

博士論文

Data-Driven Analysis of Dynamical Systems Based on
the Koopman Operator: A Machine Learning Perspective

(クープマン作用素に基づく力学系の
データによる解析：機械学習の視点から)

東京大学大学院 工学系研究科 航空宇宙工学専攻
武石 直也

平成29年12月1日

Data-Driven Analysis of Dynamical Systems Based on the Koopman Operator: A Machine Learning Perspective

Author:

Naoya TAKEISHI

Supervisor:

Prof. Takehisa YAIRI

*A dissertation submitted to
Department of Aeronautics and Astronautics
in partial fulfillment of the requirements
for the degree of Doctor of Engineering*

at

The University of Tokyo

December 1, 2017

Abstract

Time-series data are ubiquitous in the modern society, and analyzing dynamical systems that generate them is an important branch of research for many applications in science and engineering. In this dissertation, we focus on the data-driven approach to the analysis of dynamical systems. In particular, we study the modal decomposition technique based on the Koopman operator of a dynamical system, which provides powerful tools for understanding intrinsic structures of nonlinear dynamical systems. A data-driven realization of such modal decomposition is known as dynamic mode decomposition (DMD). DMD has been effectively utilized in several fields of science, but its assumptions are sometimes too restrictive in other applications.

In this paper, we develop extensions of DMD and new methodologies based on machine learning techniques for performing the data-driven modal decomposition based on the Koopman operator. The proposed methods relax the assumptions needed by DMD and enable us to apply the Koopman-based analysis of dynamical systems to a broader range of scientific and engineering domains that cannot be dealt with existing numerical methods.

In Chapter 2, we introduce the sparse nonnegative version of DMD, which is motivated by the use in image and video processing. The proposed method is based on a reformulation of DMD, and other types of regularizations and/or constraints can be imposed based on the proposed reformulation.

In Chapter 3, we propose probabilistic and Bayesian DMDs, with which uncertainty in data are explicitly treated. Hence, they are suited to experimental datasets. Moreover, the proposed probabilistic formulation is a foundation for further extensions of the model.

In Chapter 4, we address the challenge to perform the modal decomposition based on the Koopman operator for systems where both observation noise and process noise are present. To this end, we propose subspace DMD, which is motivated by methods of the subspace system identification developed in the control theory.

In Chapter 5, we suggest a new framework to make DMD applicable to various kinds of datasets. Since DMD has to be applied to data generated from a set of observables that spans a Koopman invariant subspace, we develop a method to learn such observables from data.

Acknowledgements

本博士論文の執筆および関連研究の遂行にあたり、多くの方々のお力添えをいただきました。この場をお借りして心よりの感謝を申し上げます。

指導教員である矢入健久先生には、学部4年、修士課程、博士後期課程の計6年間にわたってご指導をいただきました。矢入先生のもとでは、興味の赴くままに研究できる自由な環境を与えていただきつつ、企業等との共同研究にも深く関わることができ、他では得られない経験ができたと思います。また、理論的な面から応用場面でのアイデアにいたる幅広い助言をいただき、当初私にとって全く未知だった機械学習という分野を勉強し研究を行ううえで大きな助けになりました。

堀浩一先生、岩崎晃先生、中谷辰爾先生、河原吉伸先生、柳澤大地先生には、本博士論文の副査として中間審査等の場で様々な助言をいただきました。特に堀先生と岩崎先生には、研究室での生活や研究会においてもお世話になりました。また、河原先生には動的モード分解という研究テーマに取り組むきっかけをいただき、具体的な研究についても議論をいただきました。

先端研や本郷での研究室のメンバーには、研究活動に関わることでなく、飲み会や研究室旅行などを通して様々なモチベーションをいただきました。研究への情熱という点では必ずしも自信の持てなかった私ですが、研究室のメンバーとの交流があったからこそ学部から博士課程の6年間で過ごせたと思っています。

末筆ながら、これまでの生活の支えとなってくれた家族へ心よりの感謝を記します。

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Data-driven analysis of dynamical systems	1
1.2 Koopman operator on dynamical systems	4
1.3 Koopman mode decomposition (KMD)	6
1.3.1 Applications of KMD	8
1.3.2 Numerical approximation of KMD	9
1.4 Dynamic mode decomposition (DMD)	11
1.4.1 Algorithm of DMD	11
1.4.2 Connection between KMD and DMD	12
1.4.3 Variants of DMD	15
1.5 Summary of remaining chapters	19
2 Sparse nonnegative DMD	23
2.1 Introduction	23
2.2 Reformulation as a block-multiconvex problem	24
2.3 Sparse nonnegative DMD	26
2.4 Experiments	28
2.4.1 Mode extraction from noisy data	28
2.4.2 Extraction of part-based dynamic modes	30

2.5	Summary	31
3	Probabilistic and Bayesian DMD	35
3.1	Introduction	35
3.2	Probabilistic DMD	36
3.2.1	Generative model	36
3.2.2	Maximum-likelihood estimator	37
3.3	Bayesian DMD	39
3.3.1	Posterior inference by Gibbs sampling	39
3.3.2	Sparsity-promoting prior	41
3.4	Related work	42
3.5	Numerical examples	43
3.5.1	Estimation from noisy observations	43
3.5.2	Automatic relevance determination	45
3.5.3	Applications	48
3.6	Summary	53
4	Subspace DMD	55
4.1	Introduction	55
4.2	Stochastic Koopman analysis with noisy observations	56
4.2.1	Process noise on dynamics	56
4.2.2	Observation noise on observables	58
4.2.3	Statistics of noisy observables on RDS	59
4.2.4	Subspace DMD	61
4.2.5	Relation to subspace system identification and extension to controlled systems	64
4.3	Numerical examples	65
4.3.1	Oscillation perturbed by noise	65

4.3.2	Noisy damping modes	67
4.3.3	Quantitative investigation of effects of noises	69
4.3.4	Low-rank high-dimensional data	73
4.3.5	Application: Cylinder wake	77
4.4	Summary	77
5	Learning Koopman invariant subspaces for DMD	79
5.1	Introduction	79
5.2	Learning Koopman invariant subspaces	80
5.2.1	Minimizing residual sum of squares of linear least-squares regression	80
5.2.2	Linear delay embedder for state space reconstruction	82
5.2.3	Reconstruction of original measurements	83
5.2.4	Implementation using neural networks	84
5.3	Related work	85
5.4	Setup of numerical examples and applications	86
5.5	Numerical examples	87
5.6	Applications	93
5.7	Summary	96
6	Conclusion	99
6.1	Summary of contributions	99
6.2	Remaining challenges	101
	Bibliography	103

Chapter 1

Introduction

1.1 Data-driven analysis of dynamical systems

Time-series data are ubiquitous in various disciplines of science and in operation of modern engineering systems because of the development of measurement instruments. Nowadays, analysis and forecasting of time-series play an indispensable role in making a scientific discovery and in achieving security and efficiency in system operations. For example, operators of complex engineering systems, such as manufacturing plants and artificial satellites, survey readings of thousands of sensors everyday (possibly using some support tools) to maintain the operation. Methods for analyzing time-series¹ have been subjected to study in statistics for years (Box et al., 2015) and are still studied intensively in multiple research areas such as statistics, control, and machine learning.

Analysis of time-series is strongly associated with the concept of *dynamical systems*. A dynamical system is a mathematical system in which the temporal evolution of a quantity is described with a function, and most of the time-series data can be regarded as being generated from a kind of dynamical systems. There exists a large body of mathematical analysis study on dynamical systems, whose results have been employed in diverse application areas such as physics, chemistry, biology, astronomy, economics, and sociology, where subjects of research are modeled in terms of dynamical systems. In this dissertation, however, we focus on the *data-driven* approach to understanding the characteristics of dynamical systems, i.e., we pursue methods to obtain knowledge about dynamical systems of interest from time-series data generated by those systems.

¹Or more generally, any types of sequence.

A data-driven approach for analyzing subjects of scientific interest, often termed data science, is increasingly recognized as a distinct paradigm of science (Hey et al., 2009), in which methodologies developed in statistics and machine learning have been intensively utilized. However, a simple application of an arbitrary method of machine learning is inappropriate. For scientific studies, it is highly desirable for those methodologies to be soundly build on theoretical rationales² and to be as simple as possible. A popular one of such methods that has been utilized in analysis of dynamical systems would be proper orthogonal decomposition (POD) (see e.g. Holmes et al. (2012)).³ POD decomposes the dynamics into orthogonal modes that optimally capture the energy of the dynamics, and it has been extensively applied in fluid dynamics to understand the patterns that govern fluid fields (Bonnet et al., 1994; Noack et al., 2003).

In the data-driven analysis of dynamical systems, the purpose of study usually lies in identifying latent structures or laws behind data. For example, in neuroscience, one would like to know the intrinsic state of a neural system that may be captured in signals of an electroencephalogram, which are often very noisy and high-dimensional. This kind of issues can be addressed using decomposition of dynamical systems / time-series like POD. However, a possible drawback of POD is that the decomposition only based on energy does not always reveal important patterns of time-series, which led to the development of other methodologies including the one discussed in this dissertation. In this dissertation, we focus on one of those decomposition methods, namely *dynamic mode decomposition* (DMD) (Rowley et al., 2009; Schmid, 2010; Kutz et al., 2016a).

The reason we focus on DMD is mainly twofold: 1) it is simple and computationally feasible, and 2) it has a sound theoretical rationale based on the notion of the Koopman operator of dynamical systems (Koopman, 1931; Mezić, 2005). If certain assumptions are satisfied in a dataset, using DMD, one can conduct a data-driven analysis of dynamical systems, which is theoretically interpretable and computationally efficient. The effectiveness of

²Of course, most data-analysis methods have their own theoretical background. Here, we specifically expect that such theoretical background is (mathematically and/or physically) appropriate for explaining aspects of research subjects to which the data-analysis methods are applied.

³In machine learning and pattern recognition, a method equivalent to data-driven POD is known as principal component analysis (PCA) and has been applied for modal decomposition and dimensionality reduction of a wide variety of numerical datasets (Jolliffe, 2002).

DMD has been confirmed in several scientific and engineering fields including computational fluid dynamics, where data are usually clean and high-dimensional. However, in many other fields of science and engineering, DMD is not always appropriate, i.e., the necessary assumptions can hardly be satisfied (even approximately). While many researchers have been working on relaxing those assumptions by improving algorithms, we find several important aspects of practical data analyses, such as uncertainty in data, have not been addressed enough yet. In this dissertation, we introduce improvements, extensions, and new methodologies to the modal decomposition technique based on the Koopman operator.

The most of the proposed techniques in this dissertation are based on the methods and concepts often utilized in machine learning researches, and we believe that they open up a new pathway to data-driven analysis of dynamical systems. Using the proposed techniques, scientists and engineers can apply the data decomposition method based on the theory of Koopman operator to datasets that could not be appropriately treated beforehand. For instance, the method introduced in Chapter 4 enables us to use the decomposition method for datasets with both observation noise and process noise, which we were not necessarily able to treat correctly with the existing methods. Moreover, based on the machine learning perspective introduced in this dissertation, one can easily consider further extensions of the method. For example, the Bayesian treatment of the decomposition in Chapter 3 enables us to transfer fruitful advances of Bayesian modeling to DMD, which will lead to a foundation of more advanced modeling of time-series data.

In the remainder of this chapter, technical preliminaries regarding dynamical systems, the Koopman operator, and modal decomposition based on it are provided. Note that, since our original motivation lies in data science applications of those concepts, some expressions may seem informal from a rigorous mathematical viewpoint for ease of discussion. For more strict explanation in terms of mathematics, readers are recommended to consult literatures, such as [Hirsch et al. \(2013\)](#), [Lasota and Mackey \(1994\)](#), [Mezić \(2005\)](#), [Budišić et al. \(2012\)](#), and [Mohr \(2014\)](#).

1.2 Koopman operator on dynamical systems

Firstly, we mention the notion of state space models, in which the evolution of a state (quantity that holds all information to describe system's status at each time) is described by a function. The state space model is a universal tool to describe subjects of interest in a wide disciplines of sciences, such as applied mathematics, physics, chemistry, biology, astronomy, economics, and sociology. Moreover, in many practices of machine learning and data science, the state-space models have been utilized to analyze series data like time-series. In both of these areas, the most simple form⁴ of the state-space model is defined in terms of a differential equation (corresponding to the continuous-time setting):

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}), \quad (1.1)$$

or in terms of a difference equation (corresponding to the discrete-time setting):

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t), \quad (1.2)$$

where $\mathbf{x} \in \mathcal{M}$ is the state of the system, \mathcal{M} is a manifold referred to as state space or phase space, and $\mathbf{f} : \mathcal{M} \rightarrow \mathcal{M}$ is a function. In the remainder of this dissertation, theories and algorithms are constructed solely on the state-space model as a difference equation, i.e., Eq. (1.2), because the discrete-time setting is usually adopted in the practice of machine learning and data science.⁵ In the discrete-time setting, time index t is often defined by $t \in \mathbb{T} = \{0\} \cup \mathbb{N}$, and we also adopt this definition.

The analysis using the state-space models becomes challenging when \mathbf{f} is highly nonlinear. While there have been great efforts on analyzing and learning such nonlinear dynamics, we depart such perspective. That is, as an alternative to the state-space representation, we take the *operator-theoretic* approach to analysis of dynamical systems. This view depends on the Koopman (composition) operator (Koopman, 1931; Lasota and Mackey, 1994; Mezić and Banaszuk, 2004; Mezić, 2005) or its adjoint, the Perron–Frobenius (Ruelle or transfer) operator (Ruelle, 1968; Lasota and Mackey, 1994). The virtue of

⁴As the most simple form, the *autonomous deterministic* dynamical systems are presented here. However, a non-autonomous system, a non-deterministic (random) system, and a non-autonomous random system can also be considered. In fact, the motivation of the work introduced in Chapter 4 is to adopt the analysis based on the Koopman operator to random dynamical systems. Also, studies on Koopman-based analysis for non-autonomous systems are introduced in Section 1.4.3.

⁵The continuous-time setting is favored in physics.

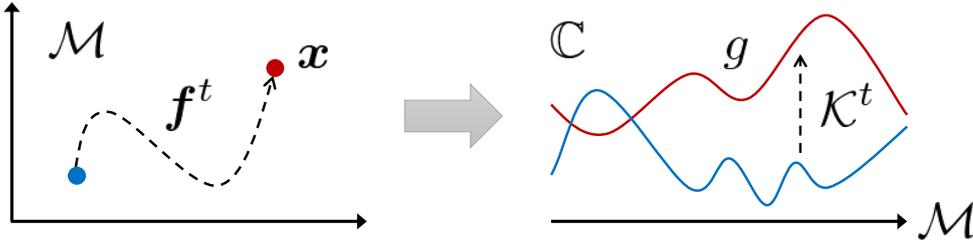


FIGURE 1.1: Schematic diagram on lifting nonlinear dynamics to a linear regime based on Koopman operator \mathcal{K} .

this view lies in the linearity of those operators; we can leverage the machinery of the linear operator theory for analyzing and learning those operators and the nonlinear dynamics behind them. At the same time, however, the compensation of the linearity must be paid: the infinite dimensionality of the representation. This is because those operators are defined in a space of functions or measures, which are infinite-dimensional in general. The idea can be summarized conceptually as follows; by introducing those operators, we *lift nonlinear dynamical systems to a linear regime* as depicted in Figure 1.1.

Let $(\mathcal{M}, \Sigma_{\mathcal{M}}, \mu_{\mathcal{M}})$ be a probability space associated with phase space \mathcal{M} and $\mathbf{f} : \mathcal{M} \rightarrow \mathcal{M}$ be a measurable function with respect to $\mu_{\mathcal{M}}$. Now consider an *observable* (observation function) $g : \mathcal{M} \rightarrow \mathbb{C}$ in a function space \mathcal{G} , which is a vector space. The choice of the function space can be arbitrary, but generally we consider $\mathcal{G} = L^2(\mathcal{M}, \mu_{\mathcal{M}})$ for theoretical and computational convenience. Then, the Koopman operator on a discrete-time dynamical system \mathbf{f} is defined as follows.

Definition 1.1 (Koopman operator (Koopman, 1931; Mezić, 2005)). Let \mathcal{M} be a state space and $\mathbf{f} : \mathcal{M} \rightarrow \mathcal{M}$ be a measurable function corresponding to a discrete-time dynamical system. Moreover, consider a (possibly infinite-dimensional) vector space of observables \mathcal{G} . *Koopman operator* (composition operator) $\mathcal{K} : \mathcal{G} \rightarrow \mathcal{G}$ is a linear operator that describes evolution of an observable, i.e.,

$$\mathcal{K}g = g \circ \mathbf{f}, \quad (1.3)$$

where \circ denotes the composition of functions.

Remark 1.1. The same sort of definition can be considered also for a continuous-time dynamical system (1.1). Instead of the Koopman operator, the *Koopman semigroup* $\{\mathcal{K}_c^t\}_{t \in \mathbb{R}^+}$ on this dynamical system is defined by

$$\mathcal{K}_c^t g(\mathbf{x}) = g(\phi(\mathbf{x}, t)), \quad (1.4)$$

where $\phi(\mathbf{x}, t)$ is the flow map that takes \mathbf{x} as the initial state and returns the state after a time interval of length $t \in \mathbb{R}$. The infinitesimal generator, \mathcal{K}_c , of the Koopman semigroup is given as

$$\mathcal{K}_c = \lim_{t \rightarrow 0} \frac{\mathcal{K}_c^t g - g}{t}. \quad (1.5)$$

The Koopman operator has been recently utilized for analysis of dynamical systems by several researchers. For example, [Lan and Mezić \(2013\)](#) proposed a linearization scheme of nonlinear systems on attractors, [Mauroy et al. \(2013\)](#) and [Sootla and Mauroy \(2017\)](#) suggested the analysis of basins of attractions using spectra of the Koopman operator, [Mauroy et al. \(2015\)](#) build a theory on differential positivity of systems, and [Mauroy and Mezić \(2016\)](#) proposed a method of global stability analysis using the Koopman operator. Moreover, applications that utilize the modal decomposition based on the Koopman operator are introduced in Section [1.3.1](#).

The adjoint of the Koopman operator is known as the Perron–Frobenius operator ([Ruelle, 1968](#); [Lasota and Mackey, 1994](#)). Now consider a distribution density of \mathbf{x} , namely, $\mathbf{x} \sim p \in L^1(\mathcal{M}, \mu_{\mathcal{M}})$, where p is nonnegative for almost everywhere and normalized to one. The Perron–Frobenius operator \mathcal{P} is defined as

$$\int_{\mathcal{A}} \mathcal{P}p \, d\mu_{\mathcal{M}} = \int_{f^{-1}(\mathcal{A})} p \, d\mu_{\mathcal{M}}, \quad (1.6)$$

where $\mathcal{A} \in \Sigma_{\mathcal{M}}$. The Perron–Frobenius operator has been utilized for applications such as construction of observability Gramian ([Vaidya, 2007](#)), global stability analysis ([Vaidya and Mehta, 2008](#)), and computation of basins of attraction ([Wang and Vaidya, 2010](#)). However, in this dissertation, we do not focus on the methods based on the Perron–Frobenius operator.

1.3 Koopman mode decomposition (KMD)

Modal decomposition of observables based on eigenvalues of the Koopman operator has been utilized in various applications, which is often termed as Koopman mode decomposition (KMD) ([Mezić and Banaszuk \(2004\)](#); [Mezić \(2005\)](#); [Budišić et al. \(2012\)](#)). In this section, we introduce the fundamental idea, applications, and numerical approximation methods of KMD.

Let $\lambda \in \mathbb{C}$ and $\varphi : \mathcal{M} \rightarrow \mathbb{C}$ respectively denote the eigenvalue and the eigenfunction of the Koopman operator, i.e.,

$$\mathcal{K}\varphi_i = \lambda_i\varphi_i, \quad (1.7)$$

where we use subscript $i \in \mathbb{N}$ for numbering eigenvalues, eigenfunctions, and other corresponding quantities.

Definition 1.2 (Koopman modes). Let $\{\varphi_i \mid i = 1, \dots, N\}$ (N may be infinite) be a set of eigenfunctions of the Koopman operator. *Koopman modes* of an observable g with regard to $\{\varphi_i\}$ are the coefficients of projection of g to $\text{span}\{\varphi_i\}$.

Here, assume that observable g is in the span of the (possibly infinite number of) eigenfunctions, i.e.,

$$g \in \text{span}\{\varphi_i \mid i = 1, \dots, N\}, \quad N \in \mathbb{N} \quad \text{or} \quad N = \infty. \quad (1.8)$$

Then, g can be exactly expressed in terms of the eigenfunctions and Koopman modes $\{v_i\}$:

$$g(\mathbf{x}) = \sum_{i=1}^N \varphi_i(\mathbf{x})v_i. \quad (1.9)$$

Since φ is the eigenfunction of \mathcal{K} , applying \mathcal{K} to both sides of Eq. (1.9) yields

$$g(\mathbf{f}(\mathbf{x})) = \sum_{i=1}^N \lambda_i\varphi_i(\mathbf{x})v_i. \quad (1.10)$$

Similarly, starting at some initial condition $\mathbf{x} = \mathbf{x}_0$, after t applications of \mathcal{K} ,

$$g(\mathbf{x}_t) = \sum_{i=1}^N \lambda_i^t w_i, \quad w_i = \varphi_i(\mathbf{x}_0)v_i. \quad (1.11)$$

Since snapshots of time-series data are multidimensional in general, instead of a scalar-valued observable g , we often define KMD for a vector-valued observable $\mathbf{g} : \mathcal{M} \rightarrow \mathbb{C}^n$, i.e., concatenation of n observables

$$\mathbf{g} = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \end{bmatrix}^\top. \quad (1.12)$$

Analogously to Eq. (1.11), KMD for g is defined as

$$\mathbf{g}(\mathbf{x}_t) = \sum_{i=1}^N \lambda_i^t \mathbf{w}_i, \quad \mathbf{w}_i = \varphi_i(\mathbf{x}_0) \mathbf{v}_i, \quad (1.13)$$

where the Koopman modes are now defined with a n -dimensional vector \mathbf{v} .

Expression (1.13) is a standard form referred to as KMD. In Eq. (1.13), the value of observable at time t is transformed into the summation of N coefficients $\mathbf{w}_1, \dots, \mathbf{w}_N$, each of which is multiplied by λ_i^t . Since λ is a complex value in general, the magnitude of λ_i represents the growth/decay rate (along time) of \mathbf{w}_i , and the angular component of λ_i represents the (angular) frequency of \mathbf{w}_i . In other words, the original observable is decomposed into a set of vectors $\{\mathbf{w}_i\}$ that evolve along time according to $|\lambda_i|$ and $\angle \lambda_i$. Though \mathbf{v} should be called Koopman modes following the original definition, Definition 1.2, sometimes \mathbf{w} is also called Koopman modes.

Note that, in the above discussion, KMD is defined relying on the premise (1.8), which is not very trivial. Speaking from the viewpoint of spectral decomposition of a linear operator, one can regard that this view ignores continuous spectra of the Koopman operator, which may be too restrictive for complex systems. See literatures such as Budišić et al. (2012) for decomposition considering the continuous spectra. Moreover, computing KMD considering continuous spectra of the Koopman operator is an active area of research; see e.g., Korda et al. (2017). However, this direction is out of scope of this dissertation.

1.3.1 Applications of KMD

Numerical methods that are connected to KMD are utilized in various applications. Especially, there is a vast amount of literature on applications of KMD-like methods in fluid dynamics, partly because a popular numerical approximation of KMD was first invented in the community of fluid dynamics (Schmid and Sesterhenn, 2008; Rowley et al., 2009; Schmid, 2010). They utilize the modal decomposition for reduced-order modeling of fluids and for inspecting structures of a fluid field using the Koopman modes. We do not provide a whole reference on those applications, but readers can consult review articles such as Mezić (2013) and Rowley and Dawson (2017).

The modal decomposition based on the Koopman operator is also utilized in various areas other than fluid dynamics, which include analysis of power systems (Susuki and Mezić, 2011, 2012, 2014; Barocio et al., 2015; Raak et al., 2016), analysis of building systems (Eisenhower et al., 2010; Georgescu et al., 2012), analysis of acoustic phenomena (Jourdain et al., 2013), simulation of a medical procedure (Bourantas et al., 2014), analysis of meteorological data (Giannakis et al., 2015), background separation of video streams (Kutz et al., 2015; Erichson et al., 2016), computation of dynamic textures of video streams (Surana, 2015), fusion of multiple datasets (Williams et al., 2015c), epidemiology (Proctor and Eckhoff, 2015), robotics (Berger et al., 2015), neuroscience (Brunton et al., 2016a), financial trading (Mann and Kutz, 2016; Hua et al., 2016), analysis of human locomotion (Boudali et al., 2017), analysis of chaotic systems (Brunton et al., 2017), prediction of high-dimensional time-series (Hua et al., 2017), edge detection from images (Bi et al., 2017), and analysis of sport plays (Fujii et al., 2017). In those applications, they examine spatial distribution of the Koopman modes or profiles of eigenfunctions of the Koopman operator, which can further be utilized for understanding the mechanics of phenomena and for designing appropriate procedures to manipulate them.

The KMD-like methods are also used in the context of control: nonlinear system identification (Mauroy and Goncalves, 2016), construction of an observer form and Kalman filter (Surana, 2016), model-based control of robots (Abraham et al., 2017), investigation of controllability of network (de Badyn et al., 2017), and analysis of self-organization in network (Caro-Ruiz et al., 2017).

Note that, in those applications, Koopman operator or its spectral components (eigenvalues, eigenfunctions, and modes) are approximated from numerical data because computing Koopman operator exactly is impossible unless one knows the exact form of underlying dynamical systems (and it is often almost impossible even if one knows the exact form of dynamics). Such approximation is executed by methods reviewed in Section 1.3.2.

1.3.2 Numerical approximation of KMD

There have been proposed several methodologies to approximate the spectra and corresponding modes of the Koopman operator.

If one has a sufficient amount of trajectories of observables generated by the dynamical system of interest, Fourier/Laplace average (Budišić et al., 2012; Mauroy and Mezić, 2012; Mauroy et al., 2013; Mohr, 2014), sometimes termed generalized Laplace analysis (GLA), can be utilized. This method is theoretically sound, but the usage is limited because computation must be performed for every value of eigenvalue candidates.

Ulam’s method (see e.g., Bollt and Santitissadeekorn (2013); Froyland and González-Tokman (2014); Froyland et al. (2014)), sometimes referred to as Ulam–Galerkin projection, can also be used to approximate the Koopman operator. Ulam’s method had been originally used as a method to approximate the Perron–Frobenius operator, but it can be used to approximate its adjoint as pointed out by researchers such as Froyland and González-Tokman (2014) Williams et al. (2015a) and Klus et al. (2016). Ulam’s method utilizes the partition of the state space into small spatial bins and thus is very intuitive. However, the computational load tends to be heavy if the state space is high-dimensional.

There is another line of research by Berry et al. (2015); Giannakis et al. (2015); Giannakis (2017); Giannakis et al. (2017), where they approximate the spectra of the Koopman operator using a smooth orthonormal basis determined by the diffusion map algorithm (Coifman and Lafon, 2006) with the variable bandwidth diffusion kernel (Berry and Harlim, 2016). Also, Shnitzer et al. (2017) utilizes diffusion map to construct a Kalman filter whose state is defined by eigenfunctions of the backward Fokker-Planck operator of Langevin dynamics, which is strongly related to the stochastic Koopman operator.

We end this section by mentioning a method referred to as dynamic mode decomposition (DMD). DMD was first invented in the area of fluid mechanics (Schmid and Sesterhenn, 2008), and its connection to KMD was pointed out by Rowley et al. (2009). Since then, DMD has been utilized as a realization of KMD in a vast amount of applications. Since DMD is the main concern of this dissertation, its algorithmic details and theoretical connection to KMD are explained in Section 1.4.

1.4 Dynamic mode decomposition (DMD)

DMD was first introduced as a method to extract dynamically relevant structures of fluid flow from numerical data (Schmid and Sesterhenn, 2008; Schmid, 2010). Afterward, the connection to the Koopman operator theory was first pointed out by Rowley et al. (2009). In this section, we firstly describe the core functionality of the algorithm of DMD and then explain the connection between KMD and DMD. Afterward, several variants of DMD including popular implementations often used in recent studies are introduced.

1.4.1 Algorithm of DMD

Suppose we have time-series $(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_m)$ with m snapshots, where $\mathbf{y}_t \in \mathbb{C}^n$ denotes the t -th snapshot, and consider the following two matrices build from them:

$$\begin{aligned} \mathbf{Y}_0 &= \begin{bmatrix} \mathbf{y}_0 & \cdots & \mathbf{y}_{m-1} \end{bmatrix} \in \mathbb{C}^{n \times m} \quad \text{and} \\ \mathbf{Y}_1 &= \begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_m \end{bmatrix} \in \mathbb{C}^{n \times m}. \end{aligned} \quad (1.14)$$

The fundamental functionality of DMD is just an eigendecomposition of the coefficient matrix of a linear model on time-series as follows.⁶

Algorithm 1.1 (DMD (Rowley et al., 2009; Schmid, 2010; Tu et al., 2014b)).

1. Build a pair of data matrices $(\mathbf{Y}_0, \mathbf{Y}_1)$ as in Eq. (1.14).
2. Compute eigenvalues λ , eigenvectors \mathbf{w} and left-eigenvectors \mathbf{z} of matrix $\mathbf{A} = \mathbf{Y}_1 \mathbf{Y}_0^\dagger$.
3. Normalize \mathbf{w}_i and \mathbf{z}_i so that $\mathbf{w}_i^\dagger \mathbf{z}_i = \delta_{i'i}$ ($\delta_{i'i}$ is 1 if $i' = i$ and 0 otherwise), for $i, i' = 1, \dots, n$.
4. Return λ_i , \mathbf{w}_i , and \mathbf{z}_i , for $i = 1, \dots, n$, with regard to nonzero λ_i .

Remark 1.2. Note that matrix \mathbf{A} is a least-squares estimator of a linear model between \mathbf{Y}_0 and \mathbf{Y}_1 , i.e.,

$$\mathbf{A} = \arg \min_{\mathbf{A}'} \|\mathbf{Y}_1 - \mathbf{A}' \mathbf{Y}_0\|_2^2 = \mathbf{Y}_1 \mathbf{Y}_0^\dagger. \quad (1.15)$$

⁶The original definition of DMD (Schmid and Sesterhenn, 2008; Rowley et al., 2009; Schmid, 2010) is based on the projection of the coefficient matrix to Krylov subspace and computation of empirical Ritz values/vectors to avoid explicit storing and eigendecomposition of a large matrix, but we introduced only the core functionality of DMD here.

If the multiplicities of the nonzero eigenvalues of \mathbf{A} are one (i.e., all the nonzero eigenvalues are distinct), the original snapshots can be decomposed as

$$\mathbf{y}_t = \sum_{i=1}^n \mathbf{w}_i \varphi_{i,t}, \quad (1.16)$$

and

$$\mathbf{y}_{t+1} = \sum_{i=1}^n \mathbf{w}_i \varphi_{i,t+1} \approx \sum_{i=1}^r \lambda_i \mathbf{w}_i \varphi_{i,t}, \quad (1.17)$$

where $\varphi_{i,t} = \mathbf{z}_i^H \mathbf{y}_t$, and the approximation (\approx) is in the least-squares sense. In this way, we can consider an approximation expression

$$\mathbf{y}_t \approx \sum_{i=1}^n \lambda_i^t \mathbf{w}_i \varphi_{i,0}, \quad (1.18)$$

which resembles KMD, Eq. (1.13). Vector \mathbf{w} is referred to as *dynamic modes*.

We derived the decomposition conducted by DMD as an approximation form in Eqs. (1.17) and (1.18), but if \mathbf{Y}_0 and \mathbf{Y}_1 are *linearly consistent* (Tu et al., 2014b), i.e., $\text{Ker}(\mathbf{Y}_0) \in \text{Ker}(\mathbf{Y}_1)$, Eqs. (1.17) and (1.18) hold exactly. Note that this condition of linear consistency is less restrictive than the linear system condition, i.e., $\mathbf{y}_{t+1} = \mathbf{A}\mathbf{y}_t$; in other words, time-series can be (exactly) decomposed by DMD if sufficient data are provided even if it is not generated by a linear system.

1.4.2 Connection between KMD and DMD

Actually, the output of DMD asymptotically coincides with the quantities that appear in KMD under several conditions. While there would be multiple ways to specify those conditions, we utilize the notion of Koopman invariant subspaces.

Definition 1.3 (Koopman invariant subspace). Consider G such that $G \subset \mathcal{G}$, where \mathcal{G} is the space of observables. G is called *Koopman invariant subspace* if

$$\mathcal{K}g \in G, \quad \forall g \in G. \quad (1.19)$$

Suppose that there exists a Koopman invariant subspace G for the dynamical system and the space of observables of interest. Let us consider the restriction of \mathcal{K} to G and denote it by K . If G is finite-dimensional, then K

also becomes a finite-dimensional linear operator. Here, suppose that we have a set of observables $\{g_1, \dots, g_n\}$ ($n < \infty$) that spans G , and let \mathbf{K} be the representation of K with regard to $\{g_1, \dots, g_n\}$, i.e.,

$$\begin{bmatrix} K g_1 & \cdots & K g_n \end{bmatrix}^\top = \mathbf{K} \mathbf{g}, \quad (1.20)$$

where $\mathbf{g} = [g_1 \ \cdots \ g_n]^\top$. Moreover, let φ be the eigenfunction of K corresponding to an eigenvalue λ . Then, φ with regard to $\{g_1, \dots, g_n\}$ is expressed as

$$\varphi(\mathbf{x}) = \mathbf{z}^\mathbf{H} \mathbf{g}(\mathbf{x}), \quad (1.21)$$

where \mathbf{z} is the left-eigenvector of \mathbf{K} corresponding to eigenvalue λ , because

$$K(\mathbf{z}^\mathbf{H} \mathbf{g}(\mathbf{x})) = \mathbf{z}^\mathbf{H} \mathbf{K} \mathbf{g}(\mathbf{x}) = \lambda \mathbf{z}^\mathbf{H} \mathbf{g}(\mathbf{x}).$$

Let \mathbf{w}_i and \mathbf{z}_i respectively be the right- and the left-eigenvector of \mathbf{K} corresponding to an eigenvalue λ_i for $i = 1, \dots, n$. In the sequel, without loss of generality, we assume that \mathbf{w} and \mathbf{z} are normalized so that $\mathbf{w}_i^\mathbf{H} \mathbf{z}_i = \delta_{i' i}$.

Assumption 1.1 (Distinct eigenvalues). All the nonzero eigenvalues of K are distinct, i.e., their multiplicities are one.

Since we assumed that the eigenvalues of the Koopman operator are distinct (Assumption 1.1), decomposition of observables as in Eq. (1.13) is obtained as follows. First, any values of \mathbf{g} can be expressed by

$$\mathbf{g}(\mathbf{x}) = \sum_{i=1}^n \mathbf{w}_i \mathbf{z}_i^\mathbf{H} \mathbf{g}(\mathbf{x}) = \sum_{i=1}^n \mathbf{w}_i \varphi_i(\mathbf{x}), \quad (1.22)$$

where φ_i is the eigenfunction of K corresponding to eigenvalue λ_i . Applying K on both sides of Eq. (1.22) repeatedly starting at $\mathbf{x} = \mathbf{x}_0$, we obtain the modal decomposition of the values of the observables, i.e.,

$$\mathbf{g}(\mathbf{x}_t) = \sum_{i=1}^n \lambda_i^t \mathbf{w}_i \varphi_i(\mathbf{x}_0). \quad (1.23)$$

The convergence of Algorithm 1.1 in the large sample limit can be shown with the assumption that the time-average asymptotically coincides with the space-average of a measurable function on the state space (Assumption 1.2), and also with the assumption that data generated from a Koopman invariant

subspace (Assumption 1.3). Note that Assumption 1.2 can be originally characterized by the notion of invariant measure of a measure-preserving system or the notion of physical measure of a dissipative system. We omit their details here, but in the context of DMD, see e.g., [Arbabi and Mezić \(2017\)](#).

Assumption 1.2. The time average of a measurable function $\phi : \mathcal{M} \rightarrow \mathbb{C}$ converges to its space average, i.e.,

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{j=0}^{m-1} \phi(\mathbf{x}_j) = \mathbb{E}_{\mathcal{M}}[\phi(\mathbf{x})] = \int_{\mathcal{M}} \phi(\mathbf{x}) d\mu_{\mathcal{M}},$$

for almost all $\mathbf{x}_0 \in \mathcal{M}$.

Assumption 1.3. The snapshots in time-series $(\mathbf{y}_0, \dots, \mathbf{y}_m)$ are generated with a set of observables that spans a Koopman invariant subspace, i.e.,

$$\mathbf{y}_t = \begin{bmatrix} g_1(\mathbf{x}_t) & \cdots & g_n(\mathbf{x}_t) \end{bmatrix}^{\top}, \quad \text{for } t = 0, \dots, m, \quad (1.24)$$

where $\{g_1, \dots, g_n\}$ spans a Koopman invariant subspace G .

Theorem 1.1. Suppose Assumptions 1.1, 1.2, and 1.3 hold. If all the modes are sufficiently excited in the data (i.e., $\text{rank}(\mathbf{Y}_0) = \dim(G)$) and all the nonzero eigenvalues of $\mathbf{A} = \mathbf{Y}_1 \mathbf{Y}_0^{\dagger}$ are distinct, then the dynamic modes calculated by Algorithm 1.1 converge to the eigenvectors of \mathbf{K} corresponding to its nonzero eigenvalues in $m \rightarrow \infty$ with probability one.

Proof. Taking the inner product of both sides of Eq. (1.20) with \mathbf{g} , we have

$$\mathbf{K} \mathbf{G}_0 = \mathbf{G}_1, \quad \mathbf{G}_0 = \mathbb{E}_{\mathcal{M}} [\mathbf{g} \mathbf{g}^{\text{H}}], \quad \mathbf{G}_1 = \mathbb{E}_{\mathcal{M}} [(\mathbf{g} \circ \mathbf{f}) \mathbf{g}^{\text{H}}],$$

and thus the minimum-norm solution for \mathbf{K} is given as $\mathbf{K} = \mathbf{G}_1 \mathbf{G}_0^{\dagger}$. In contrast, from the definition of \mathbf{A} ,

$$\mathbf{A} = \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_0^{\dagger}, \quad \hat{\mathbf{G}}_0 = \frac{1}{m} \mathbf{Y}_0 \mathbf{Y}_0^{\text{H}}, \quad \hat{\mathbf{G}}_1 = \frac{1}{m} \mathbf{Y}_1 \mathbf{Y}_0^{\text{H}},$$

and by Assumption 1.2, empirical matrices $\hat{\mathbf{G}}_0$ and $\hat{\mathbf{G}}_1$ converge to \mathbf{G}_0 and \mathbf{G}_1 , respectively, in $m \rightarrow \infty$ with probability one. Moreover, because $\text{rank}(\mathbf{Y}_0) = \dim(G)$, the rank of $\hat{\mathbf{G}}_0$ is always $\dim(G)$ and thus $\hat{\mathbf{G}}_0^{\dagger}$ converges to \mathbf{G}_0^{\dagger} ([Rakočević, 1997](#)). Further, because Algorithm 1.1 returns the eigenvectors corresponding to all the nonzero eigenvalues of \mathbf{A} , and because of Assumption 1.1, the outputs of Algorithm 1.1 are continuous with respect to \mathbf{A} . Therefore,

the dynamic modes converge to the eigenvectors of \mathbf{K} corresponding to its nonzero eigenvalues with probability one. \square

Remark 1.3. We have shown the convergence utilizing the assumption of ergodicity (Assumption 1.2). One can also prove the convergence (in probability) using the law of large numbers if one assumes the snapshots in \mathbf{Y}_0 are independently sampled from $\mu_{\mathcal{M}}$. In fact, an algorithm defined by Tu et al. (2014b) does not require the sequential sampling as in Eq. (1.14) as long as the corresponding columns of the data matrices are sampled with a fixed temporal interval. This definition of DMD must rely on the law of large numbers to show the convergence.

Remark 1.4. The way of constructing a Koopman invariant subspace is not discussed here. Several solutions for approximating it have been proposed, which are introduced in Section 1.4.3. Also, note that the approximation of a Koopman invariant subspace is one of the central issues of this dissertation (Chapter 5).

We have shown the convergence of DMD to KMD using the assumption of data generated from observables that span a Koopman invariant subspace. However, the connection between KMD and DMD can be shown from slightly different perspectives. In fact, a theory by Tu et al. (2014b) depends on the assumption that the observables in hand are in the space spanned by eigenfunctions of the Koopman operator, which is similar to the assumption mentioned earlier, Eq. (1.8). Also, Arbabi and Mezić (2017) showed the convergence of DMD based on the Hankel matrix build with an observable *contained in* (not spanning) a Koopman invariant subspace.

1.4.3 Variants of DMD

The concept of DMD was introduced in Section 1.4.1, but Algorithm 1.1 describes only the core functionality of DMD. In this section, we introduce a popular working algorithm of DMD, referred to as exact DMD (Tu et al., 2014b). Moreover, several variants and extensions of DMD are explained.

Exact DMD

When DMD was first discussed with that name in the community of fluid dynamics (Schmid and Sesterhenn, 2008; Rowley et al., 2009), it was implemented by Arnoldi algorithm (Arnoldi, 1951). There, DMD's matrix, \mathbf{A} in Eq. (1.15), was projected to a data-driven Krylov space, and the eigendecomposition was performed in that space. Afterward, Schmid (2010) suggested that projecting \mathbf{A} to the principal space of \mathbf{Y}_0 using its singular value decomposition (SVD) was more computationally stable. These projections to a certain space (whose dimensionality is smaller than that of the original data) enable us to avoid computing eigendecomposition in a high-dimensional space. However, Tu et al. (2014b) pointed out the back-projection of the output of the eigendecomposition was not necessarily appropriate in those algorithms and proposed a modified algorithm, which they termed *exact DMD*.

Algorithm 1.2 (Exact DMD (Tu et al., 2014b)).

1. Build a pair of data matrices $(\mathbf{Y}_0, \mathbf{Y}_1)$ as in Eq. (1.14).
2. Compute the compact SVD as $\mathbf{Y}_0 = \mathbf{U}_r \mathbf{S}_r \mathbf{V}_r^H$ with $\mathbf{U}_r \in \mathbb{C}^{n \times r}$, $\mathbf{S}_r \in \mathbb{C}^{r \times r}$ and $\mathbf{V}_r \in \mathbb{C}^{m \times r}$, where $r = \text{rank}(\mathbf{Y}_0)$.
3. Define matrix $\tilde{\mathbf{A}} = \mathbf{U}_r^H \mathbf{Y}_1 \mathbf{V}_r \mathbf{S}_r^{-1}$.
4. Compute eigenvalues λ , eigenvectors $\tilde{\mathbf{w}}$ and left-eigenvectors $\tilde{\mathbf{z}}$ of $\tilde{\mathbf{A}}$, for $\lambda \neq 0$.
5. Compute $\mathbf{w}_i = \lambda_i^{-1} \mathbf{Y}_1 \mathbf{V}_r \mathbf{S}_r^{-1} \tilde{\mathbf{w}}_i$ and $\mathbf{z}_i = \mathbf{U}_r^H \tilde{\mathbf{z}}_i$, for $i = 1, \dots, r$.
6. Normalize $\{\mathbf{w}_i\}$ and $\{\mathbf{z}_i\}$ so that $\mathbf{w}_i^H \mathbf{z}_i = \delta_{ii}$.
7. Return λ_i , \mathbf{w}_i , and \mathbf{z}_i , for $i = 1, \dots, r$.

For $\lambda \neq 0$, λ , \mathbf{w} , and \mathbf{z} coincide with the eigenvalue, eigenvector, and left-eigenvector of matrix $\mathbf{A} = \mathbf{Y}_1 \mathbf{Y}_0^\dagger$, respectively (Tu et al., 2014b). In the remainder of this dissertation, we use this exact DMD as the standard implementation of DMD.

Extensions for approximating a Koopman invariant subspace

As stated in Section 1.4.2, the key point of DMD as a numerical realization of KMD lies in preparing a set of observables that spans a subspace invariant

to \mathcal{K} . Several researchers have worked on this issue; Williams et al. (2015a) proposed a method using a user-defined dictionary of observables to adopt DMD to highly nonlinear systems, which is called extended DMD, and Klus et al. (2016) and Korda and Mezić (2017) showed the convergence of extended DMD to KMD. Moreover, Brunton et al. (2016b) utilized an identification technique based on a sparse regression (Brunton et al., 2016c) to identify the dynamic-specific observables to be used. In addition, Kawahara (2016) defined the Koopman analysis for observables in reproducing kernel Hilbert spaces to build a theory of DMD based on the reproducing kernels. DMD using kernel trick was also mentioned by Williams et al. (2015b), but they did not manifest the theory based on the notion of reproducing kernel Hilbert spaces.

Another option, especially for deterministic systems, is to use delay coordinates, i.e., stacking observations at neighboring timestamps in each column of the data matrices. In general, a Krylov-like sequence of observables $\{g, \mathcal{K}g, \mathcal{K}^2g, \dots\}$ rapidly becomes almost linearly dependent, and thus can be used to obtain a subspace that is *approximately* invariant to \mathcal{K} . Based on the delayed measurements, we obtain a data matrix as a Hankel matrix. The use of delay coordinates for DMD was first discussed by Tu et al. (2014b), and Brunton et al. (2017) mentioned DMD based on Hankel matrices, referring to the well-known Taken's theorem (Takens, 1981). Susuki and Mezić (2015) and Raak et al. (2016) defined an approximation of the Koopman analysis using Prony's method, which also uses Hankel matrices. Arbabi and Mezić (2017) showed the convergence of DMD on Hankel matrices build with an observable contained in a Koopman invariant subspace. However, since delay coordinates with a linear monomial cannot span a Koopman invariant subspace of nonlinear systems exactly, one should use a combination of the nonlinear observables and the delay coordinates.

Extensions for noisy data

When applying DMD to datasets obtained by experiments, one must be aware of observation noise in the dataset. There are several styles to make DMD adapted to noisy data. Chen et al. (2012) suggested to formulate DMD as an optimization problem (optimized DMD) and solved it with a derivative-free optimization method. Guéniat et al. (2015) extended the optimized DMD to nonuniformly-sampled datasets. Moreover, there is a series of researches

on low-rank approximation of DMD's matrix (Wynn et al., 2013; Dicle et al., 2016; Héas and Herzet, 2017), in which spurious dynamic modes due to observation noise can be automatically eliminated. Such methods put in a low-rank regularization/constraint in their optimization problems. Jovanović et al. (2014) suggested a slightly different approach, sparsity-promoting DMD, where they ran a postprocessing to reduce the number of dynamic modes based on lasso.

Now remember that the matrix whose eigendecomposition is computed by original DMD, \mathbf{A} in Eq. (1.15), is the result of (ordinary) least-squares between \mathbf{Y}_0 and \mathbf{Y}_1 . Hence, in this formulation, only noises in the dependent variable, i.e., \mathbf{Y}_1 , are considered, and noises in \mathbf{Y}_0 are not taken into account. This asymmetry causes a problem when dataset is contaminated with observation noise. Dawson et al. (2016) and Hemati et al. (2017) proposed total-least-squares DMD, in which noises in both \mathbf{Y}_0 and \mathbf{Y}_1 are considered explicitly.

In addition, Duke et al. (2012) and Pan et al. (2015) conducted error analyses of DMD algorithms with different implementations, which are helpful to choose an implementation according to each dataset.

Extensions for non-autonomous systems

Most of DMD-related algorithms premise that the system behind data is autonomous, i.e., f is time-invariant and no external signal is concerned. However, one may be interested in analyzing time-series generated by non-autonomous systems including controlled systems. In fact, there have been proposed DMD for controlled systems (Proctor et al., 2016; Annoni et al., 2016). Moreover, Mezić and Surana (2016) proposed a computation method for non-autonomous periodic system, and Maćešić et al. (2017) considered switching or eigenvalue-changing systems. In addition, there is an attempt to combine DMD with clustering of time-series (Narasingam and Kwon, 2017).

Extensions for efficient computation

For efficient computation of DMD on streaming data, Hemati et al. (2014) proposed an online algorithm. For memory and processing efficiency, Brunton et al. (2015) and Erichson et al. (2016) suggested a combination of DMD

and compressed sensing. Also, [Klus and Schütte \(2016\)](#) proposed to use the tensor-train decomposition for speeding up computation of the pseudo-inverse. Moreover, there is a thread of studies on randomized DMD ([Erichson and Donovan, 2016](#); [Bistrrian and Navon, 2017](#); [Erichson et al., 2017](#)). In addition, [Le Clainche and Vega \(2017\)](#) mentioned higher-order DMD, and [Ohmichi \(2017\)](#) proposed preconditioned DMD. Also, [Kutz et al. \(2015, 2016b\)](#) proposed multi-resolution DMD ([Kutz et al., 2015, 2016b](#)), and [Tu et al. \(2014a\)](#) suggested a method to compute dynamic modes even with sub-Nyquist rate ([Tu et al., 2014a](#)).

Connection to other methods

The correspondences between DMD and other numerical methods, such as time-lagged independent component analysis (TICA), variational approach of conformation dynamics (VAC), and resolvent analysis, have been pointed out by several researchers ([Sharma et al., 2016](#); [Klus et al., 2017](#); [Towne et al., 2017](#); [Wu et al., 2017](#)).

1.5 Summary of remaining chapters

Although there have been proposed many algorithmic variants of DMD as introduced in Section [1.4.3](#), the assumptions they explicitly and implicitly make are often too restrictive for applying them to practical time-series datasets. For example, no existing method is appropriate for datasets that are generated from a random dynamical system and contaminated with observation noise, resulting in incorrect approximation of the spectra of the Koopman operator. The central motivation in this dissertation lies in broadening the scope of dynamical systems / time-series data that can be analyzed based on the theory of the Koopman operator. To this end, we developed several techniques that are based on the methodologies often utilized in machine learning researches, such as nonnegative decomposition, Bayesian modeling, and neural networks. In the remaining chapters of this dissertation, we introduce four distinct topics, as summarized in the following.

In Chapter [2](#), we suggest to impose nonnegativity and sparsity on the estimated dynamic modes. This is motivated by application of DMD to

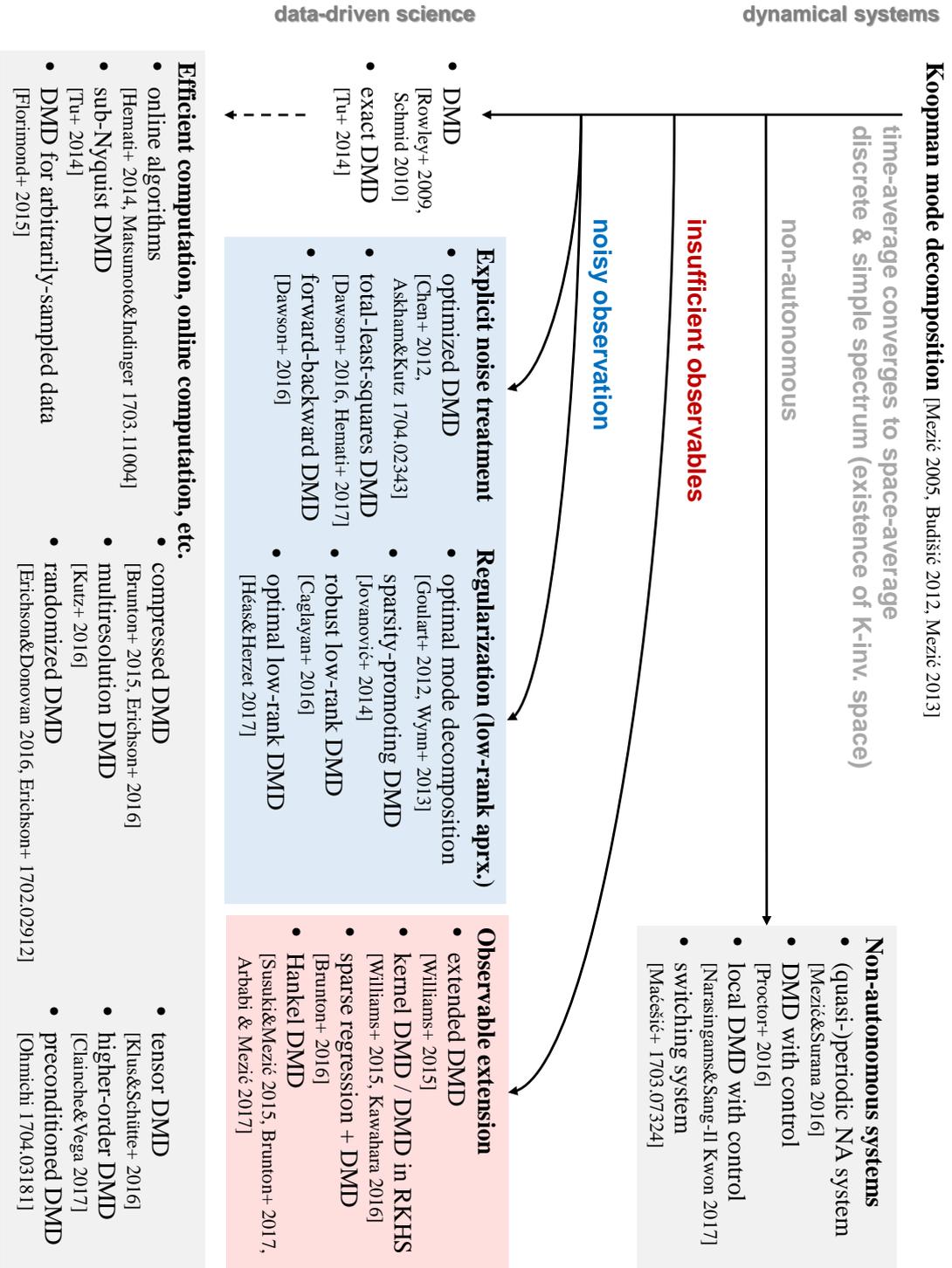


FIGURE 1.2: Diagram summarizing variants of DMD.

video streams, where signals are inherently nonnegative, and thus nonnegative dynamic modes can be inspected more easily. We achieve such constrained/regularized DMD by reformulating it as an optimization problem.

In Chapter 3, we propose a probabilistic model corresponding to DMD and develop a procedure for Bayesian posterior inference on dynamic modes, values of eigenfunctions, and eigenvalues. This probabilistic/Bayesian formulation enables us not only to conduct the posterior inference, but also to consider DMD's extension in the unified manner of Bayesian. We introduce an example of such extension, in which a sparsity-promoting prior is put on dynamic modes to determine the number of dynamic modes automatically from data.

In Chapter 4, we expand the scope of time-series data which DMD can deal with. In particular, we focus on datasets generated from a random dynamical system and contaminated with observation noise. Existence of either of the process noise and the observation noise does not matter in existing DMD algorithms. However, if both of the process and observation noises are present, no existing method can correctly approximate the spectra of the stochastic Koopman operator. We therefore propose a method that considers process and observation noises explicitly.

In Chapter 5, we tackle the fundamental but most nontrivial assumption of DMD, i.e., the data generated from a Koopman invariant subspace (Assumption 1.3). Our idea is to learn a set of observables that spans a Koopman invariant subspace from data. We also show the implementation of this idea using neural networks.

This dissertation is concluded in Chapter 6 with discussion on remaining technical challenges to be elaborated.

Chapter 2

Sparse nonnegative DMD

2.1 Introduction

Image and video processing is one of the most active areas of signal processing research. DMD can be a useful tool for video processing because it can extract a set of modes from time-series according to their dynamical characteristics, i.e., temporal decay rates and frequencies. In fact, [Kutz et al. \(2015\)](#) utilized DMD for background/foreground separation of video streams by extracting static low-frequency dynamic modes as the background. In addition, [Erichson et al. \(2016\)](#) proposed the use of DMD with compressed sensing for the fast background separation of video streams.

For video processing, DMD becomes more attractive if it can compute dynamic modes that take *nonnegative* values because such modes can be easily inspected and understood due to the inherent nonnegativity of video data. Moreover, obtaining *sparse* part-based dynamic modes is important for a meaningful representation of video streams. While there have been proposed many algorithmic variants of DMD that impose a low-rank constraint or approximation ([Chen et al., 2012](#); [Wynn et al., 2013](#); [Jovanović et al., 2014](#); [Dicle et al., 2016](#); [Héas and Herzet, 2017](#)), none of them can explicitly impose other constraints and/or regularizations such as nonnegativity and sparsity on the structures of the estimated dynamic modes.

In this chapter, we propose to reformulate DMD as a block-multiconvex optimization problem so as to impose the nonnegativity constraint and the sparsity regularization on dynamic modes. Using the proposed sparse nonnegative version of DMD, we can decompose a video stream into part-based modes as shown in Section 2.4. This is analogous to the well-known results of the part-based image decomposition by nonnegative matrix factorization

(NMF) (Lee and Seung, 1999), but an important difference of DMD from NMF is that DMD extracts modes related to the *dynamics* behind data.

The remainder of this chapter is organized as follows. In Sections 2.2 and 2.3, the details of the proposed method are described. In Section 2.4, the experimental results with the proposed method are shown. A summary of this chapter is provided in Section 2.5.

2.2 Reformulation as a block-multiconvex problem

We reformulate DMD as a block-multiconvex optimization problem, which enables us to impose constraints and regularizations directly on the structures of the estimated dynamic modes. To this end, we use the polar-coordinate expression for dynamic modes $\mathbf{w} \in \mathbb{C}^n$ and eigenvalues $\lambda \in \mathbb{C}$, i.e.,

$$\begin{aligned} [\mathbf{w}_i]_d &= q_{i,d} \exp(j\theta_{i,d}), \quad q_{i,d}, \theta_{i,d} \in \mathbb{R}, \\ \lambda_i &= r_i \exp(j\phi_i), \quad r_i, \phi_i \in \mathbb{R}, \end{aligned} \quad (2.1)$$

where j is the imaginary unit, and $[\mathbf{w}_i]_d$ denotes the d -th element of \mathbf{w}_i . Using the polar-coordinate expression, we can perform an optimization for real numbers. Moreover, suppose that we prepare n' dynamic modes for approximating the original dataset. Then, in order to construct an optimization problem corresponding to DMD, we use the following matrix-form representation:

$$\begin{aligned} \mathbf{W} &= \mathbf{Q} \odot \mathbf{\Theta}, \quad [\mathbf{Q}]_{i,d} = q_{i,d}, \quad [\mathbf{\Theta}]_{i,d} = \exp(j\theta_{i,d}), \\ \mathbf{\Lambda}_v &= \mathbf{R} \odot \mathbf{\Phi}, \quad [\mathbf{R}]_{i,t} = r_i^t, \quad [\mathbf{\Phi}]_{i,t} = \exp(jt\phi_i), \\ &\text{for } i = 1, \dots, n', \quad d = 1, \dots, n, \quad t = 0, \dots, m, \end{aligned} \quad (2.2)$$

where \odot denotes the element-wise multiplication. Here,

$$\begin{aligned} \mathbf{W} &= \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_{n'} \end{bmatrix} \in \mathbb{C}^{n \times n'}, \quad \text{and} \\ \mathbf{\Lambda}_v &= \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^m \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{n'} & \lambda_{n'}^2 & \cdots & \lambda_{n'}^m \end{bmatrix}. \end{aligned}$$

Note that $\mathbf{\Lambda}_v$ is called a Vandermonde matrix because of its structure.

Now remember that the decomposition of DMD should take the form of Eq. (1.18). Consequently, the objective function to be minimized here is

$$\begin{aligned}\ell(q, \theta, r, \phi; \mathbf{Y}) &= \frac{1}{2} \|\mathbf{Y} - \mathbf{W}\Lambda_v\|_F^2 \\ &= \frac{1}{2} \text{tr}(\mathbf{Y}^H \mathbf{Y} - \mathbf{Y}^H \mathbf{W}\Lambda_v - \Lambda_v^H \mathbf{W}^H \mathbf{Y} + \Lambda_v^H \mathbf{W}^H \mathbf{W}\Lambda_v),\end{aligned}\quad (2.3)$$

where \mathbf{Y} is the full-data matrix defined as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_0 & \cdots & \mathbf{y}_m \end{bmatrix} \in \mathbb{C}^{n \times (m+1)}.\quad (2.4)$$

Since ℓ is a block-multiconvex function (Xu and Yin, 2013) with regard to q , θ , r , and ϕ , its local minimum is easily obtained by a block coordinate descent. The pseudocode is shown in Algorithm 2.1. The update of each block (at Lines 3–6) can be performed with any solver, and we used the L-BFGS method (Liu and Nocedal, 1989) with gradients of ℓ as follows:

$$\begin{aligned}\frac{\partial h}{\partial q_{i,d}} &= -\text{Re} \left[\mathbf{D}\Lambda_v^H [\bar{\Theta}]_{i,d} \right]_{i,d}, \\ \frac{\partial h}{\partial \theta_{i,d}} &= -\text{Im} \left[\mathbf{D}\Lambda_v^H [\bar{\Theta}]_{i,d} \right]_{i,d} [\mathbf{Q}]_{i,d}, \\ \frac{\partial h}{\partial r_i} &= -\text{Re} \left[\text{tr}(\mathbf{D}^T \bar{\mathbf{W}} (\mathbf{O}_i \odot \bar{\Phi})) \right], \\ \frac{\partial h}{\partial \phi_i} &= -\text{Im} \left[\text{tr}(\mathbf{D}^T \bar{\mathbf{W}} (\mathbf{R} \odot \mathbf{M}_i \odot \bar{\Phi})) \right],\end{aligned}$$

where $\mathbf{D} = \mathbf{Y} - \mathbf{W}\Lambda_v$. \mathbf{O}_i and \mathbf{M}_i are matrices whose elements except those in the i -th rows are zero. That is,

$$\mathbf{O}_i = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 2r_i & \cdots & mr_i^{m-1} \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n' \times (m+1)},\quad (2.5)$$

Algorithm 2.1: DMD as a block-multiconvex problem

Data: full-data matrix defined in Eq. (2.4) and the number of modes n'
Result: dynamic modes q, θ and eigenvalues r, ϕ

- 1 $q_0, \theta_0, r_0, \phi_0 \leftarrow$ initialization using standard DMD;
- 2 **for** $k = 1, 2, \dots$ **do**
- 3 $q_k \leftarrow \arg \min_q \ell(q, \theta_{k-1}, r_{k-1}, \phi_{k-1}; \mathbf{Y});$
- 4 $\theta_k \leftarrow \arg \min_\theta \ell(q_k, \theta, r_{k-1}, \phi_{k-1}; \mathbf{Y});$
- 5 $r_k \leftarrow \arg \min_r \ell(q_k, \theta_k, r, \phi_{k-1}; \mathbf{Y});$
- 6 $\phi_k \leftarrow \arg \min_\phi \ell(q_k, \theta_k, r_k, \phi; \mathbf{Y});$
- 7 **if converge then return** $q_k, \theta_k, r_k, \phi_k;$
- 8 **end**

and

$$M_i = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & j & 2j & \cdots & mj \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n' \times (m+1)}. \quad (2.6)$$

2.3 Sparse nonnegative DMD

Based on the reformulation above, we can impose constraints and/or regularizations directly on variables $q, \theta, r,$ and ϕ . Our purpose is formulating *sparse nonnegative DMD* (SN-DMD) by imposing the nonnegativity constraint and the L1 regularization¹ on the dynamic modes, i.e., q is constrained to be nonnegative, θ is fixed to be zero, and a regularization term $\gamma|q|$ is introduced into the objective function. Formally, the new objective function $\hat{\ell}$ is given as

$$\hat{\ell}(q, r, \phi; \mathbf{Y}) = \ell(q, 0, r, \phi; \mathbf{Y}) + \gamma \sum_{i=1}^{n'} \sum_{d=1}^n |q_{i,d}| + I_{q \geq 0}(q), \quad (2.7)$$

¹It is known that the nonnegativity itself encourages the sparsity in the setting of NMF (Lee and Seung, 1999), but in our case, we need the explicit L1 regularization because of the presence of the imaginary part.

Algorithm 2.2: Sparse nonnegative DMD

Data: full-data matrix defined in Eq. (2.4) and the number of modes n'
Result: dynamic modes q and eigenvalues r, ϕ

```

1  $q_0, \theta_0, r_0, \phi_0 \leftarrow$  initialization using standard DMD;
2  $\theta_0 \leftarrow 0$ ;
3 for  $k = 1, 2, \dots$  do
4    $q_k \leftarrow \arg \min_q \hat{\ell}(q, r_{k-1}, \phi_{k-1}; \mathbf{Y})$ ;
5    $r_k \leftarrow \arg \min_r \hat{\ell}(q_k, r, \phi_{k-1}; \mathbf{Y})$ ;
6    $\phi_k \leftarrow \arg \min_\phi \hat{\ell}(q_k, r_k, \phi; \mathbf{Y})$ ;
7   if converge then return  $q_k, r_k, \phi_k$ ;
8 end

```

where ℓ is defined in Eq. (2.3), γ is a regularization parameter, and $I_{q \geq 0}(q)$ is an indicator function such that

$$I_{q \geq 0}(q) = \begin{cases} 0 & (q \geq 0) \\ +\infty & (q < 0). \end{cases}$$

The pseudocode for SN-DMD using the block coordinate descent is shown in Algorithm 2.2. To solve the update step at Line 4, we utilize the proximal Newton-type method with an L-BFGS Hessian approximation (Lee et al., 2014). The updates of the other quantities (at Lines 5–6) are computed in the same manner as in Algorithm 2.1.

Note that there is an interesting similarity between the formulation of SN-DMD and that of the complex NMF (Kameoka et al., 2009) as follows. First, let us rewrite the decomposition performed by DMD using the polar expression in Eq. (2.1), with the nonnegativity constraint on the dynamic modes ($q \geq 0, \theta = 0$):

$$[\mathbf{y}_t]_d = \sum_{i=1}^{n'} q_{i,d} r_i^t e^{jt\phi_i}, \quad q_{i,d} \geq 0.$$

On other hand, the formulation of the complex NMF is like

$$[\mathbf{y}_t]_d = \sum_{i=1}^{n'} h_{i,d} u_{t,i} e^{j\varphi_{d,t,i}}, \quad h_{i,d} \geq 0, \quad u_{t,i} \geq 0,$$

where h and u are some nonnegative bases. Comparing the above two formulations, one could notice that the proposed method is regarded as a variant of the complex NMF, with special structures $u_{t,i} = r_i^t$ (while the nonnegativity of u is lost) and $\varphi_{d,t,i} = t\phi_i$. Because the nonnegativity of u is lost in

SN-DMD, the explicit sparse regularization becomes more important than in the complex NMF.

2.4 Experiments

In this section, we show the results of two experiments: one is with a synthetic dataset and the other is with a real-world dataset.

2.4.1 Mode extraction from noisy data

To quantitatively investigate the performance of SN-DMD, we conducted an experiment with synthetic data as follows. First, we generated a sequence of noisy images $\{\mathbf{y}_t \in \mathbb{R}^{64 \times 64}\}$ by the following equations for $t = 0, \dots, 15$:

$$\mathbf{z}_t = 0.99^t \mathbf{w}_1 + 0.9^t \mathbf{w}_2, \quad \mathbf{y}_t = \mathbf{z}_t + \mathbf{e}_t,$$

where $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^{64 \times 64}$ were basis images shown in Figure 2.1f (\mathbf{w}_1 on the left and \mathbf{w}_2 on the right), and $\{\mathbf{e}_t\}$ was a noise sequence whose element was generated independently by a zero-mean Gaussian with variance 10^{-2} . Obviously the dynamic modes of the noise-free sequence $\{\mathbf{z}_t\}$ are \mathbf{w}_1 and \mathbf{w}_2 with eigenvalues 0.99 and 0.9, respectively. However, it is not trivial how accurate we can estimate these dynamic modes and eigenvalues from noisy sequence $\{\mathbf{y}_t\}$.

We input the noisy sequence to standard DMD (Algorithm 1.2, referred to as `standard`), total-least-squares DMD ((Dawson et al., 2016), referred to as `total-ls`), DMD as the block-multiconvex problem (Algorithm 2.1, referred to as `optimization`), and the proposed method (Algorithm 2.2) with $n' = 2$. The proposed method is applied with two settings: one with only the nonnegativity (referred to as `nonneg`) and the other with both the sparsity and the nonnegativity (referred to as `sparse-nonneg`). We set $\gamma = 1$ without any intensive search.

The estimation results are shown in Figure 2.1 and Table 2.2. The proposed method, `sparse-nonneg`, gives the best estimation among the methods listed above, in the sense that the estimated dynamic modes (Figure 2.1e) are not contaminated with much noise and the estimated eigenvalues are the most accurate. Comparing the results of `nonneg` and those of `sparse-nonneg`,

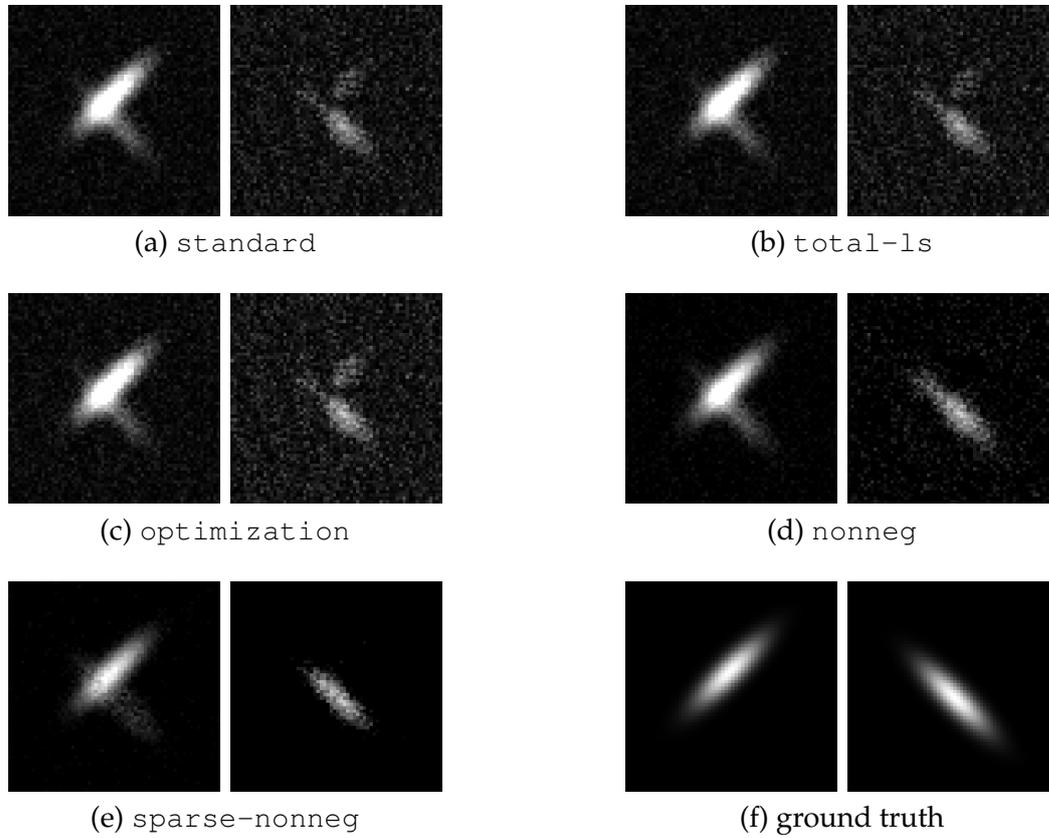


FIGURE 2.1: (a–e) Estimated and (f) true dynamic modes (best viewed on a display). The left of each panel corresponds to w_1 and the right corresponds to w_2 .

FIGURE 2.2: Estimated and true eigenvalues.

	λ_1	λ_2
standard	0.967	0.805
total-ls	0.969	0.825
optimization	0.967	0.805
nonneg	0.977	0.770
sparse-nonneg	0.996	0.891
ground truth	0.990	0.900

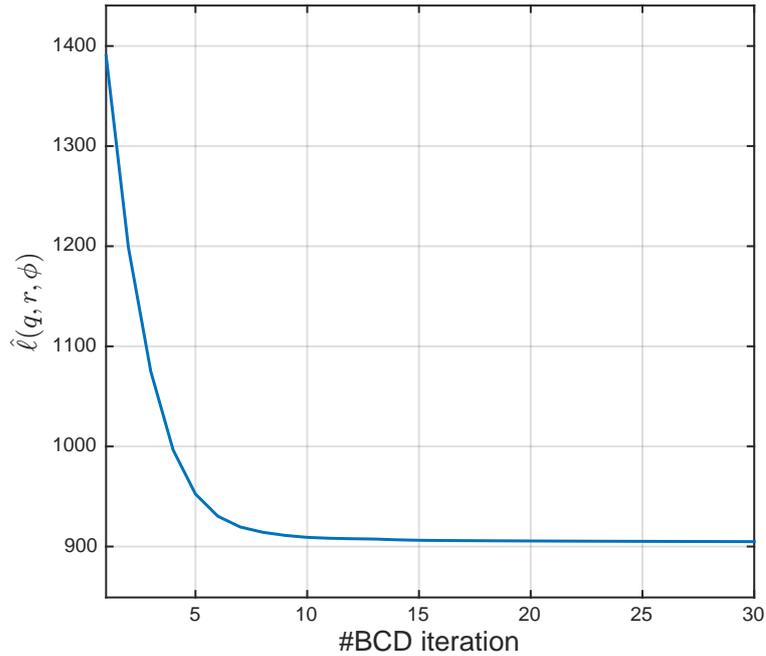


FIGURE 2.3: Values of the objective function (2.7) along the iterations of the block coordinate descent.

we can suppose that the sparsity plays an important role for estimation accuracy. Moreover, the values of the objective function (2.7) along the number of iterations in Algorithm 2.2 are plotted in Figure 2.3. We can confirm that the objective function rapidly decreased and converged within the first ~ 10 iterations in the block coordinate descent.

2.4.2 Extraction of part-based dynamic modes

We conducted another experiment using a real-world video dataset. If we apply DMD to a video stream, we can expect that it will be decomposed into multiple modes with different temporal frequencies, wherein the “zero-frequency” mode corresponds to the background and the other modes to the foreground. The background/foreground separation has been studied intensively so far. One of the popular solutions is the low-rank/sparse decomposition such as robust PCA (RPCA) (Candés et al., 2011). Moreover, Kutz et al. (Kutz et al., 2015) directly utilized DMD for this task. Here, note that the advantage of DMD is that it can distinguish not only the background and foreground, but also the components *within the foreground* according to their temporal frequencies. We address the task of foreground decomposition by SN-DMD.

We used AVSS AB Hard dataset ([IEEE International Conference on Advanced Video and Signal based Surveillance, 2007](#)), which was a surveillance video of a platform with moving people and trains. As a preprocessing, we trimmed the first 251 frames of the original sequence that were showing a title credit, resized each frame into 192×240 pixels and extracted the first 150 frames as a dataset. The example frames are shown in Figure 2.4. To this dataset, we applied RPCA by fast principal component pursuit ([Rodriguez and Wohlberg, 2013](#)), standard DMD (Algorithm 1.2), and SN-DMD (Algorithm 2.2). The DMDs are computed with the specified number of modes $n' = 7$, and the regularization parameter was set $\gamma = 1$ without any intensive search.

The results are shown in Figure 2.5. The time-averages of the low-rank and sparse components extracted by RPCA are shown in Figure 2.5a, and the dynamic modes and temporal frequencies² extracted by DMDs are shown in Figures 2.5b and 2.5c. From the viewpoint of background/foreground separation, the background was successfully extracted as the low-rank component by RPCA and the zero-frequency dynamic modes by DMDs. In contrast, the other modes (the sparse component by RPCA and the nonzero-frequency dynamic modes) correspond to non-static parts of the video, i.e. the foreground. As can be seen, peaks (white regions) of the foreground modes correspond to the regions where some moving objects passed. In Figure 2.5c, the nonzero-frequency dynamic modes by SN-DMD are *part-based* in the sense that, for example, 0.094Hz-mode represents a standing person whereas 0.283Hz-mode represents the train and some other people. On the other hand, in Figure 2.5b, the nonzero-frequency dynamic modes by standard DMD have less distinctive spatial features; e.g., the region corresponding to the train is activated both in 0.094Hz-mode and in 0.283Hz-mode.

2.5 Summary

In this chapter, we have reformulated DMD as an optimization problem and proposed sparse nonnegative DMD (SN-DMD), which directly imposes the sparsity regularization and the nonnegativity constraint on the structures of the estimated dynamic modes. In particular, we exemplified that SN-DMD

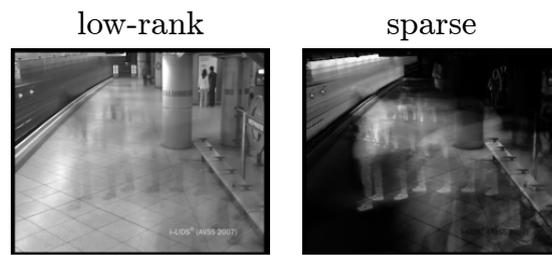
²Frequency $\hat{\ell}$ [Hz] is calculated by $f = \text{Im}(\log \lambda)/(2\pi\Delta t)$ with (discrete-time) eigenvalue λ and time interval Δt [sec] between frames.

could decompose a video with multiple moving objects into a set of *part-based* dynamic modes, which is analogous to the well-known results of NMF (Lee and Seung, 1999).

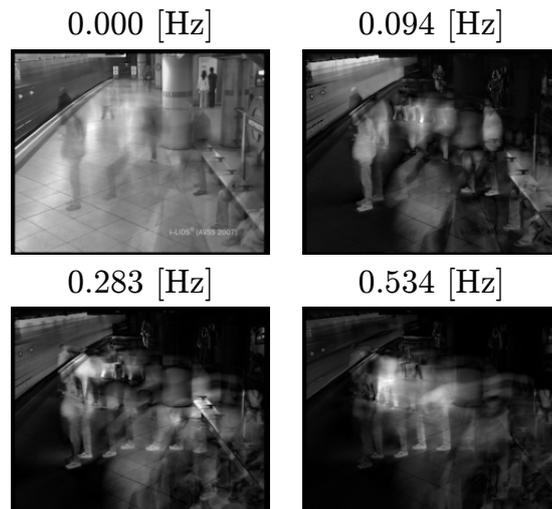
There remain several points to be elaborated or improved. One is the optimization procedure. We relied on the block coordinate descent whose blocks use (proximal) L-BFGS in this work, but more efficient methods can be applied according to the property of the loss function. Also, based on the proposed formulation, it is possible to impose more complex regularizations and constraints. For example, structured regularizations like fused-lasso and group-lasso will be effective for highly-structured data.



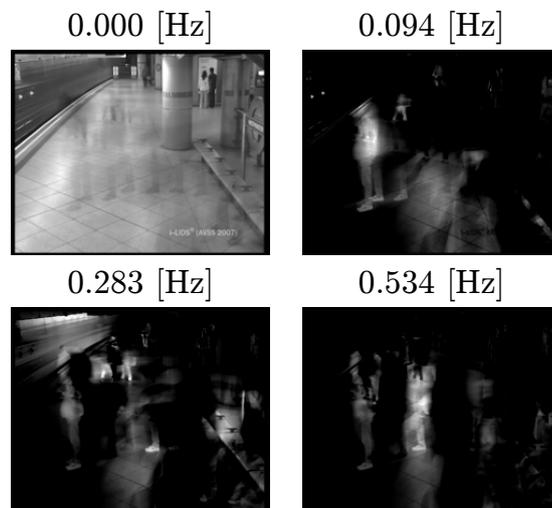
FIGURE 2.4: Summary of the first 180 frames (7.2 secs) of AVSS AB Hard dataset ([IEEE International Conference on Advanced Video and Signal based Surveillance, 2007](#)). Displayed every 15 frames, i.e., 0.6 secs.



(a) RPCA



(b) standard DMD



(c) SN-DMD

FIGURE 2.5: Modal decomposition of AVSS AB Hard dataset video (*IEEE International Conference on Advanced Video and Signal based Surveillance, 2007*). The upper panel of (a) and the upper-left panel of (b,c) correspond to the static part (i.e., background) of the video. The other panels correspond to the moving part (i.e., foreground).

Chapter 3

Probabilistic and Bayesian DMD

3.1 Introduction

Dynamic mode decomposition (DMD) (Rowley et al., 2009; Schmid, 2010; Kutz et al., 2016a) has been attracting attention in various fields of science and engineering as an approach for the above purpose that works without explicit knowledge on the governing equations. DMD generates modes that are directly related to the *dynamics* behind the data, and thus, these modes are a useful tool for the diagnostics of complex dynamic phenomena. Insofar, several algorithmic variants of DMD have been utilized according to given datasets or purposes. However, most of these variants are deterministic (i.e., lack corresponding probabilistic models), and thus it could be difficult to incorporate uncertainty in data into the analysis. Building a probabilistic model for DMD enables us to treat the data statistically and consider observation noises explicitly, as well as to enrich the DMD techniques systematically by modifying the involved probabilistic distributions.

In this chapter, we propose extensions of DMD, which provide a principled way to transfer the advantages of the probabilistic and Bayesian formulations into DMD. To this end, we first develop a probabilistic model corresponding to DMD (*probabilistic DMD*), whose maximum-likelihood estimator coincides with the solution of DMD. Then, we provide the formulation of *Bayesian DMD* and the corresponding Gibbs sampler for posterior inference. Also, we can consider variants of DMD within the unified Bayesian framework by modifying the priors of Bayesian DMD. In particular, we discuss the case of using a sparsity-promoting prior for dynamic modes, which allows us to automatically determine the number of modes in the light of data. Note that the ideas of these extensions of DMD are similar to those of probabilistic

PCA (Tipping and Bishop, 1999), Bayesian PCA (Bishop, 1999), and sparse PCA (Guan and Dy, 2009) as extensions of PCA.

The remainder of this chapter is organized as follows. The probabilistic model for DMD is described in Section 3.2, and based on that model, Bayesian DMD is introduced in Section 3.3. In Section 3.4, several related studies are briefly introduced. In Section 3.5, we show the experimental results with synthetic and real-world datasets. A summary of this chapter is provided in Section 3.6.

3.2 Probabilistic DMD

We develop a probabilistic model motivated by DMD, i.e., Eqs. (1.16) and (1.17). The maximum-likelihood estimator (MLE) of this model coincides with the solution of DMD in the no-noise limit. As will be described in Section 3.3, this probabilistic model forms the foundation for Bayesian DMD.

3.2.1 Generative model

Let $\mathbf{y}_{\ell,j} \in \mathbb{C}^n$ be the j -th column of \mathbf{Y}_ℓ in Eq. (1.14) plus observation noise, for $\ell = 0, 1$.¹ Following the relations in Eqs. (1.16) and (1.17), the probabilistic DMD model for such data can be given by

$$\begin{aligned} \mathbf{y}_{0,j} \mid \varphi_{1,j}, \dots, \varphi_{k,j} &\sim \mathcal{CN} \left(\sum_{i=1}^k \varphi_{i,j} \mathbf{w}_i, \sigma^2 \mathbf{I} \right), \\ \mathbf{y}_{1,j} \mid \varphi_{1,j}, \dots, \varphi_{k,j} &\sim \mathcal{CN} \left(\sum_{i=1}^k \lambda_i \varphi_{i,j} \mathbf{w}_i, \sigma^2 \mathbf{I} \right), \end{aligned} \quad (3.1)$$

where we assume that the observation noise is Gaussian, and $\mathcal{CN}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ is the complex Gaussian distribution (Goodman, 1963) whose density is defined as

$$\mathcal{CN}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}) = \frac{1}{\pi^n \sigma^{2n}} \exp \left(-\frac{1}{\sigma^2} (\mathbf{y} - \boldsymbol{\mu})^H (\mathbf{y} - \boldsymbol{\mu}) \right). \quad (3.2)$$

¹The usage of letter \mathbf{y} here may violate the original definition in Eq. (1.14), but we use $\mathbf{y}_{\ell,j}$ to denote the noise-contaminated snapshots. Also note that we use j instead of t for indexing the snapshots.

Here, $\mathbf{w}_{1:k}$, $\lambda_{1:k}$ and σ^2 are the parameters to be estimated ($\lambda_{1:k}$ denotes a set $\{\lambda_1, \dots, \lambda_k\}$), and k is the tunable hyperparameter that determines the number of modes (usually $k \leq n$). In addition, we treat $\varphi_{i,j}$ as a latent variable with the standard Gaussian prior

$$\varphi_{i,j} \sim \mathcal{CN}(0, 1). \quad (3.3)$$

3.2.2 Maximum-likelihood estimator

To derive the MLE of probabilistic DMD, let us rewrite likelihood (3.1) in a matrix form, i.e.,

$$\mathbf{y}_j \mid \boldsymbol{\varphi}_j \sim \mathcal{CN}(\mathbf{B}\boldsymbol{\varphi}_j, \sigma^2\mathbf{I}), \quad (3.4)$$

where we use notations as follows:

$$\mathbf{y}_j = \begin{bmatrix} \mathbf{y}_{0,j} \\ \mathbf{y}_{1,j} \end{bmatrix}, \quad (3.5)$$

$$\boldsymbol{\varphi}_j = [\varphi_{1,j} \ \dots \ \varphi_{k,j}]^\top, \quad (3.6)$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{W} \\ \mathbf{W}\boldsymbol{\Lambda} \end{bmatrix}, \quad (3.7)$$

$$\mathbf{W} = [\mathbf{w}_1 \ \dots \ \mathbf{w}_k], \quad (3.8)$$

$$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_k). \quad (3.9)$$

Marginalizing out $\boldsymbol{\varphi}$ with prior (3.3), we have

$$\mathbf{y}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{B}\mathbf{B}^\text{H} + \sigma^2\mathbf{I}). \quad (3.10)$$

In the following, we describe the relationship between probabilistic model (3.10), total-least-squares DMD (Dawson et al., 2016), which is a “noise-aware” variant of DMD, and standard DMD (Algorithm 1.1 or 1.2). In short, their estimation results coincide in the no-noise limit.

Theorem 3.1. *Suppose we have a dataset that is possibly contaminated by observation noises \mathbf{E} :*

$$\hat{\mathbf{Y}}_\ell = \bar{\mathbf{Y}}_\ell + \mathbf{E}_\ell = [\mathbf{y}_{\ell,1} \ \dots \ \mathbf{y}_{\ell,m}], \quad \text{for } \ell = 0, 1,$$

and let $\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\mathbf{Y}}_0^\top & \hat{\mathbf{Y}}_1^\top \end{bmatrix}^\top$ and $\Sigma_y = m^{-1} \hat{\mathbf{Y}} \hat{\mathbf{Y}}^\text{H}$. In addition, let $(\sigma^2)^*$, \mathbf{W}^* and Λ^* be the MLEs of Eq. (3.10) given $\hat{\mathbf{Y}}$. If $k = n$, then the columns of \mathbf{W}^* and the elements of $\text{diag}(\Lambda^*)$ coincide with the dynamic modes and eigenvalues obtained by total-least-squares DMD, respectively.

Proof. Following the derivations in [Tipping and Bishop \(1999\)](#), the MLEs for probabilistic model (3.10) are given as

$$(\sigma^2)^* = \frac{1}{2n - k} \sum_{i=k+1}^{2n} \mu_i, \quad (3.11)$$

$$\mathbf{B}^* = \begin{bmatrix} \mathbf{W}^* \\ \mathbf{W}^* \Lambda^* \end{bmatrix} = \mathbf{U}_k (\mathbf{M}_k - (\sigma^2)^* \mathbf{I})^{\frac{1}{2}} \mathbf{R}, \quad (3.12)$$

with $\mathbf{U}_k = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_k \end{bmatrix}$ and $\mathbf{M}_k = \text{diag}(\mu_1, \dots, \mu_k)$, where μ_i is the i -th largest eigenvalue of Σ_y with corresponding eigenvector \mathbf{u}_i , and \mathbf{R} is an arbitrary unitary matrix. If $k = n$, we have

$$\mathbf{W}^* \Lambda^* (\mathbf{W}^*)^{-1} = \mathbf{U}_{1,n} \mathbf{U}_{0,n}^{-1}, \quad (3.13)$$

where $\mathbf{U}_{0,n}$ comprises the first n rows and $\mathbf{U}_{1,n}$ comprises the last n rows of \mathbf{U}_n . Hence the columns of \mathbf{W}^* and the elements of $\text{diag}(\Lambda^*)$ are obtained by the eigendecomposition of $\mathbf{U}_{1,n} \mathbf{U}_{0,n}^{-1}$, which is exactly the same procedure with the one in total-least-squares DMD ([Dawson et al., 2016](#)). \square

Theorem 3.2. *If \mathbf{Y}_0 and \mathbf{Y}_1 are linearly consistent ([Tu et al., 2014b](#)), and there is no observation noise (i.e., $\mathbf{E} = 0$), then the estimation results of total-least-squares DMD coincides with those of standard DMD ([Algorithm 1.1](#) or [1.2](#)).*

Proof. From the definition of the linear consistency ([Tu et al., 2014b](#)), when there is no observation noise, $\text{rank}(\Sigma_y) = n$. Hence, $m^{-\frac{1}{2}} \hat{\mathbf{Y}} = \mathbf{U}_n \mathbf{M}_n^{\frac{1}{2}} \mathbf{V}_n^\text{H}$ (\mathbf{V}_n is comprising first n right singular vectors of $m^{-\frac{1}{2}} \hat{\mathbf{Y}}$). Consequently,

$$\hat{\mathbf{Y}}_1 \hat{\mathbf{Y}}_0^\dagger = \mathbf{U}_{1,n} \mathbf{M}_n^{\frac{1}{2}} \mathbf{V}_n^\text{H} \left(\mathbf{V}_n \mathbf{M}_n^{-\frac{1}{2}} \mathbf{U}_{0,n}^{-1} \right) = \mathbf{U}_{1,n} \mathbf{U}_{0,n}^{-1}, \quad (3.14)$$

which shows the equivalence of the outputs of total-least-squares DMD and standard DMD. \square

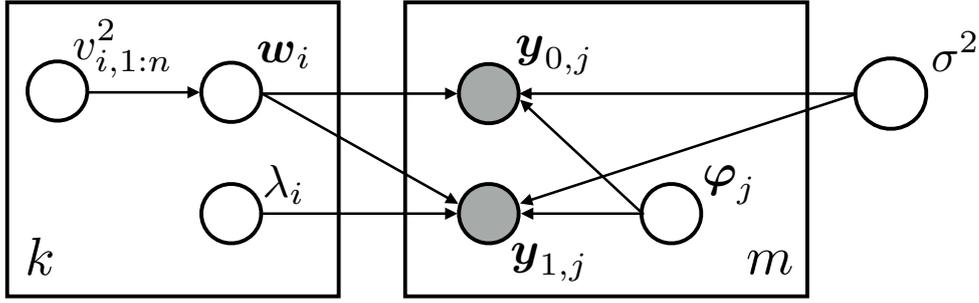


FIGURE 3.1: Graphical model of Bayesian DMD.

3.3 Bayesian DMD

For the Bayesian treatment of DMD, we consider the following priors on the parameters in probabilistic model (3.1). First, we put a Gaussian prior on $\mathbf{w}_{1:k}$:

$$\mathbf{w}_i \mid v_{i,1:n}^2 \sim \mathcal{CN}(\mathbf{0}, \text{diag}(v_{i,1}^2, \dots, v_{i,n}^2)) \quad (3.15)$$

with an inverse gamma hyperprior on $v_{i,d}^2$ ($d = 1, \dots, n$):

$$v_{i,d}^2 \sim \text{InvGamma}(\alpha_v, \beta_v), \quad (3.16)$$

whose shape parameter is α_v and rate parameter is β_v . Moreover, we consider priors on $\lambda_{1:k}$ and σ^2 as

$$\lambda_i \sim \mathcal{CN}(0, 1), \quad (3.17)$$

$$\sigma^2 \sim \text{InvGamma}(\alpha_\sigma, \beta_\sigma). \quad (3.18)$$

A graphical model of the resulting Bayesian DMD is shown in Figure 3.1.

3.3.1 Posterior inference by Gibbs sampling

The conditional probabilistic distribution on each latent variable in the above model becomes a complex Gaussian or an inverse gamma distribution and thus is easy to sample. Consequently, we can develop a Gibbs sampler for inferring the latent variables. In the sequel, we use the following notations:

$$\boldsymbol{\xi}_{-i,j} = \mathbf{y}_{0,j} - \sum_{i' \neq i} \varphi_{i',j} \mathbf{w}_{i'}, \quad (3.19)$$

$$\boldsymbol{\eta}_{-i,j} = \mathbf{y}_{1,j} - \sum_{i' \neq i} \lambda_{i'} \varphi_{i',j} \mathbf{w}_{i'}. \quad (3.20)$$

The sampling procedures of the Gibbs sampler are summarized in Algorithm 3.1. In the following, $\bar{\lambda}$ denotes the complex conjugate of λ , and $[\mathbf{w}_i]_d$ denotes the d -th element of \mathbf{w}_i . In our implementation, the hyperparameters α and β were set to 10^{-3} .

Algorithm 3.1 (Gibbs sampling for Bayesian DMD).

(1) Sample $\mathbf{w}_{1:k}$ from $\mathcal{CN}(\mathbf{m}_{\mathbf{w}_i}, \mathbf{P}_{\mathbf{w}_i}^{-1})$ with

$$\begin{aligned} \mathbf{P}_{\mathbf{w}_i} &= \text{diag}(v_{i,1}^{-2}, \dots, v_{i,n}^{-2}) + \frac{(1 + |\lambda_i|^2) \sum_j |\varphi_{i,j}|^2}{\sigma^2} \mathbf{I}, \\ \mathbf{m}_{\mathbf{w}_i} &= \mathbf{P}_{\mathbf{w}_i}^{-1} \frac{1}{\sigma^2} \sum_j \bar{\varphi}_{i,j} (\boldsymbol{\xi}_{-i,j} + \bar{\lambda}_i \boldsymbol{\eta}_{-i,j}). \end{aligned}$$

(2) Sample $v_{1:k,1:n}^2$ from $\text{InvGamma}(a_{v_{i,d}^2}, b_{v_{i,d}^2})$ with

$$a_{v_{i,d}^2} = \alpha_v + 1, \quad b_{v_{i,d}^2} = \beta_v + |[\mathbf{w}_i]_d|^2.$$

(3) Sample $\lambda_{1:k}$ from $\mathcal{CN}(m_{\lambda_i}, p_{\lambda_i}^{-1})$ with

$$p_{\lambda_i} = 1 + \frac{\mathbf{w}_i^H \mathbf{w}_i}{\sigma^2} \sum_j |\varphi_{i,j}|^2, \quad m_{\lambda_i} = \frac{\mathbf{w}_i^H}{p_{\lambda_i} \sigma^2} \sum_j \bar{\varphi}_{i,j} \boldsymbol{\eta}_{-i,j}.$$

(4) Sample $\boldsymbol{\varphi}_{1:m}$ from $\mathcal{CN}(\mathbf{m}_{\boldsymbol{\varphi}_j}, \mathbf{P}_{\boldsymbol{\varphi}_j}^{-1})$ with

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\varphi}_j} &= \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{W}^H \mathbf{W} + \bar{\Lambda} \mathbf{W}^H \mathbf{W} \Lambda), \\ \mathbf{m}_{\boldsymbol{\varphi}_j} &= \mathbf{P}_{\boldsymbol{\varphi}_j}^{-1} \frac{1}{\sigma^2} (\mathbf{W}^H \mathbf{y}_{0,j} + \bar{\Lambda} \mathbf{W}^H \mathbf{y}_{1,j}). \end{aligned}$$

(5) Sample σ^2 from $\text{InvGamma}(a_{\sigma^2}, b_{\sigma^2})$ with

$$\begin{aligned} a_{\sigma^2} &= \alpha_\sigma + 2mn, \\ b_{\sigma^2} &= \beta_\sigma + \sum_j (\mathbf{y}_{0,j} - \mathbf{W} \boldsymbol{\varphi}_j)^H (\mathbf{y}_{0,j} - \mathbf{W} \boldsymbol{\varphi}_j) \\ &\quad + \sum_j (\mathbf{y}_{1,j} - \mathbf{W} \Lambda \boldsymbol{\varphi}_j)^H (\mathbf{y}_{1,j} - \mathbf{W} \Lambda \boldsymbol{\varphi}_j). \end{aligned}$$

(6) Repeat (1)–(5) for a sufficient number of iterations.

3.3.2 Sparsity-promoting prior

One of the difficulties when applying DMD to noisy data in practice lies in determination of the effective number of dynamic modes. [Jovanović et al. \(2014\)](#) proposed sparsity-promoting DMD, in which the number of dynamic modes are determined by a lasso-like post-processing. In this work, we develop a Bayesian approach for automatic determination of the number of dynamic modes using a sparsity-promoting prior. Differently from the method based on a lasso-like post-processing ([Jovanović et al., 2014](#)), this approach works without manual tuning of hyperparameters through the empirical Bayes technique.

Following the approach proposed by [Park and Casella \(2008\)](#), we incorporate the two-level Laplacian prior on dynamic modes $\mathbf{w}_{1:k}$. That is, we replace prior (3.15) and hyperprior (3.16) respectively by

$$\mathbf{w}_i \mid v_{i,1:n}^2 \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \text{diag}(v_{i,1}^2, \dots, v_{i,n}^2)), \quad \text{and} \quad (3.21)$$

$$v_{i,d}^2 \sim \text{Exponential}\left(\frac{\gamma_i^2}{2}\right) \quad (3.22)$$

with new hyperparameters $\gamma_{1:k}$. They change the parameters of the conditional distributions for $\mathbf{w}_{1:k}$ and σ^2 (at Steps (1) and (5) in Algorithm 3.1) as follows:

$$\mathbf{P}_{\mathbf{w}_i} = \frac{1}{\sigma^2} \text{diag}(v_{i,1}^{-2}, \dots, v_{i,n}^{-2}) + \frac{(1 + |\lambda_i|^2) \sum_j |\varphi_{i,j}|^2}{\sigma^2} \mathbf{I}, \quad (3.23)$$

$$a_{\sigma^2} = \alpha_\sigma + 2mn + \frac{1}{2}kn, \quad (3.24)$$

$$b_{\sigma^2} = \beta_\sigma + \sum_j (\mathbf{y}_{0,j} - \mathbf{W} \boldsymbol{\varphi}_j)^H (\mathbf{y}_{0,j} - \mathbf{W} \boldsymbol{\varphi}_j) + \sum_j (\mathbf{y}_{1,j} - \mathbf{W} \boldsymbol{\Lambda} \boldsymbol{\varphi}_j)^H (\mathbf{y}_{1,j} - \mathbf{W} \boldsymbol{\Lambda} \boldsymbol{\varphi}_j) + \sum_{i,d} \frac{|[\mathbf{w}_i]_d|^2}{2v_{i,d}^2}. \quad (3.25)$$

Further, the distribution for $v_{1:k,1:n}^2$ (at Step 2) becomes the generalized inverse

Gaussian distribution (see e.g. [Devroye \(2014\)](#)) with the following parameters:

$$a_{v_{i,d}^2} = \gamma_i^2, \quad (3.26)$$

$$b_{v_{i,d}^2} = \frac{|[\mathbf{w}_i]_d|^2}{\sigma^2}, \quad (3.27)$$

$$p_{v_{i,d}^2} = \frac{1}{2}. \quad (3.28)$$

To draw a sample from the generalized inverse Gaussian distribution, we used an efficient sampler proposed by [Devroye \(2014\)](#).

Empirical Bayes for hyperparameter The set of hyperparameters $\gamma_{1:k}$ needs to be chosen appropriately for successful model selection. We determine it by maximizing the marginal likelihood, since we empirically found that this was more stable than using gamma distribution as a hyperprior for $\gamma_{1:k}$. We use a Monte Carlo EM algorithm ([Casella, 2001](#)), which comprises iterations between the Gibbs sampling with the modified parameters (E-step) and the maximization of the marginal likelihood (M-step) by

$$\gamma_i^{(Q)} = \sqrt{2n \left(\sum_d \mathbb{E}_{\gamma_i^{(Q-1)}} [v_{i,d}^2] \right)^{-1}},$$

where $\gamma_i^{(Q)}$ denotes the hyperparameter at the Q -th iteration of the EM, and $\mathbb{E}_{\gamma_i^{(Q-1)}}[\cdot]$ denotes the expectation under the hyperparameter at the previous iteration.

3.4 Related work

While no previous work incorporates the probabilistic and Bayesian point of view to DMD, several studies elaborated on the effects of the observation noise; [Duke et al. \(2012\)](#) and [Pan et al. \(2015\)](#) conducted error analyses on the outputs of DMD, and there is a line of research on low-rank approximation of DMD ([Chen et al., 2012](#); [Wynn et al., 2013](#); [Jovanović et al., 2014](#); [Dicle et al., 2016](#); [Héas and Herzet, 2017](#)), with which we can mitigate the noise by ignoring insignificant components of data. In addition, [Dawson et al. \(2016\)](#) proposed total-least-squares DMD, which explicitly considered the presence of observation noise in datasets by formulating DMD as a total least-squares

problem. Note that, in Theorem 3.1, we have shown that the MLE of probabilistic DMD coincides with the solution of total-least-squares DMD. Also, Bagheri (2014) investigated effects of process noise (rather than observation noise) on the output of DMD.

3.5 Numerical examples

We conducted experiments to demonstrate the performance of Bayesian DMD (termed BDMD in this section) regarding the tolerance to noise, the posterior inference, and the automatic determination of the number of modes. In addition, we examined the applications of BDMD to dimensionality reduction and time-series denoising tasks.

3.5.1 Estimation from noisy observations

We validated the performance of BDMD on two types of noisy datasets: one was obtained from a limit cycle, and the other was generated from a system with damping modes.

Limit cycle

We generated data from the discrete-time Stuart–Landau equation in polar-coordinates:

$$\begin{aligned} r_{t+1} &= r_t + \Delta t(\mu r_t - r_t^3), \\ \theta_{t+1} &= \theta_t + \Delta t(\gamma - \beta r_t^2), \end{aligned} \quad (3.29)$$

and the noisy observable (i is the imaginary unit here):

$$\mathbf{y}_t = \begin{bmatrix} e^{-2i\theta_t} & e^{-i\theta_t} & 1 & e^{i\theta_t} & e^{2i\theta_t} \end{bmatrix}^T + \mathbf{e}_t, \quad (3.30)$$

where each element of \mathbf{e}_t was sampled independently from zero-mean Gaussian with variance 10^{-4} . The Stuart–Landau equation contains a limit cycle at $r = \sqrt{\mu}$. We set the parameters by $\mu = 1$, $\gamma = 1$, $\beta = 0$, $\Delta t = 0.01$, $r_0 = \sqrt{\mu}$, and $\theta_0 = 0$, generated 10,000 snapshots, and fed them into standard DMD (Algorithm 1.2), total-least-squares DMD (TLS-DMD) (Dawson et al., 2016), and BDMD (with $k = 5$). The estimated eigenvalues are plotted in Figure 3.2

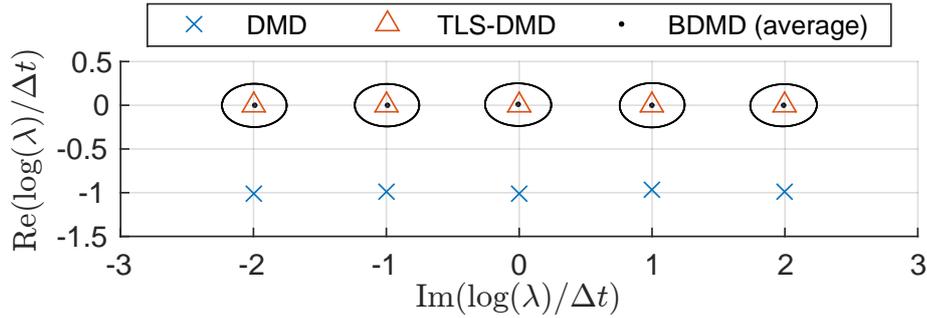


FIGURE 3.2: Eigenvalues estimated for the Stuart–Landau equation. Following a convention, $\log(\lambda)/\Delta t$ is plotted. In this scale, the true eigenvalues lie on the imaginary axis since the data are periodic. The ellipse denotes the 95% credible interval of the samples generated from the Gibbs sampler of BDMD, for each eigenvalue.

wherein the ellipses denote the 95% credible interval of the samples generated from the Gibbs sampler of BDMD, for each eigenvalue. While there is the bias on the estimation by standard DMD due to the observation noise, the estimations by TLS-DMD and BDMD coincide, which agrees with Theorem 3.1. Note that one of the advantages of BDMD is that it returns the *posterior distribution* of the parameters, instead of the point estimation like TLS-DMD.

Damping modes

We also investigated the performance for identifying damping modes, i.e., modes that decay rapidly over time. Generally, it is more difficult to identify damping modes than to identify modes in a limit cycle. The dataset was generated by

$$\mathbf{y}_t = \lambda_1^t \begin{bmatrix} 2 & 2 \end{bmatrix}^\top + \lambda_2^t \begin{bmatrix} 2 & -2 \end{bmatrix}^\top + \mathbf{e}_t, \quad (3.31)$$

where \mathbf{e}_t was zero-mean Gaussian noise with different variances σ^2 ($\sigma = 0, 0.05, 0.1, 0.15, 0.2, 0.25$), and we set $\lambda_1 = 0.9$ and $\lambda_2 = 0.8$ as the eigenvalues. We compared the performances of standard DMD, TLS-DMD, and BDMD (with $k = 2$). A typical instance of the results is depicted in Figure 3.3 wherein the box plots show the statistics of the samples generated from the Gibbs sampler of BDMD. The sample medians of BDMD and the estimations by TLS-DMD lie near, and both are more accurate than the estimations by standard DMD. In addition, we ran 100 trials on the same type of datasets

generated with different random seeds. In Table 3.1, the averages of the absolute errors of estimated eigenvalues are listed. We can observe that the point-estimate performance of BDMD is comparable to that of TLS-DMD.

3.5.2 Automatic relevance determination

We conducted an experiment to investigate how well BDMD can determine the number of modes automatically. We generated a dataset by

$$\mathbf{y}_t = A \begin{bmatrix} 0.9^t \\ 0.7^t \\ 0 \\ 0 \end{bmatrix} + \mathbf{e}_t, \quad (3.32)$$

$$A = \begin{bmatrix} 0 & -5 & 0 & 0 \\ 2 & -4 & 0 & 0 \\ 3 & -3 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix},$$

where \mathbf{e}_t was zero-mean Gaussian noise with variance 10^{-4} . For determining the number of modes given noisy datasets, standard DMD (Algorithm 1.2) may utilize the truncation of small singular values at the SVD step, but the truncation threshold is not trivial in practice. Sparsity-promoting DMD (SP-DMD) (Jovanović et al., 2014) uses a lasso-like post-processing for automatic determination of the number of modes, but it still requires to tune the regularization parameter. However, BDMD with the sparsity-promoting prior (termed BDMD-sp hereafter) can automatically determine the number of modes and the hyperparameter in the light of data inherently, without any need for manual tuning.

We applied standard DMD, SP-DMD (with $\gamma = 10$ tuned to give the best results), and BDMD-sp (with $k = 4$) to the above-mentioned data. As for BDMD-sp, we adopted the medians of the samples generated from the Gibbs sampler as the point estimation values. A typical instance of the results is depicted in Figure 3.4 wherein the structures of the true modes (matrix A) and the estimated modes are shown. In this case, SP-DMD and BDMD-sp successfully recover the structure of dynamic modes. Furthermore, we ran 100 trials with the same type of datasets generated with different random seeds, varying the number of snapshots fed into the algorithms from $m = 4$ to

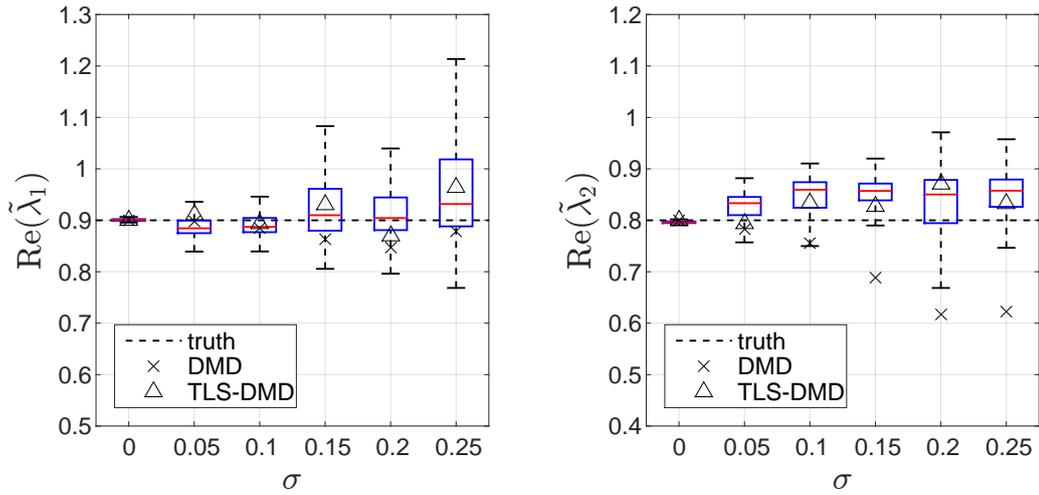


FIGURE 3.3: Estimated values of (left) λ_1 and (right) λ_2 , for each noise magnitude σ . The box plots show the statistics of the samples from the Gibbs sampler of BDMD (the red lines denote the sample medians).

TABLE 3.1: Averages (and the standard deviations) of the absolute errors of estimated (upper) λ_1 and (lower) λ_2 over 100 trials for each noise magnitude σ . As for BDMD, the medians of the samples from the Gibbs sampler were adopted as point estimation values.

$ \Delta\lambda_1 $	σ					
	.00	.05	.10	.15	.20	.25
DMD	.00 (.00)	.01 (.01)	.03 (.01)	.04 (.01)	.05 (.01)	.06 (.02)
TLS-DMD	.00 (.00)	.01 (.01)	.02 (.01)	.03 (.03)	.04 (.03)	.07 (.05)
BDMD	.00 (.00)	.02 (.01)	.02 (.01)	.01 (.01)	.01 (.01)	.02 (.02)

$ \Delta\lambda_2 $	σ					
	.00	.05	.10	.15	.20	.25
DMD	.00 (.00)	.03 (.02)	.09 (.04)	.20 (.09)	.27 (.11)	.38 (.15)
TLS-DMD	.00 (.00)	.02 (.01)	.03 (.02)	.05 (.03)	.06 (.03)	.06 (.05)
BDMD	.00 (.00)	.04 (.02)	.04 (.02)	.04 (.02)	.03 (.02)	.07 (.10)

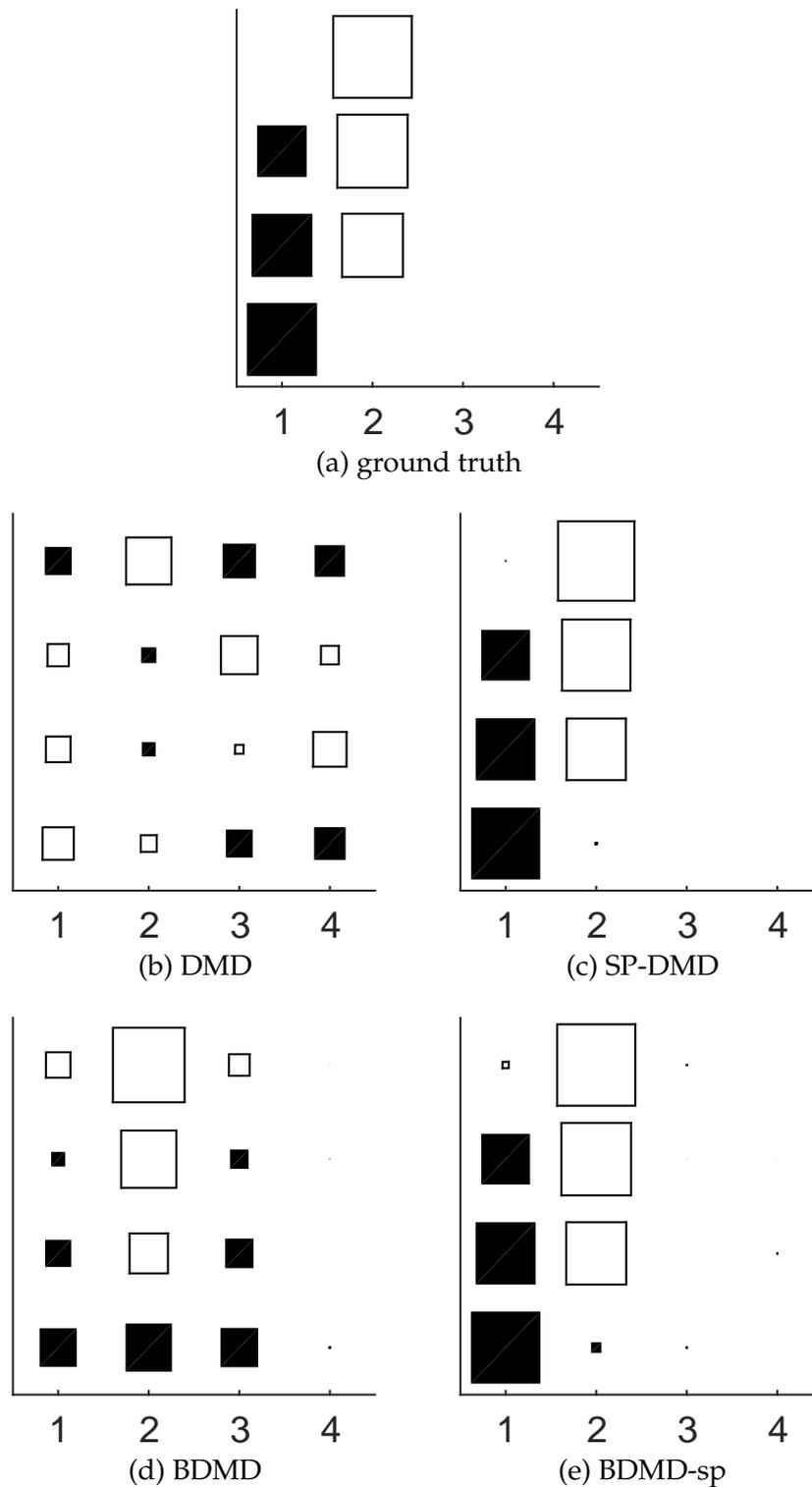


FIGURE 3.4: True and estimated dynamic modes in each column. The filled square denotes a positive value, and the empty denotes a negative value. The size of the square corresponds to the absolute value of each element.

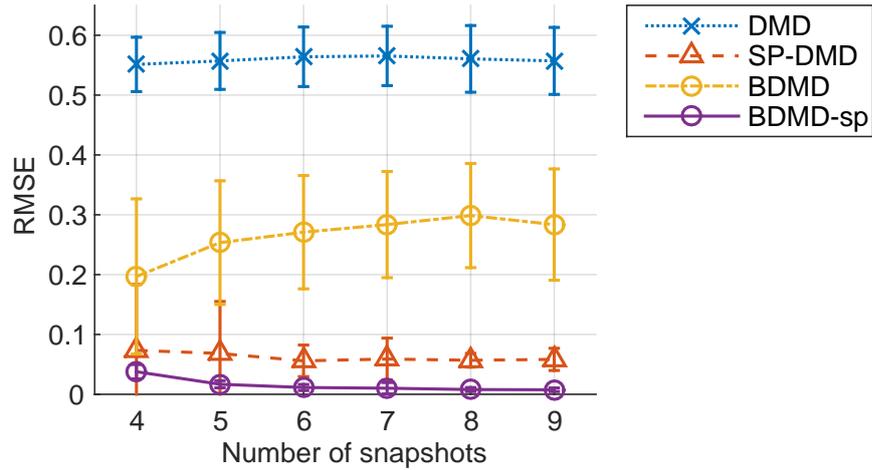


FIGURE 3.5: Averages of the RMSEs of estimated dynamic modes over 100 trials for each number of snapshots.

9. We investigated root-mean-square errors (RMSEs) between the estimated and the true modes, which were calculated after normalizing the maximum absolute values and sorting the order of the modes. The results are summarized in Figure 3.5 wherein the averages (and the standard deviations) of the RMSEs are plotted. We can see that BDMD-sp achieves smaller errors than SP-DMD does.

3.5.3 Applications

We show two examples of BDMD applications: the dimensionality reduction and the time-series denoising.

Dimensionality reduction

BDMD-sp provides a way for dimensionality reduction of time-series data, since it can concentrate their information on a small number of dynamic modes. To demonstrate the performance, we address the task of data visualization using BDMD-sp on the motion capture data of human activities.² We chose locomotion data of three subjects (Subjects #2, #16 and #35), for which both “walk” and “run/jog” motions were recorded. We concatenated the recordings of “walk” and “run/jog” of the three subjects and subsampled them by 1/4, finally obtaining 62-dimensional 421 measurements.

²Downloaded from <http://mocap.cs.cmu.edu/> (retrieved 20-Nov-2017).

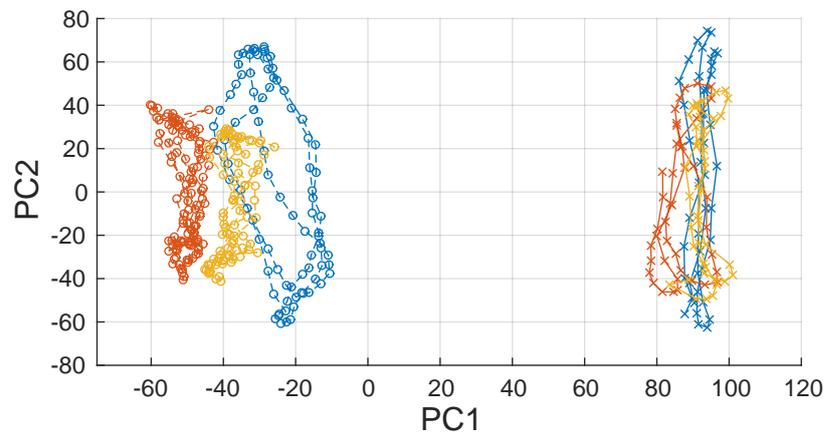
The results of the dimensionality reduction by PCA, t-SNE (van der Maaten and Hinton, 2008), and BDMD-sp (with $k = 32$) are plotted in Figure 3.6. As for PCA, we plot only the first and the second principal scores, since the characteristics of the first eight principal scores were all similar. As for BDMD-sp, we focus on latent variable φ corresponding to the dynamic mode of the largest magnitude and plot the medians of its samples generated from the Gibbs sampler. Now let us elaborate on the features of the results in Figure 3.6. The distinction between “walk” and “run/jog” is clearly observed as the different distributions of the trajectories in every plot of Figure 3.6. The distinction between Subjects #2, #16 and #35 is less obvious, while the distribution of the trajectories implies the difference of the locomotive behavior of Subject #2 from those of the other two. On this point, BDMD-sp (Figure 3.6c) shows the most consistent result wherein the trajectories of Subject #2 are consistently distributed in the upper part of the plot. This difference would stem from the different natures of the methods; BDMD-sp captures the information related to *dynamics* behind the data, while PCA and t-SNE do not.

Furthermore, the modes estimated by the standard DMD and BDMD-sp are shown in Figures 3.7a and 3.7b, respectively. We can confirm that the sparsity-promoting prior works well also on a real-world dataset, in which the magnitudes are concentrated on a few modes.

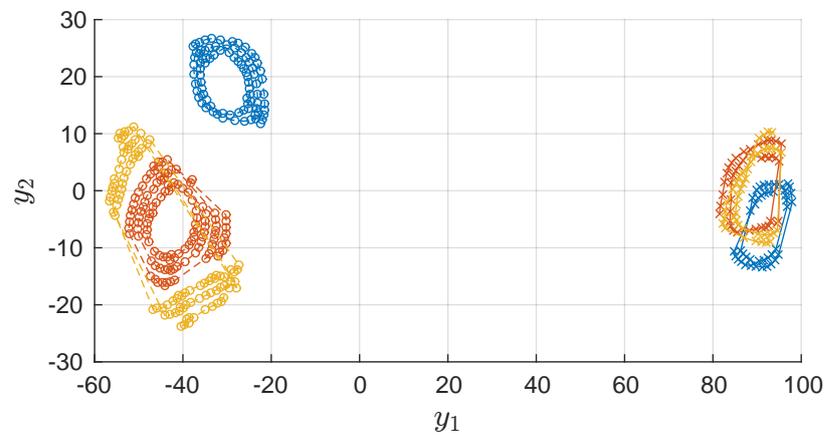
Univariate time-series denoising

We prepared a time-series dataset by extracting single series $\{x\}$ from the Lorenz attractor (Lorenz, 1963) (with $\rho = 28$, $\sigma = 10$ and $\beta = 8/3$) and contaminated them with zero-mean Gaussian noise of variance 16. The task was recovering the original series from the noisy series. We applied BDMD (with $k = 1$) on the noisy series and reconstructed them using samples generated by the Gibbs sampler.

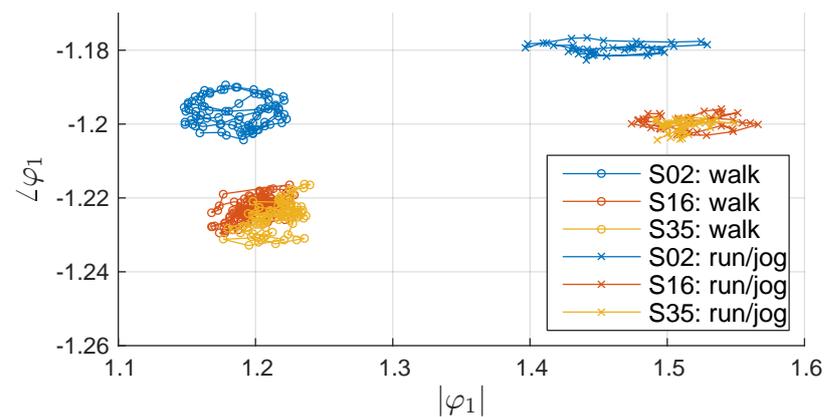
The original and the reconstructed series are plotted in Figure 3.8. The RMSE decreased from 3.2 to 2.3 by the denoising. A simple moving average as a baseline achieved RMSE 2.5 at the best, but note that we cannot necessarily obtain such performance by moving average since it needs to tune the window size.



(a) PCA



(b) t-SNE



(c) BDMD-sp

FIGURE 3.6: Results of dimensionality reduction (best viewed in colors). The first and the second principal scores are plotted for PCA, whereas the magnitudes and the angles of φ_1 are plotted for BDMD-sp. The distinction between “walk” and “run/jog” are clearly observed in every plot. The trajectories of Subject #2 are consistently distributed only in the upper part of (c).

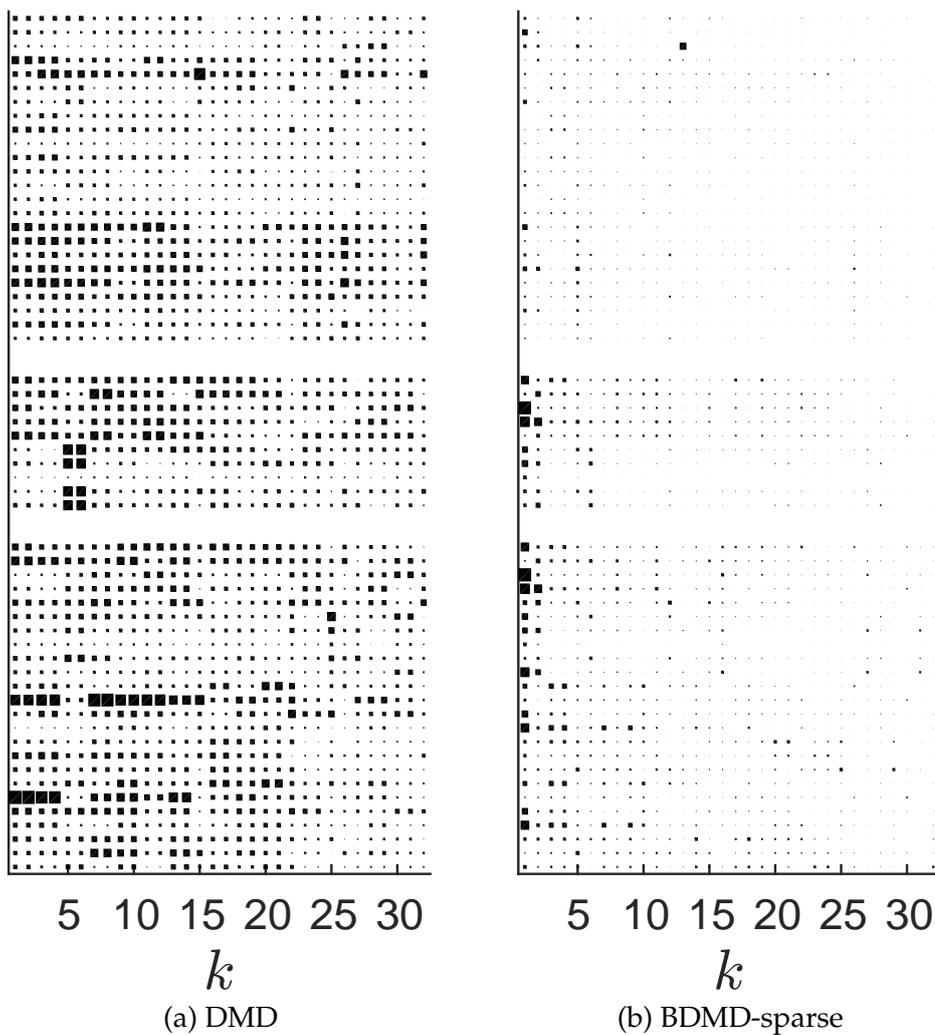


FIGURE 3.7: Absolute values of the modes estimated by (a) DMD and (b) BDMD-sp on the motion capture data. For BDMD-sp, the median of the posterior samples is shown. The size of the square denotes the magnitude of the values.

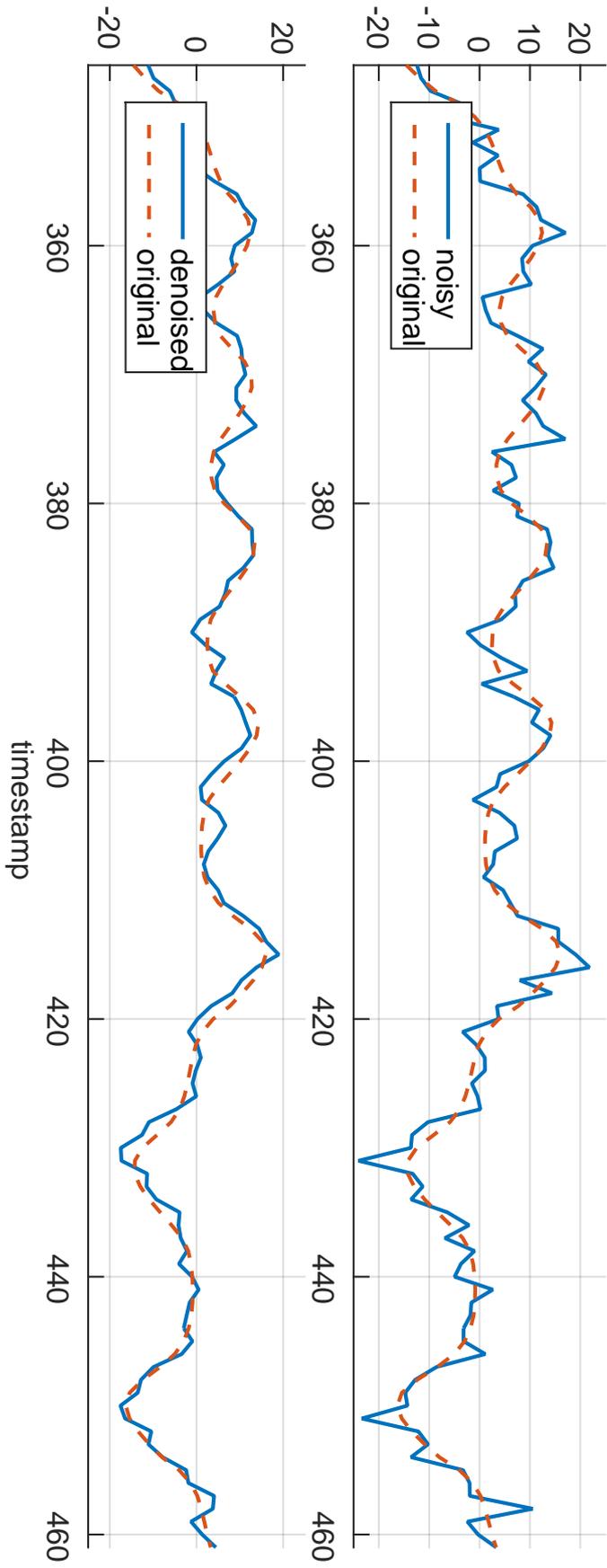


FIGURE 3.8: A part of (upper) the noisy and (lower) the denoised time-series. The RMSE decreased from 3.2 to 2.3.

3.6 Summary

In this chapter, we have introduced the probabilistic model corresponding to DMD and based on that model, proposed *Bayesian DMD* to conduct posterior inference on the DMD parameters and to enrich the DMD techniques systematically in the unified Bayesian framework. We have shown that the MLE of the proposed probabilistic model coincides with the solution of the standard DMD algorithm in the no-noise limit. Moreover, we have provided the Gibbs sampler for the posterior inference in Bayesian DMD. We have also discussed the case of using the sparsity-promoting prior for automatic determination of the effective number of dynamic modes. Finally, we have presented the results of the experiments with the synthetic and the real-world datasets, which show the effectiveness of Bayesian DMD.

Based on the Bayesian framework proposed in this study, there would be various possible extensions of DMD. One of the promising extensions would be the use of structured priors on dynamic modes. For example, the dynamic modes modeled with Markov random fields fit for images, and applications in natural language processing are possible with discrete probability distributions as prior. Then a challenge would be an efficient inference; we relied on the simple Gibbs sampler in this study, but developing more fast and efficient ways is of great importance.

Chapter 4

Subspace DMD

4.1 Introduction

Dynamic mode decomposition (DMD) (Rowley et al., 2009; Schmid, 2010; Kutz et al., 2016a) is a data-driven method that can be utilized for Koopman spectral analysis. In practice, popular implementations of DMD (e.g., Schmid (2010); Tu et al. (2014b)) suffer from *observation noise*. Several researchers have addressed this issue; Duke et al. (2012) and Pan et al. (2015) conducted error analyses on the DMD algorithms, and Dawson et al. (2016) and Hemati et al. (2017) proposed reformulating DMD as a total-least-squares problem to treat the observation noise explicitly. Moreover, there is a line of research on the low-rank approximation of dynamics, including optimized DMD (Chen et al., 2012), optimal mode decomposition (Wynn et al., 2013), sparsity promoting DMD (Jovanović et al., 2014), and the closed-form solution for a low-rank constrained problem (Héas and Herzet, 2017). In addition, we proposed a Bayesian formulation of DMD to incorporate uncertainties (Chapter 3). Those studies provide clear perspectives on the treatment of the observation noise. However, they focus on deterministic dynamical systems, i.e., they do not explicitly deal with *process noise*, which limits their applicability to situations where the underlying dynamics contain random effects.

In fact, the Koopman analysis can also be applied to dynamical systems with process noise via the *stochastic Koopman operator* (Mezić, 2005). The spectra of the stochastic Koopman operator may convey information on the process noise; Bagheri (2014) investigated the effects of weak noise on the spectra of the Koopman operator for oscillating flows. The DMD algorithms are applicable even to stochastic systems (Williams et al., 2015a), *unless observation noise is present*. However, the existing variants of DMD do not explicitly

consider *both* observation and process noise, and, in fact, most of them cannot compute the spectra of the stochastic Koopman operator accurately from noisy observations, which is partly demonstrated in Section 4.3 using numerical examples.

In this chapter, we present an algorithm based on the stochastic Koopman operator for decomposing nonlinear random dynamical systems from noisy observations. The proposed algorithm is referred to as *subspace DMD* because it has a strong connection to the subspace system identification methods developed in control theory. Subspace DMD is aware of both the observation noise and process noise at the same time, and we show its validity with numerical examples.

The remainder of this chapter is organized as follows. In Section 4.2, we introduce the main results of this chapter, the algorithm of subspace DMD. In Section 4.3, we provide numerical examples to show the empirical performance of subspace DMD. A summary of this chapter is provided in Section 4.4.

4.2 Stochastic Koopman analysis with noisy observations

Stochasticity often comprises an essential part of a variety of physical phenomena and sensing. In this section, we introduce the notions of process noise on dynamics and observation noise on observables, and discuss Koopman analysis and DMD for stochastic noisy systems.

4.2.1 Process noise on dynamics

Instead of deterministic dynamical system like Eq. (1.2), consider a discrete-time random dynamical system (RDS) (Arnold, 1998)

$$\mathbf{x}_{t+1} = \mathbf{f}_\Omega(\mathbf{x}_t, \omega_t), \quad \mathbf{x} \in \mathcal{M}, \quad \omega \in \Omega \quad (4.1)$$

with a measure-preserving base flow $\vartheta : \Omega \rightarrow \Omega$, where $(\Omega, \Sigma_\Omega, \mu_\Omega)$ is a probability space of process noise. We assume that ω_t is independent from $\mathbf{x}_0, \dots, \mathbf{x}_t$. A one-step evolution of observables g with regard to the RDS can

be characterized by *stochastic Koopman operator* \mathcal{K}_Ω (Mezić, 2005), defined as

$$\mathcal{K}_\Omega g(\mathbf{x}) := \mathbb{E}_\Omega [g(\mathbf{f}_\Omega(\mathbf{x}, \omega))], \quad (4.2)$$

where $\mathbb{E}_\Omega[\cdot]$ denotes expectation in sample space Ω . Note that deterministic Koopman operator \mathcal{K} can be regarded as a special case of \mathcal{K}_Ω . Now let K_Ω be the restriction of \mathcal{K}_Ω to its invariant subspace G , suppose that a set of observables $\{g_1, \dots, g_n\}$ spans G , and let $\mathbf{g} = [g_1 \ \dots \ g_n]^\top$. In addition, let $\mathbf{K}_\Omega \in \mathbb{C}^{n \times n}$ be the representation of K_Ω with regard to the components of \mathbf{g} . Then, we have

$$\mathbf{g}(\mathbf{x}_{t+1}) = \mathbf{K}_\Omega \mathbf{g}(\mathbf{x}_t) + \mathbf{e}_t, \quad (4.3)$$

where

$$\mathbf{e}_t := \mathbf{g}(\mathbf{f}(\mathbf{x}_t, \omega_t)) - \mathbb{E}_\Omega [\mathbf{g}(\mathbf{f}(\mathbf{x}_t, \omega_t))]. \quad (4.4)$$

Given \mathbf{x}_0 , the solution of (4.3) then becomes

$$\mathbf{g}(\mathbf{x}_t) = \mathbf{K}_\Omega^t \mathbf{g}(\mathbf{x}_0) + \sum_{k=0}^{t-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k. \quad (4.5)$$

The modal decomposition of \mathbf{g} via K_Ω can be obtained likewise, as shown in Section 1.3. Regarding the characteristics of the spectra of the stochastic Koopman operator, Bagheri (2014) elaborated on the effects of weak noise in a limit cycle, and Williams et al. (2015a) applied a variant of DMD to the data obtained from a stochastic differential equation.

The standard DMD (Algorithm 1.1 or 1.2) is also applicable to the RDS and \mathcal{K}_Ω if there is no observation noise and the assumptions mentioned earlier (Assumptions 1.1, 1.2, and 1.3), i.e. discrete eigenvalues, ergodicity, and data from a Koopman invariant subspace, also hold for the RDS and the dataset at hand. This can be shown in a manner similar to the one in Theorem 1.1, except for the definition of \mathbf{G}_0 and \mathbf{G}_1 , as follows. Let \mathbf{Y}_0 and \mathbf{Y}_1 be the data matrices generated from RDS \mathbf{f}_Ω and observable \mathbf{g} as in Eqs. (1.14) and (1.24), and let us assume the whiteness on process noise as follows:

Assumption 4.1. Process noise ω is independently and identically distributed in time, i.e., for all $t', t \in \mathbb{T}$,

$$\mathbb{E}_\Omega [\mathbf{e}_{t'} \mathbf{e}_t^H] = \mathbf{P} \delta_{t't} \quad (4.6)$$

for some $\mathbf{P} \in \mathbb{C}^{n \times n}$.

Then, from the law of large numbers and the assumption of ergodicity, the empirical matrices

$$\hat{\mathbf{G}}_0 = \frac{1}{m} \mathbf{Y}_0 \mathbf{Y}_0 = \frac{1}{m} \sum_{j=0}^{m-1} \mathbf{g}(\mathbf{x}_j) \mathbf{g}(\mathbf{x}_j)^H \quad \text{and}$$

$$\hat{\mathbf{G}}_1 = \frac{1}{m} \mathbf{Y}_1 \mathbf{Y}_0 = \frac{1}{m} \sum_{j=0}^{m-1} \mathbf{g}(\mathbf{f}_\Omega(\mathbf{x}_j, \omega_j)) \mathbf{g}(\mathbf{x}_j)^H$$

respectively converge to

$$\mathbf{G}_0 = \mathbb{E}_{\mathcal{M}} [\mathbf{g}(\mathbf{x}) \mathbf{g}(\mathbf{x})^H] \quad \text{and}$$

$$\mathbf{G}_1 = \mathbb{E}_{\mathcal{M}} [\mathbb{E}_{\Omega} [\mathbf{g}(\mathbf{f}_\Omega(\mathbf{x}, \omega))] \mathbf{g}(\mathbf{x})^H]$$

$$= \int_{\mathcal{M} \times \Omega} \mathbf{g}(\mathbf{f}_\Omega(\mathbf{x}, \omega)) \mathbf{g}(\mathbf{x})^H d\mu_{\mathcal{M}} d\mu_{\Omega}$$

with probability one. One can use this convergence property to show the applicability of the standard DMD for the RDS, as in the proof of Theorem 1.1.

4.2.2 Observation noise on observables

In addition to the process noise, let us take the observation noise into account. Consider a new (noisy) observable $\mathbf{h} : \mathcal{M} \times S \rightarrow \mathbb{C}^n$:

$$\mathbf{h}(\mathbf{x}_t, s_t) := \mathbf{g}(\mathbf{x}_t) + \mathbf{w}(s_t), \quad \mathbf{x} \in \mathcal{M}, \quad s \in S, \quad (4.7)$$

where $\mathbf{w} : S \rightarrow \mathbb{C}^n$ is a random variable on a probability space (S, Σ_S, μ_S) of the observation noise. Hereafter, we denote $\mathbf{w}(s_t)$ by \mathbf{w}_t for notational simplicity. Now assume that s is independent from \mathbf{x} and that \mathbf{w} is a white noise, i.e.,

Assumption 4.2. Observation noise \mathbf{w} is zero-mean, has time-invariant finite variance, and is temporally uncorrelated, i.e., for all $t', t \in \mathbb{T}$,

$$\mathbb{E}_S [\mathbf{w}_t] = 0, \quad (4.8)$$

$$\mathbb{E}_S [\mathbf{w}_{t'} \mathbf{w}_t^H] = \mathbf{Q} \delta_{t't}, \quad (4.9)$$

$$\mathbb{E}_{\Omega, S} [\mathbf{e}_{t'} \mathbf{w}_t^H] = \mathbf{R} \delta_{t't}, \quad (4.10)$$

for some $\mathbf{Q}, \mathbf{R} \in \mathbb{C}^{n \times n}$.

Note that under the presence of observation noise, an output of the standard DMD no longer converges to the spectra of the Koopman operator. An output of total-least-squares DMD (Dawson et al., 2016; Hemati et al., 2017) is unbiased even for noisy observations as long as the dynamics are deterministic, but it is biased as a realization of K_Ω for the RDS. These inconsistencies in the existing methods are partly revealed in the numerical examples in Section 4.3.

4.2.3 Statistics of noisy observables on RDS

We would like to develop a DMD algorithm for stochastic Koopman analysis that is always aware of *both* the process noise and observation noise. To this end, we summarize the statistics of noisy observable \mathbf{h} on RDS \mathbf{f}_Ω .

First, assume that \mathbf{g} is quasi-stationary (see Ljung (1999) for details), i.e.,

Assumption 4.3. For almost all $\mathbf{x}_0 \in \mathcal{M}$ and all $t', t \in \mathbb{T}$,

$$\mathbb{E}_\Omega [\mathbf{g}(\mathbf{x}_t)] = \mathbf{m}_t, \quad |\mathbf{m}_t| < \infty, \quad (4.11)$$

$$\mathbb{E}_\Omega [\mathbf{g}(\mathbf{x}_{t'})\mathbf{g}(\mathbf{x}_t)^\mathbf{H}] = \mathbf{G}_{t',t}, \quad \|\mathbf{G}_{t',t}\|_F < \infty, \quad (4.12)$$

$$\mathbb{E}_\mathcal{M} [\mathbf{G}_{t,t}] = \mathbf{G}, \quad (4.13)$$

for some $\mathbf{G} \in \mathbb{C}^{n \times n}$.

Then, the second-order moment of \mathbf{g} , $\mathbf{G}_{t',t}$, satisfies the following properties.

Lemma 4.1. *If Assumption 4.1 holds, then $\mathbf{G}_{t',t}$ is expressed as*

$$\mathbf{G}_{t',t} = \begin{cases} \mathbf{K}_\Omega^{t'-t} \mathbf{G}_{t,t}, & t' \geq t, \\ \mathbf{G}_{t',t'} (\mathbf{K}_\Omega^{t-t'})^\mathbf{H}, & t' < t. \end{cases} \quad (4.14)$$

Proof. From Eq. (4.5), for the case of $t' > t$, we have

$$\begin{aligned}
\mathbf{G}_{t',t} &= \mathbb{E}_\Omega \left[\left(\mathbf{K}_\Omega^{t'} \mathbf{g}(\mathbf{x}_0) + \sum_{k=0}^{t'-1} \mathbf{K}_\Omega^{t'-k-1} \mathbf{e}_k \right) \left(\mathbf{K}_\Omega^t \mathbf{g}(\mathbf{x}_0) + \sum_{k=0}^{t-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k \right)^\mathbf{H} \right] \\
&= \mathbf{K}_\Omega^{t'-t} \mathbb{E}_\Omega \left[\left(\mathbf{K}_\Omega^t \mathbf{g}(\mathbf{x}_0) + \sum_{k=0}^{t-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k + \sum_{k=t}^{t'-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k \right) \right. \\
&\quad \left. \left(\mathbf{K}_\Omega^t \mathbf{g}(\mathbf{x}_0) + \sum_{k=0}^{t-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k \right)^\mathbf{H} \right] \\
&= \mathbf{K}_\Omega^{t'-t} \mathbf{G}_{t,t} \\
&\quad + \mathbf{K}_\Omega^{t'-t} \mathbb{E}_\Omega \left[\left(\sum_{k=t}^{t'-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k \right) \left(\mathbf{K}_\Omega^t \mathbf{g}(\mathbf{x}_0) + \sum_{k=0}^{t-1} \mathbf{K}_\Omega^{t-k-1} \mathbf{e}_k \right)^\mathbf{H} \right] \\
&= \mathbf{K}_\Omega^{t'-t} \mathbf{G}_{t,t} \\
&\quad + \sum_{k=t}^{t'-1} \mathbf{K}_\Omega^{t'-k-1} \mathbb{E}_\Omega [\mathbf{e}_k] \mathbf{g}(\mathbf{x}_0)^\mathbf{H} (\mathbf{K}_\Omega^t)^\mathbf{H} \\
&\quad + \sum_{k=0}^{t-1} \sum_{k'=t}^{t'-1} \mathbf{K}_\Omega^{t'-k'-1} \mathbb{E}_\Omega [\mathbf{e}_{k'} \mathbf{e}_k^\mathbf{H}] (\mathbf{K}_\Omega^{t-k-1})^\mathbf{H} \\
&= \mathbf{K}_\Omega^{t'-t} \mathbf{G}_{t,t},
\end{aligned}$$

where the last equality holds because \mathbf{e} is zero-mean and because of Assumption 4.1. For the case of $t' > t$, from the definition of $\mathbf{G}_{t',t}$ and the above equation, we have

$$\mathbf{G}_{t',t} = \mathbf{G}_{t,t'}^\mathbf{H} = \mathbf{G}_{t',t'} \left(\mathbf{K}_\Omega^{t-t'} \right)^\mathbf{H}.$$

□

Corollary 4.1. Denote $\mathbb{E}_{\mathcal{M}} [\mathbf{G}_{t+\tau,t}]$ by \mathbf{G}_τ . If Assumption 4.3 holds, then

$$\mathbf{G}_\tau = \mathbf{K}_\Omega^\tau \mathbf{G}. \quad (4.15)$$

Now let us define $\mathbf{H}_{t',t} := \mathbb{E}_{\Omega,S} [\mathbf{h}(\mathbf{x}_{t'}) \mathbf{h}(\mathbf{x}_t)^\mathbf{H}]$, where we have dropped argument s of \mathbf{h} for ease of notation. Then, $\mathbf{H}_{t',t}$ satisfies the following properties.

Lemma 4.2. *If Assumption 4.2 holds, then $\mathbf{H}_{t',t}$ is expressed as*

$$\mathbf{H}_{t',t} = \begin{cases} \mathbf{K}_\Omega^{t'-t-1} (\mathbf{K}_\Omega \mathbf{G}_{t,t} + \mathbf{R}), & t' > t, \\ \mathbf{G}_{t,t} + \mathbf{Q}, & t' = t, \\ \mathbf{H}_{t,t'}^H, & t' < t. \end{cases} \quad (4.16)$$

Proof. From Eqs. (4.5) and (4.7), for the case of $t' > t$, we have

$$\begin{aligned} \mathbf{H}_{t',t} &= \mathbb{E}_{\Omega,S} \left[(\mathbf{g}(\mathbf{x}_{t'}) + \mathbf{w}_{t'}) (\mathbf{g}(\mathbf{x}_t) + \mathbf{w}_t)^H \right] \\ &= \mathbb{E}_{\Omega,S} \left[\mathbf{g}(\mathbf{x}_{t'}) \mathbf{g}(\mathbf{x}_t)^H + \mathbf{g}(\mathbf{x}_{t'}) \mathbf{w}_t^H + \mathbf{w}_{t'} \mathbf{g}(\mathbf{x}_t)^H + \mathbf{w}_{t'} \mathbf{w}_t^H \right] \\ &= \mathbf{K}_\Omega^{t'-t} \mathbf{G}_{t,t} + \sum_{k=0}^{t'-1} \mathbf{K}_\Omega^{t'-k-1} \mathbb{E}_{\Omega,S} [\mathbf{e}_k \mathbf{w}_t^H] + \left(\sum_{k=0}^{t-1} \mathbf{K}_\Omega^{t-k-1} \mathbb{E}_{\Omega,S} [\mathbf{e}_k \mathbf{w}_{t'}^H] \right)^H \\ &= \mathbf{K}_\Omega^{t'-t} \mathbf{G}_{t,t} + \mathbf{K}_\Omega^{t'-t-1} \mathbf{R} \\ &= \mathbf{K}_\Omega^{t'-t-1} (\mathbf{K}_\Omega \mathbf{G}_{t,t} + \mathbf{R}), \end{aligned}$$

where the third and the fourth equalities are from Assumption 4.2. When $t' = t$, from Assumption 4.2, we have

$$\begin{aligned} \mathbf{H}_{t,t} &= \mathbb{E}_{\Omega,S} \left[(\mathbf{g}(\mathbf{x}_t) + \mathbf{w}_t) (\mathbf{g}(\mathbf{x}_t) + \mathbf{w}_t)^H \right] \\ &= \mathbf{G}_{t,t} + \mathbf{Q}. \end{aligned}$$

□

Corollary 4.2. *Denote $\mathbb{E}_{\mathcal{M}} [\mathbf{H}_{t+\tau,t}]$ by \mathbf{H}_τ . If Assumption 4.3 holds, then*

$$\mathbf{H}_\tau = \begin{cases} \mathbf{K}_\Omega^{\tau-1} (\mathbf{K}_\Omega \mathbf{G} + \mathbf{R}), & \tau > 0, \\ \mathbf{G} + \mathbf{Q}, & \tau = 0. \end{cases} \quad (4.17)$$

4.2.4 Subspace DMD

Finally, we introduce a numerical method to compute an instance of the stochastic Koopman operator given noisy observations, namely, *subspace DMD*. Analogously to Eq. (1.14), let us define the data matrix as a concatenation of m observations starting at time t , i.e.,

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{h}(\mathbf{x}_t) & \cdots & \mathbf{h}(\mathbf{x}_{t+m-1}) \end{bmatrix} \in \mathbb{C}^{n \times m}. \quad (4.18)$$

Then, using a data quadruple $(\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3)$, we can obtain a calculation for K_Ω using the following theorem.

Theorem 4.1. Define $\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2$, and \mathbf{Y}_3 by Eq. (4.18), and let $\mathbf{Y}_p, \mathbf{Y}_f \in \mathbb{C}^{2n \times m}$ be

$$\begin{aligned} \mathbf{Y}_p &= \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_1 \end{bmatrix}, \quad \text{and} \\ \mathbf{Y}_f &= \begin{bmatrix} \mathbf{Y}_2 \\ \mathbf{Y}_3 \end{bmatrix}. \end{aligned} \tag{4.19}$$

Moreover, let $\mathbf{O} = \mathbf{Y}_f \mathbb{P}_{\mathbf{Y}_p^H} \in \mathbb{C}^{2n \times m}$ be the orthogonal projection of rows of \mathbf{Y}_f onto the row space of \mathbf{Y}_p . Here, consider a compact SVD

$$\mathbf{O} = \mathbf{U}_q \mathbf{S}_q \mathbf{V}_q^H \tag{4.20}$$

with $\mathbf{U}_q \in \mathbb{C}^{2n \times q}$, $\mathbf{S}_q \in \mathbb{C}^{q \times q}$, and $\mathbf{V}_q \in \mathbb{C}^{m \times q}$, where $q = \text{rank}(\mathbf{O})$. Moreover, let \mathbf{U}_{q1} be the first n rows and \mathbf{U}_{q2} be the last n rows of \mathbf{U}_q . Suppose that Assumptions 1.1 and 1.2 hold for dynamics of interest \mathbf{f}_Ω and corresponding stochastic Koopman operator \mathcal{K}_Ω , and that noiseless observables \mathbf{g} in Eq. (4.7) span a Koopman invariant subspace like those in Assumption 1.3. Also, suppose Assumptions 4.1, 4.2, and 4.3 hold. If $\text{rank}(\mathbf{Y}_p) = 2n$, and $\text{rank}(\mathbf{K}_\Omega \mathbf{G} + \mathbf{R}) = n$, then in $m \rightarrow \infty$,

$$\mathbf{U}_{q2} \mathbf{U}_{q1}^\dagger \rightarrow \mathbf{K}_\Omega \tag{4.21}$$

with probability one.

Proof. Let $\hat{\mathbf{H}}$ be the empirical matrix such that

$$\hat{\mathbf{H}}_{t+\tau, t} = \frac{1}{m} \mathbf{Y}_{t+\tau} \mathbf{Y}_t^H.$$

In $m \rightarrow \infty$, $\hat{\mathbf{H}}_{t+\tau, t}$ converges to \mathbf{H}_τ with probability one for all $t \in \mathbb{T}$ and $\tau \geq 0$, because, from the law of large numbers and the assumption of ergodicity,

$$\begin{aligned} \frac{1}{m} \mathbf{Y}_{t+\tau} \mathbf{Y}_t^H &= \frac{1}{m} \sum_{j=t}^{t+m-1} \mathbf{h}(\mathbf{x}_{j+\tau}) \mathbf{h}(\mathbf{x}_j)^H \\ &\rightarrow \int_{\mathcal{M} \times \Omega \times \mathcal{S}} \mathbf{h}(\mathbf{x}_{t+\tau}) \mathbf{h}(\mathbf{x}_t)^H d\mu_{\mathcal{M}} d\mu_\Omega d\mu_{\mathcal{S}} \\ &= \mathbf{H}_\tau. \end{aligned}$$

Because we have assumed $\text{rank}(\mathbf{Y}_p) = 2n$, in $m \rightarrow \infty$,

$$\begin{aligned}
\mathbf{O} &= \mathbf{Y}_f \mathbf{Y}_p^H (\mathbf{Y}_p \mathbf{Y}_p^H)^{-1} \mathbf{Y}_p \\
&= \begin{bmatrix} \hat{\mathbf{H}}_{2,0} & \hat{\mathbf{H}}_{2,1} \\ \hat{\mathbf{H}}_{3,0} & \hat{\mathbf{H}}_{3,1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{H}}_{0,0} & \hat{\mathbf{H}}_{0,1} \\ \hat{\mathbf{H}}_{1,0} & \hat{\mathbf{H}}_{1,1} \end{bmatrix}^{-1} \mathbf{Y}_p \\
&\rightarrow \begin{bmatrix} \mathbf{H}_2 & \mathbf{H}_1 \\ \mathbf{H}_3 & \mathbf{H}_2 \end{bmatrix} \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1^H \\ \mathbf{H}_1 & \mathbf{H}_0 \end{bmatrix}^{-1} \mathbf{Y}_p \\
&= \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_\Omega \end{bmatrix} \begin{bmatrix} \mathbf{K}_\Omega \mathbf{D} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{G} + \mathbf{Q} & \mathbf{D}^H \\ \mathbf{D} & \mathbf{G} + \mathbf{Q} \end{bmatrix}^{-1} \mathbf{Y}_p \\
&= \mathbf{O}_1 \mathbf{O}_2
\end{aligned} \tag{4.22}$$

with probability one, where $\mathbf{D} = \mathbf{K}_\Omega \mathbf{G} + \mathbf{R} \in \mathbb{C}^{n \times n}$ and

$$\begin{aligned}
\mathbf{O}_1 &= \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_\Omega \end{bmatrix}, \\
\mathbf{O}_2 &= \mathbf{D}^\top \begin{bmatrix} \mathbf{K}_\Omega \\ \mathbf{I} \end{bmatrix}^\top \begin{bmatrix} \mathbf{G} + \mathbf{Q} & \mathbf{D}^H \\ \mathbf{D} & \mathbf{G} + \mathbf{Q} \end{bmatrix}^{-1} \mathbf{Y}_p.
\end{aligned}$$

Because we have assumed $\text{rank}(\mathbf{D}) = n$, the rank of both \mathbf{O}_1 and \mathbf{O}_2 is n . Hence, in $m \rightarrow \infty$, q also becomes n . Remember that by compact SVD (4.20), we have the decomposition of \mathbf{O} into two rank- n matrices, i.e.,

$$\mathbf{O} = (\mathbf{U}_q \mathbf{S}_q^{1/2}) (\mathbf{S}_q^{1/2} \mathbf{V}_q^H).$$

Therefore, from Eq. (4.22), in $m \rightarrow \infty$, we have

$$\mathbf{U}_q \mathbf{S}_q^{1/2} \rightarrow \mathbf{O}_1 \mathbf{T} = \begin{bmatrix} \mathbf{T} \\ \mathbf{K}_\Omega \mathbf{T} \end{bmatrix}$$

with probability one, where $\mathbf{T} \in \mathbb{C}^{n \times n}$ is an arbitrary unitary matrix. Consequently, \mathbf{U}_{q1} and \mathbf{U}_{q2} become \mathbf{T} and $\mathbf{K}_\Omega \mathbf{T}$ respectively, and Eq. (4.21) holds. \square

Based on Theorem 4.1, we present a *subspace DMD* algorithm as follows.

Algorithm 4.1 (Subspace DMD).

1. Build matrices \mathbf{Y}_p and \mathbf{Y}_f by Eq. (4.19).
2. Compute orthogonal projection $\mathbf{O} = \mathbf{Y}_f \mathbb{P}_{\mathbf{Y}_p^H}$.

3. Compute compact SVD $\mathbf{O} = \mathbf{U}_q \mathbf{S}_q \mathbf{V}_q^H$ and define \mathbf{U}_{q1} and \mathbf{U}_{q2} by the first and the last n rows of \mathbf{U}_q , respectively.
4. Compute compact SVD $\mathbf{U}_{q1} = \mathbf{U} \mathbf{S} \mathbf{V}^H$ and define $\tilde{\mathbf{A}} = \mathbf{U}^H \mathbf{U}_{q2} \mathbf{V} \mathbf{S}^{-1}$.
5. Compute the eigenvalues λ and eigenvectors $\tilde{\mathbf{w}}$ of $\tilde{\mathbf{A}}$.
6. Return dynamic modes $\mathbf{w} = \lambda^{-1} \mathbf{U}_{q2} \mathbf{V} \mathbf{S}^{-1} \tilde{\mathbf{w}}$ and corresponding eigenvalues λ .

Remark 4.1. With subspace DMD, we can naturally conduct a low-rank approximation of dynamics by replacing the compact SVD in Step 3 with a truncated SVD. In contrast, in the standard DMD (Algorithm 1.2) and total-least-squares DMD (Dawson et al., 2016; Hemati et al., 2017), the low-rank approximation is achieved via the truncated proper orthogonal decomposition (POD). Note that there is also a line of research on the low-rank approximation of DMD, such as (Chen et al., 2012; Wynn et al., 2013; Jovanović et al., 2014; Héas and Herzet, 2017).

4.2.5 Relation to subspace system identification and extension to controlled systems

Subspace DMD has a strong connection to the methods called *subspace system identification* (see, e.g., Van Overschee and De Moor (1996); Katayama (2005)) in their computational methodology. Subspace system identification is a series of methods mainly for the identification of linear time-invariant systems, whereas in this chapter, we present a similar methodology for nonlinear dynamical systems involving the observables and the stochastic Koopman operator.

Subspace system identification has been studied from the viewpoint of control theory and admits the presence of *input signals* distinguished from the process noise. Therefore, an extension of subspace DMD to controlled dynamical systems would be straightforward following the methodologies developed in the research of subspace system identification. Also, one may take a closed-loop controlled system into consideration. In the context of DMD, Proctor et al. (2016) have discussed a variant of DMD for data obtained from the controlled systems.

4.3 Numerical examples

We present numerical examples for the application of subspace DMD to several types of dynamical systems to show its empirical performance. When describing target dynamical systems in the following examples, we denote Gaussian white process noise by e and Gaussian white observation noise by w . The standard deviation of the process noise is referred to as σ_p and that of the observation noise as σ_o . Moreover, we denote the number of snapshots fed into algorithms by m and the dimensionality of the data by n .

4.3.1 Oscillation perturbed by noise

The stochastic Stuart–Landau equation on a complex-valued function $z(t) = r(t) \exp(i\theta(t))$ is defined as

$$\frac{dz}{dt} = (\mu + i\gamma)z - (1 + i\beta)|z|^2z + \sigma_p e(t), \quad (4.23)$$

where $e(t)$ is Gaussian white noise with unit variance, and i denotes the imaginary unit. The solution of this equation evolves on a limit cycle at $|z| = \sqrt{\mu}$ in the absence of process noise (i.e., $\sigma_p = 0$). [Bagheri \(2014\)](#) has analyzed the effects of weak process noise on the Koopman eigenvalues of the limit cycle of the Stuart–Landau equation, which can be summarized as follows; the continuous-time eigenvalues lie on the imaginary axis if process noise is absent because the data are completely periodic, but in contrast, when perturbation (phase diffusion, as shown in [Figure 4.1a](#)) is present owing to the process noise, a line of the eigenvalues is “bent” as shown in [Figure 4.1c](#). Hence, by investigating the distribution of the eigenvalues, one can anticipate the presence and magnitude of the phase diffusion. To this end, we must eliminate the effects of observation noise if any, which produces an extra bias on the eigenvalues, leaving the effects of process noise.

Following the scheme in [Dawson et al. \(2016\)](#), we generated data using the following discretized Stuart–Landau equation in polar-coordinates with process noise:

$$\begin{bmatrix} r_{t+1} \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} r_t + (\mu r_t - r_t^3) \Delta t \\ \theta_t + (\gamma - \beta r_t^2) \Delta t \end{bmatrix} + \begin{bmatrix} \Delta t & 0 \\ 0 & \Delta t / r_t \end{bmatrix} e_t, \quad (4.24)$$

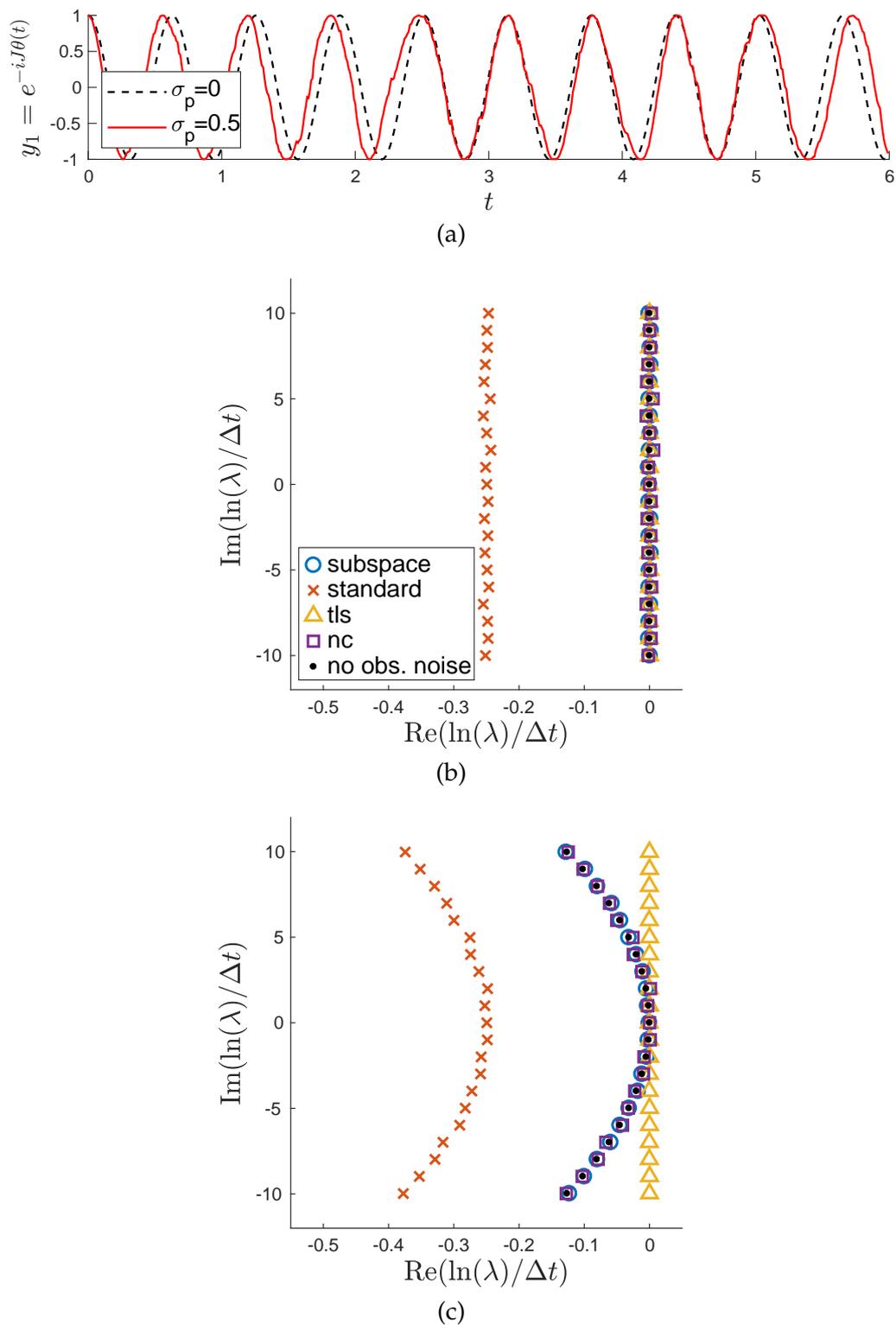


FIGURE 4.1: (a) Data generated by the noise-free/noisy Stuart–Landau equation and trigonometric observables, which show the phase diffusion when $\sigma_p > 0$ (the solid red line). (b–c) The estimated continuous-time eigenvalues, with (b) $\sigma_p = 0$ and (c) $\sigma_p = 0.5$. Subspace DMD eliminates the effects of the observation noise, keeping the effects of the process noise.

and a set of noisy trigonometric observables:

$$\mathbf{y}_t = \begin{bmatrix} e^{-10i\theta t} & e^{-9i\theta t} & \dots & e^{9i\theta t} & e^{10i\theta t} \end{bmatrix} + \mathbf{w}_t, \quad (4.25)$$

where the magnitude of the observation noise was fixed to $\sigma_o = 0.05$. We estimated the continuous-time eigenvalues using subspace DMD (Algorithm 4.1), standard DMD (Algorithm 1.2), total-least-squares DMD (tls-DMD) (Dawson et al., 2016; Hemati et al., 2017), and noise-corrected DMD (nc-DMD) (Dawson et al., 2016).

Figure 4.1b shows the eigenvalues without any process noise (i.e., $\sigma_p = 0$), and Figure 4.1c shows the ones with process noise of $\sigma_p = 0.5$. In both plots, we also show the “clean” eigenvalues computed with the data without the observation noise. When the process noise is present (in Figure 4.1c), while the eigenvalues estimated by tls-DMD differ from the clean ones (as reported in Dawson et al. (2016)), subspace DMD successfully estimates them. Note that the estimation by nc-DMD also coincides with the clean eigenvalues, but nc-DMD needs a precise estimation of magnitude of observation noise, which is often difficult to obtain. Subspace DMD can eliminate the effects of observation noise without such information while *keeping the effects of the process noise*.

4.3.2 Noisy damping modes

Let us consider the stochastic Burger’s equation

$$\partial_t u(x, t) + u \partial_x u = k \partial_x^2 u + \sigma_p e(x, t) \quad (4.26)$$

with $k > 0$ and Gaussian space-time white noise $e(x, t)$. In fact, the eigenvalues of the Koopman operator on Burger’s equation (without process noise, i.e., $\sigma_p = 0$) can be analytically obtained via the Cole–Hopf transformation and they correspond to the decaying modes of the solution (Budišić et al., 2012; Kutz et al., 2016c). When process noise is present ($\sigma_p > 0$), the solution of Burger’s equation becomes “rough,” but its global appearance remains similar to the case of no process noise, as shown in Figure 4.2a.

We approximated the solution of the stochastic Burger’s equation with $k = 0.01$ and $\sigma_p = 0.01$ using Crank–Nicolson–Maruyama method (see, e.g., Hausenblas (2003)) with initial condition $u(x, 0) = \sin(2\pi x)$ and Dirichlet

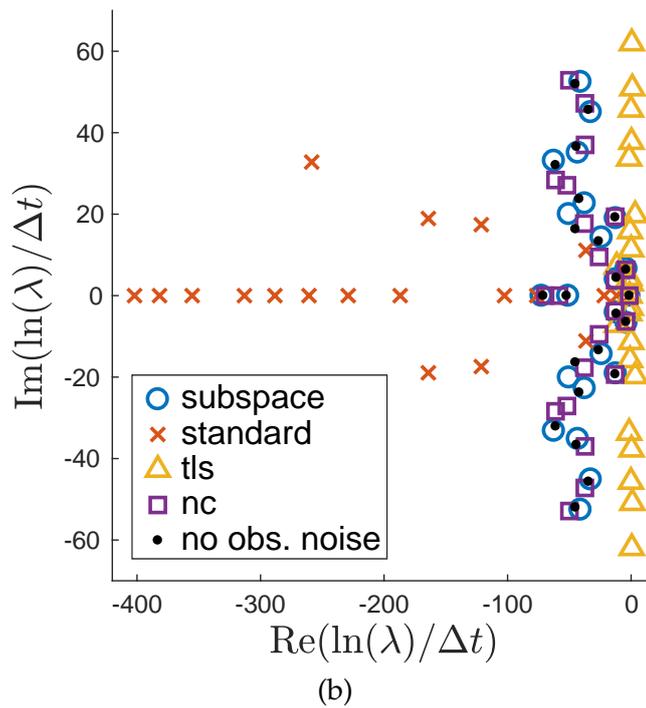
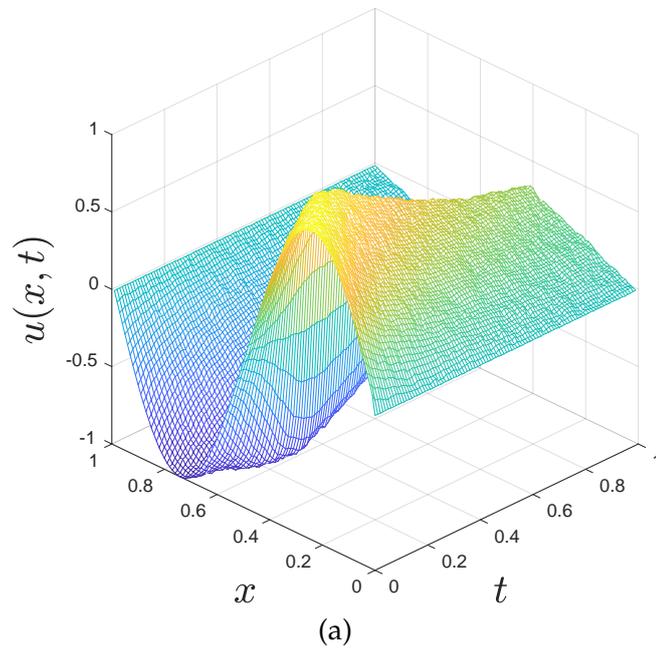


FIGURE 4.2: (a) Data generated by the stochastic Burger's equation with observation noise and (b) the estimated continuous-time eigenvalues. The eigenvalues estimated by subspace DMD agree well with the ones computed with the clean data.

boundary condition $u(0, t) = u(1, t) = 0$, setting the ranges by $x \in [0, 1]$ and $t \in [0, 1]$ and the discretization step sizes by $\Delta x = 1 \times 10^{-2}$ and $\Delta t = 5 \times 10^{-5}$. Based on the approximated solution, we finally generated data with observation noise

$$\mathbf{y}_t = \left[u(0, t) \quad u(\Delta x, t) \quad u(2\Delta x, t) \quad \dots \quad u(1, t) \right]^\top + \mathbf{w}_t, \quad (4.27)$$

where the magnitude of the observation noise was set $\sigma_o = 0.001$. The estimated eigenvalues are plotted in Figure 4.2b. While the eigenvalues obtained by tls-DMD lie approximately on the imaginary axis, the estimation by subspace DMD agrees well with the eigenvalues computed with data that contain no observation noise. Again note that, though the estimation by nc-DMD also aligns with the clean eigenvalues, it requires a precise estimation of observation noise magnitude.

4.3.3 Quantitative investigation of effects of noises

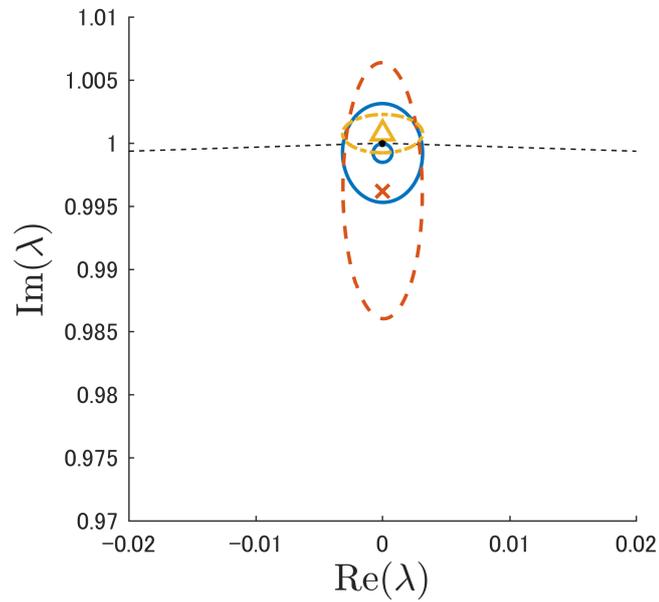
Let us investigate the performance of subspace DMD quantitatively using a simple linear system. We generated data using a linear time-invariant system

$$\mathbf{x}_t = \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{e}_t, \quad \lambda = ri, \quad (4.28)$$

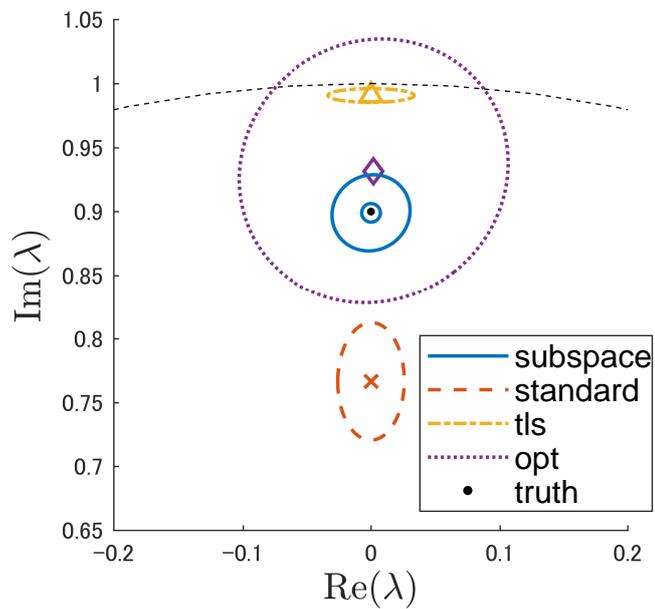
whose Koopman eigenvalues obviously contain λ and $\bar{\lambda}$. Moreover, we used the identity observable with observation noise

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{w}_t. \quad (4.29)$$

We fixed the standard deviation of the process noise to $\sigma_p = 0.1$, the eigenvalue to $\lambda = ri$ with $r = 1.0$ or 0.9 , and the initial state to $\mathbf{x}_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$. Hence, this system exhibits oscillation perpetuated by the process noise when $r = 1.0$ and is damped while being excited by the process noise when $r = 0.9$. We applied subspace DMD, standard DMD, tls-DMD, and optimized DMD (opt-DMD) (Chen et al., 2012) to multiple datasets generated with different random seeds. In those experiments, we have found that opt-DMD is unstable when $r = 1.0$ and it does not output much reasonable results because it needs to compute exponentials of eigenvalues. Hence, the results of opt-DMD when $r = 1.0$ are not plotted in Figures 4.3, 4.4, and 4.5.

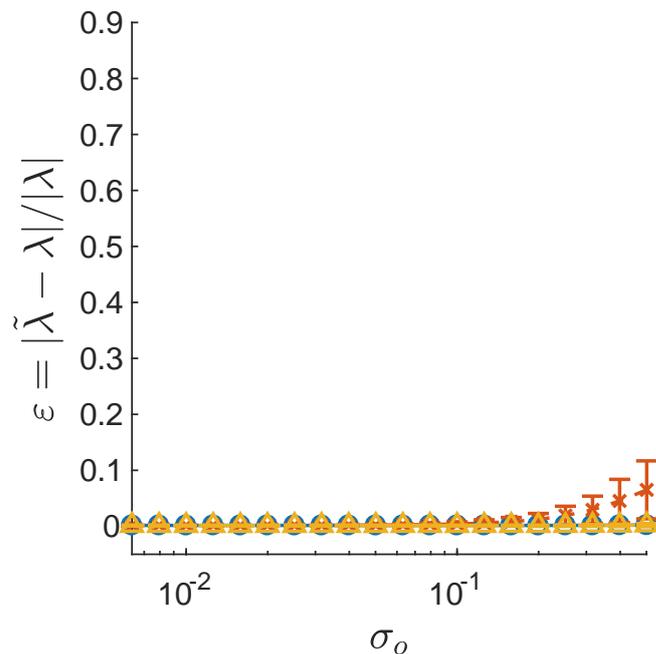


(a)

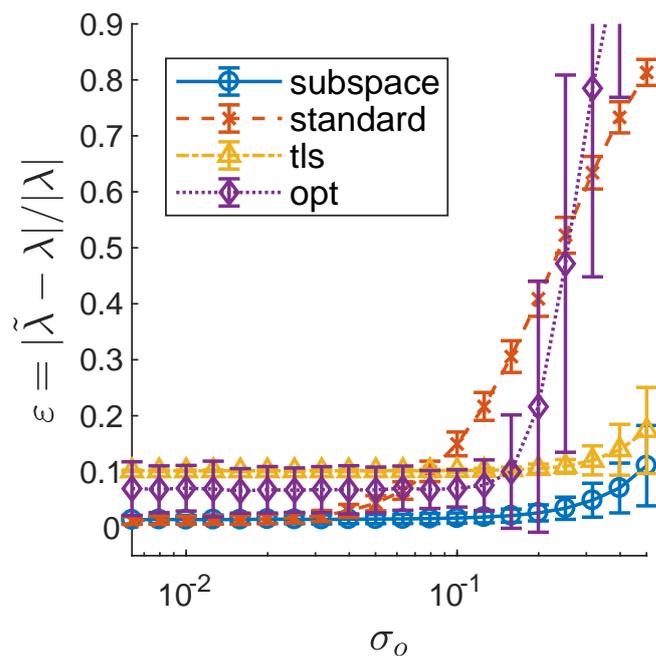


(b)

FIGURE 4.3: The 95% confidence intervals and averages of the eigenvalues estimated by subspace DMD, standard DMD, tls-DMD, and opt-DMD on the linear systems, (a) $r = 1.0$ and (b) $r = 0.9$, with process and observation noises for 1,000 random trials. When $r = 0.9$, only subspace DMD shows consistent results.



(a)



(b)

FIGURE 4.4: Relative errors of the eigenvalues estimated by subspace DMD, standard DMD, tls-DMD, and opt-DMD on the linear systems, (a) $r = 1.0$ and (b) $r = 0.9$, with process and observation noises against different magnitudes of observation noise σ_o . When $r = 0.9$, subspace DMD produces much smaller errors compared to the other two methods.

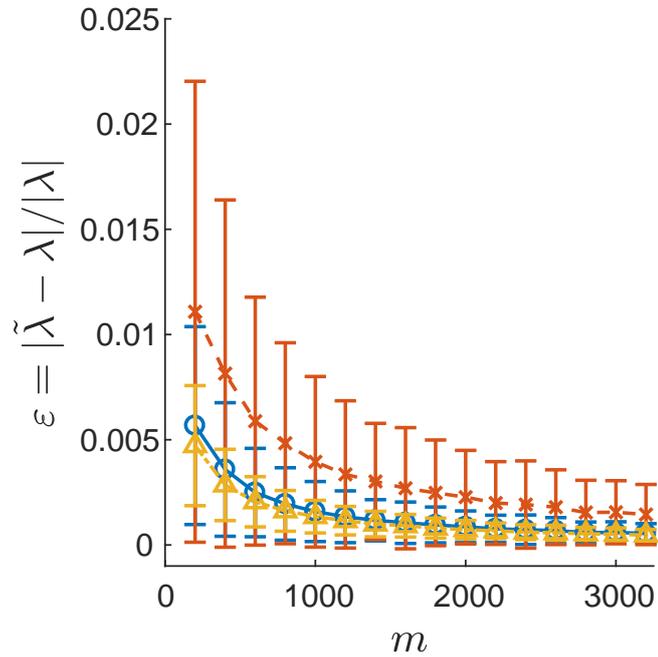
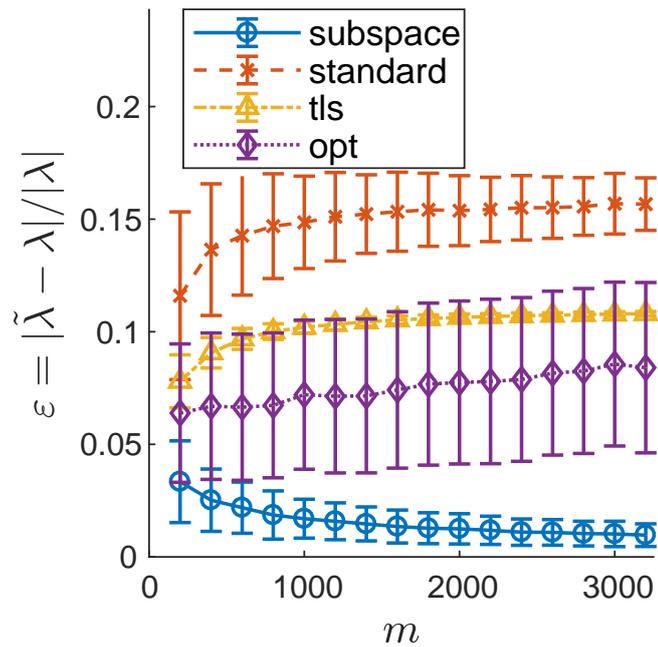
(a) $r = 1.0$ (b) $r = 0.9$

FIGURE 4.5: Relative errors of the eigenvalues estimated by subspace DMD, standard DMD, tls-DMD, and opt-DMD on the linear systems, (a) $r = 1.0$ and (b) $r = 0.9$, with process and observation noises against different numbers of snapshots m fed into the algorithms. When $r = 0.9$, only the output of subspace DMD converges to the true value.

In Figure 4.3, we show the 95% confidence intervals of the estimated eigenvalues for 1,000 random trials with observation noise of magnitude $\sigma_o = 0.1$ and $m = 1,000$ snapshots. When $r = 1.0$, the estimations by subspace DMD, standard DMD, and tls-DMD scatter around the true value, while the results of standard DMD deviate a little more than the others. When $r = 0.9$, the estimations by standard DMD, tls-DMD, and opt-DMD deviate from the true value; only the outputs of subspace DMD distribute around the true value.

In Figure 4.4, the relative errors $\varepsilon = |\tilde{\lambda} - \lambda|/|\lambda|$ of estimated eigenvalues $\tilde{\lambda}$ are plotted against different magnitudes of the observation noise σ_o , with the number of snapshots fed into the algorithms being fixed by $m = 1,000$. When $r = 1.0$, subspace DMD, standard DMD, and tls-DMD work almost equally well. When $r = 0.9$, while the errors of standard DMD and opt-DMD rapidly grow and tls-DMD generates a regular bias, subspace DMD produces almost no bias and is tolerant to the observation noise.

In Figure 4.5, relative errors ε are plotted against different m with fixed $\sigma_o = 0.1$. When $r = 1.0$, subspace DMD, standard DMD, and tls-DMD converge when m becomes large. When $r = 0.9$, only subspace DMD converges, which is expected from Theorem 4.1.

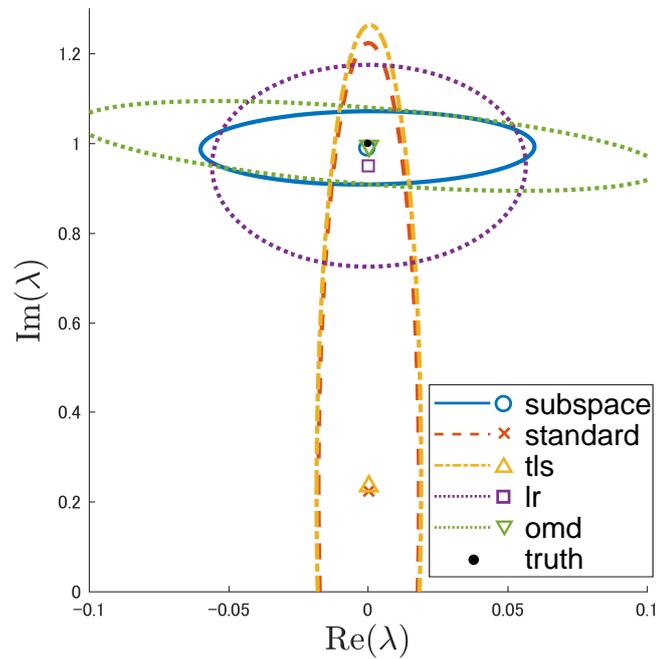
4.3.4 Low-rank high-dimensional data

We have shown the convergence of subspace DMD in the large sample limit in Theorem 4.1, but in practice, DMD is often applied in a high-dimensional setting, where the number of snapshots m is much less than dimensionality of data n . To simulate such circumstances, we generated 500-dimensional data using a linear time-invariant system:

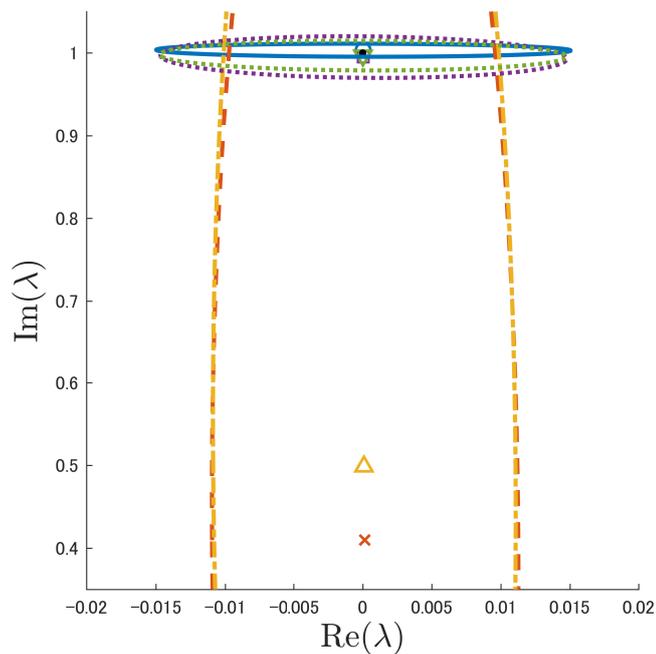
$$\mathbf{x}_t = \mathbf{L} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \mathbf{L}^\top \mathbf{x}_{t-1} + \mathbf{e}_t, \quad (4.30)$$

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{w}_t, \quad (4.31)$$

where $\mathbf{L} \in \mathbb{R}^{500 \times 2}$ satisfies $\mathbf{L}^\top \mathbf{L} = \mathbf{I}$, with $\sigma_p = 0.1$ and $\sigma_o = 0.1$. We prepared two datasets with different sizes, $m = 50$ and $m = 200$, and applied the following DMD variants: subspace DMD, standard DMD, tls-DMD, optimal low-rank DMD (lr-DMD) (Héas and Herzet, 2017), and optimal mode decomposition (OMD) (Wynn et al., 2013). For each method, we introduced



(a)



(b)

FIGURE 4.6: The 95% confidence intervals and averages of the eigenvalues estimated by subspace DMD, standard DMD, tls-DMD, lr-DMD, and OMD on the high-dimensional ($n = 500$) low-rank ($r = 2$) system for 1,000 random trials. The sample sizes are (a) $m = 50$ and (b) $m = 200$. In both cases, the outputs of subspace DMD distribute around the true values denoted by the black filled circle.

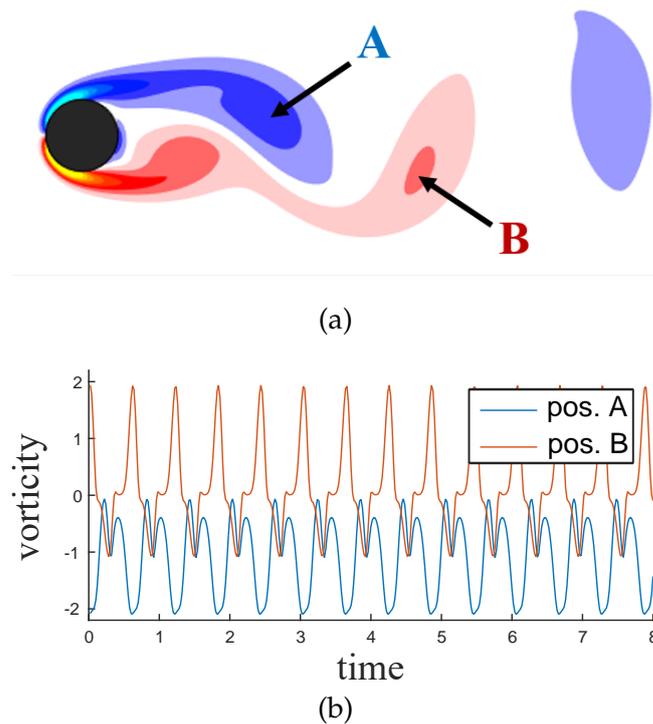


FIGURE 4.7: (a) Example snapshot of the vorticity field in the limit-cycle characterized by the Kármán vortex street. (b) Time variation of the vorticity at locations A and B. Best viewed in color.

the way to obtain a low-rank solution; only the first two POD modes were used in the standard DMD and *tls*-DMD, the rank parameter was set to two in *lr*-DMD and OMD, and only the first two columns of U_q were used in subspace DMD.

In Figure 4.6a, we show the 95% confidence intervals and averages of the estimated eigenvalues for 1,000 random trials with $m = 50$. While the estimations by standard DMD and *tls*-DMD deviate far from the true value because of the process noise (and observation noise), the estimations by subspace DMD, *lr*-DMD, and OMD distribute around the true value. In particular, the variance of the estimation by subspace DMD is smaller than that of *lr*-DMD and OMD. Figure 4.6b shows the results with $m = 200$; in this case, the distributions of the estimations by subspace DMD, *lr*-DMD, and OMD almost coincide.

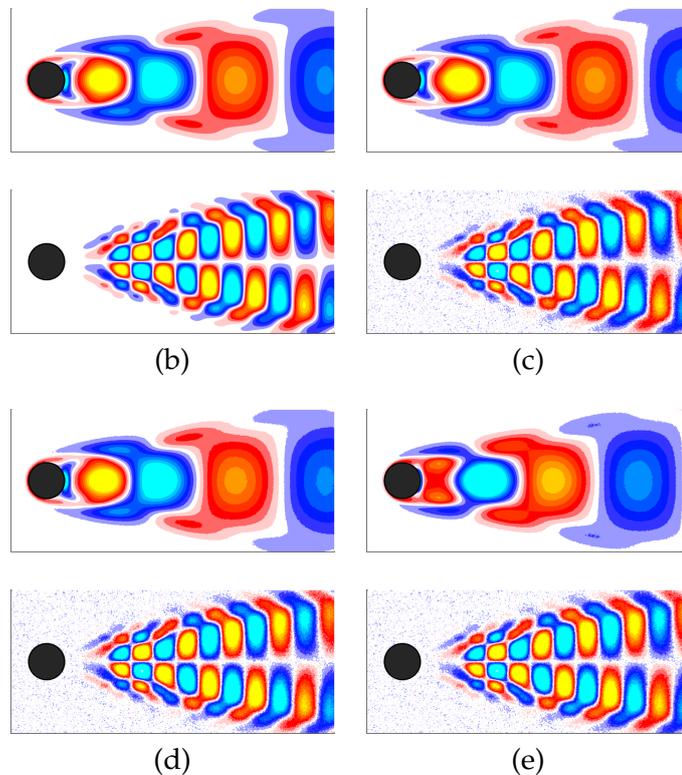
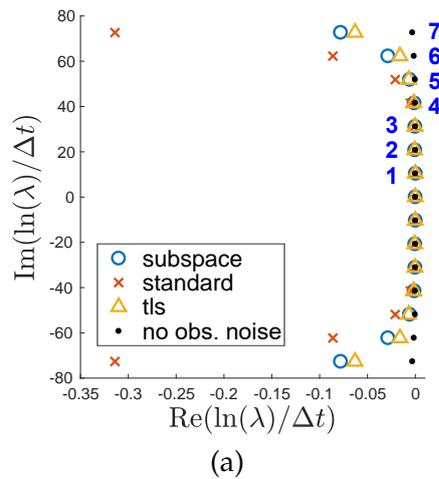


FIGURE 4.8: (a) Continuous-time eigenvalues estimated on the noise-free and noisy datasets, which are numbered from one to seven according to their frequency (i.e., imaginary part). (b) Dynamic modes computed on the noise-free dataset by standard DMD. (c–e) Dynamic modes computed on the noisy dataset by (c) subspace DMD, (d) standard DMD, and (e) tls-DMD. The upper row corresponds to eigenvalue 1 ($\sim 10i$) and the lower row corresponds to eigenvalue 4 ($\sim 40i$). No adversarial effects are present in the results of subspace DMD, even without the process noise. In (b–e), the magnitude of the dynamic modes are normalized to a common color scheme. Best viewed in color.

4.3.5 Application: Cylinder wake

As an example of an application, we applied subspace DMD to a two dimensional flow past a circular cylinder with Reynolds number $Re = 100$. We generated data using a solver based on the fast immersed boundary method with the multi-domain technique (Taira and Colonius, 2007; Colonius and Taira, 2008) with four nested grids, each of which contains 450×200 points. The diameter of the cylinder corresponds to 50 points in the finest grid. The solver uses the third-order Runge–Kutta method with time-step $\Delta t = 0.02$. We collected 400 snapshots of the vorticity fields with intervals of size $10\Delta t$ from the limit cycle characterized by the Kármán vortex street. An example of the snapshots (without observation noise) and the time-variation of vorticity at two locations (A and B) are shown in Figure 4.7. We applied subspace DMD, standard DMD, and tls-DMD to the data contaminated with observation noise of $\sigma_o = 0.1$. Every method was run with a low-rank approximation of $r = 15$ because the first 15 POD modes contained about 99.9% of the energy of the original data.

In Figure 4.8a, the eigenvalues estimated with the noisy dataset and the noise-free dataset are plotted; subspace DMD and tls-DMD generate smaller biases than standard DMD does. The eigenvalues are numbered from one to seven in Figure 4.8a, according to their frequency (the magnitude of the imaginary part). In the remainder of Figure 4.8, we show the dynamic modes corresponding to eigenvalue 1 ($\sim 10i$) in the upper row and eigenvalue 4 ($\sim 40i$) in the lower row. We confirm that no adversarial effect is present in the dynamic modes computed by subspace DMD. Note that, in this cylinder wake experiment, no process noise (except for small errors due to the numerical integration) is involved. Subspace DMD is also applicable to classical (but frequent) situations of data analysis like this, where almost no process noise is present.

4.4 Summary

In this work, we developed *subspace DMD*, an algorithm for stochastic Koopman analysis with noisy observations. We have shown that the output of the proposed algorithm converges to the spectra of the stochastic Koopman operator in the large sample limit even if both process noise and observation

noise are present. Moreover, we have shown its empirical performance with the numerical examples on different types of random dynamical systems.

Chapter 5

Learning Koopman invariant subspaces for DMD

5.1 Introduction

One of the most popular algorithms for modal decomposition based on the Koopman operator is dynamic mode decomposition (DMD) (Rowley et al., 2009; Schmid, 2010; Kutz et al., 2016a). An important premise of DMD is that the target dataset is generated from a set of observables that spans a function space invariant to the Koopman operator (referred to as Koopman invariant subspace), as stated in Section 1.4.2.

Here, an important problem in the practice of DMD arises, i.e., we often have no access to a set of observables that spans a Koopman invariant subspace. In this case, for nonlinear dynamics, we must *manually* prepare adequate observables according to the underlying dynamics. Several researchers have addressed this issue; Williams et al. (2015a) leveraged a dictionary of predefined basis functions to transform original data, and Kawahara (2016) defined Koopman spectral analysis in a reproducing kernel Hilbert space. Brunton et al. (2016b) proposed the use of observables selected in a data-driven manner (Brunton et al., 2016c) from a function dictionary. Note that, for these methods, we must select an appropriate function dictionary or kernel function according to the target dynamics. However, if we have no *a priori* knowledge about them, which is often the case, such existing methods do not have to be applied successfully to nonlinear dynamics.

In this chapter, we propose a *fully data-driven* method for modal decomposition via the Koopman operator based on the principle of *learning Koopman invariant subspaces* (LKIS) from scratch using observed data. To this end,

we estimate a set of parametric functions by minimizing the residual sum of squares (RSS) of linear least-squares regression, so that the estimated set of functions transforms the original data into a form in which the linear regression fits well. In addition to the principle of LKIS, an implementation using neural networks is described. Moreover, we introduce empirical performance of DMD based on the LKIS framework with several nonlinear dynamical systems and applications, which proves the feasibility of LKIS-based DMD as a fully data-driven method for modal decomposition via the Koopman operator.

The remaining part of this chapter is organized as follows. In Section 5.2, the main concept, learning Koopman invariant subspaces is formulated. Afterward, we review some related work in Section 5.3. Section 5.4 is a preliminary part to present setups of the experiments in the subsequent sections. In Section 5.5, numerical examples are provided to show the feasibility of the proposed framework. In Section 5.6, simple applications are shown. A summary of this chapter is provided in Section 5.7.

5.2 Learning Koopman invariant subspaces

5.2.1 Minimizing residual sum of squares of linear least-squares regression

In this chapter, we propose a method to learn a set of observables $\{g_1, \dots, g_n\}$ that spans a Koopman invariant subspace G , given a sequence of measurements as the dataset. In the following, we summarize desirable properties for such observables, upon which the proposed method is constructed.

Theorem 5.1. *Consider a set of square-integrable observables $\{g_1, \dots, g_n\}$, and define a vector-valued observable $\mathbf{g} = [g_1 \ \dots \ g_n]^\top$. In addition, define a linear operator U whose matrix form is given as $\mathbf{U} = (\int_{\mathcal{M}} (\mathbf{g} \circ \mathbf{f}) \mathbf{g}^H d\mu) (\int_{\mathcal{M}} \mathbf{g} \mathbf{g}^H d\mu)^\dagger$. Then, $\forall \mathbf{x} \in \mathcal{M}$, $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{U} \mathbf{g}(\mathbf{x})$ if and only if $\{g_1, \dots, g_n\}$ spans a Koopman invariant subspace.*

Proof. If $\forall \mathbf{x} \in \mathcal{M}$, $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{U} \mathbf{g}(\mathbf{x})$, then for any $\hat{g} = \sum_{i=1}^n a_i g_i \in \text{span}\{g_1, \dots, g_n\}$,

$$\mathcal{K} \hat{g} = \sum_{i=1}^n a_i g_i(\mathbf{f}(\mathbf{x})) = \sum_{j=1}^n \left(\sum_{i=1}^n a_i [\mathbf{U}]_{i,j} \right) g_j(\mathbf{x}) \in \text{span}\{g_1, \dots, g_n\},$$

where $[U]_{i,j}$ denotes the (i, j) -element of U ; thus, $\text{span}\{g_1, \dots, g_n\}$ is a Koopman invariant subspace. On the other hand, if $\{g_1, \dots, g_n\}$ spans a Koopman invariant subspace, there exists a linear operator K such that $\forall \mathbf{x} \in \mathcal{M}$, $\mathbf{g}(\mathbf{f}(\mathbf{x})) = K\mathbf{g}(\mathbf{x})$; thus, $\int_{\mathcal{M}}(\mathbf{g} \circ \mathbf{f})\mathbf{g}^H d\mu = \int_{\mathcal{M}} K\mathbf{g}\mathbf{g}^H d\mu$. Therefore, an instance of the matrix form of K is obtained in the form of U . \square

According to Theorem 5.1, we should obtain \mathbf{g} that makes $\mathbf{g} \circ \mathbf{f} - U\mathbf{g}$ zero. However, such problems cannot be solved with finite data because \mathbf{g} is a function. Thus, we give the corresponding empirical risk minimization problem based on the assumption of ergodicity of \mathbf{f} and the convergence property of the empirical matrix as follows.

Theorem 5.2. Define \mathbf{Y}_0 and \mathbf{Y}_1 as in Eq. (1.14) whose snapshots are generated with dynamics \mathbf{f} that suffice the ergodicity (Assumption 1.2). If all modes are sufficiently excited in the data (i.e., $\text{rank}(\mathbf{Y}_0) = n$), then matrix $\mathbf{A} = \mathbf{Y}_1\mathbf{Y}_0^\dagger$ almost surely converges to the matrix form of linear operator U in $m \rightarrow \infty$.

Proof. From Assumption 1.2, $\frac{1}{m}\mathbf{Y}_1\mathbf{Y}_0^H$ and $\frac{1}{m}\mathbf{Y}_0\mathbf{Y}_0^H$ respectively converge to $\int_{\mathcal{M}}(\mathbf{g} \circ \mathbf{f})\mathbf{g}^H d\mu$ and $\int_{\mathcal{M}}\mathbf{g}\mathbf{g}^H d\mu$ for almost all $\mathbf{x}_0 \in \mathcal{M}$. In addition, since the rank of $\mathbf{Y}_0\mathbf{Y}_0^H$ is always n , $(\frac{1}{m}\mathbf{Y}_0\mathbf{Y}_0^H)^\dagger$ converges to $(\int_{\mathcal{M}}\mathbf{g}\mathbf{g}^H d\mu)^\dagger$ in $m \rightarrow \infty$ (Rakočević, 1997). Consequently, in $m \rightarrow \infty$, $\mathbf{A} = (\frac{1}{m}\mathbf{Y}_1\mathbf{Y}_0^H)(\frac{1}{m}\mathbf{Y}_0\mathbf{Y}_0^H)^\dagger$ almost surely converges to U , which is the matrix form of linear operator U . \square

Remark 5.1. Note that, differently from previous chapters, observables \mathbf{g} that generate data matrices \mathbf{Y}_0 and \mathbf{Y}_1 do not have to span a Koopman invariant subspace here. Instead, our purpose is to *learn* \mathbf{g} that spans a Koopman invariant subspace.

Since $\mathbf{A} = \mathbf{Y}_1\mathbf{Y}_0^\dagger$ is the minimum-norm solution of the linear least-squares regression from the columns of \mathbf{Y}_0 to those of \mathbf{Y}_1 , we constitute the learning problem to estimate a set of function that transforms the original data into a form in which the linear least-squares regression fits well. In particular, we minimize RSS, which measures the discrepancy between the data and the estimated regression model (i.e., linear least-squares in this case). We define the RSS loss as follows:

$$\mathcal{L}_{\text{RSS}}(\mathbf{g}; (\mathbf{x}_0, \dots, \mathbf{x}_m)) = \left\| \mathbf{Y}_1 - (\mathbf{Y}_1\mathbf{Y}_0^\dagger)\mathbf{Y}_0 \right\|_{\mathbb{F}}^2, \quad (5.1)$$

which becomes zero when \mathbf{g} spans a Koopman invariant subspace. If we implement a smooth parametric model on \mathbf{g} , the local minima of \mathcal{L}_{RSS} can be found using gradient descent. We adopt \mathbf{g} that achieves a local minimum of \mathcal{L}_{RSS} as a set of observables that spans (approximately) a Koopman invariant subspace.

5.2.2 Linear delay embedder for state space reconstruction

In the previous subsection, we have presented an important part of the principle of LKIS, i.e., minimization of the RSS of linear least-squares regression. Note that, to define RSS loss (5.1), we need access to a sequence of the original states, i.e., $(\mathbf{x}_0, \dots, \mathbf{x}_m) \in \mathcal{M}^{m+1}$, as a dataset. In practice, however, we cannot necessarily observe full states \mathbf{x} due to limited memory and sensor capabilities. In this case, only transformed (and possibly degenerated) measurements are available, which we denote $\mathbf{y} = \psi(\mathbf{x})$ with a measurement function $\psi : \mathcal{M} \rightarrow \mathbb{R}^r$. To define RSS loss (5.1) given only degenerated measurements, we must reconstruct the original states \mathbf{x} from the actual observations \mathbf{y} .¹

Here, we utilize delay-coordinate embedding, which has been widely used for state space reconstruction in the analysis of nonlinear dynamics. Consider a univariate time-series $(\dots, y_{t-1}, y_t, y_{t+1}, \dots)$, which is a sequence of degenerated measurements $y_t = \psi(\mathbf{x}_t)$. According to the well-known Taken's theorem (Takens, 1981; Sauer et al., 1991), a faithful representation of \mathbf{x}_t that preserves the structure of the state space can be obtained by

$$\tilde{\mathbf{x}}_t = \begin{bmatrix} y_t & y_{t-\tau} & \cdots & y_{t-(d-1)\tau} \end{bmatrix}^\top \quad (5.2)$$

with some lag parameter τ and embedding dimension d if d is greater than $2 \dim(\mathbf{x})$. For a multivariate time-series, embedding with non-uniform lags provides better reconstruction (Garcia and Almeida, 2005). For example, when we have a two-dimensional time-series

$$\mathbf{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \quad (5.3)$$

¹Please be careful to not confuse observables \mathbf{g} (in terms of Koopman analysis) and measurement functions ψ (in terms of transformed and degenerated measurements).

an embedding with non-uniform lags is similar to

$$\tilde{\mathbf{x}}_t = \left[y_{1,t} \quad y_{1,t-\tau_{11}} \quad \cdots \quad y_{1,t-\tau_{1d_1}} \quad y_{2,t} \quad y_{2,t-\tau_{21}} \quad \cdots \quad y_{2,t-\tau_{2d_2}} \right]^T \quad (5.4)$$

with each value of τ and d . Several methods have been proposed for selection of τ and d (Garcia and Almeida, 2005; Hirata et al., 2006; Vlachos and Kugiumtzis, 2010); however, appropriate values may depend on the given application (attractor inspection, prediction, etc.).

We propose to surrogate the parameter selection of the delay-coordinate embedding by learning a *linear delay embedder* from data. Formally, we learn embedder ϕ such that

$$\tilde{\mathbf{x}}_t = \phi(\mathbf{y}_t^{(k)}) = \mathbf{W}_\phi \left[\mathbf{y}_t^T \quad \mathbf{y}_{t-1}^T \quad \cdots \quad \mathbf{y}_{t-k+1}^T \right]^T, \quad \mathbf{W}_\phi \in \mathbb{R}^{p \times kr}, \quad (5.5)$$

where $p = \dim(\tilde{\mathbf{x}})$, $r = \dim(\mathbf{y})$, and k is a hyperparameter of maximum lag. We estimate weight \mathbf{W}_ϕ as well as the parameters of \mathbf{g} by minimizing RSS loss (5.1), which is now defined using $\tilde{\mathbf{x}}$ instead of \mathbf{x} . Learning ϕ from data yields an embedding that is suitable for learning a Koopman invariant subspace. Moreover, we can impose L1 regularization on weight \mathbf{W}_ϕ to make it highly interpretable if necessary according to the given application.

5.2.3 Reconstruction of original measurements

Simple minimization of \mathcal{L}_{RSS} may yield trivial \mathbf{g} , such as constant values. We should impose some constraints to prevent such trivial solutions. In the proposed framework, modal decomposition is first obtained in terms of learned observables \mathbf{g} ; thus, the values of \mathbf{g} must be back-projected to the space of the original measurements \mathbf{y} to obtain a physically meaningful representation of the dynamic modes. Therefore, we modify the loss function by employing an additional term such that the original measurements \mathbf{y} can be reconstructed from the values of \mathbf{g} by a *reconstructor* \mathbf{h} , i.e., $\mathbf{y} \approx \mathbf{h}(\mathbf{g}(\tilde{\mathbf{x}}))$. Such term is given as follows:

$$\mathcal{L}_{\text{rec}}(\mathbf{h}, \mathbf{g}; (\tilde{\mathbf{x}}_0, \dots, \tilde{\mathbf{x}}_m)) = \sum_{j=0}^m \|\mathbf{y}_j - \mathbf{h}(\mathbf{g}(\tilde{\mathbf{x}}_j))\|^2, \quad (5.6)$$

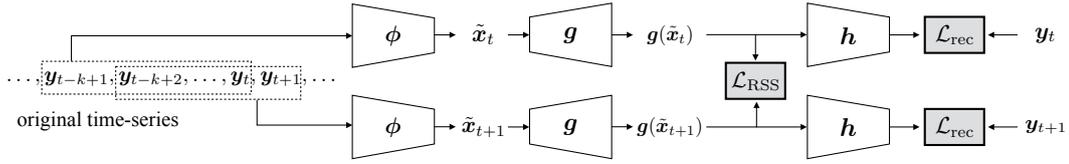


FIGURE 5.1: An instance of LKIS framework, in which g and h are implemented by MLPs.

and, if h is a smooth parametric model, this term can also be reduced using gradient descent. Finally, the objective function to be minimized becomes

$$\begin{aligned} \mathcal{L}(\phi, \mathbf{g}, \mathbf{h}; (\mathbf{y}_0, \dots, \mathbf{y}_m)) \\ = \mathcal{L}_{\text{RSS}}(\mathbf{g}, \phi; (\tilde{\mathbf{x}}_{k-1}, \dots, \tilde{\mathbf{x}}_m)) + \alpha \mathcal{L}_{\text{rec}}(\mathbf{h}, \mathbf{g}; (\tilde{\mathbf{x}}_{k-1}, \dots, \tilde{\mathbf{x}}_m)), \end{aligned} \quad (5.7)$$

where α is a parameter that controls the balance between \mathcal{L}_{RSS} and \mathcal{L}_{rec} . We empirically found that the quality of obtained observables was not so sensitive to the value of α .

5.2.4 Implementation using neural networks

In Sections 5.2.1–5.2.3, we introduced the main concepts for the LKIS framework, i.e., RSS loss minimization, learning the linear delay embedder, and reconstruction of the original measurements. Here, we demonstrate an implementation of the LKIS framework using neural networks.

Figure 5.1 shows a schematic diagram of the implementation of the framework. We model g and h using multi-layer perceptrons (MLPs) with a parametric ReLU activation function (He et al., 2015). Here, the sizes of the hidden layer of MLPs are defined by the arithmetic means of the sizes of the input and output layers of the MLPs. Thus, the remaining tunable hyperparameters are k (maximum delay of ϕ), p (dimensionality of $\tilde{\mathbf{x}}$), and n (dimensionality of g). To obtain g with dimensionality much greater than that of the original measurements, we found that it was useful to set $k > 1$ even when full-state measurements (e.g., $\mathbf{y} = \mathbf{x}$) were available. Empirically, we found that setting $n \gg p$ often resulted in unstable optimization.

After estimating the parameters of ϕ , g , and h , DMD can be performed normally by using the values of the learned g , defining the data matrices analogously to Eq. (1.14) (with values of g), and computing the eigendecomposition of $\mathbf{A} = \mathbf{Y}_1 \mathbf{Y}_0^\dagger$; the dynamic modes are obtained by \mathbf{w} , and the values

of the eigenfunctions are obtained by $\varphi = z^H g$, where w and z are the right- and left-eigenvectors of A .

In the numerical experiments described in Sections 5.5 and 5.6, we performed optimization using first-order gradient descent. To stabilize optimization, batch normalization (Ioffe and Szegedy, 2015) was imposed on the inputs of hidden layers. Note that, since RSS loss function (5.1) is *not* decomposable with regard to data points, convergence of stochastic gradient descent (SGD) cannot be shown straightforwardly. However, we empirically found that the non-decomposable RSS loss was often reduced successfully, even with mini-batch SGD. Let us show an example; the full-batch RSS loss (denoted $\mathcal{L}_{\text{RSS}}^*$) under the updates of the mini-batch SGD are plotted in Figure 5.4e. Here, $\mathcal{L}_{\text{RSS}}^*$ decreases rapidly and remains small. For SGD on non-decomposable losses, Kar et al. (2014) provided guarantees for some cases; however, examining the behavior of more general non-decomposable losses under mini-batch updates remains an open problem.

5.3 Related work

The proposed framework is motivated by the operator-theoretic view of nonlinear dynamical systems. In contrast, learning a generative (state-space) model for nonlinear dynamical systems directly has been actively studied in machine learning and optimal control communities, on which we mention a few examples. A classical but popular method for learning nonlinear dynamical systems is using an expectation-maximization algorithm with Bayesian filtering/smoothing (see, e.g., (Ghahramani and Roweis, 1999)). Recently, using approximate Bayesian inference with the variational autoencoder (VAE) technique (Kingma and Welling, 2014) to learn generative dynamical models has been actively researched. Chung et al. (2015) proposed a recurrent neural network with random latent variables, Gao et al. (2016) utilized VAE-based inference for neural population models, and Johnson et al. (2016) and Krishnan et al. (2017) developed inference methods for structured models based on inference with a VAE. In addition, Karl et al. (2017) proposed a method to obtain a more consistent estimation of nonlinear state space models. Moreover, Watter et al. (2015) proposed a similar approach in the context of optimal control. Since generative models are intrinsically aware of process and observation noises, incorporating methodologies developed in such

studies to the operator-theoretic perspective is an important open challenge to explicitly deal with uncertainty.

We would like to mention some studies closely related to our method. After the first submission of the manuscript of this chapter (in May 2017), several similar approaches to learning data transform for Koopman analysis have been proposed [Li et al. \(2017\)](#); [Yeung et al. \(2017\)](#); [Mardt et al. \(2017\)](#); [Otto and Rowley \(2017\)](#); [Lusch et al. \(2017\)](#). The relationships and relative advantages of these methods should be elaborated in the future.

5.4 Setup of numerical examples and applications

In this section, the experiment configurations engaged in the numerical examples (Section 5.5) and applications (Section 5.6) are described.

Hyperparameters

In each experiment, parameter α was fixed at 0.01. Note that the quality of the results was not sensitive to the values of α . We modeled \mathbf{g} and \mathbf{h} with multi-layer perceptrons by setting the number of hidden nodes (denoted n_h) as the arithmetic means of the input and output sizes, i.e., $n_h = \text{round}((p + n)/2)$ for \mathbf{g} and $n_h = \text{round}((n + r)/2)$ for \mathbf{h} , where $r = \dim(\mathbf{y})$, $p = \dim(\tilde{\mathbf{x}})$, and $n = \dim(\mathbf{g})$. Therefore, the remaining hyperparameters to be tuned were k (maximum lag), p , and n . However, unless otherwise noted, we fixed p by $p = kr$. Consequently, the independent hyperparameters were k and n .

Preprocessing

One *must not subtract the mean* from the original data because subtracting something from the data may change the spectra of the underlying dynamical systems (see, e.g., [Chen et al. \(2012\)](#)). If the absolute values of the data were too large, we simply divided the data by the maximum absolute value for each series.

Optimization

In optimization, we found that the adaptive learning rate by SMORMS3 (Funk, 2015) achieved fast convergence compared to a fixed learning rate and other adaptation techniques. The maximum learning rate of SMORMS3 was selected from 10^{-3} to 10^{-2} in each experiment according to the amount of data. In some cases, optimization was performed in two stages: the parameters of ϕ , g , and h were updated in the first stage, and, in the second stage, the parameters of ϕ and g were fixed and only h was updated. This two-stage optimization was particularly useful for the application of prediction, where a precise reconstruction of the original measurements was necessary. Moreover, when the original states \mathbf{x} of the dynamical system were available and used without delay (i.e., $k = 1$ and $p = r$), parameter \mathbf{W}_ϕ of the linear embedder was fixed to be an identity matrix (i.e., no embedder was used). Also, we set the mini-batch size from 100 to 500 because smaller mini-batches often led to an unstable computation of pseudo-inverse.

5.5 Numerical examples

In this section, we provide numerical examples of DMD based on the LKIS framework (LKIS-DMD) implemented using neural networks. We conducted experiments on three typical nonlinear dynamical systems: a fixed-point attractor, a limit-cycle attractor, and a system with multiple basins of attraction. We show the results of comparisons with other recent DMD algorithms, i.e., Hankel DMD (Arbabi and Mezić, 2017; Susuki and Mezić, 2015), extended DMD (Williams et al., 2015a), and DMD with reproducing kernels (Kawahara, 2016).

Fixed-point attractor

Consider a two-dimensional nonlinear map on $\mathbf{x}_t = \begin{bmatrix} x_{1,t} & x_{2,t} \end{bmatrix}^\top$:

$$\begin{aligned} x_{1,t+1} &= \lambda x_{1,t}, \\ x_{2,t+1} &= \mu x_{2,t} + (\lambda^2 - \mu)x_{1,t}^2, \end{aligned} \tag{5.8}$$

which has a stable equilibrium at the origin if $\lambda, \mu < 1$. The Koopman eigenvalues of system (5.8) include λ and μ , and the corresponding eigenfunctions

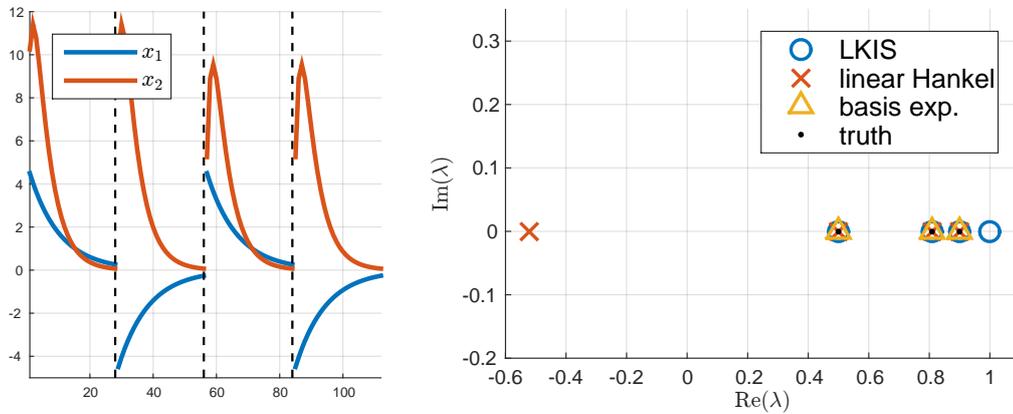


FIGURE 5.2: (left) Data generated from system (5.8) and (right) the estimated Koopman eigenvalues. While linear Hankel DMD produces an inconsistent eigenvalue, LKIS-DMD successfully identifies λ , μ , λ^2 , and $\lambda^0\mu^0 = 1$.

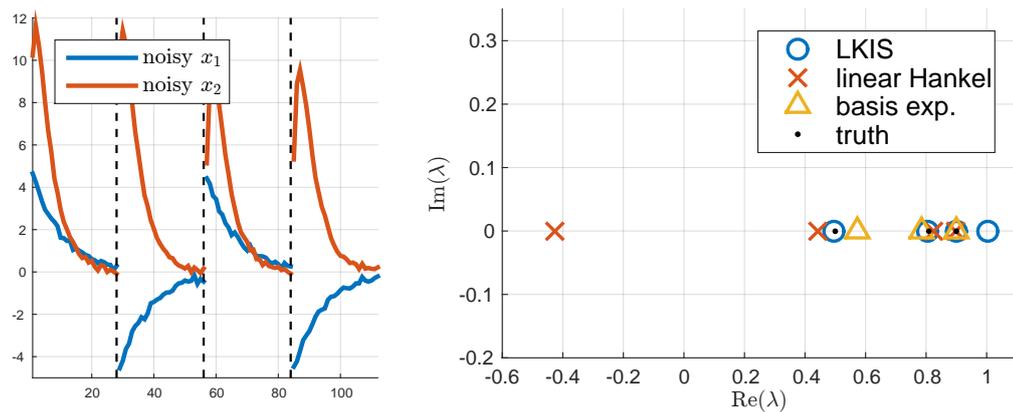


FIGURE 5.3: (left) Data generated from system (5.8) and white Gaussian observation noise and (right) the estimated Koopman eigenvalues. LKIS-DMD successfully identifies the eigenvalues even with the observation noise.

are $\varphi_\lambda(\mathbf{x}) = x_1$ and $\varphi_\mu(\mathbf{x}) = x_2 - x_1^2$, respectively. $\lambda^i \mu^j$ is also an eigenvalue with corresponding eigenfunction $\varphi_\lambda^i \varphi_\mu^j$. A minimal Koopman invariant subspace of system (5.8) is $\text{span}\{x_1, x_2, x_1^2\}$, and the eigenvalues of the Koopman operator restricted to such subspace include λ , μ and λ^2 .

We generated a dataset using system (5.8) with $\lambda = 0.9$ and $\mu = 0.5$ and applied LKIS-DMD ($n = 4$), linear Hankel DMD (Arbabi and Mezić, 2017; Susuki and Mezić, 2015) (delay 2), and DMD with basis expansion by $\{x_1, x_2, x_1^2\}$, which corresponds to extended DMD (Williams et al., 2015a) with a right and minimal observable dictionary. The estimated Koopman eigenvalues are shown in Figure 5.2, wherein LKIS-DMD successfully identifies the eigenvalues of the target invariant subspace. In Figure 5.3, we show eigenvalues estimated using data contaminated with white Gaussian observation noise ($\sigma = 0.1$). The eigenvalues estimated by LKIS-DMD coincide with the true values even with the observation noise, whereas the results of DMD with basis expansion (i.e., extended DMD) are directly affected by the observation noise.

The more detailed setup in this experiment is as follows. The data were generated with four initial values: $[5 \ 5]^\top$, $[-5 \ 5]^\top$, $[5 \ -5]^\top$, and $[-5 \ 5]^\top$, with the length of each episode being 30. In the case of noisy dataset, the standard deviation of the observation noise was set to 0.1. In both experiments (with and without observation noise), we set $k = 2$ and $n = 4$ to cover the minimal three-dimensional Koopman invariant subspace.

Limit-cycle attractor

We generated data from the limit cycle of the FitzHugh–Nagumo equation

$$\begin{aligned} \dot{x}_1 &= x_1^3/3 + x_1 - x_2 + I, \\ \dot{x}_2 &= c(x_1 - bx_2 + a), \end{aligned} \tag{5.9}$$

where $a = 0.7$, $b = 0.8$, $c = 0.08$, and $I = 0.8$. Since trajectories in a limit-cycle are periodic, the (discrete-time) Koopman eigenvalues should lie near the unit circle. Figures 5.4a, 5.4b, 5.4c, and 5.4d show the eigenvalues estimated by LKIS-DMD ($n = 16$), linear Hankel DMD (Arbabi and Mezić, 2017; Susuki and Mezić, 2015) (delay 8), and DMDs with reproducing kernels (Kawahara, 2016) (polynomial kernel of degree 4 and RBF kernel of width 1), respectively. The eigenvalues produced by LKIS-DMD agree well with those produced by

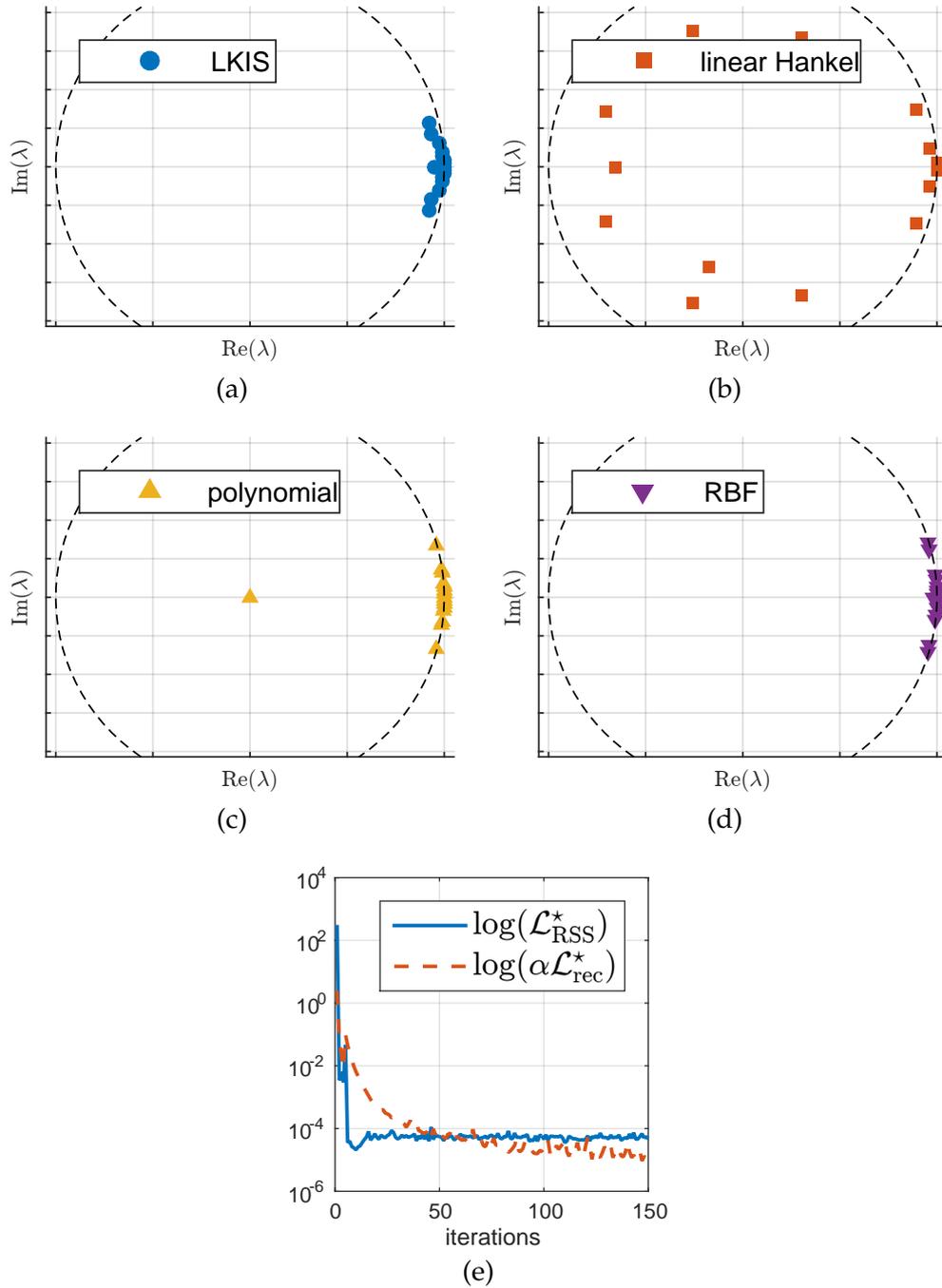


FIGURE 5.4: (a)–(d) Estimated Koopman eigenvalues on the limit-cycle of the FitzHugh-Nagumo equation by LKIS-DMD, linear Hankel DMD, and kernel DMDs with polynomial and RBF kernels. The hyperparameters of each DMD are set to produce 16 eigenvalues. (e) Full-batch (size 2,000) loss under mini-batch (size 200) SGD updates along iterations. Non-decomposable part $\mathcal{L}_{\text{RSS}}^*$ decreases rapidly and remains small, even by SGD.

kernel DMDs, whereas linear Hankel DMD produces eigenvalues that would correspond to rapidly decaying modes.

The more detailed setup in this experiment is as follows. The data were generated using MATLAB's `ode45` function (Shampine and Reichelt, 1997), which was run with time-step $\Delta t = 0.1$ and initial value $\mathbf{x}_0 = [1 \ 1.6]^\top$ for 2,000 steps. The hyperparameters of LKIS-DMD, linear Hankel DMD, and kernel DMDs were set such that they produced 16 eigenvalues, i.e., $k = 8$ and $n = 16$ for LKIS-DMD, and POD modes whose singular value was less than ε were disposed in kernel DMDs ($\varepsilon = 0.0001$ for the polynomial kernel and $\varepsilon = 0.05$ for the RBF kernel).

Multiple basins of attraction

Consider the unforced Duffing equation

$$\ddot{x} = -\delta\dot{x} - x(\beta + \alpha x^2), \quad \mathbf{x} = \begin{bmatrix} x & \dot{x} \end{bmatrix}^\top, \quad (5.10)$$

where $\alpha = 1$, $\beta = -1$, and $\delta = 0.5$. States \mathbf{x} following (5.10) evolve toward $\begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ or $\begin{bmatrix} -1 & 0 \end{bmatrix}^\top$ depending on which basin of attraction the initial value belongs to unless the initial state is on the stable manifold of the saddle. Generally, a Koopman eigenfunction whose continuous-time eigenvalue is zero takes a constant value in each basin of attraction (Williams et al., 2015a); thus, the contour plot of such an eigenfunction shows the boundary of the basins of attraction.

We generated 1,000 episodes of time-series starting at different initial values uniformly sampled from $[-2, 2]^2$. The top plot in Figure 5.5 shows the continuous-time Koopman eigenvalues estimated by LKIS-DMD ($n = 100$), all of which correspond to decaying modes (i.e., negative real parts) and agree with the property of the data. The middle plot in Figure 5.5 shows the true basins of attraction of (5.10), and the bottom plot shows the estimated values of the eigenfunction corresponding to the eigenvalue of the smallest magnitude. The surface of the estimated eigenfunction agrees qualitatively with the true boundary of the basins of attractions, which indicates that LKIS-DMD successfully identifies the Koopman eigenfunction.

The more detailed setup in this experiment is as follows. The data were generated using the settings provided in the literature (Williams et al., 2015a);

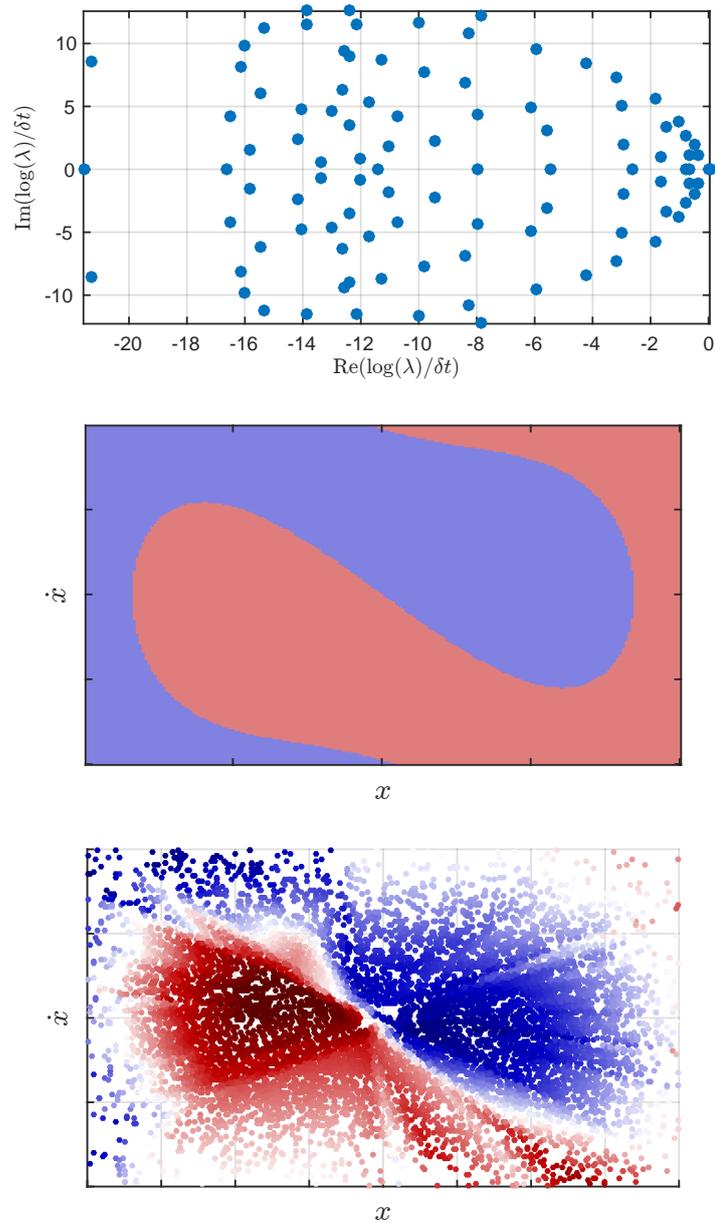


FIGURE 5.5: (*top*) The continuous-time Koopman eigenvalues estimated by LKIS-DMD on the Duffing equation. (*middle*) The true basins of attraction of the Duffing equation, wherein points in the blue region evolve toward $(1, 0)$ and points in the red region evolve toward $(-1, 0)$. Note that the stable manifold of the saddle point is not drawn precisely. (*bottom*) The values of the Koopman eigenfunction with a nearly zero eigenvalue computed by LKIS-DMD, whose level sets should correspond to the basins of attraction. There is rough agreement between the true boundary of the basins of attraction and the numerically computed boundary. The bottom two plots are best viewed in color.

1,000 initial values were drawn from the uniform distribution on $[-2, 2] \times [-2, 2]$ and each initial value was proceeded in time for 11 steps with $\Delta t = 0.25$. We used MATLAB's `ode45` function for numerical integration. For LKIS-DMD, we set $k = 1$ and $n = 100$. Note that the values of the estimated eigenfunction were evaluated and plotted in consideration of each data point.

5.6 Applications

The numerical experiments in the previous section demonstrated the feasibility of the proposed method as a fully data-driven method for Koopman spectral analysis. Here, we introduce practical applications of LKIS-DMD.

Chaotic time-series prediction

Prediction of a chaotic time-series has received significant interest in non-linear physics. We would like to perform the prediction of a chaotic time-series using DMD, since DMD can be naturally utilized for prediction as follows. Since $\mathbf{g}(\mathbf{x}_t)$ is decomposed as $\sum_{i=1}^n \varphi_i(\mathbf{x}_t) \mathbf{c}_i$ and φ is obtained by $\varphi_i(\mathbf{x}_t) = \mathbf{z}_i^H \mathbf{g}(\mathbf{x}_t)$ where \mathbf{z}_i is a left-eigenvalue of \mathbf{K} , the next step of \mathbf{g} can be described in terms of the current step, i.e., $\mathbf{g}(\mathbf{x}_{t+1}) = \sum_{i=1}^n \lambda_i(\mathbf{z}_i^H \mathbf{g}(\mathbf{x}_t)) \mathbf{c}_i$. In addition, in the case of LKIS-DMD, the values of \mathbf{g} must be back-projected to \mathbf{y} using the learned \mathbf{h} .

We generated two types of univariate time-series by extracting the $\{x\}$ series of the Lorenz attractor (Lorenz, 1963) and the Rossler attractor (Rössler, 1976). We simulated 25,000 steps for each attractor and used the first 10,000 steps for training, the next 5,000 steps for validation, and the last 10,000 steps for testing prediction accuracy. We examined the prediction accuracy of LKIS-DMD, a simple LSTM network, and linear Hankel DMD (Arbabi and Mezić, 2017; Susuki and Mezić, 2015), all of whose hyperparameters were tuned using the validation set.² The prediction accuracy of every method and an example of the predicted series on the test set by LKIS-DMD are shown in Figure 5.6. As can be seen, the proposed LKIS-DMD achieves the smallest root-mean-square (RMS) errors in the 30-step prediction.

²LSTM was applied because it had been utilized for various nonlinear time-series, and Hankel DMD was used because it had been successfully utilized for analysis of chaotic systems (Brunton et al., 2017).

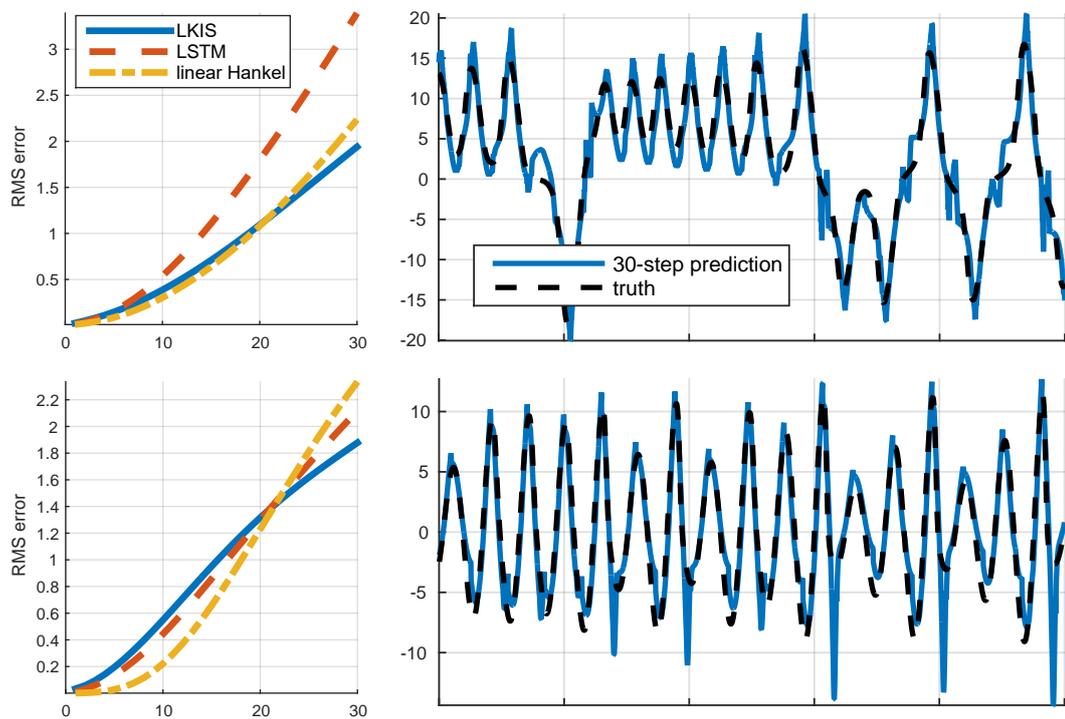


FIGURE 5.6: The left plot shows RMS errors from 1- to 30-step predictions, and the right plot shows a part of the 30-step prediction obtained by LKIS-DMD on (*upper*) the Lorenz- x series and (*lower*) the Rossler- x series.

The more detailed setup in this experiment is as follows. The data were generated from the Lorenz attractor (Lorenz, 1963) (parameters $\beta = 8/3$, $\sigma = 10$, and $\rho = 28$) and the Rossler attractor (Rössler, 1976) (parameters $a = 0.2$, $b = 0.2$, and $c = 5.7$). We generated 25,000 steps for each attractor and divided them into training, validation, and test sets. For all methods, the delay dimension was fixed at 7, i.e., $k = 7$ for LKIS-DMD and linear Hankel DMD, and backpropagation was truncated to length 7 to learn the LSTM network. We tuned n of LKIS-DMD and the dimensionality of LSTM's hidden state (denoted n_h) according to the 30-step prediction accuracies obtained using the validation set. Here, we obtained $n = 5$ and $n_h = 5$ for the Lorenz data and $n = 6$ and $n_h = 3$ for the Rossler data.

Unstable phenomena detection

One of the most popular applications of DMD is the investigation of the global characteristics of dynamics by inspecting the spatial distribution of the dynamic modes. In addition to the spatial distribution, we can investigate the temporal profiles of mode activations by examining the values of corresponding eigenfunctions. For example, assume there is an eigenfunction $\varphi_{\lambda \ll 1}$ that corresponds to a discrete-time eigenvalue λ whose magnitude is considerably smaller than one. Such a small eigenvalue indicates a rapidly decaying (i.e., unstable) mode; thus, we can detect occurrences of unstable phenomena by observing the values of $\varphi_{\lambda \ll 1}$.

We applied LKIS-DMD ($n = 10$) to a time-series generated by a far-infrared laser, which was obtained from the Santa Fe Time Series Competition Data (Weigend and Gershenfeld, 1993). We investigated the values of eigenfunction $\varphi_{\lambda \ll 1}$ corresponding to the eigenvalue of the smallest magnitude. The original time-series and values of $\varphi_{\lambda \ll 1}$ obtained by LKIS-DMD are shown in Figure 5.7. As can be seen, the activations of $\varphi_{\lambda \ll 1}$ coincide with sudden decays of the pulsation amplitudes. For comparison, we applied the novelty/change-point detection technique using one-class support vector machine (OC-SVM) (Canu and Smola, 2006) and direct density-ratio estimation by relative unconstrained least-squares importance fitting (RuLSIF) (Liu et al., 2013).³ We computed AUC, defining the sudden decays of the amplitudes as the points to be detected, which were 0.924, 0.799, and 0.803 for

³OC-SVM was applied because it was a kind of *de facto* standard for novelty/change-point detection, and RuLSIF was used because it had achieved the best performance among methods based on density-ratio estimation (Liu et al., 2013).

LKIS, OC-SVM, and RuLSIF, respectively.

The more detailed setup of this experiment is as follows. The dataset was obtained from the Santa Fe Time Series Competition Data (Weigend and Gershenfeld, 1993). Note that the author's (Weigend and Gershenfeld, 1993) original web page was not available in May 2017 (the date of the submission of the original paper of this chapter); however, the dataset itself was still available online. The length of delay (or sliding window) was fixed to 10 for all methods applied in this experiment. In addition, no intensive tuning of the other hyperparameters was conducted because the purpose was qualitative. The default settings of `libsvm` (Chang and Lin, 2011) were used for the one-class SVM (except for $\nu = 0.05$). For the density-ratio estimation by RuLSIF, the default values of the implementation by the authors of (Liu et al., 2013) were used.

5.7 Summary

In this chapter, we have proposed a framework for learning Koopman invariant subspaces, which is a fully data-driven numerical algorithm for Koopman spectral analysis. In contrast to existing approaches, the proposed method learns (approximately) a Koopman invariant subspace entirely from the available data based on the minimization of RSS loss. We have shown empirical results for several typical nonlinear dynamics and application examples.

We have also introduced an implementation using multi-layer perceptrons; however, one possible drawback of such an implementation is the local optima of the objective function, which makes it difficult to assess the adequacy of the obtained results. Rather than using neural networks, the observables to be learned could be modeled by a sparse combination of basis functions as in Brunton et al. (2016c) but still utilizing optimization based on RSS loss. Another possible future research direction could be incorporating approximate Bayesian inference methods, such as VAE (Kingma and Welling, 2014). The proposed framework is based on a discriminative viewpoint, but inference methodologies for generative models could be used to modify the proposed framework to explicitly consider uncertainty in data.

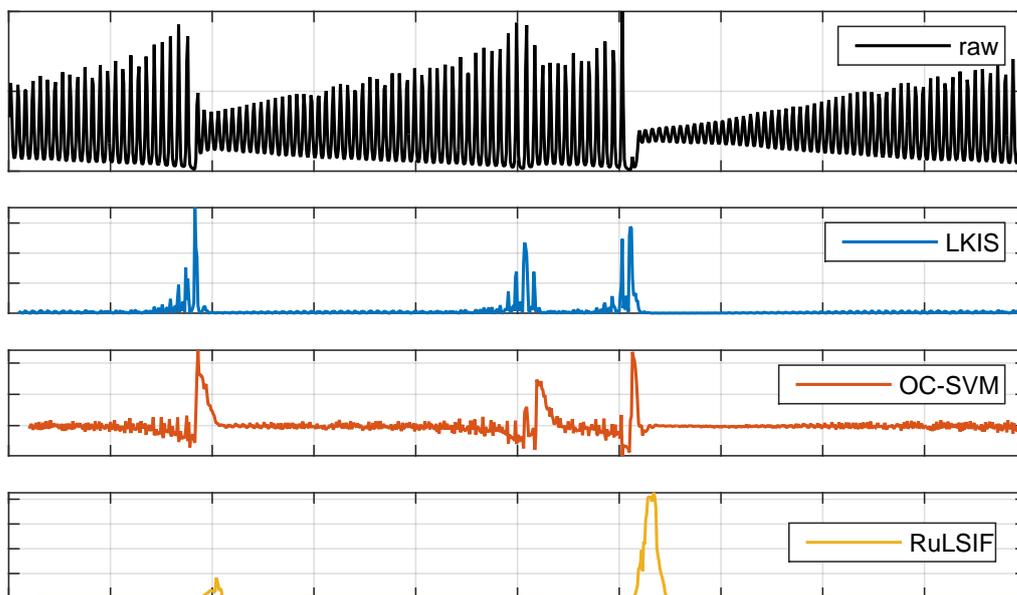


FIGURE 5.7: The top plot shows the raw time-series obtained by a far-infrared laser (Weigend and Gershenfeld, 1993). The other plots show the results of unstable phenomena detection, wherein the peaks should correspond to the occurrences of unstable phenomena.

Chapter 6

Conclusion

6.1 Summary of contributions

In this dissertation, we have introduced numerical methods based on dynamic mode decomposition (DMD) (Rowley et al., 2009; Schmid, 2010) for data-driven analysis of nonlinear dynamical systems. DMD has a strong connection to the operator-theoretic approach to analyzing dynamical systems, which depends on the Koopman operator of dynamics (Mezić, 2005), and it works as an approximation of the modal decomposition based on the Koopman operator under certain conditions. However, such conditions are often much restrictive and a simple implementation of DMD may not be appropriate for many practices of analysis. We therefore developed new techniques to conduct the analysis based on the Koopman operator even for dynamical systems / time-series data that could not be treated by the standard DMD and existing variants. These techniques were proposed from a perspective of machine learning. After describing the technical preliminaries in Chapter 1, we have introduced four distinct topics with regard to the proposed methods for Koopman analysis, which are summarized as follows.

In Chapter 2, we have introduced *sparse nonnegative* DMD, in which the nonnegative constraint and the sparsity regularization were imposed to the estimated dynamic modes. Imposing constraints and/or regularizations was enabled by the reformulation of DMD as an optimization problem wherein the dynamic modes and eigenvalues were reparameterized with polar coordinates. The objective function of the optimization problem is block multi-convex and thus can be solved using a block coordinate descent. Imposing such constraints/regularization is advantageous for processing video streams because of their inherent nonnegativity. Moreover, we can consider other

types of constraints and regularization based on the proposed reformulation framework.

In Chapter 3, we have introduced *Bayesian DMD*, with which a posterior inference on DMD-related quantities can be conducted. Bayesian DMD is based on the probabilistic formulation of DMD, which we termed *probabilistic DMD*. We have shown that the maximum-likelihood estimator of probabilistic DMD coincides with the solution of the classical DMD in the no-noise limit. The posterior inference on the dynamic modes and corresponding eigenvalues are useful when one would like to consider the uncertainty in dataset. Moreover, using Bayesian DMD, extensions of DMD can be developed in a unified framework of Bayesian modeling. As an example of such extensions, we have introduced the use of sparsity-promoting prior on dynamic modes, with which the effective number of dynamic modes can be automatically determined.

In Chapter 4, we have introduced *subspace DMD*, whose methodology has a strong connection to the techniques of subspace system identification. Subspace DMD was developed to perform Koopman spectral analysis even for noisy data generated from random dynamical systems. In other words, subspace DMD is a variant of DMD that can consider both observation and process noise. Since effects of process noise in dynamics are often of interest, only the observation noise should be eliminated. To this end, we have proposed to conduct an orthogonal projection of future snapshots to the space spanned by past snapshots and compute so-called total-least-squares DMD on those projected snapshots. With this procedure, we can compute the spectra of the stochastic Koopman operator that keeps the effects of process noise.

In Chapter 5, we have introduced the framework to *learn Koopman invariant subspaces* from data to conduct DMD in an appropriate space of observables. In the proposed framework, a set of observables that spans a Koopman invariant subspace is learned by minimizing residual sum of squares of linear least-square regression between adjacent transformed snapshots. We have introduced an implementation using neural networks to model such set of observables and have shown the effectiveness of the proposed method through several numerical examples.

6.2 Remaining challenges

In this dissertation, we have broadened the scope of dynamical systems / time-series data with which a data-driven Koopman spectral analysis can be conducted, by proposing several new techniques from a machine learning perspective. However, there remain several technical challenges that are not enough addressed in this dissertation or in other related researches yet.

Continuous spectra In most of the methods to approximate the spectra of the Koopman operator, the continuous component of the spectra is just ignored. However, we may have to consider the continuous spectra when analyzing highly complex dynamical systems. While there have been a few researches that discuss the continuous spectra of the Koopman operator, such as [Korda et al. \(2017\)](#), it is still challenging to approximate them for a general dynamical system.

Non-autonomous systems Except in a few research such as [Mezić and Surana \(2016\)](#) and [Maćešić et al. \(2017\)](#), the underlying dynamical system is assumed to be autonomous, i.e., time-invariant. However, many kinds of time-series data are generated from non-autonomous systems. One of the challenges to deal with non-autonomous systems is how to specify the form of “non-autonomousness,” and this problem may be tackled utilizing the ideas in machine learning researches.

Multi-modal data If data contain sufficient information, two datasets generated from a common dynamical system will yield the same spectral components (eigenvalues and eigenfunctions) of the Koopman operator. However, data are not sufficient in practice, and thus a common dynamical system cannot necessarily be detected from multi-modal datasets. Therefore, a technique to incorporate prior knowledge that the dynamics are the same between different datasets is of great interest.

Application to machine learning In this dissertation, we have discussed the extension of DMD from the perspective of machine learning, i.e., considered *utilizing machine learning for Koopman analysis*. Then, how about *utilizing the Koopman analysis for machine learning*? DMD has been utilized in many

data science contexts, but application to tasks that are commonly addressed in machine learning researches is still limited and will raise interesting utilities and challenges.

Bibliography

- Abraham, I., Torre, G. D. L., and Murphey, T. D. (2017). Model-based control using Koopman operators. In *Robotics: Science and Systems 2017 Proceedings*. DOI: [10.15607/RSS.2017.XIII.052](https://doi.org/10.15607/RSS.2017.XIII.052).
- Annoni, J., Seiler, P., and Jovanović, M. R. (2016). Sparsity-promoting dynamic mode decomposition for systems with inputs. In *Proceedings of the 2016 IEEE 55th Conference on Decision and Control*, pages 6506–6511. DOI: [10.1109/CDC.2016.7799270](https://doi.org/10.1109/CDC.2016.7799270).
- Arbabi, H. and Mezić, I. (2017). Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, DOI: doi.org/10.1137/17M1125236.
- Arnold, L. (1998). *Random Dynamical Systems*. Springer, DOI: [10.1007/978-3-662-12878-7](https://doi.org/10.1007/978-3-662-12878-7).
- Arnoldi, W. E. (1951). The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9:17–29, DOI: [10.1090/qam/42792](https://doi.org/10.1090/qam/42792).
- Bagheri, S. (2014). Effects of weak noise on oscillating flows: Linking quality factor, Floquet modes, and Koopman spectrum. *Physics of Fluids*, 26(9):094104, DOI: [10.1063/1.4895898](https://doi.org/10.1063/1.4895898).
- Barocio, E., Pal, B. C., Thornhill, N. F., and Messina, A. R. (2015). A dynamic mode decomposition framework for global power system oscillation analysis. *IEEE Transactions on Power Systems*, 30(6):2902–2912, DOI: [10.1109/TPWRS.2014.2368078](https://doi.org/10.1109/TPWRS.2014.2368078).
- Berger, E., Sastuba, M., Vogt, D., Jung, B., and Amor, H. B. (2015). Estimation of perturbations in robotic behavior using dynamic mode decomposition. *Advanced Robotics*, 29(5):331–343, DOI: [10.1080/01691864.2014.981292](https://doi.org/10.1080/01691864.2014.981292).

- Berry, T., Giannakis, D., and Harlim, J. (2015). Nonparametric forecasting of low-dimensional dynamical systems. *Physical Review E*, 91(3):032915, 032915, DOI: [10.1103/PhysRevE.91.032915](https://doi.org/10.1103/PhysRevE.91.032915).
- Berry, T. and Harlim, J. (2016). Variable bandwidth diffusion kernels. *Applied and Computational Harmonic Analysis*, 40(1):68–96, DOI: [10.1016/j.acha.2015.01.001](https://doi.org/10.1016/j.acha.2015.01.001).
- Bi, C., Yuan, Y., Zhang, R., Xiang, Y., Wang, Y., and Zhang, J. (2017). A dynamic mode decomposition based edge detection method for art images. *IEEE Photonics Journal*, 9(6), DOI: [10.1109/JPHOT.2017.2766881](https://doi.org/10.1109/JPHOT.2017.2766881).
- Bishop, C. M. (1999). Bayesian PCA. In *Advances in Neural Information Processing Systems*, volume 11, pages 382–388. <https://papers.nips.cc/paper/1549-bayesian-pca>.
- Bistrrian, D. A. and Navon, I. M. (2017). Randomized dynamic mode decomposition for nonintrusive reduced order modelling. *International Journal for Numerical Methods in Engineering*, 112(1):3–25, DOI: [10.1002/nme.5499](https://doi.org/10.1002/nme.5499).
- Bollt, E. M. and Santitissadeekorn, N. (2013). *Applied and Computational Measurable Dynamics*. Mathematical Modeling and Computation. SIAM, DOI: [10.1137/1.9781611972641](https://doi.org/10.1137/1.9781611972641).
- Bonnet, J.-P., Cole, D. R., Delville, J., Glauser, M. N., and Ukeiley, L. S. (1994). Stochastic estimation and proper orthogonal decomposition: Complementary techniques for identifying structure. *Experiments in Fluids*, 17(5):307–314, DOI: [10.1007/BF01874409](https://doi.org/10.1007/BF01874409).
- Boudali, A. M., Sinclair, P. J., Smith, R., and Manchester, I. R. (2017). Human locomotion analysis: Identifying a dynamic mapping between upper and lower limb joints using the Koopman operator. In *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1889–1892. DOI: [10.1109/EMBC.2017.8037216](https://doi.org/10.1109/EMBC.2017.8037216).
- Bourantas, G. C., Ghommam, M., Kagadis, G. C., Katsanos, K., Loukopoulos, V. C., Burganos, V. N., and Nikiforidis, G. C. (2014). Real-time tumor ablation simulation based on the dynamic mode decomposition method. *Medical Physics*, 41(5):053301, DOI: [10.1118/1.4870976](https://doi.org/10.1118/1.4870976).
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Controls*. Wiley Series in Probability and Statistics. Wiley, 5th edition, ISBN: [978-1-118-67502-1](https://doi.org/10.1002/9781118675021).

- Brunton, B. W., Johnson, L. A., Ojemann, J. G., and Kutz, J. N. (2016a). Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, DOI: [10.1016/j.jneumeth.2015.10.010](https://doi.org/10.1016/j.jneumeth.2015.10.010).
- Brunton, S. L., Brunton, B. W., Proctor, J. L., Kaiser, E., and Kutz, J. N. (2017). Chaos as an intermittently forced linear system. *Nature Communications*, 8:19, DOI: [10.1038/s41467-017-00030-8](https://doi.org/10.1038/s41467-017-00030-8).
- Brunton, S. L., Brunton, B. W., Proctor, J. L., and Kutz, J. N. (2016b). Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS ONE*, 11(2):e0150171, DOI: [10.1371/journal.pone.0150171](https://doi.org/10.1371/journal.pone.0150171).
- Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016c). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 113(15):3932–3937, DOI: [10.1073/pnas.1517384113](https://doi.org/10.1073/pnas.1517384113).
- Brunton, S. L., Proctor, J. L., Tu, J. H., and Kutz, J. N. (2015). Compressed sensing and dynamic mode decomposition. *Journal of Computational Dynamics*, 2(2):165–191, DOI: [10.3934/jcd.2015002](https://doi.org/10.3934/jcd.2015002).
- Budišić, M., Mohr, R., and Mezić, I. (2012). Applied Koopmanism. *Chaos*, 22:047510, DOI: [10.1063/1.4772195](https://doi.org/10.1063/1.4772195).
- Candés, E. J., Li, X., Ma, Y., and Wright, J. (2011). Robust principal component analysis? *Journal of the ACM*, 58(3):11, DOI: [10.1145/1970392.1970395](https://doi.org/10.1145/1970392.1970395).
- Canu, S. and Smola, A. (2006). Kernel methods and the exponential family. *Neurocomputing*, 69(7-9):714–720, DOI: [10.1016/j.neucom.2005.12.009](https://doi.org/10.1016/j.neucom.2005.12.009).
- Caro-Ruiz, C., Tellez-Castro, D., Pavas, A., and Mojica-Nava, E. (2017). Self-organization in networks: A data-driven Koopman approach, <https://arxiv.org/abs/1709.09576>. arXiv:1709.09576.
- Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics*, 2(4):485–500, DOI: [10.1093/biostatistics/2.4.485](https://doi.org/10.1093/biostatistics/2.4.485).

- Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27, DOI: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199).
- Chen, K. K., Tu, J. H., and Rowley, C. W. (2012). Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses. *Journal of Nonlinear Science*, 22(6):887–915, DOI: [10.1007/s00332-012-9130-9](https://doi.org/10.1007/s00332-012-9130-9).
- Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., and Bengio, Y. (2015). A recurrent latent variable model for sequential data. In *Advances in Neural Information Processing Systems*, volume 28, pages 2980–2988. <https://papers.nips.cc/paper/5653-a-recurrent-latent-variable-model-for-sequential-data>.
- Coifman, R. G. and Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, DOI: [10.1016/j.acha.2006.04.006](https://doi.org/10.1016/j.acha.2006.04.006).
- Colonus, T. and Taira, K. (2008). A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Computer Methods in Applied Mechanics and Engineering*, 197(25-28):2131–2146, DOI: [10.1016/j.cma.2007.08.014](https://doi.org/10.1016/j.cma.2007.08.014).
- Dawson, S. T., Hemati, M. S., Williams, M. O., and Rowley, C. W. (2016). Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, DOI: [10.1007/s00348-016-2127-7](https://doi.org/10.1007/s00348-016-2127-7).
- de Badyn, M. H., Alemzadeh, S., and Mesbahi, M. (2017). Controllability and data-driven identification of bipartite consensus on nonlinear signed networks, <https://arxiv.org/abs/1709.06679>. arXiv:1709.06679.
- Devroye, L. (2014). Random variate generation for the generalized inverse Gaussian distribution. *Statistics and Computing*, 24(2):239–246, DOI: [10.1007/s11222-012-9367-z](https://doi.org/10.1007/s11222-012-9367-z).
- Dicle, C., Mansour, H., Tian, D., Benosman, M., and Vetro, A. (2016). Robust low rank dynamic mode decomposition for compressed domain crowd and traffic flow analysis. In *Proceedings of 2016 IEEE International Conference on Multimedia and Expo*, pages 1–6. DOI: [10.1109/ICME.2016.7552877](https://doi.org/10.1109/ICME.2016.7552877).

- Duke, D., Soria, J., and Honnery, D. (2012). An error analysis of the dynamic mode decomposition. *Experiments in Fluids*, 52(2):529–542, DOI: [10.1007/s00348-011-1235-7](https://doi.org/10.1007/s00348-011-1235-7).
- Eisenhower, B., Maile, T., Fischer, M., and Mezić, I. (2010). Decomposing building system data for model validation and analysis using the Koopman operator. In *IBPSA-USA SimBuild 2010*, pages 434–441. <https://ibpsa-usa.org/index.php/ibpusa/article/view/318>.
- Erichson, N. B., Brunton, S. L., and Kutz, J. N. (2016). Compressed dynamic mode decomposition for background modeling. *Journal of Real-Time Image Processing*, pages 1–14, DOI: [10.1007/s11554-016-0655-2](https://doi.org/10.1007/s11554-016-0655-2).
- Erichson, N. B., Brunton, S. L., and Kutz, J. N. (2017). Randomized dynamic mode decomposition, <https://arxiv.org/abs/1702.02912>. arXiv:1702.02912.
- Erichson, N. B. and Donovan, C. (2016). Randomized low-rank dynamic mode decomposition for motion detection. *Computer Vision and Image Understanding*, 146:40–50, DOI: [10.1016/j.cviu.2016.02.005](https://doi.org/10.1016/j.cviu.2016.02.005).
- Froyland, G. and González-Tokman, C. (2014). Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. *Journal of Computational Dynamics*, 1(2):249–278, DOI: [10.3934/jcd.2014.1.249](https://doi.org/10.3934/jcd.2014.1.249).
- Froyland, G., Gottwald, G. A., and Hammerlindl, A. (2014). A computational method to extract macroscopic variables and their dynamics in multiscale systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1816–1846, DOI: [10.1137/130943637](https://doi.org/10.1137/130943637).
- Fujii, K., Inaba, Y., and Kawahara, Y. (2017). Koopman spectral kernels for comparing complex dynamics: Application to multiagent sport plays. In *Machine Learning and Knowledge Discovery in Databases*, volume 10536 of *Lecture Notes in Computer Science*, pages 127–139. DOI: [10.1007/978-3-319-71273-4_11](https://doi.org/10.1007/978-3-319-71273-4_11).
- Funk, S. (2015). RMSprop loses to SMORMS3 - beware the epsilon! available online, <http://sifter.org/~simon/journal/20150420.html>. Retrieved 29 April-2017.
- Gao, Y., Archer, E. W., Paninski, L., and Cunningham, J. P. (2016). Linear dynamical neural population models through nonlinear embeddings. In

- Advances in Neural Information Processing Systems*, volume 29, pages 163–171. <https://papers.nips.cc/paper/6430-linear-dynamical-neural-population-models-through-nonlinear-embeddings>.
- Garcia, S. P. and Almeida, J. S. (2005). Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. *Physical Review E*, 72(2):027205, 027205, DOI: [10.1103/PhysRevE.72.027205](https://doi.org/10.1103/PhysRevE.72.027205).
- Georgescu, M., Eisenhower, B., and Mezić, I. (2012). Creating zoning approximations to building energy models using the Koopman operator. In *IBPSA-USA SimBuild 2012*, pages 40–47. <https://ibpsa-usa.org/index.php/ibpusa/article/view/413>.
- Ghahramani, Z. and Roweis, S. T. (1999). Learning nonlinear dynamical systems using an EM algorithm. In *Advances in Neural Information Processing Systems*, volume 11, pages 431–437. <https://papers.nips.cc/paper/1594-learning-nonlinear-dynamical-systems-using-an-em-algorithm>.
- Giannakis, D. (2017). Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, DOI: [10.1016/j.acha.2017.09.001](https://doi.org/10.1016/j.acha.2017.09.001). in press.
- Giannakis, D., Ourmazd, A., Slawinska, J., and Zhao, Z. (2017). Spatiotemporal pattern extraction by spectral analysis of vector-valued observables, <https://arxiv.org/abs/1711.02798>. arXiv:1711.02798.
- Giannakis, D., Slawinska, J., and Zhao, Z. (2015). Spatiotemporal feature extraction with data-driven Koopman operators. In *Proceedings of the 1st International Workshop on Feature Extraction: Modern Questions and Challenges at NIPS 2015*, volume 44 of *Proceedings of Machine Learning Research*, pages 103–115. <http://proceedings.mlr.press/v44/giannakis15.html>.
- Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34(1):152–177, DOI: [10.1214/aoms/1177704250](https://doi.org/10.1214/aoms/1177704250).
- Guan, Y. and Dy, J. (2009). Sparse probabilistic principal component analysis. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 185–192. <http://proceedings.mlr.press/v5/guan09a.html>.

- Guéniat, F., Mathelin, L., and Pastur, L. R. (2015). A dynamic mode decomposition approach for large and arbitrarily sampled systems. *Physics of Fluids*, 27:025113, DOI: [10.1063/1.4908073](https://doi.org/10.1063/1.4908073).
- Hausenblas, E. (2003). Approximation for semilinear stochastic evolution equations. *Potential Analysis*, 18(2):141–186, DOI: [10.1023/A:1020552804087](https://doi.org/10.1023/A:1020552804087).
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *Proceedings of the 2015 IEEE International Conference on Computer Vision*, pages 1026–1034. DOI: [10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123).
- Héas, P. and Herzet, C. (2017). Optimal low-rank dynamic mode decomposition. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4456–4460. DOI: [10.1109/ICASSP.2017.7952999](https://doi.org/10.1109/ICASSP.2017.7952999).
- Hemati, M. S., Rowley, C. W., Deem, E. A., and 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, DOI: [10.1007/s00162-017-0432-2](https://doi.org/10.1007/s00162-017-0432-2).
- Hemati, M. S., Williams, M. O., and Rowley, C. W. (2014). Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids*, 26:111701, DOI: [10.1063/1.4901016](https://doi.org/10.1063/1.4901016).
- Hey, T., Tansley, S., and Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, ISBN: [978-0-9825442-0-4](https://doi.org/10.1007/978-0-9825442-0-4).
- Hirata, Y., Suzuki, H., and Aihara, K. (2006). Reconstructing state spaces from multivariate data using variable delays. *Physical Review E*, 74(2):026202, 026202, DOI: [10.1103/PhysRevE.74.026202](https://doi.org/10.1103/PhysRevE.74.026202).
- Hirsch, M. W., Smale, S., and Devaney, R. L. (2013). *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press, 3rd edition, DOI: [10.1016/c2009-0-61160-0](https://doi.org/10.1016/c2009-0-61160-0).
- Holmes, P., Lumley, J. L., and Berkooz, G. (2012). *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 2nd edition, DOI: [10.1017/CBO9780511919701](https://doi.org/10.1017/CBO9780511919701).

- Hua, J.-C., Noorian, F., Moss, D., Leong, P. H. W., and Gunaratne, G. H. (2017). High-dimensional time series prediction using kernel-based Koopman mode regression. *Nonlinear Dynamics*, 90(3):1785–1806, DOI: doi.org/10.1007/s11071-017-3764-y.
- Hua, J.-C., Roy, S., McCauley, J. L., and Gunaratne, G. H. (2016). Using dynamic mode decomposition to extract cyclic behavior in the stock market. *Physica A: Statistical Mechanics and its Applications*, 448:172–180, DOI: [10.1016/j.physa.2015.12.059](https://doi.org/10.1016/j.physa.2015.12.059).
- IEEE International Conference on Advanced Video and Signal based Surveillance (2007). i-Lids dataset for AVSS 2007. available online, http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html. Retrieved 24 May 2017.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456. <http://proceedings.mlr.press/v37/ioffe15.html>.
- Johnson, M., Duvenaud, D. K., Wiltschko, A., Adams, R. P., and Datta, S. R. (2016). Composing graphical models with neural networks for structured representations and fast inference. In *Advances in Neural Information Processing Systems*, volume 29, pages 2946–2954. <https://papers.nips.cc/paper/6379-composing-graphical-models-with-neural-networks-for-structured-representations-and-fast-inference>.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, 2nd edition, DOI: [10.1007/b98835](https://doi.org/10.1007/b98835).
- Jourdain, G., Eriksson, L.-E., Kim, S. H., and Sohn, C. H. (2013). Application of dynamic mode decomposition to acoustic-modes identification and damping in a 3-dimensional chamber with baffled injectors. *Journal of Sound and Vibration*, 332(18):4308–4323, DOI: [10.1016/j.jsv.2013.02.041](https://doi.org/10.1016/j.jsv.2013.02.041).
- Jovanović, M. R., Schmid, P. J., and Nichols, J. W. (2014). Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, DOI: [10.1063/1.4863670](https://doi.org/10.1063/1.4863670).

- Kameoka, H., Ono, N., Kashino, K., and Sagayama, S. (2009). Complex NMF: A new sparse representation for acoustic signals. In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3437–3440. DOI: [10.1109/ICASSP.2009.4960364](https://doi.org/10.1109/ICASSP.2009.4960364).
- Kar, P., Narasimhan, H., and Jain, P. (2014). Online and stochastic gradient methods for non-decomposable loss functions. In *Advances in Neural Information Processing Systems*, volume 27, pages 694–702. <https://papers.nips.cc/paper/5367-online-and-stochastic-gradient-methods-for-non-decomposable-loss-functions>.
- Karl, M., Soelch, M., Bayer, J., and van der Smagt, P. (2017). Deep variational Bayes filters: Unsupervised learning of state space models from raw data. In *Proceedings of the 5th International Conference on Learning Representations*. <https://openreview.net/forum?id=HyTqHL5xg>.
- Katayama, T. (2005). *Subspace Methods for System Identification*. Communications and Control Engineering. Springer, DOI: [10.1007/1-84628-158-X](https://doi.org/10.1007/1-84628-158-X).
- Kawahara, Y. (2016). Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis. In *Advances in Neural Information Processing Systems*, volume 29, pages 911–919. <https://papers.nips.cc/paper/6583-dynamic-mode-decomposition-with-reproducing-kernels-for-koopman-spectral-analysis>.
- Kingma, D. P. and Welling, M. (2014). Stochastic gradient VB and the variational auto-encoder. In *Proceedings of the 2nd International Conference on Learning Representations*. <https://arxiv.org/abs/1312.6114>.
- Klus, S., Koltai, P., and Schütte, C. (2016). On the numerical approximation of the Perron–Frobenius and Koopman operator. *Journal of Computational Dynamics*, 3(1):51–79, DOI: [10.3934/jcd.2016003](https://doi.org/10.3934/jcd.2016003).
- Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., and Noé, F. (2017). Data-driven model reduction and transfer operator approximation, <https://arxiv.org/abs/1703.10112>. arXiv:1703.10112.

- Klus, S. and 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, DOI: [10.3934/jcd.2016007](https://doi.org/10.3934/jcd.2016007).
- Koopman, B. O. (1931). Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 17(5):315–318, DOI: [10.1073/pnas.17.5.315](https://doi.org/10.1073/pnas.17.5.315).
- Korda, M. and Mezić, I. (2017). On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, DOI: [10.1007/s00332-017-9423-0](https://doi.org/10.1007/s00332-017-9423-0).
- Korda, M., Putinar, M., and Mezić, I. (2017). Data-driven spectral analysis of the koopman operator, <https://arxiv.org/abs/1710.06532>. arXiv:1710.06532.
- Krishnan, R. G., Shalit, U., and Sontag, D. (2017). Structured inference networks for nonlinear state space models. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 2101–2109. <https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14215>.
- Kutz, J. N., Brunton, S. L., Brunton, B. W., and Proctor, J. L. (2016a). *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, DOI: [10.1137/1.9781611974508](https://doi.org/10.1137/1.9781611974508).
- Kutz, J. N., Fu, X., and Brunton, S. L. (2016b). Multiresolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, DOI: [10.1137/15M1023543](https://doi.org/10.1137/15M1023543).
- Kutz, J. N., Fu, X., Brunton, S. L., and Erichson, N. B. (2015). Multi-resolution decomposition for foreground/background separation and object tracking. In *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop*, pages 921–929. DOI: [10.1109/ICCVW.2015.122](https://doi.org/10.1109/ICCVW.2015.122).
- Kutz, J. N., Proctor, J. L., and Brunton, S. L. (2016c). Koopman theory for partial differential equations, <https://arxiv.org/abs/1607.07076>. arXiv:1607.07076.
- Lan, Y. and Mezić, I. (2013). Linearization in the large of nonlinear systems and Koopman operator spectrum. *Physica D: Nonlinear Phenomena*, 242(1):42–53, DOI: [10.1016/j.physd.2012.08.017](https://doi.org/10.1016/j.physd.2012.08.017).

- Lasota, A. and Mackey, M. C. (1994). *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. Springer, 2nd edition, DOI: [10.1007/978-1-4612-4286-4](https://doi.org/10.1007/978-1-4612-4286-4).
- Le Clainche, S. and Vega, J. M. (2017). Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, DOI: [10.1137/15M1054924](https://doi.org/10.1137/15M1054924).
- Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, DOI: [10.1038/44565](https://doi.org/10.1038/44565).
- Lee, J. D., Sun, Y., and Saunders, M. A. (2014). Proximal Newton-type methods for minimizing composite functions. *SIAM Journal on Optimization*, 24(3):1420–1443, DOI: [10.1137/130921428](https://doi.org/10.1137/130921428).
- Li, Q., Dietrich, F., Bollt, E. M., and Kevrekidis, I. G. (2017). Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos*, 27:103111, DOI: [10.1063/1.4993854](https://doi.org/10.1063/1.4993854).
- Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, DOI: [10.1007/BF01589116](https://doi.org/10.1007/BF01589116).
- Liu, S., Yamada, M., Collier, N., and Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83, DOI: [10.1016/j.neunet.2013.01.012](https://doi.org/10.1016/j.neunet.2013.01.012).
- Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall, 2nd edition.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, DOI: [10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- Lusch, B., Kutz, J. N., and Brunton, S. L. (2017). Deep learning for universal linear embeddings of nonlinear dynamics, <https://arxiv.org/abs/1712.09707>. arXiv:1712.09707.
- Maćešić, S., Črnjarić-Žic, N., and Mezić, I. (2017). Koopman operator family spectrum for nonautonomous systems – part 1, <https://arxiv.org/abs/1703.07324>. arXiv:1703.07324.

- Mann, J. and Kutz, J. N. (2016). Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, 16(11):1643–1655, DOI: [10.1080/14697688.2016.1170194](https://doi.org/10.1080/14697688.2016.1170194).
- Mardt, A., Pasquali, L., Wu, H., and Noé, F. (2017). VAMPnets: Deep learning of molecular kinetics, <https://arxiv.org/abs/1710.06012>. arXiv:1710.06012.
- Mauroy, A., Forni, F., and Sepulchre, R. (2015). An operator-theoretic approach to differential positivity. In *Proceedings of the 2015 IEEE 54th Conference on Decision and Control*, pages 7028–7033. DOI: [10.1109/CDC.2015.7403327](https://doi.org/10.1109/CDC.2015.7403327).
- Mauroy, A. and Goncalves, J. (2016). Linear identification of nonlinear systems: A lifting technique based on the Koopman operator. In *Proceedings of the 2016 IEEE 55th Conference on Decision and Control*, pages 6500–6505. DOI: [10.1109/CDC.2016.7799269](https://doi.org/10.1109/CDC.2016.7799269).
- Mauroy, A. and Mezić, I. (2012). On the use of Fourier averages to compute the global isochrons of (quasi)periodic dynamics. *Chaos*, 22(3):033112, 033112, DOI: [10.1063/1.4736859](https://doi.org/10.1063/1.4736859).
- Mauroy, A. and Mezić, I. (2016). Global stability analysis using the eigenfunctions of the Koopman operator. *IEEE Transactions on Automatic Control*, 51(1):3356–3369, DOI: [10.1109/TAC.2016.2518918](https://doi.org/10.1109/TAC.2016.2518918).
- Mauroy, A., Mezić, I., and Moehls, J. (2013). Isostables, isochrons, and Koopman spectrum for the action-angle representation of stable fixed point dynamics. *Physica D: Nonlinear Phenomena*, 261:19–30, DOI: [10.1016/j.physd.2013.06.004](https://doi.org/10.1016/j.physd.2013.06.004).
- Mezić, I. (2005). Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, DOI: [10.1007/s11071-005-2824-x](https://doi.org/10.1007/s11071-005-2824-x).
- Mezić, I. (2013). Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, DOI: [10.1146/annurev-fluid-011212-140652](https://doi.org/10.1146/annurev-fluid-011212-140652).
- Mezić, I. and Banaszuk, A. (2004). Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1-2):101–133, DOI: [10.1016/j.physd.2004.06.015](https://doi.org/10.1016/j.physd.2004.06.015).

- Mezić, I. and Surana, A. (2016). Koopman mode decomposition for periodic/quasi-periodic time dependence. In *IFAC-PapersOnLine*, volume 49, pages 690–697. DOI: [10.1016/j.ifacol.2016.10.246](https://doi.org/10.1016/j.ifacol.2016.10.246).
- Mohr, R. (2014). *Spectral Properties of the Koopman Operator in the Analysis of Nonstationary Dynamical Systems*. PhD thesis, University of California, Santa Barbara, <https://alexandria.ucsb.edu/lib/ark:/48907/f3q23xc7>.
- Narasingham, A. and Kwon, J. S.-I. (2017). Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing. *Computers & Chemical Engineering*, 106:501–511, DOI: [10.1016/j.compchemeng.2017.07.002](https://doi.org/10.1016/j.compchemeng.2017.07.002).
- Noack, B. R., Afanasiev, K., Morzyński, M., Tadmor, G., and Thiele, F. (2003). A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, DOI: [10.1017/S0022112003006694](https://doi.org/10.1017/S0022112003006694).
- Ohmichi, Y. (2017). Preconditioned dynamic mode decomposition and mode selection algorithms for large datasets using incremental proper orthogonal decomposition. *AIP Advances*, 7:075318, DOI: [10.1063/1.4996024](https://doi.org/10.1063/1.4996024).
- Otto, S. E. and Rowley, C. W. (2017). Linearly-recurrent autoencoder networks for learning dynamics, <https://arxiv.org/abs/1712.01378>. arXiv:1712.01378.
- Pan, C., Xue, D., and Wang, J. (2015). On the accuracy of dynamic mode decomposition in estimating instability of wave packet. *Experiments in Fluids*, 56(8):164, DOI: [10.1007/s00348-015-2015-6](https://doi.org/10.1007/s00348-015-2015-6).
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103:681–686, DOI: [10.1198/016214508000000337](https://doi.org/10.1198/016214508000000337).
- Proctor, J. L., Brunton, S. L., and Kutz, J. N. (2016). Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, DOI: [10.1137/15M1013857](https://doi.org/10.1137/15M1013857).
- Proctor, J. L. and Eckhoff, P. A. (2015). Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health*, 7(2):139–145, DOI: [10.1093/inthealth/ihv009](https://doi.org/10.1093/inthealth/ihv009).

- Raak, F., Susuki, Y., and Mezić, I. (2016). Data-driven partitioning of power networks via Koopman mode analysis. *IEEE Transactions on Power Systems*, 31(4):2799–2808, DOI: [10.1109/TPWRS.2015.2464779](https://doi.org/10.1109/TPWRS.2015.2464779).
- Rakočević, V. (1997). On continuity of the Moore–Penrose and Drazin inverses. *Matematički Vesnik*, 49(3-4):163–172.
- Rodriguez, P. and Wohlberg, B. (2013). Fast principal component pursuit via alternating minimization. In *Proceedings of the 2013 20th IEEE International Conference on Image Processing*, pages 69–73. DOI: [10.1109/ICIP.2013.6738015](https://doi.org/10.1109/ICIP.2013.6738015).
- Rössler, O. E. (1976). An equation for continuous chaos. *Physical Letters*, 57A(5):397–398, DOI: [10.1016/0375-9601\(76\)90101-8](https://doi.org/10.1016/0375-9601(76)90101-8).
- Rowley, C. W. and Dawson, S. T. (2017). Model reduction for flow analysis and control. *Journal of Fluid Mechanics*, 49:387–417, DOI: [10.1146/annurev-fluid-010816-060042](https://doi.org/10.1146/annurev-fluid-010816-060042).
- Rowley, C. W., Mezić, I., Bagheri, S., Schlatter, P., and Henningson, D. S. (2009). Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, DOI: [10.1017/S0022112009992059](https://doi.org/10.1017/S0022112009992059).
- Ruelle, D. (1968). Statistical mechanics of a one-dimensional lattice gas. *Communications in Mathematical Physics*, 9(4):267–278, DOI: doi.org/10.1007/BF01654281.
- Sauer, T., Yorke, J. A., and Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, 65(3-4):579–616, DOI: [10.1007/BF01053745](https://doi.org/10.1007/BF01053745).
- Schmid, P. and Sesterhenn, J. (2008). Dynamic mode decomposition of numerical and experimental data. In *Bulletin of the American Physical Society*, volume 53. DOI: [2008APS..DFD.MR007S](https://doi.org/2008APS..DFD.MR007S).
- Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, DOI: [10.1017/S0022112010001217](https://doi.org/10.1017/S0022112010001217).
- Shampine, L. F. and Reichelt, M. W. (1997). The MATLAB ODE suite. *SIAM Journal on Scientific Computing*, 18(1):1–22, DOI: [10.1137/S1064827594276424](https://doi.org/10.1137/S1064827594276424).

- Sharma, A. S., Mezić, I., and 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(R), DOI: [10.1103/PhysRevFluids.1.032402](https://doi.org/10.1103/PhysRevFluids.1.032402).
- Shnitzer, T., Talmon, R., and Slotine, J.-J. (2017). Diffusion maps Kalman filter, <https://arxiv.org/abs/1711.09598>. arXiv:1711.09598.
- Sootla, A. and Mauroy, A. (2017). Geometric properties of isostables and basins of attraction of monotone systems. *IEEE Transactions on Automatic Control*, DOI: [10.1109/TAC.2017.2707660](https://doi.org/10.1109/TAC.2017.2707660). in press.
- Surana, A. (2015). Koopman operator based nonlinear dynamic textures. In *Proceedings of the 2015 American Control Conference*, pages 1333–1338. DOI: [10.1109/ACC.2015.7170918](https://doi.org/10.1109/ACC.2015.7170918).
- Surana, A. (2016). Koopman operator based observer synthesis for control-affine nonlinear systems. In *Proceedings of the 2016 IEEE 55th Conference on Decision and Control*, pages 6492–6499. DOI: [10.1109/CDC.2016.7799268](https://doi.org/10.1109/CDC.2016.7799268).
- Susuki, Y. and Mezić, I. (2011). Nonlinear Koopman modes and coherency identification of coupled swing dynamics. *IEEE Transactions on Power Systems*, 26(4):1894–1904, DOI: [10.1109/TPWRS.2010.2103369](https://doi.org/10.1109/TPWRS.2010.2103369).
- Susuki, Y. and Mezić, I. (2012). Nonlinear Koopman modes and a precursor to power system swing instabilities. *IEEE Transactions on Power Systems*, 27(3):1182–1191, DOI: [10.1109/TPWRS.2012.2183625](https://doi.org/10.1109/TPWRS.2012.2183625).
- Susuki, Y. and Mezić, I. (2014). Nonlinear Koopman modes and power system stability assessment without models. *IEEE Transactions on Power Systems*, 29(2):899–907, DOI: [10.1109/TPWRS.2013.2287235](https://doi.org/10.1109/TPWRS.2013.2287235).
- Susuki, Y. and Mezić, I. (2015). A Prony approximation of Koopman mode decomposition. In *Proceedings of the 2015 IEEE 54th Conference on Decision and Control*, pages 7022–7027. DOI: [10.1109/CDC.2015.7403326](https://doi.org/10.1109/CDC.2015.7403326).
- Taira, K. and Colonius, T. (2007). The immersed boundary method: a projection approach. *Journal of Computational Physics*, 225(2):2118–2137, DOI: [10.1016/j.jcp.2007.03.005](https://doi.org/10.1016/j.jcp.2007.03.005).

- Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. DOI: [10.1007/BFb0091924](https://doi.org/10.1007/BFb0091924).
- Tipping, M. E. and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 61(3):611–622, DOI: [10.1111/1467-9868.00196](https://doi.org/10.1111/1467-9868.00196).
- Towne, A., Schmidt, O. T., and Colonius, T. (2017). Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis, <https://arxiv.org/abs/1708.04393>. arXiv:1708.04393.
- Tu, J. H., Rowley, C. W., Kutz, J. N., and Shang, J. K. (2014a). Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. *Experiments in Fluids*, 55:1805, DOI: [10.1007/s00348-014-1805-6](https://doi.org/10.1007/s00348-014-1805-6).
- Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L., and Kutz, J. N. (2014b). On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, DOI: [10.3934/jcd.2014.1.391](https://doi.org/10.3934/jcd.2014.1.391).
- Vaidya, U. (2007). Observability gramian for nonlinear systems. In *Proceedings of the 2007 IEEE 46th Conference on Decision and Control*, pages 3357–3362. DOI: [10.1109/CDC.2007.4434828](https://doi.org/10.1109/CDC.2007.4434828).
- Vaidya, U. and Mehta, P. G. (2008). Lyapunov measure for almost everywhere stability. *IEEE Transactions on Automatic Control*, 53(1):307–323, DOI: [10.1109/TAC.2007.914955](https://doi.org/10.1109/TAC.2007.914955).
- van der Maaten, L. and Hinton, G. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- Van Overschee, P. and De Moor, B. L. (1996). *Subspace Identification for Linear Systems*. Springer, DOI: [10.1007/978-1-4613-0465-4](https://doi.org/10.1007/978-1-4613-0465-4).
- Vlachos, I. and Kugiumtzis, D. (2010). Nonuniform state-space reconstruction and coupling detection. *Physical Review E*, 82(1):016207, 016207, DOI: [10.1103/PhysRevE.82.016207](https://doi.org/10.1103/PhysRevE.82.016207).

- Wang, K. and Vaidya, U. (2010). Transfer operator approach for computing domain of attraction. In *Proceedings of the 2010 IEEE 49th Conference on Decision and Control*, pages 5390–5395. DOI: [10.1109/CDC.2010.5718147](https://doi.org/10.1109/CDC.2010.5718147).
- Watter, M., Springenberg, J., Boedecker, J., and Riedmiller, M. (2015). Embed to control: A locally linear latent dynamics model for control from raw images. In *Advances in Neural Information Processing Systems*, volume 28, pages 2746–2754. <https://papers.nips.cc/paper/5964-embed-to-control-a-locally-linear-latent-dynamics-model-for-control-from-raw-images>.
- Weigend, A. S. and Gershenfeld, N. A., editors (1993). *Time Series Prediction: Forecasting The Future And Understanding The Past*. Santa Fe Institute Series. Westview Press.
- Williams, M. O., Kevrekidis, I. G., and Rowley, C. W. (2015a). A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, DOI: [10.1007/s00332-015-9258-5](https://doi.org/10.1007/s00332-015-9258-5).
- Williams, M. O., Rowley, C. W., and Kevrekidis, I. G. (2015b). A kernel-based method for data-driven Koopman spectral analysis. *J. of Computational Dynamics*, 2(2):247–265, DOI: [10.3934/jcd.2015005](https://doi.org/10.3934/jcd.2015005).
- Williams, M. O., Rowley, C. W., Mezić, I., and Kevrekidis, I. G. (2015c). Data fusion via intrinsic dynamic variables: An application of data-driven Koopman spectral analysis. *Europhysics Letters*, 109(4):40007, 40007, DOI: [10.1209/0295-5075/109/40007](https://doi.org/10.1209/0295-5075/109/40007).
- Wu, H., Nüske, F., Paul, F., Klus, S., Koltai, P., and Noé, F. (2017). Variational koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *The Journal of Chemical Physics*, 146(15):154104, DOI: [10.1063/1.4979344](https://doi.org/10.1063/1.4979344).
- Wynn, A., Pearson, D. S., Ganapathisubramani, B., and Goulart, P. J. (2013). Optimal mode decomposition for unsteady flows. *Journal of Fluid Mechanics*, 733:473–503, DOI: [10.1017/jfm.2013.426](https://doi.org/10.1017/jfm.2013.426).
- Xu, Y. and Yin, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, DOI: [10.1137/120887795](https://doi.org/10.1137/120887795).

Yeung, E., Kundu, S., and Hodas, N. (2017). Learning deep neural network representations for Koopman operators of nonlinear dynamical systems, <https://arxiv.org/abs/1708.06850>. arXiv:1708.06850.