Bayesian Dynamic Mode Decomposition

Naoya Takeishi[§], Yoshinobu Kawahara[†,‡], Yasuo Tabei[‡], Takehisa Yairi[§]
[§]Department of Aeronautics and Astronautics, The University of Tokyo
[†]The Institute of Scientific and Industrial Research, Osaka University
[‡]RIKEN Center for Advanced Intelligence Project
{takeishi,yairi}@ailab.t.u-tokyo.ac.jp, ykawahara@sanken.osaka-u.ac.jp, yasuo.tabei@riken.jp

Abstract

Dynamic mode decomposition (DMD) is a data-driven method for calculating a modal representation of a nonlinear dynamical system, and it has been utilized in various fields of science and engineering. In this paper, we propose *Bayesian DMD*, which provides a principled way to transfer the advantages of the Bayesian formulation into DMD. To this end, we first develop a probabilistic model corresponding to DMD, and then, provide the Gibbs sampler for the posterior inference in Bayesian DMD. Moreover, as a specific example, we discuss the case of using a sparsity-promoting prior for an automatic determination of the number of dynamic modes. We investigate the empirical performance of Bayesian DMD using synthetic and real-world datasets.

1 Introduction

Analyzing nonlinear dynamical systems is fundamental for the understanding of complex phenomena in a variety of scientific and industrial fields. For example, the analysis of the Navier–Stokes equation has been one of the fundamental problems for understanding fluid flows. One of popular approaches for this purpose is decomposition of the dynamics into multiple components based on some criteria; the individual aspects of complex phenomena can be investigated by examining each decomposed component. For example, proper orthogonal decomposition (POD) (see e.g. [Holmes *et al.*, 2012]) decomposes the dynamics into orthogonal modes that optimally capture the energy of the dynamics, and it has been extensively applied in physics [Bonnet *et al.*, 1994; Noack *et al.*, 2003]. In 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].

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 (see Section 5 for some examples). Although DMD is a data-driven decomposition technique like PCA, it generates modes that are directly related to the *dynamics* behind the data; 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 paper, we propose *Bayesian DMD*, which provides a principled way to transfer the advantages of the Bayesian formulation into DMD. To this end, we first develop a probabilistic model corresponding to DMD, whose maximum-likelihood estimator coincides with the solution of DMD. Then, we provide the Gibbs sampler for the posterior inference in Bayesian DMD. Due to the Bayesian treatment, we can infer posteriors of DMD-related quantities, such as dynamic modes and eigenvalues. Moreover, we can consider variants of DMD within the unified Bayesian framework by modifying the probabilistic model. 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.

The remainder of this paper is organized as follows. We briefly review the underlying theory of DMD and its numerical procedure in Section 2. The probabilistic model for DMD is described in Section 3, and based on that model, Bayesian DMD is introduced in Section 4. In Section 5, we review the related work. In Section 6, we show the experimental results with synthetic and real-world datasets. This paper is concluded in Section 7.

2 Background

We briefly review the underlying theory of DMD, the spectral decomposition of nonlinear dynamical systems based on the Koopman operator. We recommend readers to consult papers such as [Mezić, 2005; Budišić *et al.*, 2012; Mezić, 2013] for more details on the theory of the Koopman operator and decomposition based on it.

2.1 Koopman Spectral Analysis

Our interest lies in analyzing a (possibly, nonlinear) discrete-time nonlinear dynamical system

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t), \quad \boldsymbol{x} \in \mathcal{M},$$

where $\mathcal{M}$ is the state space and $t \in \{0\} \cup \mathbb{N}$ is the time index. Let $g : \mathcal{M} \to \mathbb{C}$ be a measurement function (observable) in a function space $\mathcal{G}$. *Koopman operator* $\mathcal{K} : \mathcal{G} \to \mathcal{G}$ is defined as an *infinite-dimensional linear* operator such that

$$\mathcal{K}g(\boldsymbol{x}) = g(\boldsymbol{f}(\boldsymbol{x})), \quad g \in \mathcal{G}.$$

Defining $\mathcal{K}$, we can lift nonlinear dynamics $\boldsymbol{f}$ to a linear (but infinite-dimensional) regime. However, it is difficult to numerically compute $\mathcal{K}$ from data because of its infinite dimensionality. Nonetheless, the infinite-dimensional system can further be lifted to a finite-dimensional one as follows.

Suppose that there exists an invariant subspace of $\mathcal{K}$, i.e., a subspace $G \subset \mathcal{G}$ such that

$$\mathcal{K}g \in G, \ \forall g \in G,$$

and that a set of observables $\{g_1, \dots, g_n\}$ $(n < \infty)$ spans G. Moreover, consider the restriction of $\mathcal{K}$ to G and denote it by $K : G \to G$. Now K is a *finite-dimensional linear* operator. Note that K has a matrix-form representation $\boldsymbol{K}$ with respect to $\{g_1, \dots, g_n\}$, i.e.,

$$[Kg_1 \quad \cdots \quad Kg_n]^\mathsf{T} = \boldsymbol{K}\boldsymbol{g},$$

wherein $\boldsymbol{g} = [g_1 \quad \cdots \quad g_n]^\mathsf{T}$. In the sequel, we assume that such invariant subspace G and the set of observables that spans G exist.

Since K is linear, modal decomposition based on it can be derived as follows. Let $\varphi : \mathcal{M} \to \mathbb{C}$ be an eigenfunction of K with eigenvalue $\lambda \in \mathbb{C}$, i.e.,

$$K\varphi(\boldsymbol{x}) = \lambda\varphi(\boldsymbol{x}).$$

This eigenfunction with respect to $\{g_1, \dots, g_n\}$ is obtained by $\varphi(\boldsymbol{x}) = \boldsymbol{z}^\mathsf{H}\boldsymbol{g}(\boldsymbol{x})$, wherein $\boldsymbol{z}$ is the left-eigenvector of $\boldsymbol{K}$ corresponding to eigenvalue λ. Moreover, let $\boldsymbol{v}_i$ and $\boldsymbol{z}_i$ respectively be the right- and the left- eigenvectors of $\boldsymbol{K}$ corresponding to eigenvalue λ_i, and suppose that they are normalized so that $\boldsymbol{v}_i^\mathsf{H}\boldsymbol{z}_{i'} = \delta_{ii'}$, without loss of generality. If all the eigenvalues of K are distinct (i.e., their multiplicity is one), any values of $\boldsymbol{g}$ are expressed as

$$\boldsymbol{g}(\boldsymbol{x}_t) = \sum_{i=1}^{n} \varphi_i(\boldsymbol{x}_t)\boldsymbol{v}_i. \tag{1}$$

Applying K to both sides of Eq. (1), we have

$$\boldsymbol{g}(\boldsymbol{x}_{t+1}) = \sum_{i=1}^{n} \lambda_i\varphi_i(\boldsymbol{x}_t)\boldsymbol{v}_i. \tag{2}$$

In this way, modal decomposition of observables via the Koopman operator is given by

$$\boldsymbol{g}(\boldsymbol{x}_t) = \sum_{i=1}^{n} \lambda_i^t \boldsymbol{w}_i, \quad \boldsymbol{w}_i = \varphi_i(\boldsymbol{x}_0)\boldsymbol{v}_i, \tag{3}$$

wherein the values of $\boldsymbol{g}$ are described as a sum of *Koopman modes* $\boldsymbol{w}$, whose temporal frequency and decay rate are given by $\angle\lambda$ and $|\lambda|$, respectively. From the above, we could see that, unlike the classical modal decomposition of linear time invariant systems, this theory is applicable even to nonlinear dynamical systems.

2.2 Dynamic Mode Decomposition

DMD [Rowley *et al.*, 2009; Schmid, 2010; Kutz *et al.*, 2016a] is a numerical decomposition method, and it coincides with modal decomposition via the Koopman operator under some conditions. Suppose we have data matrices:

$$\begin{aligned} \boldsymbol{Y}_0 &= [\boldsymbol{g}(\boldsymbol{x}_0) \quad \cdots \quad \boldsymbol{g}(\boldsymbol{x}_{m-1})] \in \mathbb{C}^{n \times m} \quad \text{and} \\ \boldsymbol{Y}_1 &= [\boldsymbol{g}(\boldsymbol{x}_1) \quad \cdots \quad \boldsymbol{g}(\boldsymbol{x}_m)] \in \mathbb{C}^{n \times m}, \end{aligned} \tag{4}$$

where $\boldsymbol{g}$ is again a concatenation of n observables, and $m+1$ is the number of snapshots in the dataset. The popular implementation of DMD [Tu *et al.*, 2014] is shown in Algorithm 1, with which we finally compute the eigenvectors of a matrix $\boldsymbol{A} = \boldsymbol{Y}_1\boldsymbol{Y}_0^\dagger$ corresponding to its nonzero eigenvalues, wherein $\boldsymbol{Y}_\mathbf{0}^\dagger$ denotes the Moore–Penrose pseudoinverse of $\boldsymbol{Y}_0$. If the components of $\boldsymbol{g}$ span a subspace invariant to $\mathcal{K}$ and all modes are sufficiently excited in the data (i.e., $\operatorname{rank}(\boldsymbol{Y}_0) = \dim(G)$), then the dynamic modes computed by Algorithm 1 converge to the Koopman modes in modal decomposition (3) in the large sample limit.

Algorithm 1 (DMD [Tu *et al.*, 2014]**).**

(1) Compute the compact SVD of $\boldsymbol{Y}_0 = \boldsymbol{U}_r\boldsymbol{S}_r\boldsymbol{V}_r^\mathsf{H}$.

(2) Define a matrix $\tilde{\boldsymbol{A}} = \boldsymbol{U}_r^\mathsf{H}\boldsymbol{Y}_1\boldsymbol{V}_r\boldsymbol{S}_r^{-1}$.

(3) Calculate eigendecomposition of $\tilde{\boldsymbol{A}}$, i.e., compute $\tilde{\boldsymbol{w}}$ and λ such that $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{w}} = \lambda\tilde{\boldsymbol{w}}$.

(4) Return *dynamic modes* $\boldsymbol{w} = \lambda^{-1}\boldsymbol{Y}_1\boldsymbol{V}_r\boldsymbol{S}_r^{-1}\tilde{\boldsymbol{w}}$ and corresponding eigenvalues λ.

Note that, for DMD to obtain the theoretical rationale, the components of $\boldsymbol{g}$ need to span an (approximately) invariant subspace. There are several studies that address this issue (e.g., the use of nonlinear basis functions [Williams *et al.*, 2015], reproducing kernels [Kawahara, 2016], and delay coordinates [Susuki and Mezić, 2015; Arbabi and Mezić, 2016]). In this paper, however, we simply assume data are generated with observables that intrinsically span an (approximately) invariant subspace, as in the previous studies on DMD.

3 Probabilistic DMD

We develop a probabilistic model associated with modal decomposition via the Koopman operator (Eqs. (1) and (2)). The maximum-likelihood estimator (MLE) of this model coincides with the solution of DMD in the no-noise limit. As will be described in the next section, this probabilistic model forms the foundation for Bayesian DMD.

3.1 Generative Model

Let $\boldsymbol{y}_{\ell,j} \in \mathbb{C}^n$ be the j-th column of $\boldsymbol{Y}_\ell$ in Eq. (4) *plus observation noise*, for $\ell = 0, 1$. Following the relations in Eqs. (1) and (2), the probabilistic DMD model for such data can be given by

$$\begin{aligned} \boldsymbol{y}_{0,j} | \varphi_{1,j}, \dots, \varphi_{k,j} &\sim \mathcal{CN}\left(\sum_{i=1}^{k} \varphi_{i,j}\boldsymbol{w}_i, \ \sigma^2\boldsymbol{I}\right), \\ \boldsymbol{y}_{1,j} | \varphi_{1,j}, \dots, \varphi_{k,j} &\sim \mathcal{CN}\left(\sum_{i=1}^{k} \lambda_i\varphi_{i,j}\boldsymbol{w}_i, \ \sigma^2\boldsymbol{I}\right), \end{aligned} \tag{5}$$

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

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

Here, $\boldsymbol{w}_{1:k}$, $\lambda_{1:k}$, and σ^2 are the parameters to be estimated ($\lambda_{1:k}$ denotes a set $\{\lambda_1, \ldots, \lambda_k\}$), 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

$$\varphi_{i,j} \sim \mathcal{CN}(0,\ 1). \tag{6}$$

3.2 Maximum-likelihood Estimator

To derive the MLE of probabilistic DMD, let us rewrite likelihood (5) in a matrix form, i.e.,

$$\boldsymbol{y}_j | \boldsymbol{\varphi}_j \sim \mathcal{CN}\left(\boldsymbol{B} \boldsymbol{\varphi}_j,\ \sigma^2 \boldsymbol{I} \right),$$

where we use notations as follows:

$$\boldsymbol{y}_j = \begin{bmatrix} \boldsymbol{y}_{0,j} \\ \boldsymbol{y}_{1,j} \end{bmatrix}, \quad \boldsymbol{\varphi}_j = \begin{bmatrix} \varphi_{1,j} & \ldots & \varphi_{k,j} \end{bmatrix}^{\mathsf{T}},$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{W\Lambda} \end{bmatrix}, \ \boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1 & \ldots & \boldsymbol{w}_k \end{bmatrix}, \ \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k).$$

Marginalizing out $\boldsymbol{\varphi}$ with prior (6), we have

$$\boldsymbol{y}_j \sim \mathcal{CN}\left(\boldsymbol{0},\ \boldsymbol{B}\boldsymbol{B}^{\mathsf{H}} + \sigma^2 \boldsymbol{I} \right). \tag{7}$$

In the following, we describe the relationship between probabilistic model (7), total-least-squares DMD [Dawson *et al.*, 2016], which is a "noise-aware" variant of DMD, and standard DMD (Algorithm 1). In short, their estimation results coincide in the no-noise limit.

Proposition 1. *Suppose we have a dataset that is possibly contaminated by observation noises* $\boldsymbol{E}$*:*

$$\hat{\boldsymbol{Y}}_\ell = \boldsymbol{Y}_\ell + \boldsymbol{E}_\ell = [\boldsymbol{y}_{\ell,1} \quad \cdots \quad \boldsymbol{y}_{\ell,m}], \quad \ell = 0, 1,$$

and let $\hat{\boldsymbol{Y}} = \begin{bmatrix} \hat{\boldsymbol{Y}}_0^{\mathsf{T}} & \hat{\boldsymbol{Y}}_1^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$ *and* $\boldsymbol{\Sigma}_y = m^{-1} \hat{\boldsymbol{Y}} \hat{\boldsymbol{Y}}^{\mathsf{H}}$. *In addition, let* $(\sigma^2)^\star$, $\boldsymbol{W}^\star$ *and* $\boldsymbol{\Lambda}^\star$ *be the MLEs of Eq.* (7) *given* $\hat{\boldsymbol{Y}}$. *If* $k = n$*, then the columns of* $\boldsymbol{W}^\star$ *and the elements of* $\mathrm{diag}(\boldsymbol{\Lambda}^\star)$ *coincide with the dynamic modes and eigenvalues obtained by total-least-squares DMD, respectively.*

Proof. Following [Tipping and Bishop, 1999], the MLEs for probabilistic model (7) are given as

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

$$\boldsymbol{B}^\star = \begin{bmatrix} \boldsymbol{W}^\star \\ \boldsymbol{W}^\star \boldsymbol{\Lambda}^\star \end{bmatrix} = \boldsymbol{U}_k (\boldsymbol{M}_k - (\sigma^2)^\star \boldsymbol{I})^{\frac{1}{2}} \boldsymbol{R}$$

with $\boldsymbol{U}_k = [\boldsymbol{u}_1 \quad \ldots \quad \boldsymbol{u}_k]$ and $\boldsymbol{M}_k = \mathrm{diag}(\mu_1, \ldots, \mu_k)$, where μ_i is the i-th largest eigenvalue of $\boldsymbol{\Sigma}_y$ with corresponding eigenvector $\boldsymbol{u}_i$, and $\boldsymbol{R}$ is an arbitrary unitary matrix. If $k = n$, we have

$$\boldsymbol{W}^\star \boldsymbol{\Lambda}^\star (\boldsymbol{W}^\star)^{-1} = \boldsymbol{U}_{1,n} \boldsymbol{U}_{0,n}^{-1},$$

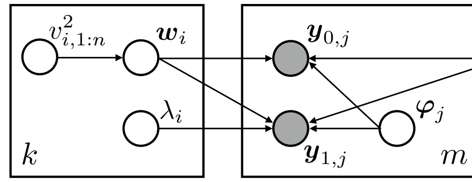


Figure 1: Graphical model of Bayesian DMD.

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

Proposition 2. *If* $\boldsymbol{Y}_0$ *and* $\boldsymbol{Y}_1$ *are linearly consistent [Tu* et al.*, 2014] and there is no observation noise (i.e.,* $\boldsymbol{E} = 0$*), then the estimation results of total-least-squares DMD coincides with those of standard DMD (Algorithm 1).*

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

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

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

4 Bayesian DMD

For the Bayesian treatment of DMD, we consider the following priors on the parameters in probabilistic model (5). First, we put a Gaussian prior on $\boldsymbol{w}_{1:k}$:

$$\boldsymbol{w}_i | v_{i,1:n}^2 \sim \mathcal{CN}\left(\boldsymbol{0},\ \mathrm{diag}\left(v_{i,1}^2, \ldots, v_{i,n}^2 \right) \right) \tag{8}$$

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

$$v_{i,d}^2 \sim \mathrm{InvGamma}\left(\alpha_v,\ \beta_v \right), \tag{9}$$

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

$$\lambda_i \sim \mathcal{CN}\left(0,\ 1\right), \quad \sigma^2 \sim \mathrm{InvGamma}\left(\alpha_\sigma,\ \beta_\sigma \right).$$

A graphical model of Bayesian DMD is shown in Figure 1.

4.1 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 notations

$$\boldsymbol{\xi}_{-i,j} = \boldsymbol{y}_{0,j} - \sum_{i' \neq i} \varphi_{i',j} \boldsymbol{w}_{i'}, \quad \text{and}$$

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

The sampling procedures of the Gibbs sampler are summarized in Algorithm 2. In the following, $\bar{\lambda}$ denotes the complex conjugate of λ, and $[\boldsymbol{w}_i]_d$ denotes the d-th element of $\boldsymbol{w}_i$. In our implementation, the hyperparameters α and β were set to 10^{-3}.

Algorithm 2 (Gibbs sampling for Bayesian DMD)**.**

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

$$\boldsymbol{P}_{\boldsymbol{w}_i} = \mathrm{diag}\left(v_{i,1}^{-2}, \ldots, v_{i,n}^{-2}\right) + \frac{(1+|\lambda_i|^2)\sum_j |\varphi_{i,j}|^2}{\sigma^2}\boldsymbol{I},$$
$$\boldsymbol{m}_{\boldsymbol{w}_i} = \boldsymbol{P}_{\boldsymbol{w}_i}^{-1}\frac{1}{\sigma^2}\sum_j \bar{\varphi}_{i,j}\left(\boldsymbol{\xi}_{-i,j} + \bar{\lambda}_i \boldsymbol{\eta}_{-i,j}\right).$$

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

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

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

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

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

$$\boldsymbol{P}_{\boldsymbol{\varphi}_j} = \boldsymbol{I} + \frac{1}{\sigma^2}\left(\boldsymbol{W}^{\mathsf{H}}\boldsymbol{W} + \bar{\boldsymbol{\Lambda}}\boldsymbol{W}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{\Lambda}\right),$$
$$\boldsymbol{m}_{\boldsymbol{\varphi}_j} = \boldsymbol{P}_{\boldsymbol{\varphi}_j}^{-1}\frac{1}{\sigma^2}\left(\boldsymbol{W}^{\mathsf{H}}\boldsymbol{y}_{0,j} + \bar{\boldsymbol{\Lambda}}\boldsymbol{W}^{\mathsf{H}}\boldsymbol{y}_{1,j}\right).$$

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

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

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

4.2 Sparsity-promoting Prior

One of the difficulties when applying DMD to noisy data in practice is how to determine 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. This approach works without manual tuning of hyperparameters through the empirical Bayes technique.

Following [Park and Casella, 2008], we incorporate the two-level Laplacian prior on $\boldsymbol{w}_{1:k}$, replacing prior (8) and hyperprior (9) by

$$\boldsymbol{w}_i | v_{i,1:n}^2 \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma^2\,\mathrm{diag}\left(v_{i,1}^2, \ldots, v_{i,n}^2\right)\right), \quad \text{and}$$
$$v_{i,d}^2 \sim \mathrm{Exponential}(\gamma_i^2/2)$$

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

$$\boldsymbol{P}_{\boldsymbol{w}_i} = \frac{1}{\sigma^2}\mathrm{diag}\left(v_{i,1}^{-2}, \ldots, v_{i,n}^{-2}\right) + \frac{(1+|\lambda_i|^2)\sum_j |\varphi_{i,j}|^2}{\sigma^2}\boldsymbol{I},$$
$$a_{\sigma^2} = \alpha_\sigma + 2mn + \frac{1}{2}kn,$$
$$b_{\sigma^2} = \beta_\sigma + \sum_j (\boldsymbol{y}_{0,j} - \boldsymbol{W}\boldsymbol{\varphi}_j)^{\mathsf{H}}(\boldsymbol{y}_{0,j} - \boldsymbol{W}\boldsymbol{\varphi}_j) + \sum_j (\boldsymbol{y}_{1,j} - \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{\varphi}_j)^{\mathsf{H}}(\boldsymbol{y}_{1,j} - \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{\varphi}_j) + \sum_{i,d}\frac{|[\boldsymbol{w}_i]_d|^2}{2v_{i,d}^2}.$$

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:

$$a_{v_{i,d}^2} = \gamma_i^2, \quad b_{v_{i,d}^2} = \frac{|[\boldsymbol{w}_i]_d|^2}{\sigma^2}, \quad \text{and} \quad p_{v_{i,d}^2} = \frac{1}{2}.$$

To draw a sample from the generalized inverse Gaussian distribution, we used an efficient sampler of [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

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

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.

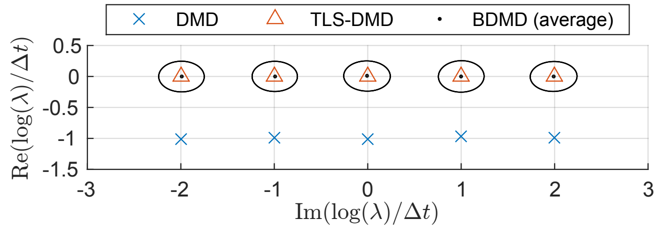


Figure 2: Eigenvalues estimated for the Stuart–Landau equation. Following a convention, $\log(\lambda)/\Delta t$ is plotted. In this scale, the true eigenvalues lie on the imaginary axis since the data are periodic. The ellipse denotes the 95% credible interval of the samples generated from the Gibbs sampler of BDMD, for each eigenvalue.

5 Related Work

DMD was originally proposed as a tool for diagnostics of fluid flows [Rowley *et al.*, 2009; Schmid, 2010], and it has been utilized in various fields of science and engineering, including fluid mechanics [Schmid *et al.*, 2011], analyses of power systems [Susuki and Mezić, 2014], epidemiology [Proctor and Eckhoff, 2015], robotic control [Berger

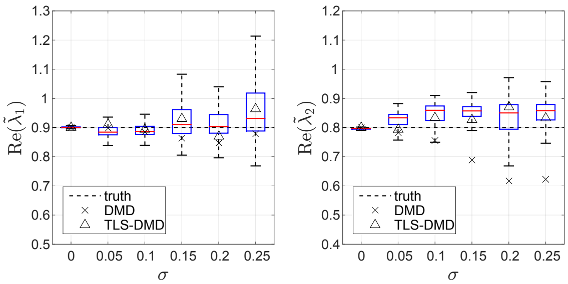


Figure 3: Estimated values of (left) λ_1 and (right) λ_2, for each noise magnitude σ. The box plots show the statistics of the samples from the Gibbs sampler of BDMD (the red lines denote the sample medians).

Table 1: Averages (and the standard deviations) of the absolute errors of estimated (left) λ_1 and (right) 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.

$\lvert\Delta\lambda_1\rvert$	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)

$\lvert\Delta\lambda_2\rvert$	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)

et al., 2015], neuroscience [Brunton *et al.*, 2016a], chaotic systems [Brunton *et al.*, 2016b], image processing [Kutz *et al.*, 2016b], and nonlinear system identification [Mauroy and Goncalves, 2016]. Moreover, there are several algorithmic variants such as the use of nonlinear basis functions [Williams *et al.*, 2015], formulation in a reproducing kernel Hilbert space [Kawahara, 2016], and consideration for controlled systems [Proctor *et al.*, 2016].

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 Proposition 1, we have shown that the MLE of probabilistic DMD coincides with the solution of total-least-squares DMD.

6 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.

6.1 Estimation with 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:

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

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

$$\boldsymbol{y}_t = \begin{bmatrix} e^{-2i\theta_t} & e^{-i\theta_t} & 1 & e^{i\theta_t} & e^{2i\theta_t} \end{bmatrix}^\mathsf{T} + \boldsymbol{e}_t,$$

where each element of $\boldsymbol{e}_t$ 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), total-least-squares DMD (TLS-DMD) [Dawson *et al.*, 2016], and BDMD (with $k = 5$). The estimated eigenvalues are plotted in Figure 2 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 Proposition 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.

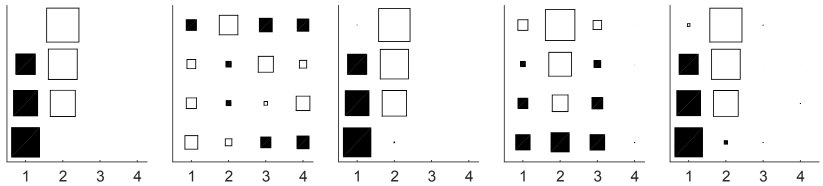
Damping modes We also investigated the performance for identifying damping modes, i.e., modes that decay rapidly over time. Generally, it is more difficult to identify damping modes than to identify modes in a limit cycle. The dataset was generated by

$$\boldsymbol{y}_t = \lambda_1^t \begin{bmatrix} 2 & 2 \end{bmatrix}^\mathsf{T} + \lambda_2^t \begin{bmatrix} 2 & -2 \end{bmatrix}^\mathsf{T} + \boldsymbol{e}_t,$$

where $\boldsymbol{e}_t$ 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 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 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.

6.2 Automatic Relevance Determination

We conducted an experiment to investigate how well BDMD can determine the number of modes automatically. We gen-



(a) ground truth (b) DMD (c) SP-DMD (d) BDMD (e) BDMD-sp

Figure 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.

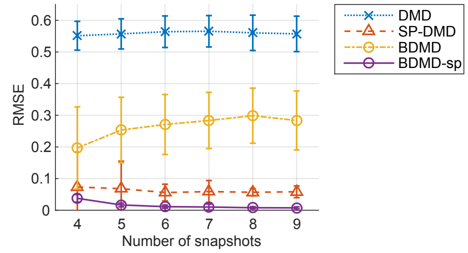
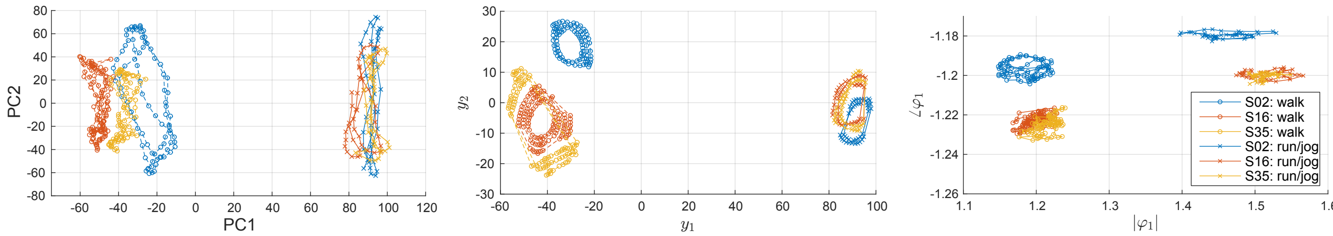


Figure 5: Averages of the RMSEs of estimated dynamic modes over 100 trials for each number of snapshots.



(a) PCA (b) t-SNE (c) BDMD-sp

Figure 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).

erated a dataset by

$$\boldsymbol{y}_t = A \begin{bmatrix} 0.9^t \\ 0.7^t \\ 0 \\ 0 \end{bmatrix} + \boldsymbol{e}_t, \quad A = \begin{bmatrix} 0 & -5 & 0 & 0 \\ 2 & -4 & 0 & 0 \\ 3 & -3 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix},$$

where $\boldsymbol{e}_t$ was zero-mean Gaussian noise with variance 10^{-4}. For determining the number of modes given noisy datasets, standard DMD (Algorithm 1) may utilize the truncation of small singular values at SVD step (Step 1), 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 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 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 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.

6.3 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.[1] 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.

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 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 6. The distinction between "walk" and "run/jog" is clearly observed as the different distributions of the trajectories in every plot of Figure 6. The distinction between Subjects #2, #16 and #35 is

[1] Downloaded from `http://mocap.cs.cmu.edu/`.

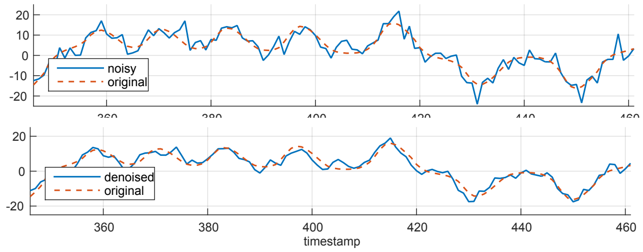


Figure 7: A part of (upper) the noisy and (lower) the denoised time-series. The RMSE decreased from 3.2 to 2.3.

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 6c) shows the most consistent result wherein the trajectories of Subject #2 are consistently distributed in the upper part of the plot.

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 7. 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.

7 Conclusions

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.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP15J09172, JP16H01548, JP26280086, JP15K12639, and JP26289320.

References

[Arbabi and Mezić, 2016] Hassan Arbabi and Igor Mezić. Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator. *arXiv:1611.06664v2*, 2016.

[Berger *et al.*, 2015] Erick Berger, Mark Sastuba, David Vogt, Bernhard Jung, and Heni B. Amor. Estimation of perturbations in robotic behavior using dynamic mode decomposition. *Advanced Robotics*, 29(5):331–343, 2015.

[Bonnet *et al.*, 1994] Jean P. Bonnet, Daniel R. Cole, Joël Delville, Mark N. Glauser, and Lawrence S. Ukeiley. Stochastic estimation and proper orthogonal decomposition: Complementary techniques for identifying structure. *Experiments in Fluids*, 17:307–314, 1994.

[Brunton *et al.*, 2016a] Bingni W. Brunton, Lise A. Johnson, Jeffrey G. Ojemann, and J. Nathan Kutz. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. of Neuroscience Methods*, 258:1–15, 2016.

[Brunton *et al.*, 2016b] Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, Eurika Kaiser, and J. Nathan Kutz. Chaos as an intermittently forced linear system. *arXiv:1608.05306v1*, 2016.

[Budišić *et al.*, 2012] Marko Budišić, Ryan M. Mohr, and Igor Mezić. Applied Koopmanism. *Chaos*, 22:047510, 2012.

[Casella, 2001] George Casella. Empirical Bayes Gibbs sampling. *Biostatistics*, 2:485–500, 2001.

[Chen *et al.*, 2012] Kelvin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses. *Nonlinear Science*, 22(6):887–915, 2012.

[Dawson *et al.*, 2016] Scott T. M. Dawson, Maziar S. Hemati, Matthew O. Williams, and Clarence W. Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, 2016.

[Devroye, 2014] Luc Devroye. Random variate generation for the generalized inverse Gaussian distribution. *Statistics and Computing*, 24:239–246, 2014.

[Dicle *et al.*, 2016] Caglayan Dicle, Hassan Mansour, Dong Tian, Mouhacine Benosman, and Anthony Vetro. Robust low rank dynamic mode decomposition for compressed domain crowd and traffic flow analysis. In *Proc. of the IEEE Int. Conf. on Multimedia and Expo*, pages 1–6, 2016.

[Duke *et al.*, 2012] Daniel J. Duke, Julio Soria, and Damon Honnery. An error analysis of the dynamic mode decomposition. *Experiments in Fluids*, 52(2):529–542, 2012.

[Goodman, 1963] Nathaniel R. Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34:152–177, 1963.

[Héas and Herzet, 2017] Patrick Héas and Cédric Herzet. Optimal low-rank dynamic mode decomposition. In *Proc. of the 42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2017.

[Holmes *et al.*, 2012] Phillip Holmes, John L. Lumley, and Gal Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 2nd edition, 2012.

[Jolliffe, 2002] Ian T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.

[Jovanović *et al.*, 2014] Mihailo R. Jovanović, Peter J. Schmid, and Joseph W. Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.

[Kawahara, 2016] Yoshinobu Kawahara. Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis. In *Advances in Neural Information Processing Systems*, volume 29, pages 911–919. 2016.

[Kutz *et al.*, 2016a] J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.

[Kutz *et al.*, 2016b] J. Nathan Kutz, Xing Fu, and Steven L. Brunton. Multiresolution dynamic mode decomposition. *SIAM J. on Applied Dynamical Systems*, 15(2):713–735, 2016.

[Lorenz, 1963] Edward N. Lorenz. Deterministic nonperiodic flow. *J. of the Atmospheric Sciences*, 20:130–141, 1963.

[Mauroy and Goncalves, 2016] Alexandre Mauroy and Jorge Goncalves. Linear identification of nonlinear systems: A lifting technique based on the Koopman operator. In *Proc. of the IEEE 55th Conf. on Decision and Control*, pages 6500–6505, 2016.

[Mezić, 2005] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41:309–325, 2005.

[Mezić, 2013] Igor Mezić. Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, 2013.

[Noack *et al.*, 2003] Bernd R. Noack, Konstantin Afanasiev, Marek Morzyński, Gilead Tadmor, and Frank Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. of Fluid Mechanics*, 497:335–363, 2003.

[Pan *et al.*, 2015] Chong Pan, Dong Xue, and Jinjun Wang. On the accuracy of dynamic mode decomposition in estimating instability of wave packet. *Experiments in Fluids*, 56(8):164, 2015.

[Park and Casella, 2008] Trevor Park and George Casella. The Bayesian lasso. *J. of the American Statistical Association*, 103:681–686, 2008.

[Proctor and Eckhoff, 2015] Joshua L. Proctor and Philip A. Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *Int. Health*, 7(2):139–145, 2015.

[Proctor *et al.*, 2016] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM J. on Applied Dynamical Systems*, 15:142–161, 2016.

[Rowley *et al.*, 2009] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. *J. of Fluid Mechanics*, 641:115–127, 2009.

[Schmid *et al.*, 2011] Peter J. Schmid, Larry Li, Matthew P. Juniper, and Oliver Pust. Applications of the dynamic mode decomposition. *Theoretical and Computational Fluid Dynamics*, 25(1):249–259, 2011.

[Schmid, 2010] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. of Fluid Mechanics*, 656:5–28, 2010.

[Susuki and Mezić, 2014] Yoshihiko Susuki and Igor Mezić. Nonlinear Koopman modes and power system stability assessment without models. *IEEE Trans. on Power Systems*, 29(2):899–907, 2014.

[Susuki and Mezić, 2015] Yoshihiko Susuki and Igor Mezić. A prony approximation of Koopman mode decomposition. In *Proc. of the IEEE 54th Conf. on Decision and Control*, pages 7022–7027, 2015.

[Tipping and Bishop, 1999] Michael E. Tipping and Christopher M. Bishop. Probabilistic principal component analysis. *J. of the Royal Statistical Society, Series B*, 61:611–622, 1999.

[Tu *et al.*, 2014] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. *J. of Computational Dynamics*, 1(2):391–421, 2014.

[van der Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing high-dimensional data using t-SNE. *J. of Machine Learning Research*, 9:2579–2605, 2008.

[Williams *et al.*, 2015] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. of Nonlinear Science*, 25(6):1307–1346, 2015.

[Wynn *et al.*, 2013] Andrew Wynn, David Pearson, Bharathram Ganapathisubramani, and Paul J. Goulart. Optimal mode decomposition for unsteady flows. *J. of Fluid Mechanics*, 733:473–503, 2013.