Bayesian Dynamic Mode Decomposition with Variational Matrix Factorization

Takahiro Kawashima,[1] **Hayaru Shouno,** [1] **Hideitsu Hino** [2]
[1] The University of Electro-Communications, Tokyo, Japan
[2] The Institute of Statistical Mathematics, Tokyo, Japan
kawashima@uec.ac.jp, shouno@uec.ac.jp, hino@ism.ac.jp

Abstract

Dynamic mode decomposition (DMD) and its extensions are data-driven methods that have substantially contributed to our understanding of dynamical systems. However, because DMD and most of its extensions are deterministic, it is difficult to treat probabilistic representations of parameters and predictions. In this work, we propose a novel formulation of a Bayesian DMD model. Our Bayesian DMD model is consistent with the procedure of standard DMD, which is to first determine the subspace of observations, and then compute the modes on that subspace. Variational matrix factorization makes it possible to realize a fully-Bayesian scheme of DMD. Moreover, we derive a Bayesian DMD model for incomplete data, which demonstrates the advantage of probabilistic modeling. Finally, both of nonlinear simulated and real-world datasets are used to illustrate the potential of the proposed method.

1 Introduction

Understanding dynamics of multivariate sequential data is an essential task in many scientific fields, including biology (Aihara and Suzuki 2010), atmospheric science (Cotton and Anthes 1992; Vallis 2017), fluid dynamics (Dowell, Hall, and Romanowski 1997), and control engineering (Levin and Narendra 1993). Such data are often complex and high-dimensional, and hence, data-driven approaches rather than complicated models are required in many cases. In fact, in recent years, several machine learning methods for analyzing dynamical systems have been proposed including dynamic-inner partial least squares (DiPLS) (Dong and Qin 2015) and dynamic-inner principal component analysis (DiPCA) (Dong and Qin 2018), neural ordinary differential equations (Chen et al. 2018), the second-order adjoint method (Ito et al. 2016), and sparse identification of nonlinear dynamics (SINDy) (Brunton, Proctor, and Kutz 2016). Proper orthogonal decomposition (POD) (Kosambi 1943; Chatterjee 2000; Jolliffe 2002) has become a commonly used method in computational fluid dynamics since it was employed for analyzing turbulent flow in (Lumley 1967). POD is essentially equivalent to principal component analysis (PCA), singular value decomposition (SVD), and the Karhunen–Loéve decomposition. Although POD is a powerful method, it cannot capture the dynamics because it does not assume any relationships among the samples.

Dynamic mode decomposition (DMD) (Schmid 2010; Kutz et al. 2016) was proposed to overcome this limitation. In short, DMD is an efficient method for approximating the eigenvalues and eigenfunctions of a Koopman operator governing multivariate sequential data. Because of the advantages of DMD, its variations have also been developed, such as sparsity-promoting DMD (spDMD) (Jovanović, Schmid, and Nichols 2014), higher-order DMD (HODMD) (Le Clainche and Vega 2017), and graph DMD (Fujii and Kawahara 2019); extended DMD (Williams, Kevrekidis, and Rowley 2015) and kernel DMD (Williams, Rowley, and Kevrekidis 2015; Kawahara 2016) are representatives of nonlinearized DMD. However, almost all extensions of DMD are deterministic. Because probabilistic perspectives provide a flexible representation in general, considering a probabilistic version of DMD would offer many benefits, including missing data imputation and estimation from incomplete data.

In this paper, we propose a novel variational matrix factorization (VMF)-based Bayesian formulation of DMD (Lim and Teh 2007; Nakajima and Sugiyama 2011; Seeger and Bouchard 2012). We refer to our proposed model as Bayesian DMD with VMF (BDMD-VMF). The learning procedure of the BDMD-VMF model is consistent with that of a standard DMD model; first determining a lower-dimensional representation of the data and then estimating the dynamics on the latent space. The likelihood function of BDMD-VMF has a projection matrix whose size depends on the input dimensions. However, we can marginalize out this projection matrix with the variational posterior of VMF. After marginalization, the number of parameters to be estimated by BDMD-VMF solely depends on the dimensionality of the latent space, which corresponds to the number of dominant modes. Moreover we derive the BDMD-VMF model for cases with missing data values.

It is remarkable that Bayesian formulation of DMD was pioneered by Takeishi et al. (2017). We refer to their model as conventional Bayesian DMD. The conventional Bayesian DMD model cannot capture the input dynamics appropriately because the degrees of freedom increases with the increase of data. We experimentally demonstrate the instability of conventional Bayesian DMD, which occurs with a

Symbol	Description	Symbol	Description
$\|z\|$	absolute value of z	$\mathrm{conj}(\boldsymbol{A})$	complex conjugate of $\boldsymbol{A}$
A_{ij}	(i,j)-th element of $\boldsymbol{A}$	$\mathrm{tr}\{\boldsymbol{A}\}$	trace of $\boldsymbol{A}$
$\boldsymbol{A}_{i:}$	i-th row of $\boldsymbol{A}$ (row vector)	$\mathrm{diag}(\{\boldsymbol{A}_i\})$	block diagonal matrix with $\{\boldsymbol{A}_i\}$
$\boldsymbol{A}_{:j}$	j-th column of $\boldsymbol{A}$ (column vector)	$\|\boldsymbol{A}\|_F$	Frobenius norm of $\boldsymbol{A}$
$\boldsymbol{A}^{-1}$	inverse of $\boldsymbol{A}$	$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$
$\boldsymbol{A}^{-}$	pseudo-inverse of $\boldsymbol{A}$	$\mathrm{vec}(\boldsymbol{A})$	vectorization of $\boldsymbol{A}$
$\boldsymbol{A}^{\top}$	transpose of $\boldsymbol{A}$	$\langle \boldsymbol{A} \rangle_{p(\boldsymbol{A})}$	expectation of $\boldsymbol{A}$ with distribution $p(\boldsymbol{A})$
$\boldsymbol{A}^{-\top}$	inverse of $\boldsymbol{A}^{\top}$	$\boldsymbol{0}_N$	N-dimensional zero vector
$\boldsymbol{A}^{*}$	Hermitian conjugate of $\boldsymbol{A}$	$\boldsymbol{O}_{N_1 N_2}$	$N_1 \times N_2$ zero matrix
$\mathrm{Re}\{\boldsymbol{A}\}$	real part of $\boldsymbol{A}$	$\boldsymbol{I}_N$	$N \times N$ identity matrix
$\det(\boldsymbol{A})$	determinant of $\boldsymbol{A}$	$\#\mathcal{S}$	number of elements in set $\mathcal{S}$

Table 1: Notation list

small sample size in Section 5.

Our main contributions can be summarized as follows:

- We propose a novel Bayesian DMD model, called BDMD-VMF, which is governed by a small number of parameters and can be employed even with missing data values.
- We derive the complex extension of VMF by extending the definition of a normal distribution to complex-valued random matrices to establish the BDMD-VMF model.
- We apply BDMD-VMF to nonlinear simulated and real-world incomplete gyroscopic data and experimentally demonstrate that the BDMD-VMF model can capture input dynamics more stably than conventional Bayesian DMD.

The remainder of this paper is organized as follows. Section 2 describes the algorithm of DMD and the relationship between Koopman operator theory and DMD; we also review the formulation of conventional Bayesian DMD. In Section 3, complex distributions are defined as a basic concept for the proposed model. Novel Bayesian DMD models (for cases of both the complete and incomplete data), BDMD-VMF models, are derived in Section 4. Section 5 shows illustrative examples of our models. Finally, Section 6 presents the concluding remarks.

Table 1 lists the symbols used in this paper.

2 Background

Koopman Spectral Analysis

Consider latent variables $\boldsymbol{x}_t \in \mathcal{X}$, which are evolved using the following nonlinear function $\boldsymbol{f} : \mathcal{X} \to \mathcal{X}$:

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

where $\mathcal{X}$ is a finite-dimensional state space. Suppose that data are obtained through the action of an observable $\mathcal{G} \ni g : \mathcal{X} \to \mathbb{C}$. Here, $\mathcal{G}$ is an appropriate space for scalar-valued observables. The Koopman operator $\mathcal{K} : \mathcal{G} \to \mathcal{G}$ mapping an observation $g(\boldsymbol{x}_t)$ to $g(\boldsymbol{x}_{t+1})$ is defined as follows:

$$(\mathcal{K}g)(\boldsymbol{x}_t) = g(\boldsymbol{x}_{t+1}) = g \circ \boldsymbol{f}(\boldsymbol{x}_t).$$

Because $\mathcal{K}$ is a linear operator regardless of the nonlinearity of $\boldsymbol{f}$, the Koopman operator $\mathcal{K}$ allows the eigendecomposition

$$\mathcal{K}\phi_k = \lambda_k \phi_k,$$

where $\lambda_k \in \mathbb{C}$ and $\phi_k : \mathcal{X} \to \mathbb{C}$ are the k-th eigenvalue and corresponding eigenfunction respectively. We consider a D-dimensional signal being observed, which is denoted as a D-dimensional observable vector $\boldsymbol{g} = (g_1, \ldots, g_D)^{\top} : \mathcal{X} \to \mathbb{C}^D$. Assuming that $\boldsymbol{g}$ is expanded in Koopman eigenfunctions ϕ_k, we can denote observations $\boldsymbol{g}(\boldsymbol{x}_t)$ using Koopman modes $\boldsymbol{w}_k \in \mathbb{C}^D$ as

$$\boldsymbol{g}(\boldsymbol{x}_t) = \sum_{k=1}^{\infty} \phi_k(\boldsymbol{x}_t)\boldsymbol{w}_k. \tag{1}$$

Then, we can transform (1) into

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{x}_t) = (\mathcal{K}\boldsymbol{g})(\boldsymbol{x}_{t-1}) &= \sum_{k=1}^{\infty} (\mathcal{K}\phi_k)(\boldsymbol{x}_{t-1})\boldsymbol{w}_k \\ &= \sum_{k=1}^{\infty} \lambda_k \phi_k(\boldsymbol{x}_{t-1})\boldsymbol{w}_k \\ &= \cdots = \sum_{k=1}^{\infty} \lambda_k^{t-1} \phi_k(\boldsymbol{x}_0)\boldsymbol{w}_k. \end{aligned} \tag{2}$$

DMD is one of the methods for finding eigenvalues $\{\lambda_k\}$ and Koopman modes $\{\boldsymbol{w}_k\}$ from the given data.

Dynamic Mode Decomposition

Let $\boldsymbol{Y} \in \mathbb{C}^{D \times T}$ be a data matrix with D-dimensional observations at time points $t = 1, \ldots, T$. We define $\boldsymbol{Y}_0, \boldsymbol{Y}_1 \in \mathbb{C}^{D \times (T-1)}$ as

$$\boldsymbol{Y}_0 = \begin{pmatrix} | & | & & | \\ \boldsymbol{Y}_{:1} & \boldsymbol{Y}_{:2} & \cdots & \boldsymbol{Y}_{:T-1} \\ | & | & & | \end{pmatrix},$$

$$\boldsymbol{Y}_1 = \begin{pmatrix} | & | & & | \\ \boldsymbol{Y}_{:2} & \boldsymbol{Y}_{:3} & \cdots & \boldsymbol{Y}_{:T} \\ | & | & & | \end{pmatrix}.$$

Then, a discrete linear time-invariant system

$$\boldsymbol{Y}_{:t} = \boldsymbol{A}\boldsymbol{Y}_{:t+1},\ t = 1, \ldots, T-1$$

can also be expressed as

$$\boldsymbol{Y}_1 = \boldsymbol{A}\boldsymbol{Y}_0.$$

DMD is an efficient method for finding the $K(\leq D)$ largest eigenvalues and the corresponding eigenvectors of

$$\bar{\boldsymbol{A}} := \arg\min_{\boldsymbol{A}} \|\boldsymbol{Y}_1 - \boldsymbol{A}\boldsymbol{Y}_0\|_F^2 = \boldsymbol{Y}_1\boldsymbol{Y}_0^- \in \mathbb{C}^{D\times D}.$$

The DMD algorithm is as follows:

1. Compute the K-th truncated SVD of $\boldsymbol{Y}_0$

   $$\boldsymbol{Y}_0 \approx \boldsymbol{U}_K\boldsymbol{L}_K\boldsymbol{V}_K^*,$$

   where $\boldsymbol{U}_K \in \mathbb{C}^{D\times K}$ and $\boldsymbol{V}_K \in \mathbb{C}^{T-1\times K}$ are composed of left and right singular vectors, respectively. The matrix $\boldsymbol{L}_K \in \mathbb{C}^{K\times K}$ is a diagonal matrix consisting of the K largest singular values. Now, $\bar{\boldsymbol{A}}$ can be approximated by

   $$\bar{\boldsymbol{A}} \approx \boldsymbol{Y}_1\boldsymbol{V}_K\boldsymbol{L}_K^{-1}\boldsymbol{U}_K^*.$$

2. Map $\bar{\boldsymbol{A}} \in \mathbb{C}^{D\times D}$ to $\tilde{\boldsymbol{A}} \in \mathbb{C}^{K\times K}$ using the K left singular vectors $\boldsymbol{U}_K$ of $\boldsymbol{Y}_0$:

   $$\tilde{\boldsymbol{A}} = \boldsymbol{U}_K^*\bar{\boldsymbol{A}}\boldsymbol{U}_K = \boldsymbol{U}_K^*\boldsymbol{Y}_1\boldsymbol{V}_K\boldsymbol{L}_K^{-1} \in \mathbb{C}^{K\times K}.$$

3. Compute the eigendecomposition of $\tilde{\boldsymbol{A}}$

   $$\tilde{\boldsymbol{A}}\tilde{\boldsymbol{w}}_k = \lambda_k\tilde{\boldsymbol{w}}_k,\ k = 1,\dots,K.$$

   This yields the K largest eigenvalues of $\bar{\boldsymbol{A}}$ as $\{\lambda_k\}$ and the corresponding eigenvectors, which are called DMD modes, as $\{\boldsymbol{U}_K\tilde{\boldsymbol{w}}_k\}$.

The above algorithm is called projected DMD. An alternative to projected DMD is exact DMD, in which each DMD mode is computed as $\boldsymbol{Y}_1\boldsymbol{V}_K\boldsymbol{L}_K^{-1}\tilde{\boldsymbol{w}}_k$ (Tu et al. 2014). Because exact DMD considers not only $\boldsymbol{Y}_0$ but also $\boldsymbol{Y}_1$, exact DMD is more widely used than projected DMD. On the other hand, projected DMD-based likelihood modeling is more tractable for further marginalization, as described in Section 4; thus, projected DMD has been mainly considered in this paper.

Conventional Bayesian DMD

Takeishi et al. (2017) proposed the first Bayesian DMD model. The likelihood function of Bayesian DMD proposed in (Takeishi et al. 2017) at the t-th timepoint is defined as the product of the following two likelihoods of $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_1$

$$\begin{aligned}
&p(\boldsymbol{Y}_{0,:t}|\{\lambda_k\},\{\phi_{k,t}\},\{\boldsymbol{w}_k\},\sigma^2)\\
&\quad= \mathcal{CN}\left(\boldsymbol{Y}_{0,:t}\middle|\sum_{k=1}^K \phi_{k,t}\boldsymbol{w}_k, \sigma^2\boldsymbol{I}_D\right), \qquad (3)\\
&p(\boldsymbol{Y}_{1,:t}|\{\lambda_k\},\{\phi_{k,t}\},\{\boldsymbol{w}_k\},\sigma^2)\\
&\quad= \mathcal{CN}\left(\boldsymbol{Y}_{1,:t}\middle|\sum_{k=1}^K \lambda_k\phi_{k,t}\boldsymbol{w}_k, \sigma^2\boldsymbol{I}_D\right), \qquad (4)
\end{aligned}$$

where $\mathcal{CN}$ indicates the complex normal distribution which is introduced in Section 3. The priors of the parameters in (3) and (4) are

$$\begin{aligned}
p(\lambda_k) &= \mathcal{CN}(\lambda_k|0,1),\\
p(\phi_{k,t}) &= \mathcal{CN}(\phi_{k,t}|0,1),\\
p(\boldsymbol{w}_k|\boldsymbol{C}_w) &= \mathcal{CN}(\boldsymbol{w}_k|\boldsymbol{0}_D,\boldsymbol{C}_w),\\
p(\sigma^2) &= \mathrm{InvGamma}(\sigma^2|\alpha_{\sigma^2},\beta_{\sigma^2}),
\end{aligned}$$

where $\boldsymbol{C}_w$ is a $D\times D$ real diagonal matrix and the hyperprior is

$$p(\boldsymbol{C}_{w,dd}) = \mathrm{InvGamma}(\boldsymbol{C}_{w,dd}|\alpha_c,\beta_c).$$

The parameter $\phi_{k,t}$ in (3) and (4) corresponds to the output of the k-th Koopman eigenfunction $\phi_k(\boldsymbol{x}_t)$ in (2). However, the Koopman eigenfunction ϕ_k does not have any degrees of freedom with respect to timepoint t because Koopman spectral analysis assumes a time-invariant system. For instance, assume that the latent variable $\boldsymbol{x}_t$ evolves linearly and let observation $\boldsymbol{g}$ be an identity function. In this case, the Koopman eigenfunction $\phi_k(\cdot)$ is specified as the inner product with the k-th left eigenvector of the transition matrix of $\boldsymbol{x}_t$ (Rowley et al. 2009). Thus, the conventional Bayesian DMD model considers the K Koopman eigenfunctions $\{\phi_k\}$ as independent $K\times T$ random variables, while it has $D-1$ free parameters. This parameter redundancy causes estimation instability, especially when $D > T$. Furthermore, from the probabilistic viewpoint, the likelihoods (3) and (4) assume conditional independence between $\boldsymbol{Y}_{0,:t+1}$ and $\boldsymbol{Y}_{1,:t}$ although $\boldsymbol{Y}_{0,:t+1} = \boldsymbol{Y}_{1,:t}$ is satisfied due to their definition. Time lags between $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_1$ could be increased to capture long-range dynamics, but such generalization still lacks an interpretation of (probabilistic) low-rank approximations. We corroborate the shortcomings of conventional Bayesian DMD through a numerical experiment reported in Section 5.

3 Preliminaries

Complex Variate Distributions

In the context of DMD, parameters are commonly defined in a complex space. Hence, we define complex normal distributions as a generalization of real distributions.

Multivariate Complex Normal Distribution The multivariate complex normal distribution of the random vector $\boldsymbol{z} \in \mathbb{C}^N$ is defined as

$$\mathcal{CN}(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \pi^{-N}|\boldsymbol{\Sigma}|^{-1}\exp(-(\boldsymbol{z}-\boldsymbol{\mu})^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})),$$

where $\boldsymbol{\mu} \in \mathbb{C}^N$ is a mean vector, and $\boldsymbol{\Sigma} \in \mathbb{C}^{N\times N}$ is a Hermitian positive definite covariance matrix (Wooding 1956). The first and second moments of the multivariate complex normal distribution are

$$\begin{aligned}
\langle\boldsymbol{z}\rangle_{\mathcal{CN}(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma})} &= \boldsymbol{\mu},\\
\langle\boldsymbol{z}\boldsymbol{z}^*\rangle_{\mathcal{CN}(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma})} &= \boldsymbol{\mu}\boldsymbol{\mu}^* + \boldsymbol{\Sigma}.
\end{aligned}$$

Matrix Variate Complex Normal Distribution We define the complex generalization of the matrix variate normal distribution (Waal 2006) of the random matrix $\boldsymbol{Z} \in \mathbb{C}^{N_1\times N_2}$ as follows:

$$\begin{aligned}
&\mathcal{MCN}(\boldsymbol{Z}|\boldsymbol{M},\boldsymbol{U},\boldsymbol{V})\\
&\quad= \pi^{-N_1N_2}|\boldsymbol{V}|^{-N_1}|\boldsymbol{U}|^{-N_2}\\
&\qquad\times\exp\left(-\mathrm{tr}\{\boldsymbol{V}^{-1}(\boldsymbol{Z}-\boldsymbol{M})^*(\boldsymbol{U}^{-\top})(\boldsymbol{Z}-\boldsymbol{M})\}\right), \qquad (5)
\end{aligned}$$

where $\boldsymbol{M} \in \mathbb{C}^{N_1\times N_2}$ is a mean parameter and $\boldsymbol{U} \in \mathbb{C}^{N_1\times N_1}$ and $\boldsymbol{V} \in \mathbb{C}^{N_2\times N_2}$ are the row-wise and column-wise Hermitian positive definite covariance matrices, respectively.

We note that (5) can be represented by the following equivalent complex normal distribution:

$$\mathcal{MCN}(\boldsymbol{Z}|\boldsymbol{M},\boldsymbol{U},\boldsymbol{V}) \equiv \mathcal{CN}(\mathrm{vec}(\boldsymbol{Z})|\mathrm{vec}(\boldsymbol{M}),\boldsymbol{U}\otimes\boldsymbol{V}). \quad (6)$$

The lower-order moments of the matrix variate complex normal distribution in (5) are

$$\langle \boldsymbol{Z} \rangle_{\mathcal{MCN}(\boldsymbol{Z}|\boldsymbol{M},\boldsymbol{U},\boldsymbol{V})} = \boldsymbol{M},$$
$$\langle \boldsymbol{Z}^*\boldsymbol{Z} \rangle_{\mathcal{MCN}(\boldsymbol{Z}|\boldsymbol{M},\boldsymbol{U},\boldsymbol{V})} = \boldsymbol{M}^*\boldsymbol{M} + \boldsymbol{V}\mathrm{tr}\{\boldsymbol{U}\}.$$

4 Methodologies

Likelihood of Bayesian DMD

DMD can be interpreted as a procedure that determines $\{\lambda_k\}$ and $\{\boldsymbol{w}_k\}$ which approximate

$$\boldsymbol{Y}_{:t} \approx \sum_{k=1}^{K} \lambda_k^{t-1}\boldsymbol{w}_k b_k,$$

where b_k is a scaling factor for the normalized vector $\boldsymbol{w}_k$. In fact, we only require a low-dimensional representation of $\boldsymbol{w}_k$, as described in Section 2. Therefore, we design the likelihood function of $\boldsymbol{Y}_{:t}$ as follows:

$$p(\boldsymbol{Y}_{:t}|\{\tilde{\boldsymbol{w}}_k\},\{\lambda_k\},\boldsymbol{U}_K,\sigma^2) = \mathcal{CN}\left(\boldsymbol{Y}_{:t}\middle|\boldsymbol{U}_K\sum_{k=1}^{K}\lambda_k^{t-1}\tilde{\boldsymbol{w}}_k, \sigma^2\boldsymbol{I}_D\right). \quad (7)$$

Here, $\boldsymbol{U}_K \in \mathbb{C}^{D\times K}$ corresponds to the left singular vectors of $\boldsymbol{Y}$ and $\tilde{\boldsymbol{w}}_k \in \mathbb{C}^K$ denotes the k-th eigenvector of the latent dynamics multiplied by a constant such that $\boldsymbol{U}_K\tilde{\boldsymbol{w}}_k = \boldsymbol{w}_k b_k$. The joint likelihood function is

$$p(\boldsymbol{Y}|\{\tilde{\boldsymbol{w}}_k\},\{\lambda_k\},\boldsymbol{U}_K,\sigma^2) = \prod_{t=1}^{T} p(\boldsymbol{Y}_{:t}|\{\tilde{\boldsymbol{w}}_k\},\{\lambda_k\},\boldsymbol{U}_K,\sigma^2) = \mathcal{MCN}(\boldsymbol{Y}|\boldsymbol{U}_K\boldsymbol{G},\boldsymbol{I}_D,\sigma^2\boldsymbol{I}_T), \quad (8)$$

where $\boldsymbol{G} \in \mathbb{C}^{K\times T}$ is defined as

$$\boldsymbol{G} = \sum_{k=1}^{K}\begin{pmatrix} | & | & & | \\ \tilde{\boldsymbol{w}}_k & \lambda\tilde{\boldsymbol{w}}_k & \cdots & \lambda^{T-1}\tilde{\boldsymbol{w}}_k \\ | & | & & | \end{pmatrix}.$$

Variational Matrix Factorization

We can directly evaluate the posterior of the parameters of (7) with appropriate priors (e.g., a probability distribution on a Stiefel manifold for $\boldsymbol{U}_K$). However, the degree of freedom depends on the dimension D, which is generally large. Fortunately, using a Gaussian-type conjugate prior, $\boldsymbol{U}_K$ can be marginalized out from (7).

To obtain a reasonable prior of $\boldsymbol{U}_K$, we apply VMF (Lim and Teh 2007; Nakajima and Sugiyama 2011; Seeger and Bouchard 2012), which is a variational Bayesian treatment of SVD, to $\boldsymbol{Y}$. If the elements of the data matrix $\boldsymbol{Y}$ are fully observed, the complex version of the VMF model is

$$p(\boldsymbol{Y}|\boldsymbol{U}_K,\boldsymbol{V}_K) = \mathcal{MCN}(\boldsymbol{Y}|\boldsymbol{U}_K\boldsymbol{V}_K^*,\boldsymbol{I}_D,s^2\boldsymbol{I}_T), \quad (9)$$
$$\boldsymbol{U}_K \sim \mathcal{MCN}(\boldsymbol{U}_K|\boldsymbol{O}_{DK},\boldsymbol{I}_D,\boldsymbol{C}_U),$$
$$\boldsymbol{V}_K \sim \mathcal{MCN}(\boldsymbol{V}_K|\boldsymbol{O}_{TK},\boldsymbol{I}_T,\boldsymbol{C}_V).$$

where $\boldsymbol{C}_U, \boldsymbol{C}_V \in \mathbb{R}^{K\times K}$ are positive-definite diagonal matrices. Variational Bayesian methods assume that the parameters of the joint posterior are independent, i.e.,

$$p(\boldsymbol{U}_K,\boldsymbol{V}_K|\boldsymbol{Y}) \approx r(\boldsymbol{U}_K)r(\boldsymbol{V}_K).$$

Then, we solve functional optimizations of $r(\boldsymbol{U}_K)$ and $r(\boldsymbol{V}_K)$ alternately to minimize the free energy of a given model. We refer to the decomposed functions r as variational posteriors. In the fully observed VMF model, the free energy F is

$$F = \left\langle \log\frac{r(\boldsymbol{U}_K)r(\boldsymbol{V}_K)}{p(\boldsymbol{Y}|\boldsymbol{U}_K,\boldsymbol{V}_K)p(\boldsymbol{U}_K)p(\boldsymbol{V}_K)}\right\rangle_{r(\boldsymbol{U}_K)r(\boldsymbol{V}_K)},$$

and the stationary conditions of r are

$$r(\boldsymbol{U}_K) \propto p(\boldsymbol{U}_K)\exp\{\langle\log p(\boldsymbol{Y}|\boldsymbol{U}_K,\boldsymbol{V}_K)\rangle_{r(\boldsymbol{V}_K)}\},$$
$$r(\boldsymbol{V}_K) \propto p(\boldsymbol{V}_K)\exp\{\langle\log p(\boldsymbol{Y}|\boldsymbol{U}_K,\boldsymbol{V}_K)\rangle_{r(\boldsymbol{U}_K)}\}.$$

Here, we treat the variance s^2 as a nuisance parameter. We also update s^2 by applying the empirical variational Bayesian (EVB) algorithm (Nakajima and Sugiyama 2011; Nakajima et al. 2013), which updates nuisance parameters by solving first-order optimality conditions for F.

VMF with complete data If the data are fully observed, the variational distributions of $\boldsymbol{U}_K$ and $\boldsymbol{V}_K$ can be simply denoted by the matrix variate complex normal distributions

$$r(\boldsymbol{U}_K) = \mathcal{MCN}(\boldsymbol{U}_K|\bar{\boldsymbol{U}}_K,\boldsymbol{I}_D,\bar{\boldsymbol{\Sigma}}_U), \quad (10)$$
$$r(\boldsymbol{V}_K) = \mathcal{MCN}(\boldsymbol{V}_K|\bar{\boldsymbol{V}}_K,\boldsymbol{I}_T,\bar{\boldsymbol{\Sigma}}_V),$$
$$\bar{\boldsymbol{\Sigma}}_U^{-1} = \boldsymbol{C}_U^{-1} + s^{-2}(T\bar{\boldsymbol{\Sigma}}_V + \bar{\boldsymbol{V}}_K^*\bar{\boldsymbol{V}}_K),$$
$$\bar{\boldsymbol{\Sigma}}_V^{-1} = \boldsymbol{C}_V^{-1} + s^{-2}(D\bar{\boldsymbol{\Sigma}}_U + \bar{\boldsymbol{U}}_K^*\bar{\boldsymbol{U}}_K),$$
$$\bar{\boldsymbol{U}}_K = s^{-2}\boldsymbol{Y}\bar{\boldsymbol{V}}_K\bar{\boldsymbol{\Sigma}}_U,\ \ \bar{\boldsymbol{V}}_K = s^{-2}\boldsymbol{Y}^*\bar{\boldsymbol{U}}_K\bar{\boldsymbol{\Sigma}}_V.$$

Moreover, the EVB update of s^2 can be derived as

$$s^2 = \frac{1}{TD}(\|\boldsymbol{Y}\|_F^2 - 2\mathrm{tr}\{\mathrm{Re}\{\bar{\boldsymbol{V}}_K\bar{\boldsymbol{U}}_K^*\boldsymbol{Y}\}\} + \mathrm{tr}\{(D\boldsymbol{\Sigma}_U + \bar{\boldsymbol{U}}_K^*\bar{\boldsymbol{U}}_K)(T\boldsymbol{\Sigma}_V + \bar{\boldsymbol{V}}_K^*\bar{\boldsymbol{V}}_K)\}).$$

VMF with incomplete data Due to its probabilistic nature, VMF can also deal with incomplete data. First, we define the following set of indices to represent the observations:

$$\mathcal{I} = \{(d,t)|\boldsymbol{Y}_{dt} \text{ is observed} \quad \text{for } d = 1,\dots,D, t = 1,\dots,T\}.$$

In this case, the likelihood of the VMF model (9) is replaced with

$$p(\mathcal{P}_{\mathcal{I}}(\boldsymbol{Y})|\boldsymbol{U}_K,\boldsymbol{V}_K) = \mathcal{MCN}(\mathcal{P}_{\mathcal{I}}(\boldsymbol{Y})|\mathcal{P}_{\mathcal{I}}(\boldsymbol{U}_K\boldsymbol{V}_K^*),\boldsymbol{I}_D,s^2\boldsymbol{I}_T),$$

where $\mathcal{P}_{\mathcal{I}} : \mathbb{C}^{D\times T} \to \mathbb{C}^{D\times T}$ is a map defined by

$$(\mathcal{P}_{\mathcal{I}}(\boldsymbol{Y}))_{dt} = \begin{cases} \boldsymbol{Y}_{dt} & \text{if } (d,t)\in\mathcal{I} \\ 0 & \text{otherwise.}\end{cases}$$

Then, the variational posteriors can be derived as

$$r(\boldsymbol{U}_K) \equiv \mathcal{CN}(\mathrm{vec}(\boldsymbol{U}_K)|\mathrm{vec}(\bar{\boldsymbol{U}}_K),\boldsymbol{\Gamma}_U) \equiv \mathcal{CN}(\mathrm{vec}(\boldsymbol{U}_K^\top)|\mathrm{vec}(\bar{\boldsymbol{U}}_K^\top),\boldsymbol{\Sigma}_U), \quad (11)$$
$$r(\boldsymbol{V}_K) \equiv \mathcal{CN}(\mathrm{vec}(\boldsymbol{V}_K)|\mathrm{vec}(\bar{\boldsymbol{V}}_K),\boldsymbol{\Gamma}_V) \equiv \mathcal{CN}(\mathrm{vec}(\boldsymbol{V}_K^\top)|\mathrm{vec}(\bar{\boldsymbol{V}}_K^\top),\boldsymbol{\Sigma}_V),$$

where the parameters are explicitly written as

$$\boldsymbol{\Sigma}_U = \mathrm{diag}(\{\boldsymbol{\Sigma}_U^{(d)}\}),\ \ \boldsymbol{\Sigma}_V = \mathrm{diag}(\{\boldsymbol{\Sigma}_V^{(t)}\}), \tag{12}$$

$$\boldsymbol{\Sigma}_U^{(d)-1} = \boldsymbol{C}_U^{-1} + s^{-2} \sum_{t:(d,t)\in\mathcal{I}} \left(\bar{\boldsymbol{V}}_{K,t:}^* \bar{\boldsymbol{V}}_{K,t:} + \boldsymbol{\Sigma}_V^{(t)} \right),$$

$$\boldsymbol{\Sigma}_V^{(t)-1} = \boldsymbol{C}_V^{-1} + s^{-2} \sum_{d:(d,t)\in\mathcal{I}} \left(\bar{\boldsymbol{U}}_{K,d:}^* \bar{\boldsymbol{U}}_{K,d:} + \boldsymbol{\Sigma}_U^{(d)} \right), \tag{13}$$

$$\mathrm{vec}(\bar{\boldsymbol{U}}_K^\top) = \begin{pmatrix} \bar{\boldsymbol{U}}_{K,1:}^\top \\ \vdots \\ \bar{\boldsymbol{U}}_{K,D:}^\top \end{pmatrix}, \mathrm{vec}(\bar{\boldsymbol{V}}_K^\top) = \begin{pmatrix} \bar{\boldsymbol{V}}_{K,1:}^\top \\ \vdots \\ \bar{\boldsymbol{V}}_{K,T:}^\top \end{pmatrix}, \tag{14}$$

$$\bar{\boldsymbol{U}}_{K,d:} = s^{-2} \sum_{t:(d,t)\in\mathcal{I}} \boldsymbol{Y}_{dt} \bar{\boldsymbol{V}}_{K,t:} \boldsymbol{\Sigma}_U^{(d)},$$

$$\bar{\boldsymbol{V}}_{K,t:} = s^{-2} \sum_{d:(d,t)\in\mathcal{I}} \boldsymbol{Y}_{dt}^* \bar{\boldsymbol{U}}_{K,d:} \boldsymbol{\Sigma}_V^{(t)}, \tag{15}$$

and

$$\boldsymbol{\Gamma}_U^{-1} = \boldsymbol{C}_{DK}^\top \boldsymbol{\Sigma}_U^{-\top} \boldsymbol{C}_{DK},$$
$$\boldsymbol{\Gamma}_V^{-1} = \boldsymbol{C}_{TK}^\top \boldsymbol{\Sigma}_V^{-\top} \boldsymbol{C}_{TK}.$$

Note that $\boldsymbol{C}_{DK} \in \{0,1\}^{DK\times DK}$ is the commutation matrix (Magnus and Neudecker 1979), which satisfies

$$\boldsymbol{C}_{DK}\mathrm{vec}(\boldsymbol{U}_K) = \mathrm{vec}(\boldsymbol{U}_K^\top),$$

and $\boldsymbol{C}_{TK}$ satisfies a similar equation. Here, $\sum_{t:(d,t)\in\mathcal{I}}$ in (12) and (14) denotes the summation over t such that $(d,t) \in \mathcal{I}$ for a given d. The summations $\sum_{d:(d,t)\in\mathcal{I}}$ in (13) and (15) are defined in the same manner. We note that if data matrix $\boldsymbol{Y}$ has missing entries, $r(\boldsymbol{U}_K)$ and $r(\boldsymbol{V}_K)$ cannot be represented by complex matrix normal distributions because the covariance matrices $\boldsymbol{\Sigma}_U$ and $\boldsymbol{\Sigma}_V$ are not decomposable into a Kronecker product of two matrices. The EVB update of s^2 is

$$s^2 = \frac{1}{\#\mathcal{I}} \sum_{(d,t)\in\mathcal{I}} [|Y_{dt}|^2 - 2\mathrm{Re}\{Y_{dt}^* \bar{\boldsymbol{U}}_{K,d:} \bar{\boldsymbol{V}}_{K,t:}^*\} + \mathrm{tr}\{(\boldsymbol{\Sigma}_U^{(d)} + \bar{\boldsymbol{U}}_{K,d:}^* \bar{\boldsymbol{U}}_{K,d:})(\boldsymbol{\Sigma}_V^{(t)} + \bar{\boldsymbol{V}}_{K,t:}^* \bar{\boldsymbol{V}}_{K,t:})\}].$$

Marginalizing Out $\boldsymbol{U}_K$

Using the variational posterior of $\boldsymbol{U}_K$ (i.e., (10) for complete data or (11) for incomplete data) as the prior of (7), we can marginalize out $\boldsymbol{U}_K$. Therefore, if the elements of a given data matrix are fully observed, we consider the joint probability

$$p(\boldsymbol{Y}, \boldsymbol{U}_K | \{\tilde{\boldsymbol{w}}_k\}, \{\lambda_k\}, \sigma^2) = \mathcal{MCN}(\boldsymbol{Y} | \boldsymbol{U}_K \boldsymbol{G}, \boldsymbol{I}_D, \sigma^2 \boldsymbol{I}_T) \times \mathcal{MCN}(\boldsymbol{U}_K | \bar{\boldsymbol{U}}_K, \boldsymbol{I}_D, \bar{\boldsymbol{\Sigma}}_U) \tag{16}$$

after convergence of the VMF algorithm and then integrate $\boldsymbol{U}_K$ from (16). Equation (7) has $2K(D+K+1)+1$ degrees of freedom (without the orthogonality condition for $\boldsymbol{U}_K$), but this is reduced to $2K(K+1)+1$ owing to the marginalization procedure. This number now only depends on the number of modes K, and not on the data dimension D. Note that the mean field approximation in VMF plays a significant role in the BDMD-VMF scheme. Because VMF assumes posterior independence from the outset, we can immediately marginalize out $\boldsymbol{U}_K$ analytically, as shown below.

Complete data case If there are no missing data, the conditional probability of $\boldsymbol{Y}_{:t}$ with the marginalization of $\boldsymbol{U}_K$ is

$$p(\boldsymbol{Y}_{:t} | \{\lambda_k\}, \{\tilde{\boldsymbol{w}}_k\}, \sigma^2) = \mathcal{CN}(\boldsymbol{Y}_{:t} | \bar{\boldsymbol{Y}}_{:t}, \sigma_t^2 \boldsymbol{I}_D),$$

where

$$\sigma_t^{-2} = \sigma^{-2}(1 - \boldsymbol{G}_{:t}^* \tilde{\boldsymbol{\Sigma}}_U^{(t)-1} \boldsymbol{G}_{:t}),$$
$$\bar{\boldsymbol{Y}}_{:t} = \sigma^{-2} \sigma_t^2 \bar{\boldsymbol{U}}_K \bar{\boldsymbol{\Sigma}}_U^{-1} \tilde{\boldsymbol{\Sigma}}_U^{(t)} \boldsymbol{G}_{t:},$$
$$\tilde{\boldsymbol{\Sigma}}_U^{(t)-1} = \sigma^{-2} \boldsymbol{G}_{:t} \boldsymbol{G}_{:t}^* + \bar{\boldsymbol{\Sigma}}_U.$$

Hence, the posterior of the BDMD-VMF model for complete data is

$$p(\{\lambda_k\}, \{\tilde{\boldsymbol{w}}_k\}, \sigma^2 | \boldsymbol{Y}) \propto p(\sigma^2) \prod_{t=1}^{T} \mathcal{CN}(\boldsymbol{Y}_{:t} | \bar{\boldsymbol{Y}}_{:t}, \sigma_t^2 \boldsymbol{I}_D) \times \prod_{k=1}^{K} p(\lambda_k) p(\tilde{\boldsymbol{w}}_k). \tag{17}$$

Incomplete data case Now, we consider the case where the data matrix $\boldsymbol{Y}$ has missing values. In this case, the variational posterior of VMF $r(\boldsymbol{U}_K)$ cannot be expressed as a matrix variate complex normal distribution. By vectorizing (8) with the equivalent representation of the matrix variate complex normal distribution (6), we obtain

$$\begin{aligned} &p(\mathrm{vec}(\boldsymbol{Y}) | \{\tilde{\boldsymbol{w}}_k\}, \{\lambda_k\}, \boldsymbol{U}_K, \sigma^2) \\ &\quad = \mathcal{CN}(\mathrm{vec}(\boldsymbol{Y}) | \mathrm{vec}(\boldsymbol{U}_K \boldsymbol{G}), \sigma^2 \boldsymbol{I}_D \otimes \boldsymbol{I}_T) \\ &\quad = \mathcal{CN}(\mathrm{vec}(\boldsymbol{Y}) | (\boldsymbol{G}^\top \otimes \boldsymbol{I}_D)\mathrm{vec}(\boldsymbol{U}_K), \sigma^2 \boldsymbol{I}_D \otimes \boldsymbol{I}_T). \end{aligned} \tag{18}$$

Therefore, the marginalized distribution of (18) is

$$p(\mathrm{vec}(\boldsymbol{Y}) | \{\tilde{\boldsymbol{w}}_k\}, \{\lambda_k\}, \sigma^2) = \mathcal{CN}(\mathrm{vec}(\boldsymbol{Y}) | \mathrm{vec}(\bar{\boldsymbol{Y}}), \boldsymbol{\Sigma}_Y), \tag{19}$$

where

$$\boldsymbol{\Sigma}_Y^{-1} = \sigma^{-2}(\boldsymbol{I}_{DT} - \sigma^{-2}(\boldsymbol{G}^\top \otimes \boldsymbol{I}_D)\hat{\boldsymbol{\Sigma}}_U(\mathrm{conj}(\boldsymbol{G}) \otimes \boldsymbol{I}_D)),$$

$$\mathrm{vec}(\bar{\boldsymbol{Y}}) = \sigma^{-2} \boldsymbol{\Sigma}_Y (\boldsymbol{G}^\top \otimes \boldsymbol{I}_D) \hat{\boldsymbol{\Sigma}}_U \boldsymbol{\Gamma}_U^{-1} \mathrm{vec}(\bar{\boldsymbol{U}}_K),$$
$$\hat{\boldsymbol{\Sigma}}_U^{-1} = \sigma^{-2} (\boldsymbol{G}\boldsymbol{G}^*)^\top \otimes \boldsymbol{I}_D + \boldsymbol{\Gamma}_U^{-1}.$$

Now, we rearrange $\mathrm{vec}(\boldsymbol{Y})$ into $(\mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}}^\top, \mathrm{vec}(\boldsymbol{Y})_{\mathrm{mis}}^\top)^\top$, where $\mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}}$ and $\mathrm{vec}(\boldsymbol{Y})_{\mathrm{mis}}$ are observed and missing elements of $\mathrm{vec}(\boldsymbol{Y})$ respectively. Then, (19) can be rewritten as

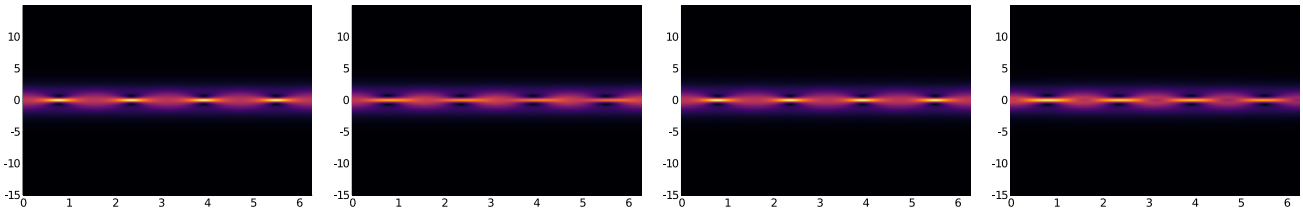
$$\mathcal{CN}\left(\begin{pmatrix} \mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}} \\ \mathrm{vec}(\boldsymbol{Y})_{\mathrm{mis}} \end{pmatrix} \middle| \begin{pmatrix} \mathrm{vec}(\bar{\boldsymbol{Y}}_1) \\ \mathrm{vec}(\bar{\boldsymbol{Y}}_2) \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{Y,11} & \boldsymbol{\Sigma}_{Y,12} \\ \boldsymbol{\Sigma}_{Y,21} & \boldsymbol{\Sigma}_{Y,22} \end{pmatrix} \right)$$

without loss of generality. Using the marginalization property of normal distributions, we obtain the following marginalized likelihood function for incomplete data:

$$\begin{aligned} &p(\mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}} | \{\tilde{\boldsymbol{w}}_k\}, \{\lambda_k\}, \sigma^2) \\ &\quad = \mathcal{CN}(\mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}} | \mathrm{vec}(\bar{\boldsymbol{Y}}_1), \boldsymbol{\Sigma}_{Y,11}). \end{aligned} \tag{20}$$

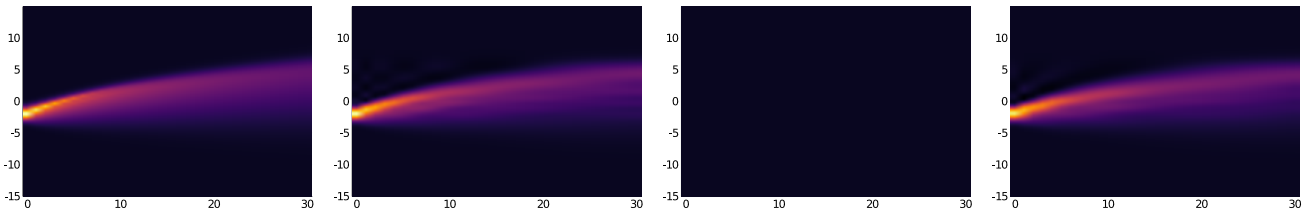
Thus, if the data matrix $\boldsymbol{Y}$ has missing values, the posterior of the BDMD-VMF model is written as

$$\begin{aligned} &p(\{\lambda_k\}, \{\tilde{\boldsymbol{w}}_k\}, \sigma^2 | \mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}}) \\ &\quad \propto \mathcal{CN}(\mathrm{vec}(\boldsymbol{Y})_{\mathrm{obs}} | \mathrm{vec}(\bar{\boldsymbol{Y}}_1), \boldsymbol{\Sigma}_{Y,11}) p(\sigma^2) \prod_{k=1}^{K} p(\lambda_k) p(\tilde{\boldsymbol{w}}_k). \end{aligned} \tag{21}$$



(a) Original (b) DMD (c) Conventional Bayesian DMD (d) BDMD-VMF

Figure 1: Reconstruction results for the nonlinear Schrödinger equation with $K = 4$.



(a) Original (b) DMD (c) Conventional Bayesian DMD (d) BDMD-VMF

Figure 2: Reconstruction results for the Burgers' equation with $K = 7$.

5 Illustrative Examples

In this section, we report the application of the proposed BDMD-VMF method to both simulated and real data. By using simulated data generated from nonlinear partial differential equations (PDEs), we confirm that BDMD-VMF yields almost the same results as DMD. In addition, unstable behavior of conventional Bayesian DMD is demonstrated. Furthermore, we apply BDMD-VMF to incomplete data through an experiment on a real-world dataset. We adopted the Metropolis algorithm, which is a Markov chain Monte Carlo (MCMC) method, to evaluate the posterior distributions (17) and (21) in all examples. Throughout the simulations reported below, we generated 7,500 MCMC samples and discarded the first 5,000 considering a burn-in periods. Unless otherwise specified, we employed appropriate weakly informative priors for any parameters. We determined the number of modes K for which DMD can capture input dynamics, through preliminary experiments.

Synthetic Data

As synthetic datasets, we used sequential datasets from two nonlinear PDEs: the nonlinear Schrödinger equation (NLSE) and Burgers' equation. The NLSE is yielded by replacing the potential of the time-dependent Schrödinger equation with the absolute square of the wave function, and the Burgers' equation is equivalent to the incompressible Navier-Stokes equation without an external force. These two nonlinear PDEs are well-studied as the nonlinear application examples of DMD (Kutz et al. 2016; Kutz, Proctor, and Brunton 2018). For each dataset, we applied DMD, conventional Bayesian DMD, and BDMD-VMF and then reconstructed the original input with their estimates.

Nonlinear Schrödinger equation The NLSE is given by

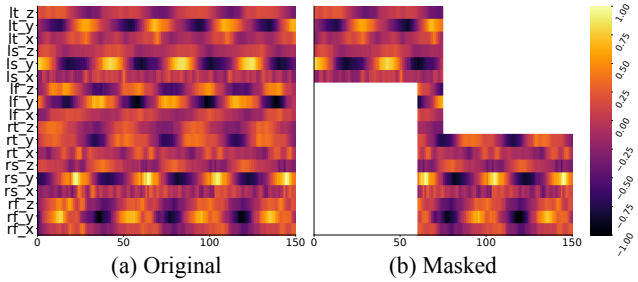
$$i\frac{\partial\psi(\xi,\tau)}{\partial\tau} = \left(\frac{1}{2}\frac{\partial^2\psi(\xi,\tau)}{\partial\xi^2} + |\psi(\xi,\tau)|^2\right)\psi(\xi,\tau),$$

where ξ and τ are continuous variables in space and time, respectively. We configured the timepoints and gridpoints as $\tau \in [0, 2\pi)$ with $t = 1, \ldots, 256$ and $\xi \in [-15, 15)$ with $d = 1, \ldots, 256$, respectively. The number of modes K was set to 4 for this experiment. Figure 1 shows the reconstruction results by DMD, conventional Bayesian DMD, and BDMD-VMF. The reconstruction RMSEs are 0.083, 0.005, and 0.057, respectively. The difference between DMD and BDMD-VMF is mainly associated with the numerical error of MCMC. Meanwhile, the conventional Bayesian DMD fits better to the given observation than DMD and BDMD-VMF. This is contributed to the fact that representation power of the conventional Bayesian DMD model significantly exceeds the standard DMD's one. As is seen from the next example, the conventional Bayesian DMD has difficulty when applied to high-dimensional and small sample data.

Burgers' equation As a second example of nonlinear PDEs, we generated data from the Burgers' equation:

$$\frac{\partial\psi(\xi,\tau)}{\partial\tau} = -\psi(\xi,\tau)\frac{\partial\psi(\xi,\tau)}{\partial\xi} + \nu\frac{\partial^2\psi(\xi,\tau)}{\partial\xi^2}, \quad \nu \geq 0,$$

with $\nu = 0.1$. The timepoints and gridpoints are $\tau \in [0, 30]$ with $t = 1, \ldots, 31$ and $\xi \in [-15, 15)$ with $d = 1, \ldots, 256$ respectively. The reconstruction results with $K = 7$ are shown in Figure 2, and the RMSEs of DMD, conventional Bayesian DMD, and BDMD-VMF are 0.009, 0.130, 0.013, respectively. As opposed to the result of the NLSE, conventional Bayesian DMD did not capture any structure of the dynamics. This phenomenon can also be attributed to the redundancy of conventional Bayesian DMD and tends to occur



(a) Original (b) Masked

Figure 3: (a) Original and (b) masked gyroscope data. The labels on the vertical-axis indicate the locations (l = left, r = right, t = thigh, s = shin, and f = foot) and directions (x-, y-, or z-axis) of the sensors.

particularly when the input dimension is much larger than the sample size.

Gyroscope Data

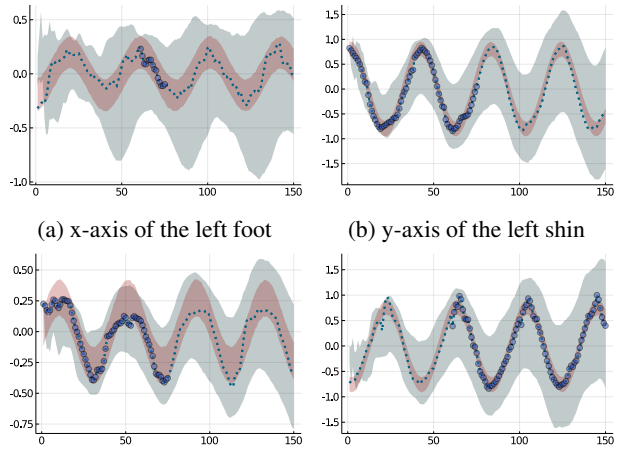
Next, as a real-world example, we applied our method to gyroscope data obtained during cycling. The HuGaDB dataset (Chereshnev and Kertész-Farkas 2018) contains signals collected from six sensors placed on the right and left thighs, shins, and feet of 18 participants. While each participant performed 12 different everyday motions, the sensors logged the accelerations and angular velocities about the x-, y-, and z-axes. We extracted 150 timesteps related to the first participant's angular velocity while cycling. Thus, the size of the data matrix was $D = 18$ and $T = 150$ [1]. Figure 3(a) shows the true clipped data, normalized to the range $[-1, 1]$. We then evaluated the posterior predictive distributions of the masked data shown in Figure 3(b) with the proposed BDMD-VMF model with $K = 2$ and the Bayesian second-order vector autoregressive (VAR(2)) model.

Figure 4 shows the posterior predictive distributions calculated using the MCMC samples. Dotted lines denote the true (un-masked) data and circular markers indicate the observed data. In Figure 4, it can be seen that the credible intervals of the BDMD-VMF model (red-shaded area) are considerably tighter than those of the VAR(2) model (gray-shaded area). In addition, the credible intervals of the BDMD-VMF model do not spread around missing data, in contrast with the case of the VAR(2) model. As opposed to most other Bayesian time-series models, our BDMD-VMF model does not assume that the data at timepoint t are generated from the data obtained so far. More precisely, $\boldsymbol{Y}_{:t}$ is independent of $\boldsymbol{Y}_{:t'}$ for $t \neq t'$ when the time-invariant parameters $\{\lambda_k\}$, $\{\boldsymbol{w}_k\}$, and σ^2 are given.

6 Concluding Remarks

In this paper, we proposed a novel Bayesian formulation of DMD called BDMD-VMF. The proposed BDMD-VMF is a probabilistic model, which allows to handle, incomplete data with a DMD-like scheme have been enabled. The learning scheme of BDMD-VMF is similar to that of standard

[1] $D = 18$ corresponds to the six three-axis gyroscopes.



(a) x-axis of the left foot (b) y-axis of the left shin

(c) z-axis of the left thigh (d) y-axis of the right shin

Figure 4: Posterior predictive distributions of the gyroscope signals about (a) x-axis of the left foot, (b) y-axis of the left shin, (c) z-axis of the left thigh, and (d) y-axis of the right shin. Dotted lines denote the true data and circular markers indicate the input data. Gray- and red-shaded areas represent the 95% credible intervals of the posterior predictive distributions of the VAR(2) and proposed BDMD-VMF models, respectively.

DMD; it first finds a lower-dimensional representation of the data matrix by VMF, and then it infers latent dynamics in the lower-dimensional space. However, in the likelihood of the DMD model (7), there are still many parameters to be estimated and the number depends on the dimension of the input data matrix. To address this problem, we marginalized out the large matrix with its VMF posterior. Using both nonlinear simulated and real-world datasets, we demonstrated reconstruction and interpolation with the proposed BDMD-VMF. Because the BDMD-VMF model is governed by a small number of parameters, it captures input dynamics more robustly than conventional Bayesian DMD.

Finally, we note some limitations of and future directions for of this work. Our BDMD-VMF model does not work appropriately if the missing data mechanism is not missing at random (NMAR), which means that whether data are missing depends on their own true values. In NMAR case, the observed data likelihood (20) becomes biased. Next, improving the computational efficiency of the proposed approach is a critical task. With the Metropolis sampling employed in Section 5, we must evaluate the (unnormalized) posterior (21) numerous times. Some variational methods that do not require conjugacy, such as black box variational inference (Ranganath, Gerrish, and Blei 2014) and automatic differentiation variational inference (Kucukelbir et al. 2017) can be used to overcome this issue. Moreover, it is worth considering a Bayesian DMD model corresponding to kernel DMD. Kernel DMD is a powerful method to handle nonlinear dynamics; however, extremely sensitive to hyperparameter selection. We conjecture that we can address this issue by incorporating the Gaussian process into BDMD-VMF.

Acknowledgments

Part of this work is supported by JSPS KAKENHI grant numbers JP17H01793, JP19K12111, and JST CREST JPMJCR1761.

Ethics Statement

In this paper, a Bayesian identifying method for dynamical systems is developed. Since dynamical systems modeling is a common method to understand real phenomena in natural sciences, our work potentially benefits a wide range of scientific areas. Especially strong impacts to fields in which uncertainty representation of predictions is significant will be expected; such as oceanography, atmospheric science, and mathematical epidemiology. The same is true to application fields where missing data are often obtained. On the other hand, our proposal just provides a new aspect of dynamic mode decomposition, therefore, there is little concern that this paper will affect society negatively.

References

Aihara, K.; and Suzuki, H. 2010. Theory of Hybrid Dynamical Systems and Its Applications to Biological and Medical Systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368(1930): 4893–4914.

Brunton, S. L.; Proctor, J. L.; and Kutz, J. N. 2016. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems. *Proceedings of the National Academy of Sciences* 113(15): 3932–3937. ISSN 0027-8424, 1091-6490.

Chatterjee, A. 2000. An Introduction to the Proper Orthogonal Decomposition. *Current Science* 78(7): 808–817. ISSN 0011-3891.

Chen, R. T. Q.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems 31*, 6571–6583. Curran Associates, Inc.

Chereshnev, R.; and Kertész-Farkas, A. 2018. HuGaDB: Human Gait Database for Activity Recognition from Wearable Inertial Sensor Networks. In *Analysis of Images, Social Networks and Texts*, Lecture Notes in Computer Science, 131–141. Cham: Springer International Publishing. ISBN 978-3-319-73013-4.

Cotton, W. R.; and Anthes, R. A. 1992. *Storm and Cloud Dynamics*. Academic Press. ISBN 978-0-08-095983-2.

Dong, Y.; and Qin, S. J. 2015. Dynamic-Inner Partial Least Squares for Dynamic Data Modeling. *IFAC-PapersOnLine* 48(8): 117–122. ISSN 2405-8963.

Dong, Y.; and Qin, S. J. 2018. A Novel Dynamic PCA Algorithm for Dynamic Data Modeling and Process Monitoring. *Journal of Process Control* 67: 1–11. ISSN 09591524.

Dowell, E. H.; Hall, K. C.; and Romanowski, M. C. 1997. Eigenmode Analysis in Unsteady Aerodynamics: Reduced Order Models. *Applied Mechanics Reviews* 50(6): 371–386. ISSN 0003-6900.

Fujii, K.; and Kawahara, Y. 2019. Dynamic Mode Decomposition in Vector-Valued Reproducing Kernel Hilbert Spaces for Extracting Dynamical Structure among Observables. *Neural Networks* 117: 94–103. ISSN 08936080.

Ito, S.; Nagao, H.; Yamanaka, A.; Tsukada, Y.; Koyama, T.; Kano, M.; and Inoue, J. 2016. Data Assimilation for Massive Autonomous Systems Based on a Second-Order Adjoint Method. *Physical Review E* 94(4): 043307.

Jolliffe, I. T. 2002. *Principal Component Analysis*. Springer Science & Business Media. ISBN 978-0-387-95442-4.

Jovanović, M. R.; Schmid, P. J.; and Nichols, J. W. 2014. Sparsity-Promoting Dynamic Mode Decomposition. *Physics of Fluids* 26(2): 024103. ISSN 1070-6631, 1089-7666.

Kawahara, Y. 2016. Dynamic Mode Decomposition with Reproducing Kernels for Koopman Spectral Analysis. In *Advances in Neural Information Processing Systems 29*, 911–919. Curran Associates, Inc.

Kosambi, D. D. 1943. Statistics in Function Space. *The Journal of the Indian Mathematical Society* 7(0): 76–88. ISSN 2455-6475.

Kucukelbir, A.; Tran, D.; Ranganath, R.; Gelman, A.; and Blei, D. M. 2017. Automatic Differentiation Variational Inference. *The Journal of Machine Learning Research* 18(1): 430–474. ISSN 1532-4435.

Kutz, J. N.; Brunton, S. L.; Brunton, B. W.; and Proctor, J. L. 2016. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM. ISBN 978-1-61197-449-2.

Kutz, J. N.; Proctor, J. L.; and Brunton, S. L. 2018. Applied Koopman Theory for Partial Differential Equations and Data-Driven Modeling of Spatio-Temporal Systems. *Complexity* 2018: 1–16. ISSN 1076-2787, 1099-0526.

Le Clainche, S.; and Vega, J. M. 2017. Higher Order Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems* 16(2): 882–925.

Levin, A.; and Narendra, K. 1993. Control of Nonlinear Dynamical Systems Using Neural Networks: Controllability and Stabilization. *IEEE Transactions on Neural Networks* 4(2): 192–206. ISSN 1941-0093.

Lim, Y. J.; and Teh, Y. W. 2007. Variational Bayesian Approach to Movie Rating Prediction. *Proceedings of KDD cup and workshop* 7: 15–21.

Lumley, J. L. 1967. The Structure of Inhomogeneous Turbulent Flows. In *Atmospheric Turbulence and Radio Wave Propagation*, 166–178. Nauka, Moskow.

Magnus, J. R.; and Neudecker, H. 1979. The Commutation Matrix: Some Properties and Applications. *The Annals of Statistics* 7(2): 381–394. ISSN 0090-5364, 2168-8966.

Nakajima, S.; and Sugiyama, M. 2011. Theoretical Analysis of Bayesian Matrix Factorization. *Journal of Machine Learning Research* 12(79): 2583–2648. ISSN 1533-7928.

Nakajima, S.; Sugiyama, M.; Babacan, S. D.; and Tomioka, R. 2013. Global Analytic Solution of Fully-Observed Variational Bayesian Matrix Factorization. *The Journal of Machine Learning Research* 14(1): 1–37. ISSN 1532-4435.

Ranganath, R.; Gerrish, S.; and Blei, D. 2014. Black Box Variational Inference. In *Artificial Intelligence and Statistics*, 814–822. ISSN 1938-7228.

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. ISSN 0022-1120, 1469-7645.

Schmid, P. J. 2010. Dynamic Mode Decomposition of Numerical and Experimental Data. *Journal of Fluid Mechanics* 656: 5–28. ISSN 0022-1120, 1469-7645.

Seeger, M.; and Bouchard, G. 2012. Fast Variational Bayesian Inference for Non-Conjugate Matrix Factorization Models. In *Artificial Intelligence and Statistics*, 1012–1018. ISSN 1938-7228.

Takeishi, N.; Kawahara, Y.; Tabei, Y.; and Yairi, T. 2017. Bayesian Dynamic Mode Decomposition. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2814–2821. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-0-3.

Tu, J. H.; Rowley, C. W.; Luchtenburg, D. M.; Brunton, S. L.; and Kutz, J. N. 2014. On Dynamic Mode Decomposition: Theory and Applications. *Journal of Computational Dynamics* 1(2): 391.

Vallis, G. K. 2017. *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press. ISBN 978-1-107-06550-5.

Waal, D. J. D. 2006. Matrix-Valued Distributions. In *Encyclopedia of Statistical Sciences*. American Cancer Society. ISBN 978-0-471-66719-3.

Williams, M. O.; Kevrekidis, I. G.; and Rowley, C. W. 2015. A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *Journal of Nonlinear Science* 25(6): 1307–1346. ISSN 1432-1467.

Williams, M. O.; Rowley, C. W.; and Kevrekidis, I. G. 2015. A Kernel-Based Method for Data-Driven Koopman Spectral Analysis. *Journal of Computational Dynamics* 2(2): 247–265.

Wooding, R. A. 1956. The Multivariate Distribution of Complex Normal Variables. *Biometrika* 43(1/2): 212–215. ISSN 0006-3444.