been used with streaming kernel regression (Giannakis et al., 2023) and for multiscale systems (Burov et al., 2021).

Diffusion forecasting (Berry et al., 2015) uses the diffusion maps algorithm (Coifman and Lafon, 2006) to construct a data-driven basis. Leveraging spectral convergence results for kernel integral operators (García Trillos et al., 2020; von Luxburg et al., 2008), this approach produces a well-conditioned and consistent approximation as both the amount of training data and the number of basis functions increase. Giannakis et al. (2015); Giannakis (2019) use the diffusion forecasting technique in a framework that approximates the generator $\mathcal{L}$ of measure-preserving ergodic flows on manifolds by an advection-diffusion operator $\mathcal{L}_\tau = \mathcal{L} - \tau\Delta$, where τ is a regularization parameter, and Δ is a Laplace-type diffusion operator. A Galerkin method was developed for the eigenvalue problem of $\mathcal{L}_\tau$, which was observed to perform efficiently for systems with a pure point spectrum, such as ergodic rotations on tori. The most straightforward case for analyzing the spectral properties of diffusion-regularized generators arises when the regularizing operator Δ commutes with $\mathcal{L}$. Das and Giannakis

(2019); Giannakis (2019) demonstrated that a commuting operator Δ can be derived from the infinite-delay limit of a family of kernel integral operators constructed using time-delay embedding.

Another prevalent method involves training *neural networks* as a suitable dictionary to construct Koopman forecasts, as demonstrated in several studies (Li et al., 2017; Takeishi et al., 2017b; Wehmeyer and Noé, 2018; Yeung et al., 2019; Azencot et al., 2020; Eivazi et al., 2021; Li and Jiang, 2021; Alford-Lago et al., 2022). This approach is typically implemented in two ways: by identifying a few key latent variables or by lifting to a higher-dimensional input space. Variational autoencoders (VAMPnets) have been employed for stochastic dynamical systems such as in molecular dynamics (Mardt et al., 2018; Wehmeyer and Noé, 2018), wherein the mapping back to the physical configuration space from the latent variables is probabilistic. The integration of Koopman analysis with graph convolutional neural networks has been explored to learn the dynamics of atoms within materials (Xie et al., 2019). Lusch et al. (2018) employ an auxiliary network to parameterize the continuously varying spectral parameter, enabling a network structure that offers both parsimony and interpretability. A notable challenge when incorporating EDMD with neural networks is the trade-off between representing data accurately and the potential for overfitting, particularly with limited data. To address this issue, Otto and Rowley (2019) proposed an architecture that combines an autoencoder with linear recurrent dynamics in the encoded space. Beyond employing neural networks for learning Koopman embeddings, Koopman theory has also been applied to understand the behavior of neural networks themselves (Manojlović et al., 2020; Dogra and Redman, 2020) and algorithms more broadly (Dietrich et al., 2020; Redman et al., 2022).

4.1.3 Convergence theory

We now outline the convergence theory for EDMD. Unfortunately, the type of convergence (in the strong operator topology) is too weak to ensure the convergence of spectral properties. To take into account the snapshot data and dictionary, we let $\mathbf{K}_{N,M}$ denote the EDMD matrix. When considering the convergence of EDMD and related methods, there are two limits of interest:

- The large-data limit which corresponds to $M \to \infty$;
- The large-subspace limit which corresponds to $N \to \infty$.

To compute the spectral properties of $\mathcal{K}$, a double limit

$$\lim_{N\to\infty} \lim_{M\to\infty} \mathbf{K}_{N,M}$$

must be considered. Generally, these limits do not commute. More broadly (and beyond the above specific interpretations of the two limits), the use of successive limits is a common occurrence in spectral problems and other areas of scientific computation and cannot be overcome regardless of the choice of algorithm (Colbrook, 2020, 2022; Colbrook and Hansen, 2022; Ben-Artzi et al., 2020).

We saw above that if the quadrature rule converges, i.e., (4.5) holds, then $\lim_{M\to\infty} \mathbf{K}_{N,M} = \mathbf{K}_N$ is a Galerkin matrix. There are essentially three options for the quadrature rule:

- **Random sampling:** We may draw $\mathbf{x}^{(m)}$ at random according to a probability measure that is absolutely continuous with respect to ω, and select the quadrature weights according to the corresponding Radon–Nikodym derivative. This was essentially observed in (Williams et al., 2015a). Convergence holds with probability one (Klus et al., 2016, Section 3.4) provided that ω is not supported on a zero level set that is a linear combination of the dictionary (Korda and Mezić, 2018b, Section 4). The convergence rate is typically $\mathcal{O}(M^{-1/2})$ (Caflisch, 1998), but is a practical approach if the state-space dimension is large. One could also consider quasi-Monte Carlo integration, which can achieve a faster rate of $\mathcal{O}(M^{-1})$ (up to logarithmic factors) under suitable conditions (Caflisch, 1998).
- **Ergodic sampling:** If the system is ergodic, then we can replace the strong law of large numbers with Birkhoff's Ergodic theorem (Birkhoff, 1931):

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} [\mathcal{K}^j g](\mathbf{x}_0) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} g(\mathbf{x}_j) = \int_\Omega g(\mathbf{x})\, \mathrm{d}\omega(\mathbf{x}) \quad \forall g \in L^1(\Omega,\omega). \tag{4.8}$$

  We may select $\mathbf{x}^{(m)} = \mathbf{x}_{m-1}$ from a single trajectory starting at ω-almost any initial condition $\mathbf{x}_0$ and $w_m = 1/M$. Often, the measures are 'physical,' meaning that the set of initial points with convergence has a positive Lebesgue measure.[17] For example, taking $g = [\mathcal{K}\psi_k]\cdot\overline{\psi_j}$ in (4.8), we obtain

$$\lim_{M\to\infty} \underbrace{\frac{1}{M}\sum_{n=0}^{M-1} \psi_k(\mathbf{x}_{n+1})\overline{\psi_j(\mathbf{x}_n)}}_{=\mathbf{A}_{jk}} = \langle \mathcal{K}\psi_k, \psi_j\rangle.$$

  Convergence in this scenario is analyzed in Arbabi and Mezić (2017a); Korda and Mezić (2018b). The convergence rate in M is problem dependent (Kachurovskii, 1996; Mezić and Sotiropoulos, 2002). For periodic and quasiperiodic attractors, the error of approximating the inner products is generally $\mathcal{O}(M^{-1})$. For strongly mixing systems, the rate of convergence slows down to $\mathcal{O}(M^{-1/2})$. However, convergence rates cannot be established for the general class of ergodic systems. For convergence rates of von Neumann's ergodic theorem in the context of Koopman operators, see Aloisio et al. (2022).

[17] There is also the notion of SRB measure, which often coincides. For a survey of these measures and their definitions, see Young (2002).

- **High-order quadrature:** If the state space dimension d is not too large and Ω is sufficiently simple, it can be effective to choose $\{(\mathbf{x}^{(m)}, w_m)\}$ according to a high-order quadrature rule. Even changing the weights $\{w_m\}$ for a fixed set of sample points $\{\mathbf{x}^{(m)}\}$ can lead to a considerable acceleration of the convergence (Colbrook and Townsend, 2023).

Colbrook et al. (2023b) provide concentration bounds on the error of the finite M EDMD matrix. Mollenhauer et al. (2022) provide a rigorous analysis of kernel autocovariance operators, including nonasymptotic error bounds under classical ergodic and mixing assumptions. Nüske et al. (2023) presented the first rigorously derived probabilistic bounds on the finite-data approximation error for the truncated Koopman generator of stochastic differential equations (SDEs) and nonlinear control systems. Two settings were analyzed: independent and identically distributed sampling and ergodic sampling, where it was assumed that the Koopman semigroup is exponentially stable for the latter. Lu and Tartakovsky (2020b) provide bounds for parabolic PDEs.

Suppose the quadrature rule converges and we have passed to the limit $M \to \infty$. Korda and Mezić (2018b) show that under a natural density assumption of V_N as $N \to \infty$, $\mathbf{K}_N$ converges strongly to $\mathcal{K}$ for bounded Koopman operators. This means that

$$\lim_{N\to\infty} \|\mathcal{K}g - \mathbf{\Psi}\mathbf{K}_N\mathcal{P}_{V_N}g\|_{L^2(\Omega,\omega)} = 0 \quad \forall g \in L^2(\Omega,\omega), \tag{4.9}$$

where, with an abuse of notation, $\mathcal{P}_{V_N}g$ denotes the vector of coefficients of $\mathcal{P}_{V_N}g$. It is straightforward to drop the assumption that $\mathcal{K}$ is bounded by making natural assumptions on the dictionary and considering g in the domain of $\mathcal{K}$ (Colbrook and Townsend, 2023). Unfortunately, strong convergence is insufficient to ensure that the spectral properties of $\mathbf{K}_N$ converge to that of $\mathcal{K}$ - Mezić (2022) provides an explicit example. We also saw an example of this effect in Section 2.3.3. In Section 4.3, we will show how Residual DMD provides convergence and error control in the final limit $N \to \infty$.

4.1.4 Infinitesimal generators

Several methods have also been proposed for continuous-time systems and approximating the Koopman infinitesimal generator defined in (2.8). For example, *generator EDMD* (gEDMD) (Klus et al., 2020b) uses time derivatives of the dictionary to extend EDMD to compute the generator, see also Klus et al. (2020a); Rosenfeld et al. (2022). Other methods include computing the matrix logarithm of the Koopman operator (Mauroy and Goncalves, 2020; Drmač et al., 2021), approximating the Koopman operator family, and using finite-differences to compute the Lie derivative of the Koopman operator (Giannakis, 2021; Sechi et al., 2021). Finally, Das and Giannakis (2019); Giannakis (2019); Giannakis and Das (2020) approach the problem of approximating both the Koopman and its generator as a manifold-learning problem on a space-time manifold. This

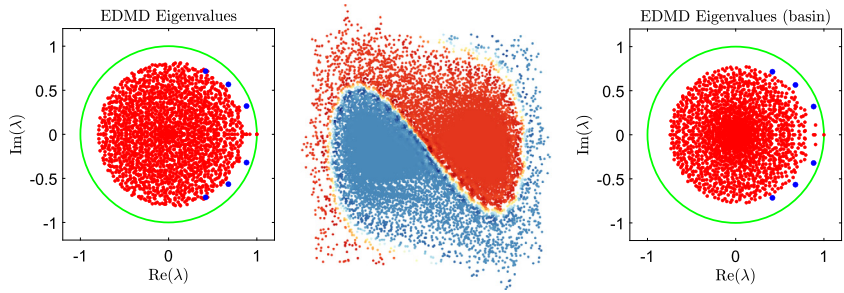

FIGURE 18 Left: EDMD eigenvalues computed over the entire state space. Middle: Eigenfunction that parametrizes the basins of attraction. Right: EDMD eigenvalues obtained after restricting the process to one basin of attraction. The powers of the dominant damped mode are shown in blue.

challenge was successfully addressed for ergodic dynamical systems, such as those evolving on a chaotic attractor.

4.1.5 Example

As an example of EDMD, we revisit the Duffing oscillator from Section 2.3.3 but follow the experiment of Williams et al. (2015a) closely. Namely, we consider the damped system:

$$\dot{x} = y, \quad \dot{y} = -0.5y + x - x^3.$$

In this regime, there are two stable spirals at $(\pm 1, 0)$ and a saddle at the origin. Almost every initial condition, except those on the stable manifold of the saddle, is drawn to one of the spirals. We collect trajectory data and form the dictionary of observables $\{\psi_j\}_{j=1}^N$ in the same manner as before. Fig. 18 (left) shows the eigenvalues computed using EDMD with $N = 2000$. The system is now damped, and there is a lattice structure of dominant but damped modes, $\{\lambda_1^n, \overline{\lambda_1}^n : n \in \mathbb{N}, \lambda_1 \approx 0.8831 + 0.3203i\}$ shown in blue. The lattice structure can be understood as follows: if g and f are eigenfunctions of $\mathcal{K}$ corresponding to eigenvalues λ and μ, respectively, and if the product fg is within the function space that forms the domain of $\mathcal{K}$, then

$$[\mathcal{K}(fg)](\mathbf{x}) = f(\mathbf{F}(\mathbf{x}))g(\mathbf{F}(\mathbf{x})) = [\mathcal{K}f](\mathbf{x})[\mathcal{K}g](\mathbf{x}) = \lambda\mu f(\mathbf{x})g(\mathbf{x}).$$

Namely, further eigenvalues and eigenfunctions can be constructed by taking products.

For this system, the eigenspace corresponding to $\lambda = 1$ is spanned by the constant function and the indicator function of the invariant set corresponding to the two basins of attraction. This is illustrated in the middle of Fig. 18. Utilizing the level sets of this eigenfunction, we limit the data to the basin of $(-1, 0)$ and rerun the process to compute a new dictionary. The resulting EDMD eigenvalues are displayed on the right side of Fig. 18, where the eigenvalue λ_1^3 is

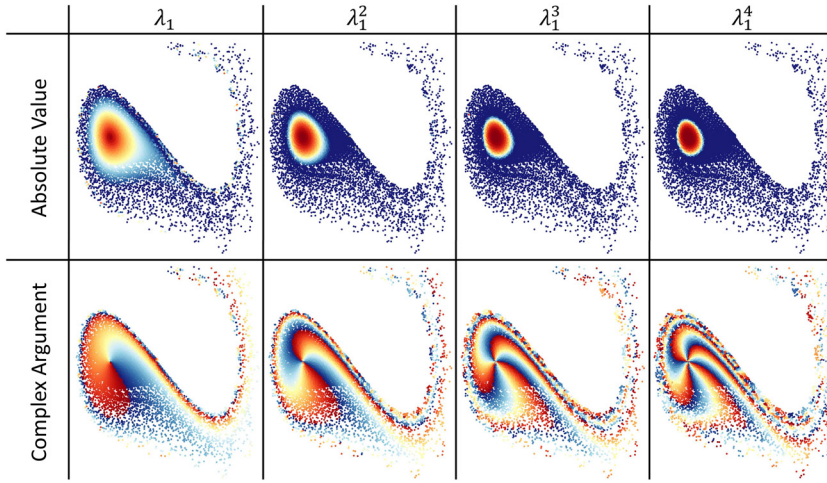

FIGURE 19 Top row: Absolute value of eigenfunction. Bottom row: Complex argument of eigenfunction. Left to right: Eigenfunctions corresponding to powers λ_1, λ_1^2, λ_1^3 and λ_1^4 of the fundamental eigenvalue λ_1.

now more distinctly observable. In Fig. 19, we plot the eigenfunctions corresponding to the powers λ_1, λ_1^2, λ_1^3, and λ_1^4 of the fundamental eigenvalue. Note that the eigenfunctions are successive powers of one another. Furthermore, the amplitude and phase of a Koopman eigenfunction are analogous to an 'action–angle' parametrization of the basin of attraction. The level sets of the absolute values of the eigenfunctions are the so-called isostables, while the level sets of the arguments of the eigenfunctions are termed isochrons (Mauroy et al., 2013).

4.2 Time-delay embedding

In many applications, only partial observations of the system are available, leading to hidden or latent variables. Additionally, the explicit construction of a robust nonlinear dictionary can be challenging, particularly when the system evolves on a low-dimensional attractor that may be unknown or fractal. Nevertheless, it is often feasible to utilize time-delayed measurements of the system to construct an augmented state vector. This approach yields an intrinsic coordinate system that is hoped to form an approximate invariant subspace. This technique was discussed in Section 2.3.2, where it was justified by Takens' embedding theorem (Takens, 2006). Mezić and Banaszuk (2004) established the connection between delay embeddings and the Koopman operator via a statistical Takens' embedding theorem.

Employing the same time step for both the delay interval and the frequency of measurements results in a data matrix with a Hankel structure. Hankel matrices have been used in system identification for decades, as seen in the eigensystem realization algorithm (Juang and Pappa, 1985) and singular spectrum

analysis (Broomhead and Jones, 1989). Although these early algorithms were initially developed for linear systems, they have frequently been applied to weakly nonlinear systems as well. The practice of computing DMD on a Hankel matrix was introduced by Tu et al. (2014b) and subsequently utilized in the field of neuroscience (B.W. Brunton et al., 2016). In this section, we focus on two prevalent methods: Hankel-DMD, which is essentially EDMD applied to a dictionary created from time-delay embedding, and the Hankel Alternative View of Koopman (HAVOK) framework, which enhances the DMD model by incorporating a forcing term.

4.2.1 Hankel dynamic mode decomposition (Hankel-DMD)

Hankel-DMD, introduced by Arbabi and Mezić (2017a) and closely related to the Prony approximation of the KMD (Susuki and Mezić, 2015), represents a specialized instance of EDMD where the dictionary is constructed through time-delay embedding. This approach is particularly effective for ergodic systems that exhibit low-dimensional attractors. We saw a slightly generalized variant of this algorithm in Section 2.3.2, where we employed distinct time steps for sampling trajectories and the lengths of the time delays. Typically, Hankel-DMD utilizes the same time steps for delay embedding and trajectory data collection.

Suppose the map $\mathbf{F}$ in (1.1) is ergodic. We can construct a dictionary by starting with an observable g and forming the *Krylov subspace*

$$V_N = \operatorname{span}\{g, \mathcal{K}g, \mathcal{K}^2 g, \dots, \mathcal{K}^{N-1} g\}.$$

Given a single trajectory of the observable, $\{g(\mathbf{x}_0), g(\mathbf{x}_1), \dots, g(\mathbf{x}_{M+N-1})\}$, the matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ in (4.2) are given explicitly by the Hankel matrices

$$\mathbf{\Psi}_X = \begin{pmatrix} g(\mathbf{x}_0) & g(\mathbf{x}_1) & \cdots & g(\mathbf{x}_{N-1}) \\ g(\mathbf{x}_1) & g(\mathbf{x}_2) & \cdots & g(\mathbf{x}_N) \\ \vdots & \vdots & \vdots & \vdots \\ g(\mathbf{x}_{M-1}) & g(\mathbf{x}_M) & \cdots & g(\mathbf{x}_{M+N-2}) \end{pmatrix},$$

$$\mathbf{\Psi}_Y = \begin{pmatrix} g(\mathbf{x}_1) & g(\mathbf{x}_2) & \cdots & g(\mathbf{x}_N) \\ g(\mathbf{x}_2) & g(\mathbf{x}_3) & \cdots & g(\mathbf{x}_{N+1}) \\ \vdots & \vdots & \vdots & \vdots \\ g(\mathbf{x}_M) & g(\mathbf{x}_{M+1}) & \cdots & g(\mathbf{x}_{M+N-1}) \end{pmatrix}. \tag{4.10}$$

Applying Birkhoff's ergodic theorem (4.8), we obtain the convergence specified in (4.5). A common simplifying assumption in Hankel-DMD is the existence of a finite-dimensional $\mathcal{K}$-invariant subspace V of $L^2(\Omega, \omega)$ generated by g. $\mathcal{K}$-invariance means that $\mathcal{K}V \subset V$ and allows us to study a portion of the spectral properties of $\mathcal{K}$ by restricting to the finite-dimensional subspace V. Suppose such a subspace exists and has dimension k, then $V_k = V$. We can identify this

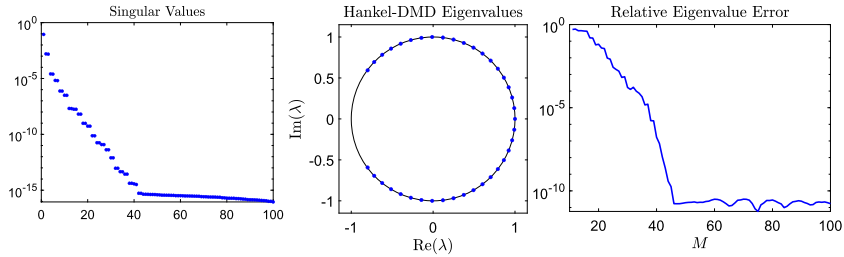

FIGURE 20 Left: Singular values of $\mathbf{\Psi}_X$ for the Hankel-DMD algorithm applied to the cylinder wake with $N = 100$ and $M = 120$. In essence, we are building a Krylov subspace by measuring the horizontal component of the velocity field at a single point. Middle: Eigenvalues computed using Hankel-DMD. Right: Convergence to the first 11 eigenvalues with increasing amount of data M.

invariant subspace as $M \to \infty$ by selecting $N = k$ and employing the aforementioned dictionary. This is proven in Arbabi and Mezić (2017a) and is derived from the ergodic theorem in conjunction with the quadrature interpretation of EDMD. These findings also apply when constructing a Krylov subspace from multiple initial observables $g_1, \ldots, g_p$. Nonetheless, the existence of such a subspace is not guaranteed, and even if it is, the dimension k is typically unknown. Practically, one postulates an *approximate invariant subspace* and truncates to $r \leq N$ modes for the basis by executing an SVD.

As an example, we revisit the cylinder wake discussed in Section 2.3.1. For the observable g, we choose the horizontal velocity at a *single point* in the middle of the channel, situated $4D$ downstream from the center of the cylinder. Initially setting $N = 100$ and $M = 120$, we plot the singular values of the data matrix $\mathbf{\Psi}_X$ on the left side of Fig. 20. It is crucial to recognize that although the spectrum is pure point in this example, g does not generate a finite-dimensional invariant subspace since g projects nontrivially onto each eigenspace. Guided by these singular values, we apply Algorithm 1 with $r = 39$, using the transposes of $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ as the snapshot matrices. The eigenvalues are illustrated in the middle of Fig. 20 and should be compared with a subset of the eigenvalues from Fig. 2. On the right side of Fig. 20, we present the relative ℓ^2 error for the first 11 eigenvalues as a function of M. The convergence is remarkable. Nonetheless, we must stress that this example is rather straightforward. Systems such as the Lorenz system, as discussed in Section 2.3.2 and tackled in Arbabi and Mezić (2017a, Section 4.1), pose a substantially more significant challenge. This is further exemplified by a slow decay of singular values in the data matrices.

4.2.2 Hankel alternative view of Koopman (HAVOK)

Consider the (truncated) SVD of the transpose of the matrix $\mathbf{\Psi}_X$ in (4.10),

$$\mathbf{\Psi}_X^\top \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \quad \mathbf{U} \in \mathbb{C}^{N\times r}, \quad \mathbf{\Sigma} \in \mathbb{R}^{r\times r}, \quad \mathbf{V} \in \mathbb{C}^{M\times r}.$$

We can view the columns of the matrix $\mathbf{V}$ as coordinates for a state $\mathbf{v} = [v_1 \;\; v_2 \;\; \cdots \;\; v_r]$. If our discrete-time dynamical system corresponds to sampling

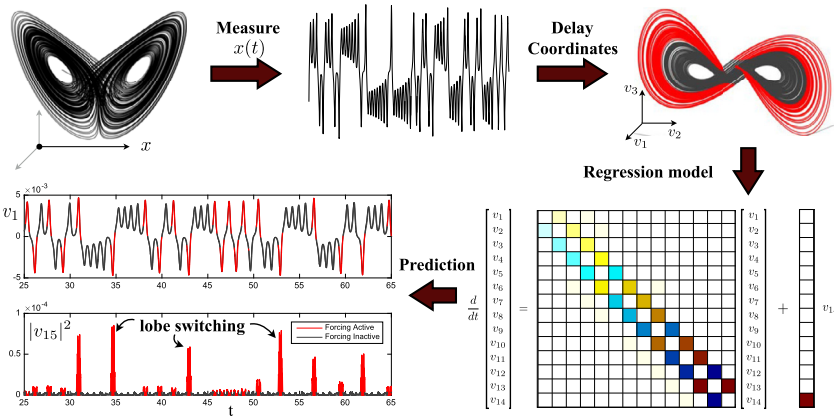

FIGURE 21 Decomposition of chaos into a linear system with forcing. A time series $x(t)$ is stacked into a Hankel matrix, the SVD of which yields a hierarchy of *eigen* time series that produce a delay-embedded attractor. A best-fit linear regression model is obtained on the delay coordinates; the linear fit for the first $r-1$ variables is excellent, but the last coordinate v_r is not well-modeled as linear. Instead, v_r is an input that forces the first $r-1$ variables. Rare forcing events correspond to lobe switching in the chaotic dynamics. Reproduced with permission from Brunton et al. (2022), copyright © 2022 Society for Industrial and Applied Mathematics, all rights reserved. This in turn was adapted from Brunton et al. (2017).

a continuous-time dynamical system, DMD/EDMD results in a linear regression model

$$\frac{d\mathbf{v}}{dt} = \hat{\mathbf{K}}\mathbf{v}, \quad \text{for some matrix} \quad \hat{\mathbf{K}} \in \mathbb{C}^{r\times r}.$$

This can be very effective for weakly nonlinear systems (Champion et al., 2019) and if r is sufficiently large to capture an almost invariant subspace (Arbabi and Mezić, 2017a). However, it can be challenging to identify a small (approximately) closed linear model for chaotic systems.

An alternative, known as the *Hankel Alternative View of Koopman* (HAVOK) framework, proposed by Brunton et al. (2017), is to build a linear model on the first $r-1$ variables $\tilde{\mathbf{v}} = [v_1 \;\; v_2 \;\; \cdots \;\; v_{r-1}]$ and impose the last variable, v_r, as a forcing term:

$$\frac{d\tilde{\mathbf{v}}}{dt} = \tilde{\mathbf{K}}\tilde{\mathbf{v}} + \mathbf{B}v_r, \quad \text{for some matrices} \quad \tilde{\mathbf{K}} \in \mathbb{C}^{r-1\times r-1}, \mathbf{B} \in \mathbb{C}^{r-1\times 1}.$$

Here, v_r acts as an input forcing to the linear dynamics of the model, which approximates the nonlinear dynamics of the original system. Typically, the statistics of v_r are non-Gaussian. For instance, in Fig. 21, we summarize the results for the Lorenz system. The long tails in the statistics of v_r correspond to rare-event forcing that drives lobe switching. For strategies on using HAVOK in systems with multiple time scales, see Champion et al. (2019). Hirsh et al. (2021) established connections between HAVOK and the Frenet–Serret frame

from differential geometry, motivating a more accurate computational modeling approach.

4.3 Controlling projection errors: residual dynamic mode decomposition (ResDMD)

We saw in Section 4.1 that EDMD builds a finite matrix approximation of the Koopman operator. In particular, for a dictionary $\{\psi_1, \ldots, \psi_N\}$ forming a finite-dimensional subspace $V_N = \text{span}\{\psi_1, \ldots, \psi_N\}$, the EDMD matrix corresponds to the projected operator $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}^*_{V_N}$. Care must be taken when discretizing or truncating an infinite-dimensional operator to a finite matrix to compute spectral properties. In general, several well-studied pitfalls include:

- **Spectral Pollution:** This term describes false eigenvalues that accumulate at points not in the spectrum as the discretization size increases.
- **Spectral Invisibility:** Discretizing an operator can cause us to miss parts of its spectrum, even as the size of the discretization increases.
- **Lack of Verification:** Even if a method converges as the discretization parameter grows, how much of the output can we trust for a finite discretization size?
- **Continuous Spectra:** Discretizing to a finite matrix results in a discrete set of eigenvalues. How can we recover continuous spectra?

We have already encountered these effects in this review (e.g., spectral pollution as discussed in Fig. 7), and they are well-known throughout the Koopman literature. In the following section, we will consider strategies to mitigate these issues, focusing on controlling projection errors when transitioning from $\mathcal{K}$ to $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}^*_{V_N}$. The algorithm that does this is *Residual DMD* (ResDMD), introduced by Colbrook and Townsend (2023).

4.3.1 The algorithm

The main idea behind ResDMD is to compute an infinite-dimensional residual. We follow the notation of Section 4.1 that described EDMD. Consider an observable $g = \boldsymbol{\Psi}\mathbf{g} \in V_N$, which we aim to be an approximate eigenfunction of $\mathcal{K}$ with an approximate eigenvalue λ. For now, the method of determining the pair (λ, g) is left unspecified. In connection with pseudospectra and the approximate point spectrum discussed in Section 2.1.2, a way to measure the suitability of the candidate pair (λ, g) is through the relative residual

$$\begin{aligned}\frac{\|(\mathcal{K}-\lambda I)g\|}{\|g\|} &= \sqrt{\frac{\int_\Omega |[\mathcal{K}g](\mathbf{x}) - \lambda g(\mathbf{x})|^2 \,\mathrm{d}\omega(\mathbf{x})}{\int_\Omega |g(\mathbf{x})|^2 \,\mathrm{d}\omega(\mathbf{x})}} \\ &= \sqrt{\frac{\langle \mathcal{K}g, \mathcal{K}g\rangle - \lambda\langle g, \mathcal{K}g\rangle - \overline{\lambda}\langle \mathcal{K}g, g\rangle + |\lambda|^2\langle g, g\rangle}{\langle g, g\rangle}}. \end{aligned} \tag{4.11}$$

For instance, if $\mathcal{K}$ is a normal operator (one that commutes with its adjoint), then

$$\operatorname{dist}(\lambda, \operatorname{Sp}(\mathcal{K})) = \inf_f \frac{\|(\mathcal{K} - \lambda I) f\|}{\|f\|} \leq \frac{\|(\mathcal{K} - \lambda I) g\|}{\|g\|}.$$

In the case of a nonnormal $\mathcal{K}$, the residual in (4.11) is closely related to the concept of pseudospectra. Adopting the quadrature interpretation of EDMD, we can define a finite data approximation of the relative residual as:

$$\operatorname{res}(\lambda, g) = \|(\mathbf{D}^{1/2}\boldsymbol{\Psi}_Y - \lambda \mathbf{D}^{1/2}\boldsymbol{\Psi}_X)\mathbf{g}\|_{\ell^2} / \|\mathbf{D}^{1/2}\boldsymbol{\Psi}_X \mathbf{g}\|_{\ell^2}.$$

We then have

$$\begin{aligned}[\operatorname{res}(\lambda, g)]^2 &= \frac{\mathbf{g}^* \left[\boldsymbol{\Psi}_Y^* \mathbf{D} \boldsymbol{\Psi}_Y - \lambda \boldsymbol{\Psi}_Y^* \mathbf{D} \boldsymbol{\Psi}_X - \overline{\lambda} \boldsymbol{\Psi}_X^* \mathbf{D} \boldsymbol{\Psi}_Y + |\lambda|^2 \boldsymbol{\Psi}_X^* \mathbf{D} \boldsymbol{\Psi}_X\right] \mathbf{g}}{\mathbf{g}^* \boldsymbol{\Psi}_X^* \mathbf{D} \boldsymbol{\Psi}_X \mathbf{g}} \\ &= \frac{\mathbf{g}\left[\boldsymbol{\Psi}_Y^* \mathbf{D} \boldsymbol{\Psi}_Y - \lambda \mathbf{A}^* - \overline{\lambda} \mathbf{A} + |\lambda|^2 \mathbf{G}\right] \mathbf{g}}{\mathbf{g}^* \mathbf{G} \mathbf{g}}, \end{aligned} \tag{4.12}$$

where $\mathbf{G}$ and $\mathbf{A}$ are the same matrices from (4.4). The right-hand side of (4.12) has an additional matrix $\mathbf{L} := \boldsymbol{\Psi}_Y^* \mathbf{D} \boldsymbol{\Psi}_Y$. Under the assumption that the quadrature rule converges, this matrix approximates $\mathcal{K}^*\mathcal{K}$:

$$\lim_{M \to \infty} \mathbf{L}_{jk} = \langle \mathcal{K}\psi_k, \mathcal{K}\psi_j \rangle. \tag{4.13}$$

Comparing (4.11) and the square-root of (4.12), we observe that

$$\lim_{M \to \infty} \operatorname{res}(\lambda, g) = \|(\mathcal{K} - \lambda I) g\| / \|g\|.$$

Note that there is no approximation or projection on the right-hand side of this equation. Consequently, we can compute an *infinite-dimensional residual* directly using finite matrices, achieving exactness in the limit of large data sets. ResDMD leverages this residual in a suite of algorithms to compute various spectral properties of $\mathcal{K}$, two of which are the focus of the subsequent discussion.

As a first approach, we can implement the EDMD algorithm (Algorithm 10) to generate candidate eigenpairs (λ, g), followed by the computation of residuals. This process is outlined in Algorithm 11. Importantly, this approach is not more computationally demanding than EDMD itself. Additionally, it is worth noting that one is not restricted to using EDMD exclusively for selecting candidate eigenpairs; any suitable method can be employed. We can avoid spectral pollution by setting a threshold to discard residuals that exceed a certain tolerance. This also serves as a validation mechanism for the computations. However, it is crucial to recognize that Algorithm 11, when relying on EDMD for computing candidate eigenpairs, does not inherently circumvent the issue of spectral

Algorithm 11 ResDMD for computing residuals (Colbrook and Townsend, 2023).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, quadrature weights $\{w_m\}_{m=1}^M$, and a dictionary of functions $\{\psi_j\}_{j=1}^N$.

1: Compute the matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ defined in (4.2) and $\mathbf{D} = \mathrm{diag}(w_1, \dots, w_M)$.
2: Compute the EDMD matrix $\mathbf{K} = (\mathbf{D}^{1/2}\mathbf{\Psi}_X)^\dagger \mathbf{D}^{1/2}\mathbf{\Psi}_Y \in \mathbb{C}^{N\times N}$.
3: Compute the eigendecomposition $\mathbf{KV} = \mathbf{V\Lambda}$. The columns of $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ are eigenvector coefficients and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues $\lambda_1, \dots, \lambda_n$.
4: For each eigenpair $(\lambda_j, \mathbf{v}_j)$ compute $\mathrm{res}(\lambda_j, \mathbf{\Psi}\mathbf{v}_j) = \|(\mathbf{D}^{1/2}\mathbf{\Psi}_Y - \lambda_j \mathbf{D}^{1/2}\mathbf{\Psi}_X)\mathbf{v}_j\|_{\ell^2} / \|\mathbf{D}^{1/2}\mathbf{\Psi}_X \mathbf{v}_j\|_{\ell^2}$.

Output: The eigenvalues $\mathbf{\Lambda}$, eigenvector coefficients $\mathbf{V} \in \mathbb{C}^{N\times N}$ and residuals $\{\mathrm{res}(\lambda_j, \mathbf{\Psi}\mathbf{v}_j)\}$.

invisibility. To address this, we need to consider approaches that approximate the pseudospectrum.

For computing pseudospectra, working in the standard ℓ^2 norm is beneficial instead of the norm induced by the matrix $\mathbf{G}$. We compute an economy QR decomposition of the data matrix

$$\mathbf{D}^{1/2}\mathbf{\Psi}_X = \mathbf{QR}, \quad \mathbf{Q} \in \mathbb{C}^{M\times N}, \mathbf{R} \in \mathbb{C}^{N\times N},$$

where $\mathbf{Q}$ has orthonormal columns and $\mathbf{R}$ is upper triangular with positive diagonals. Letting $\mathbf{w} = \mathbf{Rg}$, we have

$$\|\mathbf{D}^{1/2}\mathbf{\Psi}_X \mathbf{g}\|_{\ell^2}^2 = \mathbf{g}^*\mathbf{R}^*\mathbf{Q}^*\mathbf{QRg} = \mathbf{g}^*\mathbf{R}^*\mathbf{Rg} = \mathbf{w}^*\mathbf{w} = \|\mathbf{w}\|_{\ell^2}^2.$$

Consequently, the residual can be expressed as:

$$\mathrm{res}(z, g) = \|(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q})\mathbf{w}\|_{\ell^2} / \|\mathbf{w}\|_{\ell^2}. \tag{4.14}$$

For a given $z \in \mathbb{C}$, our objective is to minimize this residual, which corresponds to finding the smallest singular value of the matrix $(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q}) \in \mathbb{C}^{M\times N}$. Denoting the smallest singular value by $\sigma_{\inf}$, we must do this for various values of z. Given that $M > N$, a computational advantage is gained by considering the $N \times N$ matrix $(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q})^*(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q})$ and computing

$$\sqrt{\sigma_{\inf}((\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q})^*(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q}))} = \sigma_{\inf}(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q}).$$

Typically, computing singular values in this manner is not recommended due to the potential loss of precision owing to the square root. However, in most applications, the resulting error is significantly smaller than the errors inherent in

Algorithm 12 ResDMD for computing pseudospectra (Colbrook and Townsend, 2023). One can also compute the singular values directly (of a $\mathbb{C}^{M\times N}$ matrix) without the square root.

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, quadrature weights $\{w_m\}_{m=1}^M$, dictionary of functions $\{\psi_j\}_{j=1}^N$, accuracy goal $\epsilon > 0$, and grid of points $\{z_\ell\}_{\ell=1}^k \subset \mathbb{C}$.

1: Compute the matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ defined in (4.2) and $\mathbf{D} = \mathrm{diag}(w_1, \dots, w_M)$.
2: Compute an economy QR decomposition $\mathbf{D}^{1/2}\mathbf{\Psi}_X = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{C}^{M\times N}$, $\mathbf{R} \in \mathbb{C}^{N\times N}$.
3: Compute $\mathbf{C}_2 = (\mathbf{R}^*)^{-1}\mathbf{\Psi}_Y^*\mathbf{D}\mathbf{\Psi}_Y\mathbf{R}^{-1}$ and $\mathbf{C}_1 = \mathbf{Q}^*\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1}$.
4: Compute $\tau_\ell = \sigma_{\inf}(\mathbf{C}_2 - z_\ell\mathbf{C}_1^* - \overline{z_\ell}\mathbf{C}_1 + |z_\ell|^2\mathbf{I})$ for $\ell = 1, \dots, k$ ($\sigma_{\inf}$ is smallest singular value).
(If wanted, compute the corresponding right-singular vectors $\mathbf{w}_\ell$ and set $\mathbf{v}_j = \mathbf{R}^{-1}\mathbf{w}_j$.)

Output: Estimate of the pseudospectrum $\{z_\ell : \tau_\ell < \epsilon\}$ (if wanted, corresponding pseudoeigenfunctions $\{\mathbf{\Psi}\mathbf{v}_\ell : \tau_\ell < \epsilon\}$).

the data matrices or the quadrature approximation of inner products. If precision becomes a concern, $\sigma_{\inf}(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - \lambda\mathbf{Q})$ can be computed directly in subsequent algorithms. Since $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}$, we have

$$\begin{aligned}(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q})^*(\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z\mathbf{Q})\\ = (\mathbf{R}^*)^{-1}\mathbf{\Psi}_Y^*\mathbf{D}\mathbf{\Psi}_Y\mathbf{R}^{-1} - z(\mathbf{R}^*)^{-1}\mathbf{\Psi}_Y^*\mathbf{D}^{1/2}\mathbf{Q} - \overline{z}\mathbf{Q}^*\mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{R}^{-1} + |z|^2\mathbf{I}.\end{aligned}$$

The minimum singular values of this matrix are then computed across a grid of z values. This procedure is detailed in Algorithm 12, where the approximation of the ϵ-pseudospectrum is defined as the set of grid points where the minimized residual falls below ϵ. If required, the algorithm can also be extended to compute ϵ-pseudoeigenfunctions (discussed in Section 2.1.2).

When $M < N$

In the above discussion, we assumed that $M \geq N$. If $M < N$, then there are two options. One can consider two subsets of snapshot data (training and test) (Colbrook and Townsend, 2023). Or, as developed in Colbrook (2024), in particular for exact DMD and kernelized EDMD, one can compute dual residuals.

4.3.2 Convergence theory

Colbrook and Townsend (2023) present several convergence results concerning ResDMD. We have already discussed that if the quadrature rule underlying

EDMD converges,

$$\lim_{M\to\infty} \operatorname{res}(\lambda, g) = \|(\mathcal{K} - \lambda I)g\|/\|g\|.$$

Therefore, we can avoid spectral pollution in the large data limit by selecting eigenpairs with small residuals (as computed in Algorithm 11). Let $\Gamma^{\epsilon}_{N,M}$ be the output $\{z_\ell : \tau_\ell < \epsilon\}$ of Algorithm 12. With a minor modification for the boundary case where $\tau = \epsilon$,

$$\lim_{M\to\infty} \Gamma^{\epsilon}_{N,M} =: \Gamma^{\epsilon}_{N} \subset \mathrm{Sp}_{\epsilon}(\mathcal{K}).$$

In other words, ResDMD provides verified approximations of pseudospectra. Moreover, under mild conditions on the dictionary and an N-dependent grid $\{z_\ell\}_{\ell=1}^{k}$,

$$\lim_{N\to\infty} \Gamma^{\epsilon}_{N} = \operatorname{Cl}\left(\left\{\lambda \in \mathbb{C} : \exists g \in L^2(\Omega, \omega) \text{ such that } \|g\| = 1, \|(\mathcal{K} - \lambda I)g\| < \epsilon\right\}\right).$$

As $\epsilon \downarrow 0$, the set on the right-hand side converges to the approximate point spectrum $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$. Thus, ResDMD allows us to compute $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$ via a convergent algorithm. Colbrook and Townsend (2023) further discuss alterations that allow the computation of the full pseudospectrum $\mathrm{Sp}_{\epsilon}(\mathcal{K})$, and consequently the complete spectrum $\mathrm{Sp}(\mathcal{K})$. In summary, ResDMD addresses the challenges of spectral pollution and spectral invisibility, providing a method for verified spectral computations of general Koopman operators.

A careful reader will note that a few of these algorithms require us to take several parameters successively to infinity. This was also the case for EDMD, as discussed in Section 4.1.3. These limits do not generally commute, and it may be impossible to rewrite them with fewer limits or develop a different algorithm that uses fewer limits. This is a generic feature of infinite-dimensional spectral problems (Colbrook, 2020) and has given rise to the *Solvability Complexity Index* (Hansen, 2011; Ben-Artzi et al., 2020; Colbrook, 2022; Colbrook and Hansen, 2022). We do not go into the details, but there are many open questions on the foundations of computing spectral properties of Koopman operators. In particular, lower bounds on the number of successive limits needed to compute spectra of Koopman operators are an ongoing research problem (see Section 6.4).

We have yet to discuss continuous spectra, the final pitfall mentioned in the bullet point list at the start of this section. Using ResDMD, Colbrook and Townsend (2023) also provide an algorithm that computes spectral measures of Koopman operators associated with generic measure-preserving systems. This approach and others for spectral measures are discussed in Section 6.2.

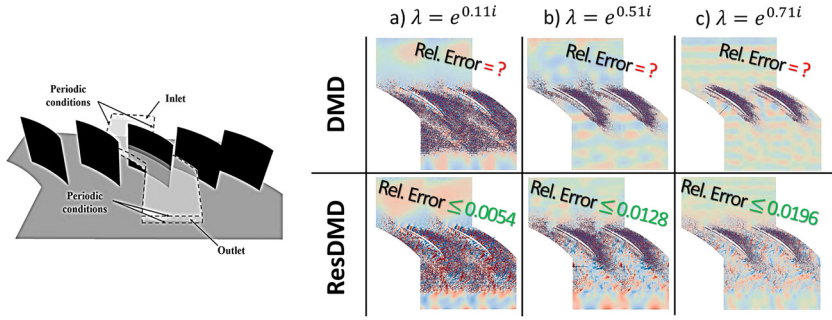

FIGURE 22 Left: Large-scale wall-resolved turbulent flow past a periodic cascade of aerofoils. Right: Comparison of computed Koopman modes using ResDMD and DMD across various frequencies. ResDMD highlights stronger acoustic waves between cascades and larger-scale turbulent fluctuations past the trailing edge, which is crucial for understanding acoustic interactions with engine turbines and nearby structures. The residuals in ResDMD underscore its capability to capture nonlinear dynamics accurately and verifiably. Reproduced with permission from Colbrook and Townsend (2023).

4.3.3 Examples

We have already seen an example of ResDMD in action in Section 2.3.3. Here, we present some examples from Colbrook and Townsend (2023); Colbrook et al. (2023a).

The first example we consider is a large-scale wall-resolved turbulent flow past a periodic cascade of aerofoils depicted on the left in Fig. 22. This setup is motivated by ongoing efforts to mitigate noise sources from aerial vehicles. The data is collected from a high-fidelity simulation solving the fully nonlinear Navier–Stokes equations (Koch et al., 2021), with a Reynolds number of 3.88×10^5 and a Mach number of 0.07. A two-dimensional slice of the pressure field is recorded at 295,122 points across trajectories of length 798 and sampled every 2×10^{-5} seconds. ResDMD can be used with kernelized EDMD, and we use $N = 250$ functions in our dictionary. Fig. 22 (right) shows the computed Koopman modes for a range of representative frequencies. We also show the corresponding Koopman modes computed using DMD. For the first column, ResDMD shows stronger acoustic waves between the cascades. Detecting these vibrations is essential as they can damage engine turbines (Parker, 1984). ResDMD shows larger-scale turbulent fluctuations past the trailing edge for the second and third columns. This can be crucial for understanding acoustic interactions with nearby structures such as subsequent blade rows (Woodley and Peake, 1999). The residuals for ResDMD are small, particularly given the enormous state-space dimension. This example demonstrates two benefits of ResDMD compared with DMD: (1) ResDMD can capture the nonlinear dynamics (just like EDMD), and (2) it computes residuals, thus providing an accuracy certificate.

With its capability to verifiably compute spectra, ResDMD can be employed for validating dictionaries in methods like EDMD and verifying the efficacy

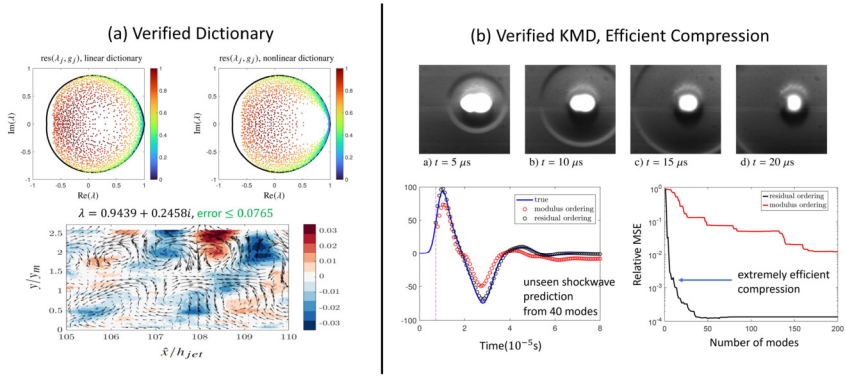

FIGURE 23 ResDMD applications in validating EDMD and KMD. Left: Comparison of two dictionaries (linear: truncated SVD, nonlinear: kernel method) for turbulent boundary layer flow analysis, with the nonlinear dictionary showing smaller residuals and revealing verified transient modes (bottom of the figure). Right: Demonstration of KMD's ability to capture a highly nonlinear shockwave, where ordering modes by residual values enables efficient data compression and precise shockwave prediction. Reproduced with permission from Colbrook et al. (2023a).

of KMD itself. We present two illustrative examples in Fig. 23, adapted from Colbrook et al. (2023a). The comparison of two dictionaries used for analyzing turbulent boundary layer flow is shown on the left. Here, the nonlinear dictionary demonstrates smaller residuals, leading to the identification of verified transient modes, as depicted at the bottom of the figure. On the right, the figure illustrates the proficiency of KMD in capturing a highly nonlinear shockwave. By ordering modes based on their residual values, we achieve efficient data compression and accurate prediction of the shockwave dynamics.

5 Variants that preserve structure

One of the most exciting recent developments in DMD is the introduction of methods that preserve structures of the underlying dynamical system in (1.1). When studying a system from a data-driven perspective, it is often the case that one possesses partial knowledge of the system's underlying physics. Methods that leverage this structure typically exhibit a greater resistance to noise, better generalization, and demand less data for training. Structure-preserving algorithms have a deep-rooted history in geometric integration (Hairer et al., 2010) and have recently gained traction in data-driven methods (Celledoni et al., 2021; Greydanus et al., 2019; Hernández et al., 2021; Hesthaven et al., 2022; Karniadakis et al., 2021; Loiseau and Brunton, 2018; Otto et al., 2023b). In the context of DMD, this area is burgeoning. We will concentrate on three methods:

- **Physics-Informed DMD:** This provides a framework for incorporating symmetries into DMD through additional constraints in the least-squares problem (2.6). The original paper focused on five fundamental physical principles:

conservation, self-adjointness, localization, causality, and shift-equivariance. The idea is far more general and has ushered in a new wave of DMD methods.

- **Measure-Preserving EDMD:** This enforces measure-preserving EDMD truncations, leading to a Galerkin method whose eigendecomposition converges to the spectral quantities of Koopman operators (including spectral measures and continuous spectra) for general measure-preserving dynamical systems. Like EDMD, it can be used with any dictionary of observables and with different data types. Preserving the measure is crucial for convergence, recovering the correct dynamical behavior, stability, robustness to noise, and improved qualitative and long-time behavior.
- **Compactification:** These methods for continuous-time measure-preserving systems are based on the compactification of the Koopman generator or its resolvent. They automatically lead to skew-adjoint approximations whose spectral properties converge to that of the Koopman generator. Additionally, approximations are expressed in a well-conditioned basis of kernel eigenvectors computed from trajectory data.

Subsequently, we will discuss additional DMD methods based on preserving structure. The methods we discuss open the door to future extensions to more general structure-preserving methods for Koopman operators and data-driven dynamical systems.

5.1 Physics-informed dynamic mode decomposition (piDMD)

5.1.1 The framework

Physics-Informed DMD (piDMD), introduced by Baddoo et al. (2023), provides an overarching framework for integrating physical principles – such as symmetries, invariances and conservation laws – into DMD. The idea is to replace the optimization problem in (2.6) by a *constrained* optimization problem

$$\min_{\mathbf{K}_{\mathrm{piDMD}} \in \mathcal{M}} \left\| \mathbf{Y} - \mathbf{K}_{\mathrm{piDMD}} \mathbf{X} \right\|_{\mathrm{F}}. \tag{5.1}$$

The matrix manifold $\mathcal{M}$ is dictated by the known physics of the system in (1.1). One selects $\mathcal{M}$ so that its members satisfy certain symmetries of the system. The optimization problem in (5.1) is known as a Procrustes problem,[18] which comprises of finding the optimal transformation between two matrices subject to certain constraints. Numerous exact solutions exist for Procrustes problems, including the notable cases of orthogonal matrices (Schönemann, 1966), and symmetric matrices (Higham, 1988). When exact solutions are not possible, algorithmic solutions can be effective (Boumal et al., 2014). Procrustes analysis finds relevance in many fields, as detailed in the monograph of Gower and Dijksterhuis (2004).

[18] In Greek mythology, Procrustes was a bandit who would stretch or amputate the limbs of his victims to force them to fit onto his bed. Herein, $\mathbf{X}$ plays the role of Procrustes' victim, $\mathbf{Y}$ is the bed, and $\mathbf{K}_{\mathrm{piDMD}}$ is the 'treatment' (stretching or amputation).

To apply piDMD, we first identify the system's known or suspected physical properties. Once the physical principles we wish to enforce are determined, these laws must be translated into the matrix manifold where the linear model will be constrained. With a defined target matrix manifold, we numerically solve the relevant Procrustes problem in (5.1). The concluding step encompasses extracting physical information from the refined model. For instance, one might analyze the spectrum, DMD modes, and the related KMD.

Baddoo et al. (2023) focus on five fundamental physical principles: conservation, self-adjointness, localization, causality, and shift-equivariance. Several closed-form solutions and efficient algorithms for the corresponding piDMD optimizations are derived. With fewer degrees of freedom, piDMD models are typically less prone to overfitting, require less training data, and are often less computationally expensive to build than standard DMD models. This reduction in the size of required training data is connected with the problem of matrix recovery from matrix-vector products, whereby enforcing structures reduces the number of queries needed (Halikias and Townsend, 2023). A fundamental issue related to the DMD algorithm is the fact that low-rank matrices are not provably recoverable from snapshot pairs (without access to adjoints) until there are at least as many pairs as state dimensions (Otto et al., 2023a, Thm. 2.5).

5.1.2 Examples

To showcase the breadth of piDMD, Fig. 24 shows six physical examples. Baddoo et al. (2023) provide full experimental details for each example. Each row corresponds to a different system, and the corresponding constraint is listed in the second column. Exact DMD (Algorithm 1) is compared to piDMD in terms of the computed matrix $\mathbf{K}$ and the eigenvalues. In general, constraining the matrix $\mathbf{K}$ to lie on the appropriate manifold $\mathcal{M}$ leads to more accurate approximations of the eigenvalues. The advantage of preserving structure is striking!

5.1.3 Future work

We have only started to tap the potential of adding constraints in the optimization problem in (2.6). This idea will likely be an active research area over the next few years. With that in mind, it is worth mentioning several challenges and directions of future work pointed out by Baddoo et al. (2023):

- **Knowing the physics:** In some scenarios where the physics is poorly understood, determining suitable physical laws to impose on the model can be challenging. Is it possible to learn symmetries and then incorporate them as constraints?
- **Complicated manifolds:** For problems with intricate geometries and multiple dimensions, interpreting the physical principle as a matrix manifold can be a roadblock, as the manifold can become exceedingly complicated.
- **Regularizers:** In many applications, such as when the data are very noisy or the physical laws and constraints are only approximately understood, it

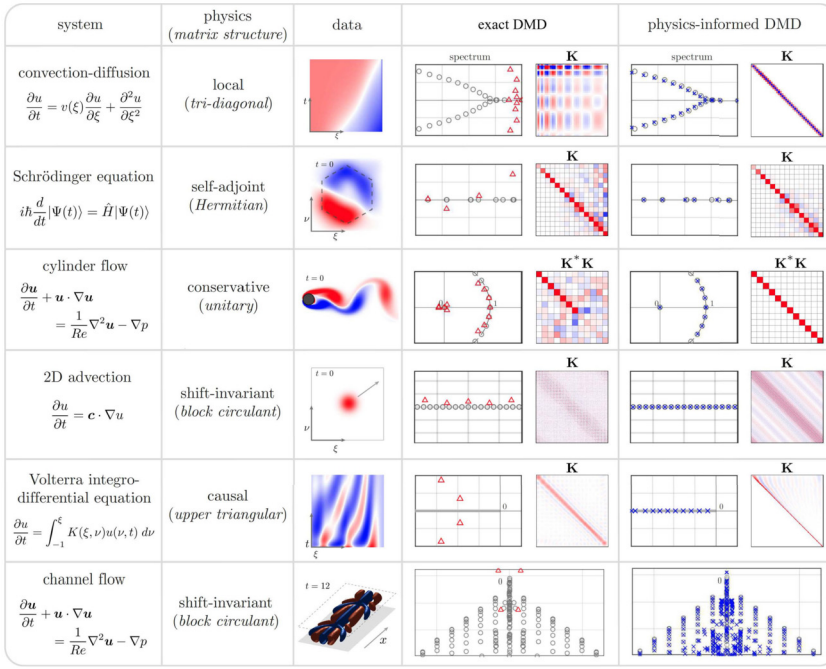

FIGURE 24 A comparison of the models learned by exact DMD (Algorithm 1) and piDMD for a range of applications. The structure of the model matrices is also illustrated. In the spectrum subplots, the true eigenvalues are shown as ○, the DMD eigenvalues as △, and the piDMD eigenvalues as ×. In each case, the eigenvalues of piDMD are more accurate than exact DMD. Reproduced with permission from Baddoo et al. (2023).

may be more appropriate to merely encourage $\mathbf{K}$ towards $\mathcal{M}$, e.g., through a regularizer.

- **Nonlinear observables:** It is not always clear how to extend the approach of piDMD to nonlinear observables and EDMD. Such an approach is crucial for strongly nonlinear systems to maintain the connection with Koopman operators. For example, if the dictionary consists of the state vector $\mathbf{x}$, an upper triangular matrix $\mathbf{K}$ can have a clear meaning in terms of causality. But how should one incorporate causality into other choices of dictionaries? The manifold can depend on the chosen dictionary in a highly complex manner.
- **Convergence:** In connection with the previous point in this list, proving the convergence of piDMD in the large data limit or large dictionary limit typically requires a Galerkin interpretation as in Section 4. This connection is not always immediate.

We expect these last two points, in particular, to lead to many exciting future works.

5.2 Measure-preserving extended dynamic mode decomposition (mpEDMD)

Measure-Preserving EDMD (mpEDMD), introduced by Colbrook (2023), enforces that the EDMD approximation is measure-preserving. The system being measure-preserving is equivalent to the Koopman operator $\mathcal{K}$ being an isometry. Namely, $\|\mathcal{K}g\| = \|g\|$ for any observable $g \in L^2(\Omega, \omega)$. The mpEDMD algorithm is simple and robust, with no tuning parameters. We outline the method, discuss its convergence properties, and end with two examples. Note that we do not need to assume the system is ergodic or invertible.

5.2.1 The algorithm

We follow the notation of Section 4.1 that described EDMD. Recall that we have a dictionary $\{\psi_1, \dots, \psi_N\}$, i.e., a list of observables, in the space $L^2(\Omega, \omega)$. These observables form a finite-dimensional subspace $V_N = \operatorname{span}\{\psi_1, \dots, \psi_N\}$. Our starting point is the observation that the Gram matrix $\mathbf{G} = \mathbf{\Psi}_X^* \mathbf{D} \mathbf{\Psi}_X$ in (4.4) provides an approximation of the inner product $\langle \cdot, \cdot \rangle$ on $L^2(\Omega, \omega)$. Namely, we have the following inner product induced by $\mathbf{G}$:

$$\mathbf{h}^* \mathbf{G} \mathbf{g} = \sum_{j,k=1}^{N} \overline{\mathbf{h}_j} \mathbf{g}_k \mathbf{G}_{j,k} \approx \sum_{j,k=1}^{N} \overline{\mathbf{h}_j} \mathbf{g}_k \langle \psi_k, \psi_j \rangle = \langle \mathbf{\Psi g}, \mathbf{\Psi h} \rangle. \tag{5.2}$$

If the convergence in (4.5) holds, then the left-hand side of (5.2) converges to the right-hand side as $M \to \infty$. Hence, if $g = \mathbf{\Psi g} \in V_N$ and we approximate the action of $\mathcal{K}$ on V_N by a matrix $\mathbf{K}$,

$$\|g\|^2 \approx \mathbf{g}^* \mathbf{G} \mathbf{g}, \quad \|\mathcal{K}g\|^2 \approx \|\mathbf{\Psi K g}\|^2 \approx \mathbf{g}^* \mathbf{K}^* \mathbf{G} \mathbf{K} \mathbf{g}.$$

Since $\mathcal{K}$ is an isometry, $\|g\|^2 = \|\mathcal{K}g\|^2$. Therefore, it is natural to enforce

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

This condition holds if and only if $\mathbf{K}^* \mathbf{G} \mathbf{K} = \mathbf{G}$. Returning to the optimization problem in (4.1), we now make two changes. First, we set $\mathbf{C} = \mathbf{G}^{1/2}$ so that $\|\mathbf{Cg}\|_{\ell^2} = \sqrt{\mathbf{g}^* \mathbf{G} \mathbf{g}} \approx \|g\|$. Second, we enforce the additional constraint $\mathbf{K}^* \mathbf{G} \mathbf{K} = \mathbf{G}$. This leads us to the optimization problem

$$\min_{\substack{\mathbf{K} \in \mathbb{C}^{N \times N} \\ \mathbf{K}^* \mathbf{G} \mathbf{K} = \mathbf{G}}} \int_\Omega \left\| \mathbf{\Psi}(\mathbf{F}(\mathbf{x})) \mathbf{G}^{-1/2} - \mathbf{\Psi}(\mathbf{x}) \mathbf{K} \mathbf{G}^{-1/2} \right\|_{\ell^2}^2 \mathrm{d}\omega(\mathbf{x}). \tag{5.3}$$

In a nutshell, we enforce that our Galerkin approximation is an isometry with respect to the learned, data-driven inner product induced by $\mathbf{G}$. After applying

the quadrature rule we used for EDMD, the discretized version of (5.3) is

$$\min_{\substack{\mathbf{K}\in\mathbb{C}^{N\times N}\\ \mathbf{K}^*\mathbf{G}\mathbf{K}=\mathbf{G}}} \sum_{m=1}^{M} w_m \left\| \mathbf{\Psi}(\mathbf{y}^{(m)})\mathbf{G}^{-1/2} - \mathbf{\Psi}(\mathbf{x}^{(m)})\mathbf{K}\mathbf{G}^{-1/2} \right\|_{\ell^2}^2. \tag{5.4}$$

Letting $\mathbf{K} = \mathbf{G}^{-1/2}\mathbf{B}\mathbf{G}^{1/2}$ for some matrix $\mathbf{B}$, the problem in (5.4) is equivalent to

$$\min_{\substack{\mathbf{B}\in\mathbb{C}^{N\times N}\\ \mathbf{B}^*\mathbf{B}=\mathbf{I}}} \left\| \mathbf{D}^{1/2}\mathbf{\Psi}_X\mathbf{G}^{-1/2}\mathbf{B} - \mathbf{D}^{1/2}\mathbf{\Psi}_Y\mathbf{G}^{-1/2} \right\|_{\mathrm{F}}^2, \tag{5.5}$$

where $\mathbf{I}$ denotes the identity matrix. The optimization problem in (5.5) is known as the *orthogonal Procrustes problem* (Schönemann, 1966; Arun, 1992). The predominant method for computing a solution is via an SVD. First, we compute an SVD of

$$\mathbf{G}^{-1/2}\mathbf{\Psi}_Y^*\mathbf{D}\mathbf{\Psi}_X\mathbf{G}^{-1/2} = \mathbf{G}^{-1/2}\mathbf{A}^*\mathbf{G}^{-1/2} = \mathbf{U}_1\mathbf{\Sigma}\mathbf{U}_2^*,$$

where $\mathbf{A} = \mathbf{\Psi}_X^*\mathbf{D}\mathbf{\Psi}_Y$ is the matrix from (4.4). A solution of (5.5) is then $\mathbf{B} = \mathbf{U}_2\mathbf{U}_1^*$ and we take $\mathbf{K} = \mathbf{G}^{-1/2}\mathbf{U}_2\mathbf{U}_1^*\mathbf{G}^{1/2}$.

Since $\mathbf{K}$ is similar to a unitary matrix, its eigenvalues lie along the unit circle. For stability purposes, the best way to compute the eigendecomposition of $\mathbf{K}$ is to do so for the unitary matrix $\mathbf{U}_2\mathbf{U}_1^*$. To numerically ensure an orthonormal basis of eigenvectors, we use MATLAB's `schur` command in the examples of this paper. It is also beneficial to replace the square root $\mathbf{G}^{1/2}$ with a suitable upper triangular matrix $\mathbf{R}$ such that $\mathbf{G} = \mathbf{R}^*\mathbf{R}$. Such an upper triangular matrix can be computed using an economy QR decomposition of the data matrix as

$$\mathbf{D}^{1/2}\mathbf{\Psi}_X = \mathbf{Q}\mathbf{R}, \quad \mathbf{Q}\in\mathbb{C}^{M\times N}, \mathbf{R}\in\mathbb{C}^{N\times N},$$

where $\mathbf{Q}$ has orthonormal columns and $\mathbf{R}$ is upper triangular with positive diagonals. This leads to a mathematically equivalent algorithm but is faster and more numerically robust in practice.[19] The computation of $\mathbf{K}$ and its eigendecomposition is summarized in Algorithm 13. Arguing as we did for EDMD, we obtain a KMD via

$$g \approx \mathbf{\Psi}\mathbf{V}\left[\mathbf{V}^{-1}(\mathbf{D}^{1/2}\mathbf{\Psi}_X)^\dagger\mathbf{D}^{1/2}\left(g(\mathbf{x}^{(1)}),\ldots,g(\mathbf{x}^{(M)})\right)^\top\right], \quad g\in L^2(\Omega,\omega).$$

Explicitly applied to the state vector $\mathbf{x}$, we have (transposed) Koopman modes

$$\mathbf{\Phi}^\top = \mathbf{V}^{-1}(\mathbf{D}^{1/2}\mathbf{\Psi}_X)^\dagger\mathbf{D}^{1/2}\left(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(M)}\right)^\top \in \mathbb{C}^{N\times d}.$$

Note that mpEDMD can be used with generic choices of dictionary that generate $\mathbf{G}$ and $\mathbf{A}$.

[19] I am indebted to Zlatko Drmač for pointing this out.

Algorithm 13 The mpEDMD algorithm (Colbrook, 2023).

Input: Snapshot data $\mathbf{X} \in \mathbb{C}^{d\times M}$ and $\mathbf{Y} \in \mathbb{C}^{d\times M}$, quadrature weights $\{w_m\}_{m=1}^M$, and a dictionary of functions $\{\psi_j\}_{j=1}^N$.

1: Compute the matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ defined in (4.2) and $\mathbf{D} = \mathrm{diag}(w_1, \dots, w_M)$.
2: Compute an economy QR decomposition $\mathbf{D}^{1/2}\mathbf{\Psi}_X = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{C}^{M\times N}$, $\mathbf{R} \in \mathbb{C}^{N\times N}$.
3: Compute an SVD of $(\mathbf{R}^{-1})^*\mathbf{\Psi}_Y^*\mathbf{D}^{1/2}\mathbf{Q} = \mathbf{U}_1\mathbf{\Sigma}\mathbf{U}_2^*$.
4: Compute the eigendecomposition $\mathbf{U}_2\mathbf{U}_1^* = \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^*$ (via a Schur decomposition).
5: Compute $\mathbf{K} = \mathbf{R}^{-1}\mathbf{U}_2\mathbf{U}_1^*\mathbf{R}$ and $\mathbf{V} = \mathbf{R}^{-1}\hat{\mathbf{V}}$.

Output: Koopman matrix $\mathbf{K}$, with eigenvectors $\mathbf{V}$ and eigenvalues $\mathbf{\Lambda}$.

Finally, the relationship between mpEDMD and piDMD is worth commenting on. For conservative systems, piDMD enforces the DMD matrix in (2.6) to be orthogonal and uses linear observables. This implicitly assumes that these linear observables (and the coordinates used) are orthonormal in $L^2(\Omega, \omega)$, an assumption that typically does not hold. In contrast, mpEDMD works in a data-driven inner product space induced by $\mathbf{G}$. The resulting Gram matrix of the observables must be included in a measure-preserving discretization; otherwise, the wrong measure may be preserved (see the example of turbulent flow in Colbrook, 2023 where mpEDMD and piDMD are contrasted). Thus, we can think of the relationship between mpEDMD and piDMD as akin to the relationship between EDMD and DMD (see the discussion in Sections 2.2.2 and 4.1.1), with an additional difference arising from the use of the inner product arising from the Gram matrix $\mathbf{G}$.

5.2.2 Convergence theory

Several convergence results for mpEDMD are proven in Colbrook (2023). First, echoing Section 4.1.3, we can consider the two limits $M \to \infty$ and $N \to \infty$. Assuming that the quadrature rule underlying EDMD converges, i.e., (4.5) holds, the EDMD matrix corresponds to $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}_{V_N}^*$. In contrast, the mpEDMD matrix corresponds to the unitary part of a *polar decomposition* of $\mathcal{P}_{V_N}\mathcal{K}\mathcal{P}_{V_N}^*$. Call this matrix $\mathbf{K}_N$. Under a natural density assumption of V_N as $N \to \infty$, $\mathbf{K}_N$ converges strongly to $\mathcal{K}$, meaning that (4.9) holds.

We can consider the spectral measures from Section 2.1.2 for a measure-preserving system. These spectral measures provide a diagonalization of the Koopman operator $\mathcal{K}$ and form the foundation of the KMD. As the dictionary $\{\psi_1, \dots, \psi_N\}$ used in EDMD becomes richer, the spectral measures computed by EDMD do not typically converge in any sense to that of $\mathcal{K}$.[20] This contrasts

[20] Even more fundamentally, the eigenvalues of EDMD typically lie within and accumulate within the unit disk. So, the measures are not even on the same space.

the convergence of the spectral measures of mpEDMD (Colbrook, 2023). The critical step in making this convergence work is that mpEDMD provides a unitary Galerkin approximation of $\mathcal{K}$.

The mpEDMD algorithm leads to the following approximations of spectral measures, where $\mathbf{K}$, $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_N]$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ are the outputs of Algorithm 13. To approximate the spectral measure $\mathcal{E}$, we consider the spectral measure, $\mathcal{E}_{N,M}$, of the matrix $\mathbf{K}$ on the Hilbert space $\mathbb{C}^N$ with the inner product in (5.2) induced by G:

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

Let $g \in L^2(\Omega, \omega)$ with $\|g\| = 1$. We approximate μ_g by $\mu_{\mathbf{g}}^{(N,M)}$, where

$$\mathrm{d}\mu_{\mathbf{g}}^{(N,M)}(\lambda) = \sum_{j=1}^{N} \delta(\lambda - \lambda_j) |\mathbf{v}_j^* \mathbf{G} \mathbf{g}|^2 \mathrm{d}\lambda$$

and $\mathbf{g}$ is normalized so that $\mathbf{g}^* \mathbf{G} \mathbf{g} = 1$. Since $\{\mathbf{G}^{1/2}\mathbf{v}_j\}_{j=1}^N$ is an orthonormal basis for $\mathbb{C}^N$, $\mu_{\mathbf{g}}^{(N,M)}$ is a probability measure on the unit circle $\mathbb{T}$.

The most natural way for measures to converge is in a *weak sense*. We say that a sequence of measures μ_n converges weakly to a measure μ on $\mathbb{T}$ if for any continuous function $\phi : \mathbb{T} \to \mathbb{C}$,

$$\lim_{n\to\infty} \int_{\mathbb{T}} \phi(\lambda)\, \mathrm{d}\mu_n(\lambda) = \int_{\mathbb{T}} \phi(\lambda)\, \mathrm{d}\mu(\lambda).$$

This convergence is captured by the so-called Wasserstein 1 metric between probability measures:

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

Under mild conditions on the dictionary as $N \to \infty$, mpEDMD has the following convergence properties. If $\lim_{N\to\infty} \mathbf{\Psi}\mathbf{g}_N = g$ and $\phi : \mathbb{T} \to \mathbb{C}$ is continuous, then

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

Moreover, for the scalar-valued spectral measures,

$$\lim_{N\to\infty} \limsup_{M\to\infty} W_1\left(\mu_g, \mu_{\mathbf{g}}^{(N,M)}\right) = 0. \tag{5.6}$$

Furthermore, if $\{g, \mathcal{K}g, \ldots, \mathcal{K}^{L_N-1}g\} \subset V_N$ (time-delay embedding), then

$$\limsup_{M\to\infty} W_1\left(\mu_g, \mu_{\mathbf{g}}^{(N,M)}\right) \lesssim \log(L_N)/L_N. \tag{5.7}$$

The bound in (5.7) provides an explicit convergence rate in the number of delays used.

Further properties of mpEDMD proven in Colbrook (2023) include respecting invariance subspace properties of $\mathcal{K}$, well-conditioning of the matrix $\mathbf{K}$ and its eigendecomposition (these do not hold for EDMD in general), convergence of KMDs, and

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

In other words, we avoid spectral invisibility and do not miss parts of the spectrum. We can also combine with the techniques of Section 4.3 to avoid spectral pollution. Finally, in connection with Section 3.1, the solution to the orthogonal Procrustes problem (5.5) is also the solution to the corresponding constrained total least squares problem (Arun, 1992). Hence, in a similar vein to tlsDMD in Section 3.1.3, mpEDMD is optimally robust when noise is present in both data matrices in (5.5) (Van Huffel and Vandewalle, 1991).

5.2.3 Examples

We consider two examples of mpEDMD. The first shows the convergence to spectral measures for a system with continuous spectra. The second shows the conservation of energy and statistics for a turbulent boundary layer flow, where the snapshots are collected experimentally.

Convergence to continuous spectra

We first revisit the Lorenz system from Section 2.3.2, but now with a discrete time step of $\Delta t = 0.1$. An arbitrary observable is chosen as

$$g(\mathbf{x}) = g(x, y, z) = c\tanh((xy - 3z)/5),$$

where c is a normalization constant ensuring $\|g\| = 1$. We employ delay-embedding to construct a Krylov subspace $V_N = \{g, \mathcal{K}g, \ldots, \mathcal{K}^{N-1}g\}$. The matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ are computed by evaluating g pointwise at the snapshot matrices of $\mathbf{x}$. A set of $M = 10^4$ snapshots is collected along a single trajectory following an initial burn-in period. It is important to recall that the spectrum of the Koopman operator is continuous, featuring an embedded trivial eigenvalue at $\lambda = 1$. Therefore, we demonstrate the convergence of the mpEDMD approximation of μ_g. For visualization purposes, we transition from variables in $\mathbb{T}$ to complex-arguments in the interval $[-\pi, \pi)$. For a probability measure μ on $\mathbb{T}$, its cumulative distribution function (cdf) on $[-\pi, \pi)$ is defined as

$$F_\mu(\theta) = \mu\left(\{\exp(it) : -\pi \le t \le \theta\}\right).$$

One can express the metric W_1 in terms of these cdfs (Hundrieser et al., 2022).

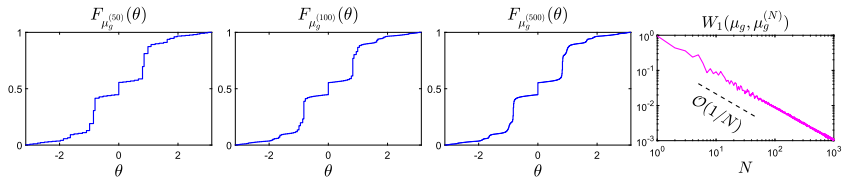

FIGURE 25 First three subplots: The cdfs computed by mpEDMD for various values of N. Far-right: The W_1 metric between the spectral measure computed by mpEDMD, $\mu_g^{(N)}$, and the spectral measure of the Koopman operator, μ_g. The W_1 distance to μ_g is computed by comparing to an approximation with larger N selected large enough to have a negligible effect on the shown errors.

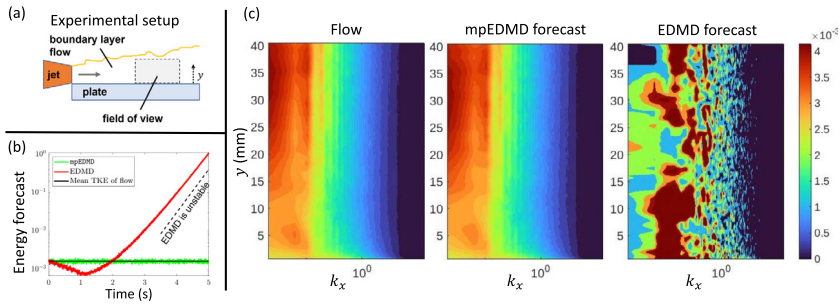

FIGURE 26 (a) Experimental wall-jet boundary layer flow setup with Reynolds number 6.4×10^4. (b) Horizontal averages of the forecasts for turbulent kinetic energy, which show the stability of mpEDMD. (c) Wavenumber spectra measure the energy content of various turbulent structures as a function of their size, thus providing an efficient measure of a flow reconstruction method's performance over various spatial scales. This demonstrates the importance of structure-preserving discretizations (mpEDMD). Reproduced with permission from Colbrook (2023), copyright © 2023 Society for Industrial and Applied Mathematics, all rights reserved.

Fig. 25 displays the cdfs of $\mu_g^{(N)} = \mu_{\mathbf{g}}^{(N,10^4)}$ for various choices of N, illustrating the convergence of spectral measures. Notably, there is a discontinuity in the cdfs at $\theta = 0$, corresponding to the eigenvalue at $\lambda = 1$. Away from this value, the cdfs show pointwise convergence. The error measured in the Wasserstein 1 metric is depicted on the far right of the plot. Consistent with (5.7), this error decreases as $\mathcal{O}(1/N)$. For similar analyses and examples regarding the projection-valued spectral measures, see Colbrook (2023).

Conservation of energy and statistics for turbulent boundary layer flow

We now examine the boundary layer formed by a thin jet injecting air onto a smooth, flat wall, as depicted in panel (a) of Fig. 26. The experiments are conducted in the wind tunnel at Virginia Tech (Szőke et al., 2021). We utilize a two-component, time-resolved particle image velocimetry system to capture 10^3 snapshots of the two-dimensional velocity field of the wall-jet flow. These snapshots are taken over a spatial grid and a period of 1 second. The jet velocity is set at $U_{\text{jet}} = 50$ m/s, corresponding to a jet Reynolds number of 6.4×10^4.

The field-of-view spans approximately 75 mm by 40 mm, with a spatial resolution of approximately $\Delta x = \Delta y \approx 0.24$ mm. This setup leads to a dimension $d = 102{,}300$ in (1.1). An SVD of the data matrix is employed to create a dictionary with $N = 1000$. The flow exhibits zero pressure gradient turbulent boundary layer characteristics within the region between the wall and the velocity profile peak at approximately $y = 15.5$ mm. Above this region, the flow is dominated by a two-dimensional shear layer characterized by large, energetic flow structures. This scenario presents a significant challenge for conventional DMD approaches due to the multiple turbulent scales within the boundary layer.

We investigate the conservation of energy and flow statistics using the KMD for future state predictions. We consider the velocity profiles predicted by mpEDMD and EDMD over 5 seconds, five times the observation window, starting from an initial state $\mathbf{x}_0$ randomly selected from the trajectory data. The results are averaged over 100 such random initializations. Panel (b) of Fig. 26 shows the turbulent kinetic energy (TKE) of the predictions, averaged in the homogeneous horizontal direction and normalized by U_{jet}^2. The instability of EDMD is evident. In contrast, mpEDMD preserves the inner product associated with the TKE.

To examine the statistics of the predictions, panel (c) of Fig. 26 presents the wavenumber spectrum. This spectrum is computed by applying the Fourier transform to the spatial autocorrelations of the predictions in the horizontal direction, as detailed in Glegg and Devenport (2017, Chapter 8). The wavenumber spectrum provides insights into the energy content of various turbulent structures based on their size. It also serves as an efficient measure of a flow-reconstruction method's performance across different spatial scales. The wavenumber spectrum derived from mpEDMD shows remarkable alignment with the actual flow, demonstrating its efficacy. In contrast, EDMD completely fails to capture the correct turbulent statistics.

5.3 Compactification methods for continuous-time systems

For continuous-time invertible measure-preserving systems, the Koopman generator $\mathcal{L}$, as defined in (2.8), is skew-adjoint. A sophisticated suite of methods exists aimed at approximating such generators through compactification. Working in continuous time presents at least two advantages. First, the generator $\mathcal{L}$ is skew-adjoint, while the Koopman operators $\mathcal{K}_{\Delta t}$ are unitary. Developing projection methods that preserve skew-adjointness is generally much more straightforward than preserving unitarity (although we have seen that mpEDMD leads to an appropriate unitary discretization). Second, by computing the spectral properties of the generator $\mathcal{L}$, we are no longer constrained by the need to select a specific discrete time step. We gain comprehensive spectral information for the entire family of Koopman operators $\{\mathcal{K}_{\Delta t} : \Delta t > 0\}$.

Das et al. (2021) developed an approach based on a one-parameter family of reproducing kernels, $\{p_\tau : \tau > 0\}$, satisfying mild regularity assumptions.

This method utilizes corresponding integral operators to perturb the Koopman generator $\mathcal{L}$ to a compact operator on the corresponding RKHS, $\mathcal{H}_\tau$. Assuming ergodic flow, Das et al. (2021) constructed a one-parameter family of skew-adjoint compact operators, $W_\tau : \mathcal{H}_\tau \to \mathcal{H}_\tau$, where $W_\tau = P_\tau \mathcal{L} P_\tau^*$ and $P_\tau : L^2(\Omega, \omega) \to \mathcal{H}_\tau$ is the integral operator defined by

$$[P_\tau g](\mathbf{x}') = \int_\Omega p_\tau(\mathbf{x}', \mathbf{x}) g(\mathbf{x}) \, \mathrm{d}\omega(\mathbf{x}).$$

The operators W_τ are unitarily equivalent to $\mathcal{L}_\tau = G_\tau^{1/2} \mathcal{L} G_\tau^{1/2}$ acting on $L^2(\Omega, \omega)$, with $G_\tau = P_\tau^* P_\tau$. The operators $\mathcal{L}_\tau$ are compact, skew-adjoint, and converge in the strong resolvent sense to the generator $\mathcal{L}$ as $\tau \to 0$. Since each $\mathcal{L}_\tau$ is compact, its spectrum can be computed by projection onto finite-dimensional subspaces without spectral pollution and without missing parts of the spectrum in the limit of infinite discretization size (Ben-Artzi et al., 2020). This procedure yields approximate Koopman eigenvalues and eigenfunctions, which have been demonstrated to lie within the ϵ-pseudospectrum of the Koopman operator, with the value of ϵ dependent on an RKHS-induced Dirichlet energy functional. In particular, approximate eigenfunctions with small Dirichlet energy as $\tau \to 0$ are approximately cyclical, slowly decorrelating observables under potentially mixing dynamics. It is important to note that two limits are implicitly involved here: the first concerns the parameter that controls the projection size used to approximate spectra of $\mathcal{L}_\tau$, and the second is as τ approaches zero. Another potential limitation of this method is its use of finite-difference schemes on time-ordered data. Although the error from these approximations can be controlled in the limit of a vanishing sampling interval via RKHS regularity, finite differencing generally reduces noise robustness.

Another approach involves the *resolvent* of the generator, $(\mathcal{L} - zI)^{-1}$, where $z \in \mathbb{C} \backslash i\mathbb{R}$. By taking the Laplace transform of the Koopman semigroup, we can observe that (Susuki et al., 2021)

$$(\mathcal{L} - zI)^{-1} = -\int_0^\infty e^{-zt} \mathcal{K}_{\Delta t} \, \mathrm{d}t, \quad \mathrm{Re}(z) > 0. \tag{5.8}$$

Valva and Giannakis (2023) combine the compactification approach from Das et al. (2021) and the integral representation of the resolvent used in Susuki et al. (2021) to construct a compact operator that acts as the resolvent of a skew-adjoint operator. The result is a family of skew-adjoint unbounded operators with compact resolvents, whose spectral measures converge weakly to those of $\mathcal{L}$. This method not only preserves skew-adjointness but also eliminates the need for finite-difference approximations of the generator by using a quadrature approximation for the integral in (5.8). It offers a flexible framework that allows for the control of approximation accuracy by varying z in relation to the sampling interval and the timespan of the training data. Additionally, the finite-rank

operators are expressed in a well-conditioned basis of kernel eigenvectors, computed from trajectory data with convergence assurances in the large-data limit (Das et al., 2021). These basis vectors are particularly well-suited to invariant measures supported on sets with complex geometries (e.g., fractal attractors) that are embedded in high-dimensional ambient spaces.

5.3.1 Example

The Rössler system (Rössler, 1976) consists of the following three coupled ordinary differential equations:

$$\dot{x} = -y - z, \quad \dot{y} = x + 0.1y, \quad \dot{z} = 0.1 + z(x - 14).$$

We consider the dynamics of $\mathbf{x} = (x, y, z)$ on the Rössler attractor. The Rössler system is often viewed as a simplified analog of the Lorenz (63) system. However, despite the simplicity of its governing equations, it exhibits complex dynamical characteristics. Theorems on the existence and measure-theoretic mixing properties of the Rössler system analogous to those for the Lorenz system have not been established. Nevertheless, the system has been studied extensively through analytical and numerical techniques, supporting the hypothesis that the Rössler system is mixing (Peifer et al., 2005). Assuming this, it follows that 0 is the only eigenvalue of $\mathcal{L}$, corresponding to a constant eigenfunction and is simple. The integral in (5.8) is approximated by truncating the domain of integration (taking advantage of the exponential decay in the integrand) and Simpson's quadrature rule. Full algorithmic details of the method are given in Valva and Giannakis (2023, Algorithm 1). Data is collected along a single trajectory of length 64,000 with time-step $\Delta t = 0.04$. We use MATLAB's `ode45` command to collect the data after an initial burn-in time to ensure that the initial point is (approximately) on the attractor. The dictionary consists of 2001 data-driven kernel eigenfunctions, and the smoothing parameter is set as $\tau = 2 \times 10^{-6}$.

Fig. 27 shows three approximate eigenfunctions along with the corresponding values of σ, so that $i\sigma$ lies in the spectrum of $\mathcal{L}$. To the figure's right, we illustrate the trajectory of these approximate eigenfunctions. The chaotic behavior of the Rössler system predominantly occurs in the $(r = \sqrt{x^2 + y^2}, z)$ coordinates, while the evolution of the azimuthal angle θ in the $z = 0$ plane proceeds at a near-constant angular frequency, approximately equal to one in natural time units. We suspect this distinction contributes to the challenge of capturing the approximate eigenfunctions in Fig. 27. The first two approximate eigenfunctions are highly coherent, predominantly functions of the azimuthal phase angle, and evolve near-periodically over several Lyapunov times. The third approximate eigenfunction (bottom row) exhibits manifest radial variability in state space and amplitude-modulated time series. Additionally, resolving the radial direction may be more challenging in a data-driven basis, as most variability in the input data occurs in the azimuthal or vertical directions.

We have also presented the Dirichlet energies, indicative of the function's variability or roughness. The approximate eigenfunction corresponding to a

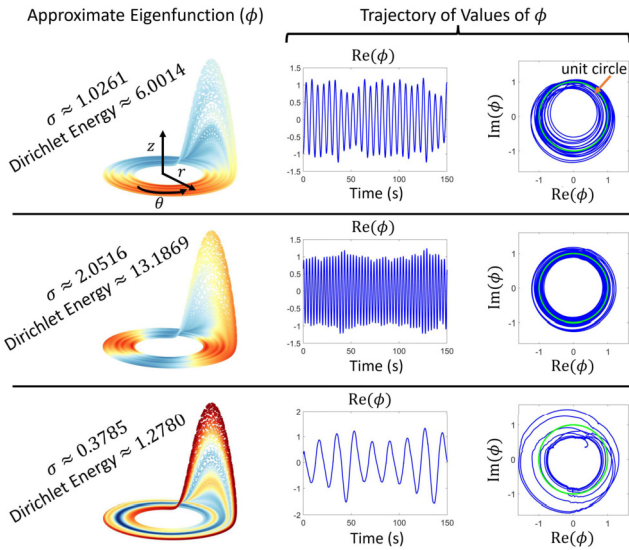

FIGURE 27 Approximate eigenfunctions of the Rössler system computed using the method of Valva and Giannakis (2023). On the right, we show the trajectories of these functions. A trajectory on the unit circle corresponds to coherent periodic behavior. The bottom row displays an approximate eigenfunction with radial variability and larger deviations from the unit circle.

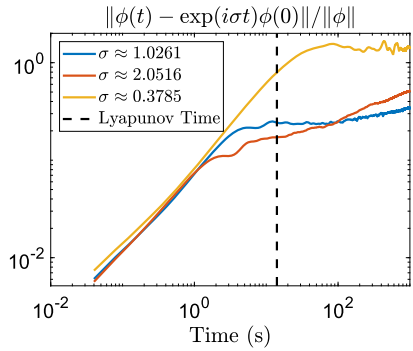

FIGURE 28 Relative residuals of the approximate eigenfunctions in Fig. 27 plotted as a function of time. The residuals increase steadily up to the characteristic Lyapunov timescale.

frequency of $\sigma = 0.3785$ demonstrates relatively low variability compared to the others. Furthermore, the increase in energy from the eigenfunction with frequency 1.0261 to its harmonic with frequency 2.0516 also mirrors this variability increase. Denoting each approximate eigenfunction's trajectory by $\phi(t)$, Fig. 28 displays the relative residual $\|\phi(t) - \exp(i\sigma t)\phi(0)\|/\|\phi\|$. These residuals steadily increase up to the characteristic Lyapunov time of the system and exhibit a larger residual for the third approximate eigenfunction. To summarize, as an a posteriori metric, Dirichlet energy is generally independent from pseu-

dospectral residuals. The information it provides can be useful in supervised learning tasks, e.g., when performing out-of-sample evaluation of the eigenfunctions in prediction problems. In practice, eigenfunctions with small Dirichlet energy also tend to have small pseudospectral residuals, though the precise ordering obtained from the two approaches may differ.

5.4 Further methods

Huang and Vaidya (2018) were among the first to enforce structure in DMD by introducing *Naturally Structured DMD*. This variant ensures positivity and offers the added option of the Markov property. In another early paper, Salova et al. (2019) investigated dynamical systems with symmetries characterized by a finite group. Utilizing representation theory, the authors demonstrated that the Koopman operator and its EDMD approximations can be block diagonalized using a symmetry basis. This basis respects the isotypic component structure related to the underlying symmetry group and the actions of its elements, providing insights into suitable dictionaries. However, the data must align with the system's symmetries to achieve an exact block-diagonal approximation matrix. In an earlier work, Sharma et al. (2016) connected the spatiotemporal symmetries of the Navier–Stokes equation with its spatial and temporal Koopman operators. Kaiser et al. (2018a) presented a method to detect conservation laws using Koopman operator approximations, which can subsequently be employed to control Hamiltonian systems.

We saw that mpEDMD and compactification methods are well-suited to measure-preserving systems. Govindarajan et al. (2019, 2021) provide another approach that is similar to the Ulam approximation of the Perron–Frobenius operator (Ulam, 1960; Li, 1976). They proposed *periodic approximations* for Koopman operators under conditions where Ω is compact, ω is absolutely continuous with respect to the Lebesgue measure, and the system is both measure-preserving and invertible (Govindarajan et al., 2019). This framework was developed into an algorithm for systems on tori and extended to continuous-time systems in Govindarajan et al. (2021). The technique hinges on constructing a periodic approximation of the dynamics via a state-space partition, thus enabling the approximation of the Koopman operator's action through a permutation. The concept of periodic approximations has roots in the works of Halmos (1944) and Lax (1971). This method yields measures that converge weakly to the spectral measures of the Koopman operator. Furthermore, periodic approximations are positive operators and uphold the multiplicative structure of the Koopman operator, i.e., $\mathcal{K}(fg) = (\mathcal{K}f)(\mathcal{K}g)$. A significant unresolved question is how these results can be generalized to handle systems that are not necessarily invariant with respect to a Lebesgue absolutely continuous measure, such as those defined on intricate domains like chaotic attractors, and how to develop efficient schemes in high dimension. Boullé and Colbrook (2024a) introduced *Multiplicative DMD* which enforces the multiplicative structure of Koopman

operators, improving the selection of observables and enabling efficient matrix approximation to better reflect spectral properties, with demonstrated robustness to noise across several examples.

DMD fails in translational problems, such as wave-like phenomena, moving interfaces, and moving shocks (Kutz et al., 2016a). This limitation can be attributed to the dominant advection behavior propagating through the entire high-dimensional domain. This propagation makes establishing a global spatiotemporal basis challenging within a low-dimensional subspace. Drawing inspiration from Lagrangian POD (Mojgani and Balajewicz, 2017), Lu and Tartakovsky (2020a) introduced *Lagrangian DMD* that constructs a reduced-order model within the Lagrangian framework. Temporally evolving characteristic lines are selected as a central observable, and a low-dimensional structure in the Lagrangian framework is identified. *Port-Hamiltonian DMD* (Morandin et al., 2023) adapts the DMD within the port-Hamiltonian systems framework to ensure the system satisfies a dissipation inequality. *Symmetric DMD* (Cohen et al., 2020) mandates the dynamics matrix to be symmetric. *Constrained DMD* (Krake et al., 2022) ensures the presence of specific frequencies by incorporating constraints into DMD. *Hermitian DMD* (Baddoo et al., 2023; Drmač, 2022b) provides a self-adjoint approximation that converges for self-adjoint Koopman operators (Boullé and Colbrook, 2024b).

On the transfer operator side, Mehta et al. (2006) address symmetries of the Perron–Frobenius operator in relation to the admissible symmetry properties of attractors. *Constrained Ulam DMD* (Goswami et al., 2018) uses a minimization problem with constraints that guarantee a positive operator with a row sum equal to one. Beyond DMD, Mardt et al. (2020) developed deep learning Markov and Koopman models with physical constraints. Pan and Duraisamy (2020b) learn continuous-time Koopman operators with deep neural networks and enforce stability by ensuring that eigenvalues have nonpositive real parts. Hirsh et al. (2021) presented a theoretical connection between time-delay embedding models and the Frenet–Serret frame (intrinsic coordinates formed by applying the Gram–Schmidt procedure to the derivatives of the trajectory) from differential geometry. This was used to develop structured HAVOK models.

6 Further topics and open problems

We conclude this review of DMD with further topics that are connected with DMD and Koopman operators, followed by outlining some future challenges in the field.

6.1 Transfer operators

Perron–Frobenius operators, also known as Ruelle (Ruelle, 1968) or transfer operators, act on measures through pullbacks. When considering appropriately chosen spaces of observables and measures, the Koopman and Perron–

Frobenius operators emerge as dual pairs, thereby offering equivalent information. In the context of ergodic dynamical systems, natural spaces of observables are typically L^2 spaces of complex-valued scalar functions associated with invariant probability measures, while natural spaces of measures involve complex measures with L^2 densities. The operational distinction between Koopman and Perron–Frobenius operators has led to the development of two distinct families of approximation techniques. However, recent works, such as Klus et al. (2016), have started to bridge this gap.

For data-driven techniques employing Perron–Frobenius operators, see the seminal work of Dellnitz and Junge (1999). A prevalent approach in these methods involves approximating the Perron–Frobenius operator's spectrum using Ulam's method (Ulam, 1960; Li, 1976). This method involves partitioning the state space into a finite set of disjoint subsets. The transition probabilities between these subsets are then estimated by counting transitions observed in extensive simulations or experimental data. The derived transition probability matrix essentially serves as a Galerkin projection of a smoothed compact transfer operator, slightly perturbed by noise. The matrix's eigenvectors, corresponding to eigenvalues at or near the unit circle, are instrumental in identifying coherent sets. Subsequent research by Dellnitz, Froyland, and colleagues (Dellnitz et al., 2000; Froyland and Dellnitz, 2003; Froyland, 2007, 2008; Froyland et al., 2014a) focused on specific system classes with quasicompact Perron–Frobenius operators. Their work rigorously demonstrated Ulam's method's efficacy in accurately approximating isolated Perron–Frobenius eigenvalues and their associated eigenfunctions.

The Perron–Frobenius operator has been used to analyze the global behavior of dynamical systems across various fields. Its applications span molecular dynamics (Schütte and Sarich, 2013; Schütte et al., 2023), fluid dynamics (Froyland et al., 2014b, 2016), meteorology and atmospheric sciences (Tantet et al., 2015, 2018; Froyland et al., 2021), as well as engineering (Vaidya et al., 2010; Ober-Blöbaum and Padberg-Gehle, 2015). Various toolboxes, such as GAIO (Dellnitz et al., 2001), can compute almost invariant sets or metastable states. These toolboxes utilize adaptive box discretizations of the state space to approximate the system's behavior efficiently. However, it is important to note that this approach is generally more suited to low-dimensional problems.

6.2 Continuous spectra and spectral measures

In Section 2.1.2, we saw how Koopman operators associated with measure-preserving systems have spectral measures that provide a KMD. In the course of this review, we have met several methods that converge to spectral measures: mpEDMD in Section 5.2 (see Fig. 25 for an example), methods based on compactification in Section 5.3, and partitioning of the state space to obtain periodic approximations in Section 5.4. There are other methods that are not based on the eigenvalues of a finite matrix. Korda et al. (2020) approximate the moments of

spectral measures of ergodic systems using (4.8), and then use the Christoffel–Darboux kernel to analyze the atomic and absolutely continuous parts of the spectrum. They also compute the spectral projection on a given segment of the unit circle. See also Arbabi and Mezić (2017b), who use harmonic averaging and Welch's method (Welch, 1967) to compute the discrete and continuous spectrum of the Koopman operator for post-transient flows. Using the resolvent operator and ResDMD, Colbrook and Townsend (2023) compute smoothed approximations of spectral measures associated with general measure-preserving dynamical systems. They prove explicit high-order convergence theorems for the computation of spectral measures in various senses, including computing the density of the continuous spectrum, spectral projections of subsets of the unit circle, and the discrete spectrum. These smoothing techniques can also be used for self-adjoint operators (Colbrook et al., 2021). *Rigged DMD* (Colbrook et al., 2024) takes this even further and computes generalized eigenfunctions of Koopman operators.

However, we end this discussion with the following warning to the reader about recovering atomic parts of spectral measures that should be kept in mind for all of the above methods. As soon as the spectral measure μ_g has atoms (i.e., $\mathcal{K}$ has eigenvalues, and g is not orthogonal to all the eigenspaces), the map $\lambda \mapsto \mu_g(\{\lambda\})$ is discontinuous. One can prove that, in general, separating the point spectrum from the rest of the spectrum, either in terms of spectral measures or spectral sets, is impossible for *any* algorithm. This holds even for simple classes of operators (Colbrook, 2021), unless we know a priori that the spectrum is discrete in a region of interest (Colbrook et al., 2021, Section 7.3). An excellent example and discussion of this point is provided on the second page of Govindarajan et al. (2019). This is one reason the above methods can only compute spectral measures in a weak or setwise sense. It also helps explain why many of these techniques involve some form of smoothing.

6.3 Stochastic dynamical systems

Stochastic dynamical systems are widely used to model and study systems that evolve under the influence of both deterministic and random effects. It is common to replace (1.1) with a discrete-time Markov process

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n, \tau_n), \qquad n = 0, 1, 2, \ldots, \tag{6.1}$$

where $\{\tau_n\} \in \Omega_s$ are independent and identically distributed random variables with distribution ρ supported on Ω_s, and $\mathbf{F} : \Omega \times \Omega_s \to \Omega$ is a function. The stochastic Koopman operator (also called the Kolmogorov operator) is the expectation:

$$[\mathcal{K}g](\mathbf{x}) = \int_{\Omega_s} g(\mathbf{F}(\mathbf{x}, \tau)) \, \mathrm{d}\rho(\tau). \tag{6.2}$$

In contrast to the deterministic case, stochastic Koopman operators typically have discrete spectra due to diffusion. A primary focus has been the challenge of noisy observables in EDMD-type methods (Takeishi et al., 2017c; Wanner and Mezić, 2022), and debiasing DMD (Hemati et al., 2017; Dawson et al., 2016; Takeishi et al., 2017c). Črnjarić-Žic et al. (2020) developed a stochastic Hankel-DMD algorithm for numerical approximations of the stochastic Koopman operator. Klus et al. (2020b) used gEDMD to derive models for SDEs with applications in control. Sinha et al. (2020) provided an explicit optimization-based approximation of stochastic Koopman operators. Wu and Noé (2020) developed a variational approach for Markov processes that finds optimal feature mappings and optimal Markovian models of the dynamics from the top singular components of the Koopman operator.

The definition in (6.2) involves an expectation. Colbrook et al. (2023b) demonstrated the benefits and necessity of going beyond expectations of trajectories. They incorporated the concept of variance into the Koopman framework, establishing its relationship with batched Koopman operators. This led to an extension of ResDMD, resulting in convergence to the spectral properties of stochastic Koopman operators, a Koopman analog of a variance-bias decomposition, and the concept of variance-pseudospectra as a measure of statistical coherency.

6.4 Some open problems

There has been substantial interest in Koopmanism over the last decade, and we expect this interest only to grow. This is an exciting time to be working in this field, which is at the crossroads of dynamical systems theory, data analysis, spectral theory, and computational analysis. We end with some open problems that the author believes will lead to important breakthroughs in the coming years:

- **Banach (and other function) spaces:** We have focused on Koopman operators defined on the Hilbert space $L^2(\Omega, \omega)$. In some cases, it is more appropriate to consider function spaces that are Banach spaces (Mohr and Mezić, 2014; Mezić, 2020). Computational tools for infinite-dimensional spectral problems on Banach spaces are less developed than those for Hilbert spaces. An exception is the transfer operator community, which has developed methods for quasicompact Perron–Frobenius operators (see Section 6.1). Another challenge is the development of the theory of the KMD in the absence of spectral theorems. These issues are expected to be particularly significant in applying the Koopman framework and DMD to transient and off-attractor dynamics.

  For some chaotic systems under appropriate conditions, the eigenvalues in the large subspace limit of EDMD correspond to the eigenvalues of transfer operators in suitable function spaces, as discussed in Slipantschuk et al. (2020); Wormell (2023); Bandtlow et al. (2023). Reconciling these general-

ized eigenfunctions, the L^2 flavor of DMD methods, and appropriate Banach spaces is a key open problem. These questions could be crucial for understanding and improving the effectiveness of EDMD, including guidance on the choice of dictionary.

- **Choice of dictionary:** One of the most significant open problems in EDMD is the selection of observables or dictionaries. At present, this process can be considered more of an art than a science. While well-conditioned bases can be constructed for some systems, as outlined in Section 5.3, this task often presents a substantial challenge. This is also true for methods based on delay-embedding, where the choice of delay itself is a classical problem with many available heuristics.
- **Foundations:** All of the convergence results for DMD and Koopman operators rely on algorithms that depend on several parameters, with successive limits of these parameters taken to achieve convergence. This is not accidental and occurs generically in infinite-dimensional spectral computations (Ben-Artzi et al., 2020). It is possible to classify the difficulty of computational problems, including data-driven ones. To date, the Koopman community has only provided upper bounds, i.e., algorithms that converge for specific classes of problems. A significant open problem is the development of lower bounds, i.e., universal impossibility results that indicate an intrinsic difficulty in a problem that cannot be overcome by any algorithm. Such results are beginning to emerge in the world of deep learning, particularly regarding the existence vs. trainability of neural networks (Colbrook et al., 2022). We expect them to be equally fruitful in the Koopman context.
  Lower bounds are essential for several reasons. First, they prevent futile searches for nonexistent algorithms. Second, they often elucidate why certain algorithms cannot exist. When combined with upper bounds, this knowledge can lead to natural assumptions about the dynamical systems or the data required to achieve our computational goals.
- **Further structure-preserving methods:** We have already outlined several open problems stimulated by piDMD in Section 5.1.3. Understanding the relationships between structures or symmetries in dynamical systems and their manifestation in the Koopman spectrum lies at the forefront of current knowledge. A crucial challenge is extending constraints applicable to DMD with linear observables to EDMD with nonlinear observables, which will be instrumental in applying structure-preserving methods to nonlinear systems effectively. Another related open problem is establishing the convergence of constrained DMD methods to the spectral properties of Koopman operators and the convergence of KMDs.
- **Verified control:** An exciting development area in modern Koopman theory is its use for controlling nonlinear systems. In Section 3.4.3, we discussed how the practice of Koopman control currently surpasses the theoretical understanding. Only a limited number of systems with a known Koopman-invariant subspace and verifiable eigenfunctions exist. Thus, devel-

oping methods to validate Koopman models for control purposes is a crucial problem. Successfully addressing this issue will likely lead to further insights and enhancements in the practice of Koopman control.

Acknowledgments

I am grateful to numerous friends for their support and stimulating discussions during the completion of this work. The following is in no particular order. Thank you Steven Brunton and Nathan Kutz, for your thoughts and perspectives on an early version of this manuscript, in particular, Table 1. Thank you Igor Mezić, for our discussions regarding spectral theory and computation. Thank you Benjamin Erichson, for running the numerical experiments behind Figs. 13 and 14 and discussing rDMD with me. Thank you Claire Valva and Dimitrios Giannakis, for discussing compactification methods with me, and thank you Claire for running the numerical experiments behind Fig. 27. Thank you Zlatko Drmač, for our discussions regarding the numerical linear algebra aspects of DMD. Thank you Julia Slipantschuk and Caroline Wormell for our discussions regarding numerical methods for transfer operators. Thank you Zlatko Drmač, Dimitrios Giannakis, Igor Mezić, and Samuel Otto, for your insightful feedbacks on an initial draft.

References

Abraham, I., de la Torre, G., Murphey, T., 2017. Model-based control using Koopman operators. In: Robotics: Science and Systems XIII. RSS2017, Robotics: Science and Systems Foundation. MIT Press Journals.

Abraham, I., Murphey, T.D., 2019. Active learning of dynamics for data-driven control using Koopman operators. IEEE Transactions on Robotics 35 (5), 1071–1083.

Adcock, B., Hansen, A.C., 2021. Compressive Imaging: Structure, Sampling, Learning. Cambridge University Press.

Ahmed, S.E., Dabaghian, P.H., San, O., Bistrian, D.A., Navon, I.M., 2022. Dynamic mode decomposition with core sketch. Physics of Fluids 34 (6).

Alexander, R., Giannakis, D., 2020. Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. Physica D. Nonlinear Phenomena 409, 132520.

Alford-Lago, D.J., Curtis, C.W., Ihler, A.T., Issan, O., 2022. Deep learning enhanced dynamic mode decomposition. Chaos 32 (3).

Alla, A., Kutz, J.N., 2017. Nonlinear model order reduction via dynamic mode decomposition. SIAM Journal on Scientific Computing 39 (5), B778–B796.

Aloisio, M., Carvalho, S.L., de Oliveira, C.R., Souza, E., 2022. On spectral measures and convergence rates in von Neumann's ergodic theorem. arXiv preprint. arXiv:2209.05290.

Andreuzzi, F., Demo, N., Rozza, G., 2023. A dynamic mode decomposition extension for the forecasting of parametric dynamical systems. SIAM Journal on Applied Dynamical Systems 22 (3), 2432–2458.

Arbabi, H., Korda, M., Mezić, I., 2018. A data-driven Koopman model predictive control framework for nonlinear partial differential equations. In: 2018 IEEE Conference on Decision and Control (CDC). IEEE, pp. 6409–6414.

Arbabi, H., Mezić, I., 2017a. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. SIAM Journal on Applied Dynamical Systems 16 (4), 2096–2126.

Arbabi, H., Mezić, I., 2017b. Study of dynamics in post-transient flows using Koopman mode decomposition. Physical Review Fluids 2 (12), 124402.

Arnold, V.I., 1989. Mathematical Methods of Classical Mechanics. Springer New York.

Arun, K.S., 1992. A unitarily constrained total least squares problem in signal processing. SIAM Journal on Matrix Analysis and Applications 13 (3), 729–745.

Askham, T., Kutz, J.N., 2018. Variable projection methods for an optimized dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems 17 (1), 380–416.

Azencot, O., Erichson, N.B., Lin, V., Mahoney, M., 2020. Forecasting sequential data using consistent Koopman autoencoders. In: International Conference on Machine Learning. PMLR, pp. 475–485.

Azencot, O., Yin, W., Bertozzi, A., 2019. Consistent dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems 18 (3), 1565–1585.

Baddoo, P.J., Herrmann, B., McKeon, B.J., Brunton, S.L., 2022. Kernel learning for robust dynamic mode decomposition: linear and nonlinear disambiguation optimization. Proceedings of the Royal Society A. Mathematical, Physical and Engineering Sciences 478 (2260), 20210830.

Baddoo, P.J., Herrmann, B., McKeon, B.J., Nathan Kutz, J., Brunton, S.L., 2023. Physics-informed dynamic mode decomposition. Proceedings of the Royal Society A. Mathematical, Physical and Engineering Sciences 479 (2271), 20220576.

Bagheri, S., 2013. Koopman-mode decomposition of the cylinder wake. Journal of Fluid Mechanics 726, 596–623.

Bagheri, S., 2014. Effects of weak noise on oscillating flows: linking quality factor, Floquet modes, and Koopman spectrum. Physics of Fluids 26 (9).

Bai, Z., Kaiser, E., Proctor, J.L., Kutz, J.N., Brunton, S.L., 2020. Dynamic mode decomposition for compressive system identification. AIAA Journal 58 (2), 561–574.

Bandtlow, O.F., Just, W., Slipantschuk, J., 2023. EDMD for expanding circle maps and their complex perturbations. arXiv preprint. arXiv:2308.01467.

Beer, G., 1993. Topologies on Closed and Closed Convex Sets, vol. 268. Springer Netherlands.

Ben-Artzi, J., Colbrook, M.J., Hansen, A.C., Nevanlinna, O., Seidel, M., 2020. Computing spectra - on the solvability complexity index hierarchy and towers of algorithms. arXiv preprint. arXiv:1508.03280.

Berkooz, G., Holmes, P., Lumley, J.L., 1993. The proper orthogonal decomposition in the analysis of turbulent flows. Annual Review of Fluid Mechanics 25 (1), 539–575.

Berry, T., Giannakis, D., Harlim, J., 2015. Nonparametric forecasting of low-dimensional dynamical systems. Physical Review E 91 (3), 032915.

Birkhoff, G.D., 1931. Proof of the ergodic theorem. Proceedings of the National Academy of Sciences 17 (12), 656–660.

Birkhoff, G.D., Koopman, B.O., 1932. Recent contributions to the ergodic theory. Proceedings of the National Academy of Sciences 18 (3), 279–282.

Bistrian, D.A., Navon, I.M., 2017. Randomized dynamic mode decomposition for nonintrusive reduced order modelling. International Journal for Numerical Methods in Engineering 112 (1), 3–25.

Böttcher, A., Silbermann, B., 1983. The finite section method for Toeplitz operators on the quarterplane with piecewise continuous symbols. Mathematische Nachrichten 110 (1), 279–291.

Boullé, N., Colbrook, M.J., 2024a. Multiplicative dynamic mode decomposition. arXiv preprint. arXiv:2405.05334.

Boullé, N., Colbrook, M.J., 2024b. On the convergence of hermitian dynamic mode decomposition. arXiv preprint. arXiv:2401.03192.

Boumal, N., Mishra, B., Absil, P.-A., Sepulchre, R., 2014. Manopt, a MATLAB toolbox for optimization on manifolds. Journal of Machine Learning Research 15 (1), 1455–1459.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and Regression Trees. Routledge.

Brockett, R.W., 1976. Volterra series and geometric control theory. Automatica 12 (2), 167–176.

Broomhead, D.S., Jones, R., 1989. Time-series analysis. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 423 (1864), 103–121.

Bruder, D., Gillespie, B., David Remy, C., Vasudevan, R., 2019. Modeling and control of soft robots using the Koopman operator and model predictive control. In: Robotics: Science and Systems XV. RSS2019, Robotics: Science and Systems Foundation.

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

Brunton, S.L., Brunton, B.W., Proctor, J.L., Kaiser, E., Kutz, J.N., 2017. Chaos as an intermittently forced linear system. Nature Communications 8 (1), 1–9.

Brunton, S.L., Brunton, B.W., Proctor, J.L., Kutz, J.N., 2016a. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. PLoS ONE 11 (2), e0150171.

Brunton, S.L., Budišić, M., Kaiser, E., Kutz, J.N., 2022. Modern Koopman theory for dynamical systems. SIAM Review 64 (2), 229–340.

Brunton, S.L., Kutz, J.N., 2022. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press.

Brunton, S.L., Proctor, J.L., Tu, J.H., Kutz, J.N., 2016b. Compressed sensing and dynamic mode decomposition. Journal of Computational Dynamics 2 (2), 165–191.

Budišić, M., Mohr, R., Mezić, I., 2012. Applied Koopmanism. Chaos 22 (4), 047510.

Burov, D., Giannakis, D., Manohar, K., Stuart, A., 2021. Kernel analog forecasting: multiscale test problems. Multiscale Modeling & Simulation 19 (2), 1011–1040.

Caflisch, R.E., 1998. Monte Carlo and quasi-Monte Carlo methods. Acta Numerica 7, 1–49.

Candes, E.J., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52 (2), 489–509.

Candes, E.J., Tao, T., 2006. Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Transactions on Information Theory 52 (12), 5406–5425.

Candes, E.J., Wakin, M.B., 2008. An introduction to compressive sampling. IEEE Signal Processing Magazine 25 (2), 21–30.

Carleman, T., 1932. Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires. Acta Mathematica 59, 63–87.

Celledoni, E., Ehrhardt, M.J., Etmann, C., McLachlan, R.I., Owren, B., Schönlieb, C.-B., Sherry, F., 2021. Structure-preserving deep learning. European Journal of Applied Mathematics 32 (5), 888–936.

Champion, K.P., Brunton, S.L., Kutz, J.N., 2019. Discovery of nonlinear multiscale systems: sampling strategies and embeddings. SIAM Journal on Applied Dynamical Systems 18 (1), 312–333.

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

Cohen, I., Azencot, O., Lifshits, P., Gilboa, G., 2020. Mode decomposition for homogeneous symmetric operators. Preprint arXiv.

Coifman, R.R., Lafon, S., 2006. Diffusion maps. Applied and Computational Harmonic Analysis 21 (1), 5–30.

Colbrook, M., Horning, A., Townsend, A., 2021. Computing spectral measures of self-adjoint operators. SIAM Review 63 (3), 489–524.

Colbrook, M.J., 2020. The Foundations of Infinite-Dimensional Spectral Computations. PhD thesis. University of Cambridge.

Colbrook, M.J., 2021. Computing spectral measures and spectral types. Communications in Mathematical Physics 384 (1), 433–501.

Colbrook, M.J., 2022. On the computation of geometric features of spectra of linear operators on Hilbert spaces. Foundations of Computational Mathematics, 1–82.

Colbrook, M.J., 2023. The mpEDMD algorithm for data-driven computations of measure-preserving dynamical systems. SIAM Journal on Numerical Analysis 61 (3), 1585–1608.

Colbrook, M.J., Antun, V., Hansen, A.C., 2022. The difficulty of computing stable and accurate neural networks: on the barriers of deep learning and Smale's 18th problem. Proceedings of the National Academy of Sciences 119 (12), e2107151119.

Colbrook, M.J., Ayton, L.J., Szőke, M., 2023a. Residual dynamic mode decomposition: robust and verified Koopmanism. Journal of Fluid Mechanics 955, A21.

Colbrook, M.J., Hansen, A.C., 2022. The foundations of spectral computations via the solvability complexity index hierarchy. Journal of the European Mathematical Society 25 (12), 4639–4728.

Colbrook, M.J., Li, Q., Raut, R.V., Townsend, A., 2023b. Beyond expectations: residual dynamic mode decomposition and variance for stochastic dynamical systems. arXiv preprint. arXiv:2308.10697.

Colbrook, M.J., Roman, B., Hansen, A.C., 2019. How to compute spectra with error control. Physical Review Letters 122 (25), 250201.

Colbrook, M.J., Townsend, A., 2023. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. Communications on Pure and Applied Mathematics 77 (1), 221–283.

Colbrook, M.J., 2024. Another look at residual dynamic mode decomposition in the regime of fewer snapshots than dictionary size. arXiv preprint. arXiv:2403.05891.

Colbrook, M.J., Drysdale, C., Horning, A., 2024. Rigged dynamic mode decomposition: data-driven generalized eigenfunction decompositions for Koopman operators. arXiv preprint. arXiv:2405.00782.

Conway, J.B., 2007. A Course in Functional Analysis, vol. 96, 2 edn. Springer New York.

Črnjarić-Žic, N., Maćešić, S., Mezić, I., 2020. Koopman operator spectrum for random dynamical systems. Journal of Nonlinear Science 30 (5), 2007–2056.

Das, S., Giannakis, D., 2019. Delay-coordinate maps and the spectra of Koopman operators. Journal of Statistical Physics 175 (6), 1107–1145.

Das, S., Giannakis, D., 2020. Koopman spectra in reproducing kernel Hilbert spaces. Applied and Computational Harmonic Analysis 49 (2), 573–607.

Das, S., Giannakis, D., Slawinska, J., 2021. Reproducing kernel Hilbert space compactification of unitary evolution groups. Applied and Computational Harmonic Analysis 54, 75–136.

Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics.

Dawson, S.T.M., Hemati, M.S., Williams, M.O., Rowley, C.W., 2016. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. Experiments in Fluids 57 (3), 1–19.

DeGennaro, A.M., Urban, N.M., 2019. Scalable extended dynamic mode decomposition using random kernel approximation. SIAM Journal on Scientific Computing 41 (3), A1482–A1499.

Dellnitz, M., Froyland, G., Junge, O., 2001. The Algorithms Behind GAIO - Set Oriented Numerical Methods for Dynamical Systems. Springer Berlin Heidelberg, pp. 145–174.

Dellnitz, M., Froyland, G., Sertl, S., 2000. On the isolated spectrum of the Perron–Frobenius operator. Nonlinearity 13 (4), 1171–1188.

Dellnitz, M., Junge, O., 1999. On the approximation of complicated dynamical behavior. SIAM Journal on Numerical Analysis 36 (2), 491–515.

Dietrich, F., Thiem, T.N., Kevrekidis, I.G., 2020. On the Koopman operator of algorithms. SIAM Journal on Applied Dynamical Systems 19 (2), 860–885.

Dogra, A.S., Redman, W., 2020. Optimizing neural networks via Koopman operator theory. Advances in Neural Information Processing Systems 33, 2087–2097.

Donoho, D.L., 2006. Compressed sensing. IEEE Transactions on Information Theory 52 (4), 1289–1306.

Drmač, Z., 2020. Dynamic Mode Decomposition - a Numerical Linear Algebra Perspective. Springer International Publishing, pp. 161–194.

Drmač, Z., 2022a. A LAPACK implementation of the dynamic mode decomposition I.

Drmač, Z., 2022b. A LAPACK implementation of the dynamic mode decomposition II. The symmetric/Hermitian DMD (xSYDMD/xHEDMD).

Drmač, Z., Mezić, I., Mohr, R., 2018. Data driven modal decompositions: analysis and enhancements. SIAM Journal on Scientific Computing 40 (4), A2253–A2285.

Drmač, Z., Mezić, I., Mohr, R., 2019. Data driven Koopman spectral analysis in Vandermonde–Cauchy form via the DFT: numerical method and theoretical insights. SIAM Journal on Scientific Computing 41 (5), A3118–A3151.

Drmač, Z., Mezić, I., Mohr, R., 2020. On least squares problems with certain Vandermonde–Khatri–Rao structure with applications to DMD. SIAM Journal on Scientific Computing 42 (5), A3250–A3284.

Drmač, Z., Mezić, I., Mohr, R., 2021. Identification of nonlinear systems using the infinitesimal generator of the Koopman semigroup - a numerical implementation of the Mauroy–Goncalves method. Mathematics 9 (17), 2075.

Dubrovin, B.A., Fomenko, A.T., Novikov, S.P., 1984. Modern Geometry - Methods and Applications Part I. The Geometry of Surfaces, Transformation Groups, and Fields, vol. 104. Springer New York.

Duke, D., Soria, J., Honnery, D., 2012. An error analysis of the dynamic mode decomposition. Experiments in Fluids 52 (2), 529–542.

Eckmann, J.-P., Ruelle, D., 1985. Ergodic theory of chaos and strange attractors. Reviews of Modern Physics 57 (3), 617–656.

Eisner, T., Farkas, B., Haase, M., Nagel, R., 2015. Operator Theoretic Aspects of Ergodic Theory, vol. 272. Springer International Publishing.

Eivazi, H., Guastoni, L., Schlatter, P., Azizpour, H., Vinuesa, R., 2021. Recurrent neural networks and Koopman-based frameworks for temporal predictions in a low-order model of turbulence. International Journal of Heat and Fluid Flow 90, 108816.

Erichson, N.B., Brunton, S.L., Kutz, J.N., 2019a. Compressed dynamic mode decomposition for background modeling. Journal of Real-Time Image Processing 16 (5), 1479–1492.

Erichson, N.B., Donovan, C., 2016. Randomized low-rank dynamic mode decomposition for motion detection. Computer Vision and Image Understanding 146, 40–50.

Erichson, N.B., Mathelin, L., Kutz, J.N., Brunton, S.L., 2019b. Randomized dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems 18 (4), 1867–1891.

Folkestad, C., Pastor, D., Mezić, I., Mohr, R., Fonoberova, M., Burdick, J., 2020. Extended dynamic mode decomposition with learned Koopman eigenfunctions for prediction and control. In: 2020 American Control Conference (ACC). IEEE, pp. 3906–3913.

Foucart, S., Rauhut, H., 2013. A Mathematical Introduction to Compressive Sensing. Springer New York.

Froyland, G., 2007. On Ulam approximation of the isolated spectrum and eigenfunctions of hyperbolic maps. Discrete and Continuous Dynamical Systems 17 (3), 671–689.

Froyland, G., 2008. Unwrapping eigenfunctions to discover the geometry of almost-invariant sets in hyperbolic maps. Physica D. Nonlinear Phenomena 237 (6), 840–853.

Froyland, G., Dellnitz, M., 2003. Detecting and locating near-optimal almost-invariant sets and cycles. SIAM Journal on Scientific Computing 24 (6), 1839–1863.

Froyland, G., Giannakis, D., Lintner, B.R., Pike, M., Slawinska, J., 2021. Spectral analysis of climate dynamics with operator-theoretic approaches. Nature Communications 12 (1).

Froyland, G., González-Tokman, C., Quas, A., 2014a. Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. Journal of Computational Dynamics 1 (2), 249–278.

Froyland, G., González-Tokman, C., Watson, T.M., 2016. Optimal mixing enhancement by local perturbation. SIAM Review 58 (3), 494–513.

Froyland, G., Koltai, P., 2023. Detecting the birth and death of finite-time coherent sets. Communications on Pure and Applied Mathematics 76 (12), 3642–3684.

Froyland, G., Stuart, R.M., van Sebille, E., 2014b. How well-connected is the surface of the global ocean? Chaos 24 (3), 033126.

Fujii, K., Kawahara, Y., 2019. Dynamic mode decomposition in vector-valued reproducing kernel Hilbert spaces for extracting dynamical structure among observables. Neural Networks 117, 94–103.

García Trillos, N., Gerlach, M., Hein, M., Slepčev, D., 2020. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. Foundations of Computational Mathematics 20 (4), 827–887.

Gavish, M., Donoho, D.L., 2014. The optimal hard threshold for singular values is $4/\sqrt{3}$. IEEE Transactions on Information Theory 60 (8), 5040–5053.

Giannakis, D., 2019. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. Applied and Computational Harmonic Analysis 47 (2), 338–396.

Giannakis, D., 2021. Delay-coordinate maps, coherence, and approximate spectra of evolution operators. Research in the Mathematical Sciences 8 (1), 1–33.

Giannakis, D., Das, S., 2020. Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis. Physica D. Nonlinear Phenomena 402, 132211.

Giannakis, D., Henriksen, A., Tropp, J.A., Ward, R., 2023. Learning to forecast dynamical systems from streaming data. SIAM Journal on Applied Dynamical Systems 22 (2), 527–558.

Giannakis, D., Ourmazd, A., Slawinska, J., Zhao, Z., 2019. Spatiotemporal pattern extraction by spectral analysis of vector-valued observables. Journal of Nonlinear Science 29 (5), 2385–2445.

Giannakis, D., Slawinska, J., Zhao, Z., 2015. Spatiotemporal feature extraction with data-driven Koopman operators. In: Feature Extraction: Modern Questions and Challenges. PMLR, pp. 103–115.

Glegg, S., Devenport, W., 2017. Aeroacoustics of Low Mach Number Flows: Fundamentals, Analysis, and Measurement. Academic Press, London, England.

Golub, G.H., Pereyra, V., 1973. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. SIAM Journal on Numerical Analysis 10 (2), 413–432.

Goswami, D., Paley, D.A., 2017. Global bilinearization and controllability of control-affine nonlinear systems: a Koopman spectral approach. In: 2017 IEEE 56th Annual Conference on Decision and Control (CDC). IEEE, pp. 6107–6112.

Goswami, D., Thackray, E., Paley, D.A., 2018. Constrained Ulam dynamic mode decomposition: approximation of the Perron–Frobenius operator for deterministic and stochastic systems. IEEE Control Systems Letters 2 (4), 809–814.

Govindarajan, N., Mohr, R., Chandrasekaran, S., Mezić, I., 2019. On the approximation of Koopman spectra for measure preserving transformations. SIAM Journal on Applied Dynamical Systems 18 (3), 1454–1497.

Govindarajan, N., Mohr, R., Chandrasekaran, S., Mezic, I., 2021. On the approximation of Koopman spectra of measure-preserving flows. SIAM Journal on Applied Dynamical Systems 20 (1), 232–261.

Gower, J.C., Dijksterhuis, G.B., 2004. Procrustes Problems, vol. 30. Oxford University Press.

Greydanus, S., Dzamba, M., Yosinski, J., 2019. Hamiltonian neural networks. Advances in Neural Information Processing Systems 32.

Gu, M., 2015. Subspace iteration randomization and singular value problems. SIAM Journal on Scientific Computing 37 (3), A1139–A1173.

Guéniat, F., Mathelin, L., Pastur, L.R., 2015. A dynamic mode decomposition approach for large and arbitrarily sampled systems. Physics of Fluids 27 (2).

Haggerty, D.A., Banks, M.J., Kamenar, E., Cao, A.B., Curtis, P.C., Mezić, I., Hawkes, E.W., 2023. Control of soft robots with inertial dynamics. Science Robotics 8 (81), eadd6864.

Hairer, E., Lubich, C., Wanner, G., 2010. Geometric Numerical Integration, second edition, first softcover print edn. Springer Series in Computational Mathematics, vol. 31. Springer, Berlin.

Halikias, D., Townsend, A., 2023. Structured matrix recovery from matrix-vector products. Numerical Linear Algebra with Applications 31 (1).

Halko, N., Martinsson, P.G., Tropp, J.A., 2011. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Review 53 (2), 217–288.

Halmos, P.R., 1944. Approximation theories for measure preserving transformations. Transactions of the American Mathematical Society 55 (1), 1–18.

Halmos, P.R., 1963. What does the spectral theorem say? The American Mathematical Monthly 70 (3), 241–247.

Halmos, P.R., 2017. Lectures on Ergodic Theory. Dover Books on Mathematics. Unabridged republication of the work originally published in 1956 edn, Dover Publications, Mineola, New York.

Han, Y., Hao, W., Vaidya, U., 2020. Deep learning of Koopman representation for control. In: 2020 59th IEEE Conference on Decision and Control (CDC). IEEE, pp. 1890–1895.

Hansen, A., 2011. On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators. Journal of the American Mathematical Society 24 (1), 81–124.

Hansen, P.C., Nagy, J.G., O'Leary, D.P., 2006. Deblurring Images: Matrices, Spectra, and Filtering. Society for Industrial and Applied Mathematics.

Haseli, M., Cortés, J., 2023. Modeling nonlinear control systems via Koopman control family: universal forms and subspace invariance proximity. arXiv preprint. arXiv:2307.15368.

Hasnain, A., Boddupalli, N., Balakrishnan, S., Yeung, E., 2020. Steady state programming of controlled nonlinear systems via deep dynamic mode decomposition. In: 2020 American Control Conference (ACC). IEEE, pp. 4245–4251.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer New York.

Hemati, M., Deem, E., Williams, M., Rowley, C.W., Cattafesta, L.N., 2016. Improving separation control with noise-robust variants of dynamic mode decomposition. In: 54th AIAA Aerospace Sciences Meeting. American Institute of Aeronautics and Astronautics, p. 1103.

Hemati, M., Yao, H., 2017. Dynamic mode shaping for fluid flow control: new strategies for transient growth suppression. In: 8th AIAA Theoretical Fluid Mechanics Conference. American Institute of Aeronautics and Astronautics, p. 3160.

Hemati, M.S., Rowley, C.W., Deem, E.A., Cattafesta, L.N., 2017. De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets. Theoretical and Computational Fluid Dynamics 31 (4), 349–368.

Hemati, M.S., Williams, M.O., Rowley, C.W., 2014. Dynamic mode decomposition for large and streaming datasets. Physics of Fluids 26 (11).

Hernández, Q., Badías, A., González, D., Chinesta, F., Cueto, E., 2021. Structure-preserving neural networks. Journal of Computational Physics 426, 109950.

Herrmann, B., Baddoo, P.J., Semaan, R., Brunton, S.L., McKeon, B.J., 2021. Data-driven resolvent analysis. Journal of Fluid Mechanics 918, A10.

Hesthaven, J.S., Pagliantini, C., Rozza, G., 2022. Reduced basis methods for time-dependent problems. Acta Numerica 31, 265–345.

Hey, A.J.G., Tansley, S., Tolle, K.M. (Eds.), 2009. The Fourth Paradigm, vol. 1. Microsoft Research, Redmond, Wash.

Higham, N.J., 1988. The symmetric procrustes problem. BIT 28 (1), 133–143.

Higham, N.J., 2008. Functions of Matrices: Theory and Computation. Society for Industrial and Applied Mathematics.

Hill, T.L., 1986. An Introduction to Statistical Thermodynamics. Courier Corporation.

Hirsh, S.M., Harris, K.D., Kutz, J.N., Brunton, B.W., 2020. Centering data improves the dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems 19 (3), 1920–1955.

Hirsh, S.M., Ichinaga, S.M., Brunton, S.L., Nathan Kutz, J., Brunton, B.W., 2021. Structured time-delay models for dynamical systems with connections to Frenet–Serret frame. Proceedings of the Royal Society A. Mathematical, Physical and Engineering Sciences 477 (2254), 20210097.

Huang, B., Ma, X., Vaidya, U., 2018. Feedback stabilization using Koopman operator. In: 2018 IEEE Conference on Decision and Control (CDC). IEEE, pp. 6434–6439.

Huang, B., Ma, X., Vaidya, U., 2020. Data-Driven Nonlinear Stabilization Using Koopman Operator. Springer International Publishing, pp. 313–334.

Huang, B., Vaidya, U., 2018. Data-driven approximation of transfer operators: naturally structured dynamic mode decomposition. In: 2018 Annual American Control Conference (ACC). IEEE, pp. 5659–5664.

Huhn, Q.A., Tano, M.E., Ragusa, J.C., Choi, Y., 2023. Parametric dynamic mode decomposition for reduced order modeling. Journal of Computational Physics 475, 111852.

Hundrieser, S., Klatt, M., Munk, A., 2022. The Statistics of Circular Optimal Transport. Springer Nature Singapore, pp. 57–82.

Ikeda, M., Ishikawa, I., Schlosser, C., 2022. Koopman and Perron–Frobenius operators on reproducing kernel Banach spaces. Chaos 32 (12).

Jackson, C.P., 1987. A finite-element study of the onset of vortex shedding in flow past variously shaped bodies. Journal of Fluid Mechanics 182 (1), 23–45.

Jiang, L., Liu, N., 2022. Correcting noisy dynamic mode decomposition with Kalman filters. Journal of Computational Physics 461, 111175.

Jovanović, M.R., Schmid, P.J., Nichols, J.W., 2014. Sparsity-promoting dynamic mode decomposition. Physics of Fluids 26 (2).

Józsa, T., Szőke, M., Teschner, T.-R., Könözsy, L.Z., Moulitsas, I., 2016. Validation and verification of a 2D lattice Boltzmann solver for incompressible fluid flow. In: Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2016). ECCOMAS Congress 2016, Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens (NTUA) Greece.

Juang, J.-N., Pappa, R.S., 1985. An eigensystem realization algorithm for modal parameter identification and model reduction. Journal of Guidance, Control, and Dynamics 8 (5), 620–627.

Kachurovskii, A.G., 1996. The rate of convergence in ergodic theorems. Russian Mathematical Surveys 51 (4), 653–703.

Kaiser, E., Kutz, J.N., Brunton, S.L., 2018a. Discovering conservation laws from data for control. In: 2018 IEEE Conference on Decision and Control (CDC). IEEE, pp. 6415–6421.

Kaiser, E., Kutz, J.N., Brunton, S.L., 2018b. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proceedings of the Royal Society A. Mathematical, Physical and Engineering Sciences 474 (2219), 20180335.

Kaiser, E., Kutz, J.N., Brunton, S.L., 2021. Data-driven discovery of Koopman eigenfunctions for control. Machine Learning: Science and Technology 2 (3), 035023.

Kamb, M., Kaiser, E., Brunton, S.L., Kutz, J.N., 2020. Time-delay observables for Koopman: theory and applications. SIAM Journal on Applied Dynamical Systems 19 (2), 886–917.

Kantz, H., Schreiber, T., 2006. Nonlinear Time Series Analysis, second edition edn. Cambridge Nonlinear Science Series. Cambridge Univ. Press, Cambridge.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nature Reviews Physics 3 (6), 422–440.

Katznelson, Y., 2004. An Introduction to Harmonic Analysis. Cambridge University Press.

Kawahara, Y., 2016. Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis. Advances in Neural Information Processing Systems 29.

Khosravi, M., 2023. Representer theorem for learning Koopman operators. IEEE Transactions on Automatic Control 68 (5), 2995–3010.

Klus, S., Bittracher, A., Schuster, I., Schütte, C., 2018a. A kernel-based approach to molecular conformation analysis. Journal of Chemical Physics 149 (24), 244109.

Klus, S., Gelß, P., Peitz, S., Schütte, C., 2018b. Tensor-based dynamic mode decomposition. Nonlinearity 31 (7), 3359–3380.

Klus, S., Koltai, P., Schütte, C., 2016. On the numerical approximation of the Perron-Frobenius and Koopman operator. Journal of Computational Dynamics 3 (1), 1–12.

Klus, S., Nüske, F., Hamzi, B., 2020a. Kernel-based approximation of the Koopman generator and Schrödinger operator. Entropy 22 (7), 722.

Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., Noé, F., 2018c. Data-driven model reduction and transfer operator approximation. Journal of Nonlinear Science 28 (3), 985–1010.

Klus, S., Nüske, F., Peitz, S., Niemann, J.-H., Clementi, C., Schütte, C., 2020b. Data-driven approximation of the Koopman generator: model reduction, system identification, and control. Physica D. Nonlinear Phenomena 406, 132416.

Klus, S., Schuster, I., Muandet, K., 2020c. Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces. Journal of Nonlinear Science 30 (1), 283–315.

Klus, S., Schütte, C., 2016. Towards tensor-based methods for the numerical approximation of the Perron–Frobenius and Koopman operator. Journal of Computational Dynamics 3 (2), 139–161.

Koch, R., Sanjosé, M., Moreau, S., 2021. Large-eddy simulation of a linear compressor cascade with tip gap: aerodynamic and acoustic analysis. In: AIAA Aviation 2021 Forum. American Institute of Aeronautics and Astronautics, p. 2312.

Koopman, B.O., 1931. Hamiltonian systems and transformation in Hilbert space. Proceedings of the National Academy of Sciences 17 (5), 315–318.

Koopman, B.O., von Neumann, J., 1932. Dynamical systems of continuous spectra. Proceedings of the National Academy of Sciences 18 (3), 255–263.

Korda, M., Mezić, I., 2018a. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. Automatica 93, 149–160.

Korda, M., Mezić, I., 2018b. On convergence of extended dynamic mode decomposition to the Koopman operator. Journal of Nonlinear Science 28 (2), 687–710.

Korda, M., Mezić, I., 2020. Optimal construction of Koopman eigenfunctions for prediction and control. IEEE Transactions on Automatic Control 65 (12), 5114–5129.

Korda, M., Putinar, M., Mezić, I., 2020. Data-driven spectral analysis of the Koopman operator. Applied and Computational Harmonic Analysis 48 (2), 599–629.

Korda, M., Susuki, Y., Mezić, I., 2018. Power grid transient stabilization using Koopman model predictive control. IFAC-PapersOnLine 51 (28), 297–302.

Kostic, V., Novelli, P., Maurer, A., Ciliberto, C., Rosasco, L., Pontil, M., 2022. Learning dynamical systems via Koopman operator regression in reproducing kernel Hilbert spaces. Advances in Neural Information Processing Systems 35, 4017–4031.

Kou, J., Zhang, W., 2017. An improved criterion to select dominant modes from dynamic mode decomposition. European Journal of Mechanics. B, Fluids 62, 109–129.

Krake, T., Klötzl, D., Eberhardt, B., Weiskopf, D., 2022. Constrained dynamic mode decomposition. IEEE Transactions on Visualization and Computer Graphics 29 (1), 1–11.

Krener, A.J., 1974. Linearization and bilinearization of control systems. In: Proceedings of the 1974 Allerton Conference on Circuit and Systems Theory, Urbana III.

Kryloff, N., Bogoliouboff, N., 1937. La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. Annals of Mathematics 38 (1), 65–113.

Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L., 2016a. Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems. Society for Industrial and Applied Mathematics.

Kutz, J.N., Fu, X., Brunton, S.L., 2016b. Multiresolution dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems 15 (2), 713–735.

Lax, P.D., 1971. Approximation of measure preserving transformations. Communications on Pure and Applied Mathematics 24 (2), 133–135.

Le Clainche, S., Vega, J.M., 2017. Higher order dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems 16 (2), 882–925.

Leroux, R., Cordier, L., 2016. Dynamic mode decomposition for non-uniformly sampled data. Experiments in Fluids 57 (5), 94.

Lewin, M., Séré, É., 2010. Spectral pollution and how to avoid it. Proceedings of the London Mathematical Society 100 (3), 864–900.

Li, M., Jiang, L., 2021. Deep learning nonlinear multiscale dynamic problems using Koopman operator. Journal of Computational Physics 446, 110660.

Li, Q., Dietrich, F., Bollt, E.M., Kevrekidis, I.G., 2017. Extended dynamic mode decomposition with dictionary learning: a data-driven adaptive spectral decomposition of the Koopman operator. Chaos 27 (10), 103111.

Li, T.-Y., 1976. Finite approximation for the Frobenius-Perron operator. A solution to Ulam's conjecture. Journal of Approximation Theory 17 (2), 177–186.

Li, Y., He, H., Wu, J., Katabi, D., Torralba, A., 2019. Learning compositional Koopman operators for model-based control. arXiv preprint. arXiv:1910.08264.

Liu, Z., Kundu, S., Chen, L., Yeung, E., 2018. Decomposition of nonlinear dynamical systems using Koopman Gramians. In: 2018 Annual American Control Conference (ACC). IEEE, pp. 4811–4818.

Loiseau, J.-C., Brunton, S.L., 2018. Constrained sparse Galerkin regression. Journal of Fluid Mechanics 838, 42–67.

Loparo, K., Blankenship, G., 1978. Estimating the domain of attraction of nonlinear feedback systems. IEEE Transactions on Automatic Control 23 (4), 602–608.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20 (2), 130–141.

Lu, H., Tartakovsky, D.M., 2020a. Lagrangian dynamic mode decomposition for construction of reduced-order models of advection-dominated phenomena. Journal of Computational Physics 407, 109229.

Lu, H., Tartakovsky, D.M., 2020b. Prediction accuracy of dynamic mode decomposition. SIAM Journal on Scientific Computing 42 (3), A1639–A1662.

Lusch, B., Kutz, J.N., Brunton, S.L., 2018. Deep learning for universal linear embeddings of nonlinear dynamics. Nature Communications 9 (1), 1–10.

Luzzatto, S., Melbourne, I., Paccaut, F., 2005. The Lorenz attractor is mixing. Communications in Mathematical Physics 260 (2), 393–401.

Maćešić, S., Črnjarić-Žic, N., Mezić, I., 2018. Koopman operator family spectrum for nonautonomous systems. SIAM Journal on Applied Dynamical Systems 17 (4), 2478–2515.

Mamakoukas, G., Castano, M., Tan, X., Murphey, T., 2019. Local Koopman operators for data-driven control of robotic systems. In: Robotics: Science and Systems XV. RSS2019, Robotics: Science and Systems Foundation.

Mamakoukas, G., Castano, M.L., Tan, X., Murphey, T.D., 2021. Derivative-based Koopman operators for real-time control of robotic systems. IEEE Transactions on Robotics 37 (6), 2173–2192.

Mañé, R., 1987. Ergodic Theory and Differentiable Dynamics. Springer Berlin Heidelberg.

Manohar, K., Kaiser, E., Brunton, S.L., Kutz, J.N., 2019. Optimized sampling for multiscale dynamics. Multiscale Modeling & Simulation 17 (1), 117–136.

Manojlović, I., Fonoberova, M., Mohr, R., Andrejčuk, A., Drmač, Z., Kevrekidis, Y., Mezić, I., 2020. Applications of Koopman mode analysis to neural networks. arXiv preprint. arXiv:2006.11765.

Mardt, A., Pasquali, L., Noé, F., Wu, H., 2020. Deep learning Markov and Koopman models with physical constraints. In: Mathematical and Scientific Machine Learning. PMLR, pp. 451–475.

Mardt, A., Pasquali, L., Wu, H., Noé, F., 2018. VAMPnets for deep learning of molecular kinetics. Nature Communications 9 (1), 1–11.

Martinsson, P.-G., Tropp, J.A., 2020. Randomized numerical linear algebra: foundations and algorithms. Acta Numerica 29, 403–572.

Mauroy, A., Goncalves, J., 2020. Koopman-based lifting techniques for nonlinear systems identification. IEEE Transactions on Automatic Control 65 (6), 2550–2565.

Mauroy, A., Mezić, I., 2016. Global stability analysis using the eigenfunctions of the Koopman operator. IEEE Transactions on Automatic Control 61 (11), 3356–3369.

Mauroy, A., Mezić, I., Moehlis, J., 2013. Isostables, isochrons, and Koopman spectrum for the action–angle representation of stable fixed point dynamics. Physica D. Nonlinear Phenomena 261, 19–30.

Mauroy, A., Susuki, Y., Mezić, I., 2020. Koopman Operator in Systems and Control. Lecture Notes in Control and Information Sciences Ser., vol. 484. Springer International Publishing AG, Cham.

Mehta, P.G., Hessel-von Molo, M., Dellnitz, M., 2006. Symmetry of attractors and the Perron–Frobenius operator. Journal of Difference Equations and Applications 12 (11), 1147–1178.

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

Mezić, I., 2013. Analysis of fluid flows via spectral properties of the Koopman operator. Annual Review of Fluid Mechanics 45 (1), 357–378.

Mezić, I., 2015. On applications of the spectral theory of the Koopman operator in dynamical systems and control theory. In: 2015 54th IEEE Conference on Decision and Control (CDC). IEEE, pp. 7034–7041.

Mezić, I., 2020. Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry. Journal of Nonlinear Science 30 (5), 2091–2145.
Mezić, I., 2021. Koopman operator, geometry, and learning of dynamical systems. Notices of the American Mathematical Society 68 (07), 1.
Mezić, I., 2022. On numerical approximations of the Koopman operator. Mathematics 10 (7), 1180.
Mezić, I., Banaszuk, A., 2004. Comparison of systems with complex behavior. Physica D. Nonlinear Phenomena 197 (1–2), 101–133.
Mezić, I., Sotiropoulos, F., 2002. Ergodic theory and experimental visualization of invariant sets in chaotically advected flows. Physics of Fluids 14 (7), 2235–2243.
Mezić, I., Surana, A., 2016. Koopman mode decomposition for periodic/quasi-periodic time dependence. IFAC-PapersOnLine 49 (18), 690–697.
Mezić, I., Wiggins, S., 1999. A method for visualization of invariant sets of dynamical systems based on the ergodic partition. Chaos 9 (1), 213–218.
Mohr, R., Mezić, I., 2014. Construction of eigenfunctions for scalar-type operators via Laplace averages with connections to the Koopman operator. arXiv preprint. arXiv:1403.6559.
Mohri, M., Rostamizadeh, A., Talwalkar, A., 2018. Foundations of Machine Learning, Adaptive Computation and Machine Learning, second edition edn. The MIT Press, Cambridge, Massachusetts.
Mojgani, R., Balajewicz, M., 2017. Lagrangian basis method for dimensionality reduction of convection dominated nonlinear flows. arXiv preprint. arXiv:1701.04343.
Mollenhauer, M., Klus, S., Schütte, C., Koltai, P., 2022. Kernel autocovariance operators of stationary processes: estimation and convergence. Journal of Machine Learning Research 23 (327), 1–34.
Morandin, R., Nicodemus, J., Unger, B., 2023. Port-Hamiltonian dynamic mode decomposition. SIAM Journal on Scientific Computing 45 (4), A1690–A1710.
Narasingam, A., Kwon, J.S.-I., 2019. Koopman Lyapunov-based model predictive control of nonlinear chemical process systems. AIChE Journal 65 (11), e16743.
Needell, D., Tropp, J.A., 2009. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Applied and Computational Harmonic Analysis 26 (3), 301–321.
Netto, M., Mili, L., 2018. A robust data-driven Koopman Kalman filter for power systems dynamic state estimation. IEEE Transactions on Power Systems 33 (6), 7228–7237.
Neumann, J.v., 1932. Proof of the quasi-ergodic hypothesis. Proceedings of the National Academy of Sciences 18 (1), 70–82.
Noack, B.R., Afanasiev, K., Morzynski, M., Tadmor, G., Thiele, F., 2003. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. Journal of Fluid Mechanics 497, 335–363.
Noack, B.R., Eckelmann, H., 1994. A global stability analysis of the steady and periodic cylinder wake. Journal of Fluid Mechanics 270, 297–330.
Noack, B.R., Stankiewicz, W., Morzyński, M., Schmid, P.J., 2016. Recursive dynamic mode decomposition of transient and post-transient wake flows. Journal of Fluid Mechanics 809, 843–872.
Noé, F., Nüske, F., 2013. A variational approach to modeling slow processes in stochastic dynamical systems. Multiscale Modeling & Simulation 11 (2), 635–655.
Nonomura, T., Shibata, H., Takaki, R., 2018. Dynamic mode decomposition using a Kalman filter for parameter estimation. AIP Advances 8 (10).
Nonomura, T., Shibata, H., Takaki, R., 2019. Extended-Kalman-filter-based dynamic mode decomposition for simultaneous system identification and denoising. PLoS ONE 14 (2), e0209836.
Nüske, F., Gelß, P., Klus, S., Clementi, C., 2021. Tensor-based computation of metastable and coherent sets. Physica D. Nonlinear Phenomena 427, 133018.
Nüske, F., Keller, B.G., Pérez-Hernández, G., Mey, A.S.J.S., Noé, F., 2014. Variational approach to molecular kinetics. Journal of Chemical Theory and Computation 10 (4), 1739–1752.
Nüske, F., Peitz, S., Philipp, F., Schaller, M., Worthmann, K., 2023. Finite-data error bounds for Koopman-based prediction and control. Journal of Nonlinear Science 33 (1), 14.

Nyquist, H., 1928. Certain topics in telegraph transmission theory. Transactions of the American Institute of Electrical Engineers 47 (2), 617–644.

Ober-Blöbaum, S., Padberg-Gehle, K., 2015. Multiobjective optimal control of fluid mixing. PAMM 15 (1), 639–640.

Otto, S.E., Padovan, A., Rowley, C.W., 2023a. Model reduction for nonlinear systems by balanced truncation of state and gradient covariance. SIAM Journal on Scientific Computing 45 (5), A2325–A2355.

Otto, S.E., Rowley, C.W., 2019. Linearly recurrent autoencoder networks for learning dynamics. SIAM Journal on Applied Dynamical Systems 18 (1), 558–593.

Otto, S.E., Rowley, C.W., 2021. Koopman operators for estimation and control of dynamical systems. Annual Review of Control, Robotics, and Autonomous Systems 4 (1), 59–87.

Otto, S.E., Zolman, N., Kutz, J.N., Brunton, S.L., 2023b. A unified framework to enforce, discover, and promote symmetry in machine learning. arXiv preprint. arXiv:2311.00212.

Pan, C., Xue, D., Wang, J., 2015. On the accuracy of dynamic mode decomposition in estimating instability of wave packet. Experiments in Fluids 56 (8), 1–15.

Pan, S., Arnold-Medabalimi, N., Duraisamy, K., 2021. Sparsity-promoting algorithms for the discovery of informative Koopman-invariant subspaces. Journal of Fluid Mechanics 917, A18.

Pan, S., Duraisamy, K., 2020a. On the structure of time-delay embedding in linear models of nonlinear dynamical systems. Chaos 30 (7), 073135.

Pan, S., Duraisamy, K., 2020b. Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability. SIAM Journal on Applied Dynamical Systems 19 (1), 480–509.

Parker, R., 1984. Acoustic resonances and blade vibration in axial flow compressors. Journal of Sound and Vibration 92 (4), 529–539.

Pazy, A., 2010. Semigroups of Linear Operators and Applications to Partial Differential Equations, 3 edn. Applied Mathematical Sciences, vol. 44. Springer, New York.

Peifer, M., Schelter, B., Winterhalder, M., Timmer, J., 2005. Mixing properties of the Rössler system and consequences for coherence and synchronization analysis. Physical Review E 72 (2), 026213.

Peitz, S., 2018. Controlling nonlinear PDEs using low-dimensional bilinear approximations obtained from data. arXiv preprint. arXiv:1801.06419.

Peitz, S., Klus, S., 2019. Koopman operator-based model reduction for switched-system control of PDEs. Automatica 106, 184–191.

Peitz, S., Klus, S., 2020. Feedback Control of Nonlinear PDEs Using Data-Efficient Reduced Order Models Based on the Koopman Operator. Springer, pp. 257–282.

Peitz, S., Otto, S.E., Rowley, C.W., 2020. Data-driven model predictive control using interpolated Koopman generators. SIAM Journal on Applied Dynamical Systems 19 (3), 2162–2193.

Pereyra, V., Scherer, G., 2010. Exponential data fitting and its applications. Bentham eBooks, Sharjah, UAE.

Philipp, F., Schaller, M., Worthmann, K., Peitz, S., Nüske, F., 2023. Error bounds for kernel-based approximations of the Koopman operator. arXiv preprint. arXiv:2301.08637.

Poincaré, H., 1899. Les méthodes nouvelles de la mécanique céleste. Il Nuovo Cimento 10 (1), 128–130.

Proctor, J.L., Brunton, S.L., Kutz, J.N., 2016. Dynamic mode decomposition with control. SIAM Journal on Applied Dynamical Systems 15 (1), 142–161.

Proctor, J.L., Brunton, S.L., Kutz, J.N., 2018. Generalizing Koopman theory to allow for inputs and control. SIAM Journal on Applied Dynamical Systems 17 (1), 909–930.

Redman, W.T., 2021. On Koopman mode decomposition and tensor component analysis. Chaos 31 (5).

Redman, W.T., Fonoberova, M., Mohr, R., Kevrekidis, I.G., Mezić, I., 2022. Algorithmic (semi-) conjugacy via Koopman operator theory. In: 2022 IEEE 61st Conference on Decision and Control (CDC). IEEE.

Redman, W.T., Huang, D., Fonoberova, M., Mezić, I., 2023. Koopman learning with episodic memory. arXiv preprint. arXiv:2311.12615.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite SST analysis for climate. Journal of Climate 15 (13), 1609–1625.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. Journal of Climate 20 (22), 5473–5496.

Roch, S., Silbermann, B., 1996. C^*-algebra techniques in numerical analysis. Journal of Operator Theory, 241–280.

Rokhlin, V., Szlam, A., Tygert, M., 2010. A randomized algorithm for principal component analysis. SIAM Journal on Matrix Analysis and Applications 31 (3), 1100–1124.

Rosenfeld, J.A., Kamalapurkar, R., Gruss, L.F., Johnson, T.T., 2022. Dynamic mode decomposition for continuous time systems with the Liouville operator. Journal of Nonlinear Science 32 (1), 1–30.

Rössler, O.E., 1976. An equation for continuous chaos. Physics Letters A 57 (5), 397–398.

Rowley, C.W., 2005. Model reduction for fluids, using balanced proper orthogonal decomposition. International Journal of Bifurcation and Chaos 15 (03), 997–1013.

Rowley, C.W., Dawson, S.T.M., 2017. Model reduction for flow analysis and control. Annual Review of Fluid Mechanics 49 (1), 387–417.

Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.S., 2009. Spectral analysis of nonlinear flows. Journal of Fluid Mechanics 641, 115–127.

Rowley, C.W., Williams, D.R., Colonius, T., Murray, R.M., Macmynowski, D.G., 2006. Linear models for control of cavity flow oscillations. Journal of Fluid Mechanics 547 (1), 317–330.

Ruelle, D., 1968. Statistical mechanics of a one-dimensional lattice gas. Communications in Mathematical Physics 9 (4), 267–278.

Salova, A., Emenheiser, J., Rupe, A., Crutchfield, J.P., D'Souza, R.M., 2019. Koopman operator and its approximations for systems with symmetries. Chaos 29 (9).

Sashidhar, D., Kutz, J.N., 2022. Bagging, optimized dynamic mode decomposition for robust, stable forecasting with spatial and temporal uncertainty quantification. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 380 (2229), 20210199.

Scherl, I., Strom, B., Shang, J.K., Williams, O., Polagye, B.L., Brunton, S.L., 2020. Robust principal component analysis for modal decomposition of corrupt fluid flows. Physical Review Fluids 5 (5), 054401.

Schmid, P., Sesterhenn, J., 2008. Dynamic mode decomposition of numerical and experimental data. Bulletin of the American Physical Society 53.

Schmid, P.J., 2009. Dynamic mode decomposition of experimental data. In: 8th International Symposium on Particle Image Velocimetry (PIV09).

Schmid, P.J., 2010. Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics 656, 5–28.

Schmid, P.J., 2022. Dynamic mode decomposition and its variants. Annual Review of Fluid Mechanics 54 (1), 225–254.

Scholkopf, B., 2001. The kernel trick for distances. Advances in Neural Information Processing Systems, 301–307.

Schönemann, P.H., 1966. A generalized solution of the orthogonal procrustes problem. Psychometrika 31 (1), 1–10.

Schütte, C., Klus, S., Hartmann, C., 2023. Overcoming the timescale barrier in molecular dynamics: transfer operators, variational principles and machine learning. Acta Numerica 32, 517–673.

Schütte, C., Sarich, M., 2013. Metastability and Markov State Models in Molecular Dynamics, vol. 24. American Mathematical Society.

Sechi, R., Sikorski, A., Weber, M., 2021. Estimation of the Koopman generator by Newton's extrapolation. Multiscale Modeling & Simulation 19 (2), 758–774.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27 (3), 379–423.

Shargorodsky, E., 2008. On the level sets of the resolvent norm of a linear operator. Bulletin of the London Mathematical Society 40 (3), 493–504.

Sharma, A.S., Mezić, I., McKeon, B.J., 2016. Correspondence between Koopman mode decomposition, resolvent mode decomposition, and invariant solutions of the Navier-Stokes equations. Physical Review Fluids 1 (3), 032402.

Sharma, H., Vaidya, U., Ganapathysubramanian, B., 2019. A transfer operator methodology for optimal sensor placement accounting for uncertainty. Building and Environment 155, 334–349.

Shields, P.C., 1973. The Theory of Bernoulli Shifts. University of Chicago Press, Chicago.

Sinha, S., Huang, B., Vaidya, U., 2020. On robust computation of Koopman operator and prediction in random dynamical systems. Journal of Nonlinear Science 30 (5), 2057–2090.

Sinha, S., Vaidya, U., Rajaram, R., 2016. Operator theoretic framework for optimal placement of sensors and actuators for control of nonequilibrium dynamics. Journal of Mathematical Analysis and Applications 440 (2), 750–772.

Sinha, S., Vaidya, U., Yeung, E., 2019. On computation of Koopman operator from sparse data. In: 2019 American Control Conference (ACC). IEEE, pp. 5519–5524.

Slipantschuk, J., Bandtlow, O.F., Just, W., 2020. Dynamic mode decomposition for analytic maps. Communications in Nonlinear Science and Numerical Simulation 84, 105179.

Smith, T.M., Reynolds, R.W., 2005. A global merged land–air–sea surface temperature reconstruction based on historical observations (1880–1997). Journal of Climate 18 (12), 2021–2036.

Son, S.H., Narasingam, A., Kwon, J.S.-I., 2020. Handling plant-model mismatch in Koopman Lyapunov-based model predictive control via offset-free control framework. arXiv preprint. arXiv:2010.07239.

Sootla, A., Mauroy, A., 2016. Properties of isostables and basins of attraction of monotone systems. In: 2016 American Control Conference (ACC). IEEE, pp. 7365–7370.

Surana, A., 2016. Koopman operator based observer synthesis for control-affine nonlinear systems. In: 2016 IEEE 55th Conference on Decision and Control (CDC). IEEE, pp. 6492–6499.

Surana, A., Banaszuk, A., 2016. Linear observer synthesis for nonlinear systems using Koopman operator framework. IFAC-PapersOnLine 49 (18), 716–723.

Susuki, Y., Mauroy, A., Mezić, I., 2021. Koopman resolvent: a Laplace-domain analysis of nonlinear autonomous dynamical systems. SIAM Journal on Applied Dynamical Systems 20 (4), 2013–2036.

Susuki, Y., Mezić, I., 2015. A Prony approximation of Koopman mode decomposition. In: 2015 54th IEEE Conference on Decision and Control (CDC). IEEE.

Szőke, M., Jozsa, T.I., Koleszár, Á., Moulitsas, I., Könözsy, L., 2017. Performance evaluation of a two-dimensional lattice Boltzmann solver using CUDA and PGAS UPC based parallelisation. ACM Transactions on Mathematical Software 44 (1), 1–22.

Szőke, M., Nurani Hari, N., Devenport, W.J., Glegg, S.A., Teschner, T.-R., 2021. Flow field analysis around pressure shielding structures. In: AIAA Aviation 2021 Forum. American Institute of Aeronautics and Astronautics, p. 2293.

Taira, K., Brunton, S.L., Dawson, S.T.M., Rowley, C.W., Colonius, T., McKeon, B.J., Schmidt, O.T., Gordeyev, S., Theofilis, V., Ukeiley, L.S., 2017. Modal analysis of fluid flows: an overview. AIAA Journal 55 (12), 4013–4041.

Taira, K., Hemati, M.S., Brunton, S.L., Sun, Y., Duraisamy, K., Bagheri, S., Dawson, S.T.M., Yeh, C.-A., 2020. Modal analysis of fluid flows: applications and outlook. AIAA Journal 58 (3), 998–1022.

Takeishi, N., Kawahara, Y., Tabei, Y., Yairi, T., 2017a. Bayesian dynamic mode decomposition. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. IJCAI-2017, International Joint Conferences on Artificial Intelligence Organization, pp. 2814–2821.

Takeishi, N., Kawahara, Y., Yairi, T., 2017b. Learning Koopman invariant subspaces for dynamic mode decomposition. Advances in Neural Information Processing Systems 30.

Takeishi, N., Kawahara, Y., Yairi, T., 2017c. Subspace dynamic mode decomposition for stochastic Koopman analysis. Physical Review E 96 (3), 033310.

Takens, F., 2006. Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80. Springer, pp. 366–381.

Tantet, A., Lucarini, V., Lunkeit, F., Dijkstra, H.A., 2018. Crisis of the chaotic attractor of a climate model: a transfer operator approach. Nonlinearity 31 (5), 2221–2251.

Tantet, A., van der Burgt, F.R., Dijkstra, H.A., 2015. An early warning indicator for atmospheric blocking events using transfer operators. Chaos 25 (3).

Tissot, G., Cordier, L., Benard, N., Noack, B.R., 2014. Model reduction using dynamic mode decomposition. Comptes Rendus. Mécanique 342 (6–7), 410–416.

Towne, A., Schmidt, O.T., Colonius, T., 2018. Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. Journal of Fluid Mechanics 847, 821–867.

Trefethen, L.N., Embree, M., 2005. Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators. Princeton University Press.

Trefethen, L.N., Trefethen, A.E., Reddy, S.C., Driscoll, T.A., 1993. Hydrodynamic stability without eigenvalues. Science 261 (5121), 578–584.

Tu, J.H., Rowley, C.W., Kutz, J.N., Shang, J.K., 2014a. Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. Experiments in Fluids 55 (9), 1–13.

Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., Nathan Kutz, J., 2014b. On dynamic mode decomposition: theory and applications. Journal of Computational Dynamics 1 (2), 391–421.

Tufillaro, N.B., Abbot, T., Reilly, J., Hickey, F.R., 1993. An experimental approach to nonlinear dynamics and chaos. American Journal of Physics 61 (10), 958–959.

Udell, M., Townsend, A., 2019. Why are big data matrices approximately low rank? SIAM Journal on Mathematics of Data Science 1 (1), 144–160.

Ulam, S.M., 1960. Problems in Modern Mathematics. Science Editions.

Vaidya, U., 2007. Observability Gramian for nonlinear systems. In: 2007 46th IEEE Conference on Decision and Control. IEEE, pp. 3357–3362.

Vaidya, U., Mehta, P.G., Shanbhag, U.V., 2010. Nonlinear stabilization via control Lyapunov measure. IEEE Transactions on Automatic Control 55 (6), 1314–1328.

Valva, C., Giannakis, D., 2023. Consistent spectral approximation of Koopman operators using resolvent compactification. arXiv preprint. arXiv:2309.00732.

Van Huffel, S., Vandewalle, J., 1991. The Total Least Squares Problem: Computational Aspects and Analysis. Society for Industrial and Applied Mathematics.

Vega, J.M., Le Clainche, S., 2021. Spatio-Temporal Koopman Decomposition, vol. 28. Elsevier, pp. 121–157.

von Luxburg, U., Belkin, M., Bousquet, O., 2008. Consistency of spectral clustering. The Annals of Statistics 36 (2), 555–586.

Walters, P., 2000. An Introduction to Ergodic Theory, 1 softcover printing edn. Graduate Texts in Mathematics, vol. 79. Springer, New York.

Wanner, M., Mezić, I., 2022. Robust approximation of the stochastic Koopman operator. SIAM Journal on Applied Dynamical Systems 21 (3), 1930–1951.

Wehmeyer, C., Noé, F., 2018. Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. Journal of Chemical Physics 148 (24).

Welch, P., 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics 15 (2), 70–73.

Williams, M.O., Hemati, M.S., Dawson, S.T.M., Kevrekidis, I.G., Rowley, C.W., 2016. Extending data-driven Koopman analysis to actuated systems. IFAC-PapersOnLine 49 (18), 704–709.

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

Williams, M.O., Rowley, C.W., Kevrekidis, I.G., 2015b. A kernel-based method for data-driven Koopman spectral analysis. Journal of Computational Dynamics 2 (2), 247–265.

Woodley, B.M., Peake, N., 1999. Resonant acoustic frequencies of a tandem cascade. Part 2. Rotating blade rows. Journal of Fluid Mechanics 393, 241–256.

Wormell, C.L., 2023. Orthogonal polynomial approximation and extended dynamic mode decomposition in chaos. arXiv preprint. arXiv:2305.08074.

Wu, H., Noé, F., 2020. Variational approach for learning Markov processes from time series data. Journal of Nonlinear Science 30 (1), 23–66.

Wu, H., Nüske, F., Paul, F., Klus, S., Koltai, P., Noé, F., 2017. Variational Koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations. Journal of Chemical Physics 146 (15).

Wu, Z., Brunton, S.L., Revzen, S., 2021. Challenges in dynamic mode decomposition. Journal of the Royal Society Interface 18 (185), 20210686.

Wynn, A., Pearson, D.S., Ganapathisubramani, B., Goulart, P.J., 2013. Optimal mode decomposition for unsteady flows. Journal of Fluid Mechanics 733, 473–503.

Xie, T., France-Lanord, A., Wang, Y., Shao-Horn, Y., Grossman, J.C., 2019. Graph dynamical networks for unsupervised learning of atomic scale dynamics in materials. Nature Communications 10 (1), 2667.

Yeung, E., Kundu, S., Hodas, N., 2019. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems. In: 2019 American Control Conference (ACC). IEEE, pp. 4832–4839.

Yeung, E., Liu, Z., Hodas, N.O., 2018. A Koopman operator approach for computing and balancing Gramians for discrete time nonlinear systems. In: 2018 Annual American Control Conference (ACC). IEEE, pp. 337–344.

Young, L.-S., 2002. What are SRB measures, and which dynamical systems have them? Journal of Statistical Physics 108 (5/6), 733–754.

Zaslavsky, G.M., 2002. Chaos, fractional kinetics, and anomalous transport. Physics Reports 371 (6), 461–580.

Zebib, A., 1987. Stability of viscous flow past a circular cylinder. Journal of Engineering Mathematics 21 (2), 155–165.

Zhang, H., Rowley, C.W., Deem, E.A., Cattafesta, L.N., 2019. Online dynamic mode decomposition for time-varying systems. SIAM Journal on Applied Dynamical Systems 18 (3), 1586–1609.

Zhao, Z., Giannakis, D., 2016. Analog forecasting with dynamics-adapted kernels. Nonlinearity 29 (9), 2888–2939.