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, ykaharaha@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., 2016] 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
\[ x_{t+1} = f(x_t), \quad x \in \mathcal{M}, \]
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 \( K : \mathcal{G} \to \mathcal{G} \) is defined as an infinite-dimensional linear operator such that
\[ Kg(x) = g(f(x)), \quad g \in \mathcal{G}. \]

Defining \( K \), we can lift nonlinear dynamics \( f \) to a linear (but infinite-dimensional) regime. However, it is difficult to numerically compute \( K \) 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 \( K \), i.e., a subspace \( G \subset \mathcal{G} \) such that
\[ Kg \in G, \quad \forall g \in G, \]
and that a set of observables \( \{g_1, \ldots, g_n\} \) \( (n < \infty) \) spans \( G \). Moreover, consider the restriction of \( K \) 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 \( \tilde{K} \) with respect to \( \{g_1, \ldots, g_n\} \), i.e.,
\[ [Kg_1 \cdots Kg_n]^T = Kg, \]
wherein \( g = [g_1 \cdots g_n]^T \). 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.,
\[ K \varphi(x) = \lambda \varphi(x). \]

This eigenfunction with respect to \( \{g_1, \ldots, g_n\} \) is obtained by \( \varphi(x) = z^H g(x) \), wherein \( z \) is the left-eigenvector of \( K \) corresponding to eigenvalue \( \lambda \). Moreover, let \( v_i \) and \( z_i \) respectively be the right- and the left-eigenvectors of \( K \) corresponding to eigenvalue \( \lambda_i \), and suppose that they are normalized so that \( v_i^H z_i = \delta_{ii} \), without loss of generality. If all the eigenvalues of \( K \) are distinct (i.e., their multiplicity is one), any values of \( g \) are expressed as
\[ g(x_t) = \sum_{i=1}^n \varphi_i(x_t) v_i. \tag{1} \]

Applying \( K \) to both sides of Eq. (1), we have
\[ g(x_{t+1}) = \sum_{i=1}^n \lambda_i \varphi_i(x_t) v_i. \tag{2} \]

In this way, modal decomposition of observables via the Koopman operator is given by
\[ g(x_t) = \sum_{i=1}^n \lambda_i^T w_i, \quad w_i = \varphi_i(x_0) v_i, \tag{3} \]
wherein the values of \( g \) are described as a sum of Koopman modes \( w_i \), 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:
\[ Y_0 = [g(x_0) \cdots g(x_{m-1})] \in \mathbb{C}^{n \times m} \quad \text{and} \quad Y_1 = [g(x_1) \cdots g(x_m)] \in \mathbb{C}^{n \times m}, \tag{4} \]
where \( g \) 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 \( A = Y_1 Y_0^T \) corresponding to its nonzero eigenvalues, wherein \( Y_0^\dagger \) denotes the Moore–Penrose pseudoinverse of \( Y_0 \). If the components of \( g \) span a subspace invariant to \( K \) and all modes are sufficiently excited in the data (i.e., \( \text{rank}(Y_0) = \text{dim}(G) \)), 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 \( Y_0 = U_r S_r V_r^H \).
2. Define a matrix \( \tilde{A} = U_r^H Y_1 V_r S_r^{-1} \).
3. Calculate eigendecomposition of \( \tilde{A} \), i.e., compute \( \tilde{w} \) and \( \lambda \) such that \( \tilde{A} \tilde{w} = \lambda \tilde{w} \).
4. Return dynamic modes \( w = \lambda^{-1} Y_1 V_r S_r^{-1} \tilde{w} \) and corresponding eigenvalues \( \lambda \).

Note that, for DMD to obtain the theoretical rationale, the components of \( g \) need to span an (approximately) invariant subspace. There are several studies that address this issue (e.g., a use of nonlinear basis functions [Williams et al., 2015], reproducing kernels [Kawahara, 2016], and delay coordinates [Susuki and Mezic, 2015; Arbabi and Mezic, 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 \( y_{i,j} \in \mathbb{C} \) be the \( j \)-th column of \( Y_\ell \) 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
\[ y_{0,j} \varphi_{1,j} \cdots \varphi_{k,j} \sim \mathcal{CN}\left( \sum_{i=1}^k \varphi_{i,j} w_i, \sigma^2 I \right), \tag{5} \]

\[ y_{1,j} \varphi_{1,j} \cdots \varphi_{k,j} \sim \mathcal{CN}\left( \sum_{i=1}^k \lambda_i \varphi_{i,j} w_i, \sigma^2 I \right), \]
where we assume that the observation noise is Gaussian, 
\( C\mathcal{N}(\mu, \sigma^2 I) \) is the complex Gaussian 
distribution [Goodman 1963] whose density is defined as
\[
C\mathcal{N}(\mu, \sigma^2 I) = \frac{1}{\pi n \sigma^{2n}} \exp \left( -\frac{1}{\sigma^2} (y - \mu)^H(y - \mu) \right)
\]
Here, \( w_{1:k}, \lambda_{1:k} \), and \( \sigma^2 \) 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
\[
\varphi_{i,j} \sim C\mathcal{N}(0, 1).
\]

### 3.2 Maximum-likelihood Estimator

To derive the MLE of probabilistic DMD, let us rewrite likelihood (5) in a matrix form, i.e.,
\[
y_j | \varphi_j \sim C\mathcal{N}(B \varphi_j, \sigma^2 I),
\]
where we use notations as follows:
\[
y_j = \begin{bmatrix} y_{0,j} \\ y_{1,j} \end{bmatrix}, \quad \varphi_j = [\varphi_{1,j} \ldots \varphi_{k,j}]^T,
\]
\[
B = \begin{bmatrix} W & WA \end{bmatrix}, \quad W = [w_1 \ldots w_k], \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k).
\]

Marginalizing out \( \varphi \) with prior (6), we have
\[
y_j \sim C\mathcal{N}(0, BB^H + \sigma^2 I).
\]

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 \( E \):
\[
\hat{Y}_\ell = Y + E_\ell = [y_{\ell,1} \ldots y_{\ell,m}], \quad \ell = 0, 1,
\]
and let \( \hat{Y} = \begin{bmatrix} \hat{Y}_0^T \hat{Y}_1^T \end{bmatrix}^T \) and \( \Sigma_y = m^{-1} \hat{Y}\hat{Y}^H \). In addition, let \((\sigma^2)^*, W^* \) and \( \Lambda^* \) be the MLEs of Eq. (7) given \( \hat{Y} \). If \( k = n \), then the columns of \( W^* \) and the elements of \( \text{diag}(\Lambda^*) \) 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
\[
(\sigma^2)^* = \frac{1}{2n - k} \sum_{i=k+1}^{2n} \mu_i, \quad \text{and}
\]
\[
B^* = \begin{bmatrix} W^* \\ W^* \Lambda^* \end{bmatrix} = U_k (M_k - (\sigma^2)^* I) \frac{1}{2} R
\]
with \( U_k = [u_1 \ldots u_k] \) and \( M_k = \text{diag}(\mu_1, \ldots, \mu_k) \), where \( \mu_i \) is the \( i \)-th largest eigenvalue of \( \Sigma_y \) with corresponding eigenvector \( u_i \), and \( R \) is an arbitrary unitary matrix. If \( k = n \), we have
\[
W^* \Lambda^*(W^*)^{-1} = U_{1,n} U_{0,n}^{-1},
\]
where \( U_{0,n} \) comprises the first \( n \) rows and \( U_{1,n} \) comprises the last \( n \) rows of \( U_n \). Hence the columns of \( W^* \) and the elements of \( \text{diag}(\Lambda^*) \) are obtained by the eigendecomposition of \( U_{1,n} U_{0,n}^{-1} \), which is exactly the same procedure with the one in total-least-squares DMD [Dawson et al., 2016].

**Proposition 2.** If \( Y_0 \) and \( Y_1 \) are linearly consistent [Tu et al., 2014] and there is no observation noise (i.e., \( E = 0 \)), 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, \( \text{rank}(\Sigma_y) = n \). Hence, \( m^{-1} \hat{Y} = U_n M_n^2 V_n^H V_n \) (\( V_n \) is comprising first \( n \) right singular vectors of \( m^{-1} \hat{Y} \)). Consequently,
\[
Y_1 Y_0^H = U_{1,n} M_n^2 V_n^H (V_n M_n^{-2} U_{0,n}^{-1}) = U_{1,n} U_{0,n}^{-1},
\]
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 \( w_{1:k} \):
\[
w_j | v_{1:j}^2 \sim C\mathcal{N}(0, \text{diag}(v_{1:j}^2, \ldots, v_{1:n}^2))
\]
with an inverse gamma hyperprior on \( v_{1:d}^2 \) (\( d = 1, \ldots, n \)):
\[
v_{1,d}^2 \sim \text{InvGamma}(\alpha_v, \beta_v)
\]
whose shape parameter is \( \alpha_v \) and rate parameter is \( \beta_v \). Moreover, we consider priors on \( \lambda_{1:k} \) and \( \sigma^2 \) as
\[
\lambda_i \sim C\mathcal{N}(0, 1), \quad \sigma^2 \sim \text{InvGamma}(\alpha_\sigma, \beta_\sigma).
\]
A graphical model of Bayesian DMD is shown in Figure 1.

![Figure 1: Graphical model of Bayesian DMD.](image-url)
The sampling procedures of the Gibbs sampler are summarized in Algorithm 2. In the following, $\lambda$ denotes the complex conjugate of $\lambda$, and $[w_i]_d$ denotes the d-th element of $w_i$. In our implementation, the hyperparameters $\alpha$ and $\beta$ were set to $10^{-3}$.

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

1. Sample $w_{1:k}$ from $\mathcal{CN}(m_{w_1}, P_{w_1}^{-1})$ with
   \[
   P_{w_1} = \text{diag}(v_{1,1}^2, \ldots, v_{1,n}^2) + \frac{1 + |\lambda| \sum_j |\varphi_{ij}|^2}{\sigma^2} I,
   \]
   \[
   m_{w_1} = P_{w_1}^{-1} \sum_j \varphi_{ij} (\xi_{i,j} + \lambda_i n_{i,j}).
   \]

2. Sample $v_{1:k,1:n}^2$ from $\text{InvGamma}(a_{v_{i,d}}, b_{v_{i,d}})$ with
   \[
   a_{v_{i,d}} = \alpha_v + 1, \quad b_{v_{i,d}} = \beta_v + |[w_i]_d|^2.
   \]

3. Sample $\lambda_{1:k}$ from $\mathcal{CN}(m_{\lambda_k}, p_{\lambda_k}^{-1})$ with
   \[
   p_{\lambda_k} = 1 + \frac{w_i^H w_i}{\sigma^2} \sum_j |\varphi_{ij}|^2, \quad m_{\lambda_k} = m_{\lambda_k}^H + \frac{w_i^H \varphi_{ij} n_{i,j}}{\sigma^2}.
   \]

4. Sample $\varphi_{1:m}$ from $\mathcal{CN}(m_{\varphi_j}, P_{\varphi_j}^{-1})$ with
   \[
   P_{\varphi_j} = I + \frac{1}{\sigma^2} (W^H W + \Lambda W^H W A),
   \]
   \[
   m_{\varphi_j} = P_{\varphi_j}^{-1} \frac{1}{\sigma^2} (W^H y_{0,j} + \Lambda W^H y_{1,j}).
   \]

5. Sample $\sigma^2$ from $\text{InvGamma}(a_{\sigma^2}, b_{\sigma^2})$ with
   \[
   a_{\sigma^2} = \alpha_\sigma + 2mn, \quad b_{\sigma^2} = \beta_\sigma + \sum_j (y_{0,j} - W \varphi_j)^H (y_{0,j} - W \varphi_j) + \sum_i |[w_i]_d|^2.
   \]

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 $w_{1:k}$, replacing prior (8) and hyperprior (9) by

\[
\begin{align*}
   [w_i]_{1:n}^2 &\sim \mathcal{CN}(0, \sigma^2 \text{diag}(v_{1,1}^2, \ldots, v_{1,n}^2)), \quad \text{and} \\
   v_{i,d}^2 &\sim \text{Exponential}(\gamma_i^2/2)
\end{align*}
\]

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

\[
\begin{align*}
P_{w_1} &= \frac{1}{\sigma^2} \text{diag}(v_{1,1}^2, \ldots, v_{1,n}^2) + \frac{1 + |\lambda| \sum_j |\varphi_{ij}|^2}{\sigma^2} I, \\
   a_{\sigma^2} &= \alpha_\sigma + 2mn + \frac{1}{2} \log n, \\
   b_{\sigma^2} &= \beta_\sigma + \sum_j (y_{0,j} - W \varphi_j)^H (y_{0,j} - W \varphi_j) + \sum_i |[w_i]_d|^2.
\end{align*}
\]

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:

\[
\begin{align*}
a_{v_{i,d}^2} &= \gamma_i^2, \\
b_{v_{i,d}^2} &= \frac{|[w_i]_d|^2}{\sigma^2}, \quad \text{and} \\
p_{v_{i,d}^2} &= \frac{1}{2}.
\end{align*}
\]

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

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

where $\gamma_i^{(Q)}$ denotes the hyperparameter at the Q-th iteration of the EM, and $\mathbb{E}_{\gamma_i^{(Q-1)}} [v_{i,d}^2]$ denotes the expectation under the hyperparameter at the previous iteration.

### 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...
neuroscience [Brunton et al., 2016a], chaotic systems [Brunton et al., 2016b], image processing [Kutz et al., 2016], 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 work on low-rank approximation of DMD [Chen et al., 2012; Wynn et al., 2013; Jovanović et al., 2014; Dicke 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):

\[ y_t = [e^{-2i\theta_t} \quad e^{2i\theta_t} \quad 1 \quad e^{i\theta_t} \quad e^{-i\theta_t}]^T + e_t, \]

where each element of \(e_t\) 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

\[ y_t = \lambda_1^2 [2 \quad 2]^T + \lambda_2^2 [2 \quad -2]^T + e_t, \]

where \(e_t\) was zero-mean Gaussian noise with different variances \(\sigma^2 (\sigma = 0, 0.05, 0.1, 0.15, 0.2, 0.25)\), 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-

Table 1: Averages (and the standard deviations) of the absolute errors of estimated (left) \(\lambda_1\) and (right) \(\lambda_2\) over 100 trials for each noise magnitude \(\sigma\). As for BDMD, the medians of the samples from the Gibbs sampler were adopted as point estimation values.

<table>
<thead>
<tr>
<th>(\Delta \lambda_1)</th>
<th>(\sigma)</th>
<th>(\Delta \lambda_2)</th>
<th>(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.00</td>
<td>.05</td>
<td>.10</td>
</tr>
<tr>
<td>DMD</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>TLS-DMD</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>BDMD</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

The box plots show the statistics (left) \(\lambda_1\) and (right) \(\lambda_2\) for each noise magnitude \(\sigma\). The box plots show the statistics of the samples from the Gibbs sampler of BDMD (the red lines denote the sample medians).
erated a dataset by

\[ y_t = A \begin{bmatrix} 0.9^t \\ 0.7^t \\ 0 \\ 0 \end{bmatrix} + e_t, \quad A = \begin{bmatrix} 0 & -5 & 0 & 0 \\ 2 & -4 & 0 & 0 \\ 3 & -3 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix}, \]

where \( e_t \) 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) \cite{Jovanovic2014} 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 \cite{vanDerMaaten2008}, 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 \( \varphi \) 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/.
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.

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

Acknowledgements

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

References


