

# Factorially Switching Dynamic Mode Decomposition for Koopman Analysis of Time-Variant Systems

Naoya Takeishi, Takehisa Yairi, and Yoshinobu Kawahara

**Abstract**—The modal decomposition based on the spectra of the Koopman operator has gained much attention in various areas such as data science and optimal control, and dynamic mode decomposition (DMD) has been known as a data-driven method for this purpose. However, there is a fundamental limitation in DMD and most of its variants; these methods are based on the premise that the target system is time-invariant at least within the data at hand. In this work, we aim to compute DMD on time-varying dynamical systems. To this end, we propose a probabilistic model that has *factorially switching* dynamic modes. In the proposed model, which is based on probabilistic DMD, observation at each time is expressed using a subset of dynamic modes, and the activation of the dynamic modes varies over time. We present an approximate inference method using expectation propagation and demonstrate the modeling capability of the proposed method with numerical examples of temporally-local events and transient phenomena.

## I. INTRODUCTION

Analysis of dynamical systems using the modal decomposition based on the Koopman operator has gained much attention in various fields of science and engineering (see, e.g., [1]) and also has been utilized in the context of control [2], [3], [4], [5], [6]. Dynamic mode decomposition (DMD) [7], [8] is a popular method for modal decomposition of data, which is related to the spectral decomposition of the Koopman operator under some conditions. However, there is a fundamental limitation in DMD and most of its existing variants; these methods are based on the premise that the target dynamical system can be regarded as a time-invariant system at least within the data at hand. Hence, the outputs of most DMD algorithms may not be valid if, for example, system’s parameters change because of unobserved external effects. Further, in other cases, because the amount of data is limited and thus only a limited number of dynamic modes can be identified, the outputs of DMD tend to be less meaningful when the data contain transient phenomena between unstable equilibria and attractors or temporally localized phenomena that rise and fall suddenly.

In this work, we tackle the problem of conducting DMD on data generated by dynamics that should be regarded as a time-variant system. Our approach is based on the following idea; the dynamical system varies over time, but a part of the

spectral components of the Koopman operator is common even if the system changes. Consequently, each dynamic mode should be turned “on” or “off” as time advances, and only the “on” modes would contribute to the observation at each time. With this intuition, a dataset is modeled by a switching dynamical system where the switching state is factorized into “on-off” states of each dynamic modes. To this end, we propose a model based on probabilistic DMD [9] and derive an approximative inference.

The remainder is organized as follows. In Section II, the technical background is introduced. In Sections III and IV, details of the proposed method are presented. In Section V, the relation between the proposed method and similar models is discussed. In Section VI, numerical examples are shown. Finally, Section VII concludes the paper.

## II. BACKGROUND

First, we briefly introduce the fundamental concepts of dynamical systems and the Koopman operator for completeness. See, e.g., [10], [1] for details. In this paper, we consider a discrete-time dynamical system

$$\mathbf{v}_{i+1} = \mathbf{f}_i(\mathbf{v}_i), \quad \mathbf{v} \in \mathcal{M}, \quad i \in \mathbb{N}, \quad (1)$$

where  $\mathcal{M}$  is a state space. The Koopman operator at time index  $i$ ,  $\mathcal{K}_i$ , is defined as a composition of  $\mathbf{f}_i$  and an observable  $g: \mathcal{M} \rightarrow \mathbb{C}$ , i.e.,

$$\mathcal{K}_i g(\mathbf{v}) = g(\mathbf{f}_i(\mathbf{v})). \quad (2)$$

In the analysis of dynamical systems based on the Koopman operator, we utilize its eigenvalues  $\lambda \in \mathbb{C}$  and eigenfunction  $\varphi: \mathcal{M} \rightarrow \mathbb{C}$ , i.e.,

$$\mathcal{K}_i \varphi_j = \lambda_j \varphi_j, \quad (3)$$

for  $j = 1, \dots, r'$ , where  $r'$  may be infinite.

Let  $(\mathbf{x}_i, \mathbf{y}_i) = (g(\mathbf{v}_i), g(\mathbf{v}_{i+1}))$  be a pair of snapshots, where  $g$  is an  $m$ -dimensional vector-valued observable. Now assume that the components of  $g$  span an  $r$ -dimensional function space that is invariant to the action of  $\mathcal{K}_i$ , and that all the eigenvalues of  $\mathcal{K}_i$  restricted to such space are distinct. Then,  $\mathbf{x}_i$  and  $\mathbf{y}_i$  can be decomposed as

$$\mathbf{x}_i = \sum_{j=1}^r \varphi_j(\mathbf{v}_i) \mathbf{w}_j \quad \text{and} \quad \mathbf{y}_i = \sum_{j=1}^r \lambda_j \varphi_j(\mathbf{v}_i) \mathbf{w}_j$$

with some vector coefficients  $\mathbf{w}_j \in \mathbb{C}^m$ , which are termed the Koopman modes or in DMD’s context, *dynamic modes*.

DMD [7], [8], [11] can be considered as a data-driven computation of such decomposition, and in fact, the connection between DMD and the spectral decomposition of the Koopman operator has been discussed (see, e.g., [11], [12],

N. Takeishi is with RIKEN Center for Advanced Intelligence Project, Japan. A part of this work was done while he was at the University of Tokyo, Japan [naoya.takeishi@riken.jp](mailto:naoya.takeishi@riken.jp)

T. Yairi is with the Department of Aeronautics and Astronautics, The University of Tokyo, Japan [yairi@ailab.t.u-tokyo.ac.jp](mailto:yairi@ailab.t.u-tokyo.ac.jp)

Y. Kawahara is with The Institute of Scientific and Industrial Research, Osaka University, Japan and RIKEN Center for Advanced Intelligence Project, Japan [ykawahara@sanken.osaka-u.ac.jp](mailto:ykawahara@sanken.osaka-u.ac.jp)

[13], [14]). DMD basically computes the eigendecomposition of  $\mathbf{Y}\mathbf{X}^\dagger$ , where the columns of  $\mathbf{X}$  comprise  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , and  $\mathbf{Y}$  is build similarly. Recently, many variants of DMD have been proposed. For example, extensions of DMD using nonlinear basis functions [15] or the kernel method [16] are useful for nonlinear systems. The proposed method in this work is based on probabilistic DMD [9], which is a probabilistic model whose maximum-likelihood solution coincides with the solution of DMD under some conditions.

### III. FACTORIALLY SWITCHING DMD

In this work, we developed a probabilistic model for conducting DMD on time-variant dynamical systems. We refer to the proposed model as *factorially switching DMD* (FSDMD) as each dynamic mode is adaptively turned “on” and “off,” and the “on-off” states of multiple modes determine the overall switching state of the system. In this section, we describe the generative process of data in FSDMD, which is shown as a graphical model in Figure 1. The inference procedures are described in the next section.

#### A. Observation Model

Let  $(\mathbf{x}_i \in \mathbb{C}^m, \mathbf{y}_i \in \mathbb{C}^m)$  be the  $i$ -th pair of snapshots ( $i = 1, \dots, n$ ). The observation model (likelihood) is

$$p(\mathbf{x}_i, \mathbf{y}_i | \chi_i, \psi_i) = p(\mathbf{x}_i | \chi_i)p(\mathbf{y}_i | \psi_i), \quad (4)$$

$$p(\mathbf{x}_i | \chi_i) = \mathcal{CN}_{\mathbf{x}_i}(\mathbf{W}\chi_i, \sigma^2\mathbf{I}), \quad (5)$$

$$p(\mathbf{y}_i | \psi_i) = \mathcal{CN}_{\mathbf{y}_i}(\mathbf{W}\Lambda\psi_i, \sigma^2\mathbf{I}), \quad (6)$$

where  $\chi, \psi \in \mathbb{C}^r$  are latent variables, and  $r$  denotes the total number of dynamic modes.  $\mathbf{W} \in \mathbb{C}^{m \times r}$  is a matrix whose columns comprise dynamic modes,  $\Lambda \in \mathbb{C}^{r \times r}$  is a diagonal matrix whose diagonal elements  $\lambda = \text{diag}(\Lambda)$  comprise eigenvalues, and  $\sigma^2$  is the observation noise variance. We refer to  $\mathbf{W}$ ,  $\lambda$ , and  $\sigma^2$  as the model parameters and denote their set by  $\theta$ . Moreover,  $\mathcal{CN}_{\mathbf{x}}(\mathbf{m}, \mathbf{V})$  denotes the complex normal distribution on  $\mathbf{x}$  with mean  $\mathbf{m}$  and covariance  $\mathbf{V}$ . Note that this observation model is almost identical to the one of probabilistic DMD [9]; the difference between them lies in the formulation of the latent variables,  $\chi$  and  $\psi$ .

#### B. Priors

The priors on  $\chi$  and  $\psi$  are defined using the *two-level spike-and-slab* model, i.e.,

$$p(\chi_{j,i} | \varphi_{j,i}, z_{\chi,j,i}) = (1 - z_{\chi,j,i})\delta(\chi_{j,i}) + z_{\chi,j,i}\delta(\chi_{j,i} - \varphi_{j,i}), \quad (7)$$

$$p(\psi_{j,i} | \varphi_{j,i}, z_{\psi,j,i}) = (1 - z_{\psi,j,i})\delta(\psi_{j,i}) + z_{\psi,j,i}\delta(\psi_{j,i} - \varphi_{j,i}), \quad (8)$$

$$p(\varphi_{j,i}) = \mathcal{CN}_{\varphi_{j,i}}(0, 1), \quad (9)$$

where  $\chi_{j,i}$  denotes the  $j$ -th element of  $\chi_i$  for  $j = 1, \dots, r$ , and  $\delta(\cdot)$  is the Dirac delta function. Here,  $\varphi$  is a latent variable corresponding to a value of Koopman eigenfunctions.  $z_\chi, z_\psi \in \{0, 1\}$  are also latent variables and they control the “on-off” of each mode at each timestep. For instance, if  $z_{\chi,j,i} = z_{\psi,j,i} = 1$ ,  $\mathbf{w}_j$  is a valid dynamic mode for the system at time  $i$ . If  $z_{\chi,j,i} = z_{\psi,j,i} = 0$ ,  $\mathbf{w}_j$  does not

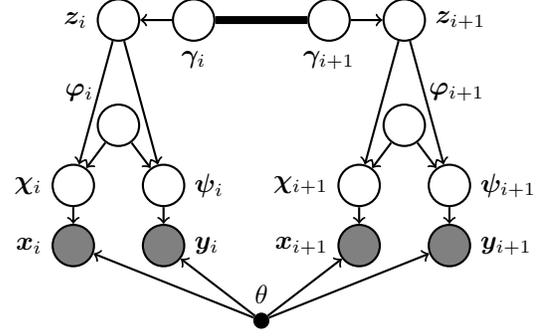


Fig. 1: Graphical model of FSDMD for two time slices  $i$  and  $i + 1$ . The gray nodes are observed random variables, the white nodes are hidden random variables, and  $\theta$  denotes the set of model parameters. The arrows between nodes denote their dependency. The thick edge between  $\gamma$  indicates the dependency modeled by GPs.

contribute as a dynamic mode at time  $i$ . Furthermore, if  $z_{\chi,j,i} = 1$  and  $z_{\psi,j,i} = 0$ , the  $j$ -th mode turns from “on” to “off” within the observation interval of the  $i$ -th snapshot pair, and vice versa.

The “on-off” latent variables,  $z_\chi$  and  $z_\psi$ , are usually not independent because snapshots in a pair are taken at a fixed interval, and the snapshot pairs are often ordered as a time-series. Therefore, the priors on  $z_\chi$  and  $z_\psi$  should be structured according to the underlying dependency. To this end, we use the formulation of Andersen *et al.* [17], [18], with which the Gaussian processes (GPs) are employed for setting structured priors on  $z_\chi$  and  $z_\psi$  as follows:

$$p(\mathbf{z}_j | \gamma_j) = \prod_{i=1}^{2n} \mathcal{B}_{z_{j,i}}(\Phi(\gamma_{j,i})), \quad (10)$$

$$p(\gamma_j) = \mathcal{N}_{\gamma_j}(\mu_j \mathbf{1}, \Sigma_j). \quad (11)$$

Here,  $\mathbf{z}_j = [z_{\chi,j,1} z_{\psi,j,1} \dots z_{\chi,j,n} z_{\psi,j,n}]^\top \in \{0, 1\}^{2n}$  and  $\gamma_j \in \mathbb{R}^{2n}$ . Moreover,  $\mathcal{B}_z(\cdot)$  represents the Bernoulli distribution on  $z$ ,  $\Phi(\cdot)$  is the CDF of standard normal, and  $\mathbf{1}$  denotes a column vector filled with ones. The mean of GP,  $\mu_j \in \mathbb{R}$ , determines the bias in  $z_{j,i}$ , i.e.,  $z$  tends to be zero if  $\mu_j < 0$  and tends to be one if  $\mu_j > 0$ . Further,  $\Sigma_j$  is the covariance matrix of GP, and its  $(i_1, i_2)$ -th element is determined with the value of a positive semidefinite kernel  $k_j(i_1, i_2)$ , where  $i_1$  and  $i_2$  denote the timestamps of two snapshots. Basically, the value of  $\mu$  and the type of  $k(i_1, i_2)$  should be chosen by the user. Meanwhile, we provide an empirical Bayes update rule of  $\mu$  in Section IV.

#### C. Assumptions for Fast GP Inference

The inference with a general covariance matrix of GP may be prohibitively slow owing to its inversion. However, in our case, it can be accelerated by adopting the following two assumptions. First, suppose that the kernel is translation invariant, i.e.,  $k_j(i_1, i_2) = k_j(i_2 - i_1)$ . Second, suppose that the snapshot pairs are from an evenly-spaced time-series, i.e.,  $\mathbf{x}_i = \mathbf{g}(\mathbf{v}_i)$  and  $\mathbf{y}_i = \mathbf{g}(\mathbf{v}_{i+1}) = \mathbf{x}_{i+1}$ , and thus  $z_{\psi,j,i} =$

$z_{\chi,j,i+1}$ . Consequently, Eqs. (10) and (11) are expressed as

$$p(z_j | \gamma_j) = p(z'_j | \gamma'_j) = \prod_{i=1}^{n+1} \mathcal{B}_{z'_{j,i}}(\Phi(\gamma'_{j,i})), \quad (10')$$

$$p(\gamma_j) = p(\gamma'_j) = \mathcal{N}_{\gamma'}(\mu_j \mathbf{1}, \Sigma'_j), \quad (11')$$

with new variables  $\mathbf{z}' = [z_{\chi,j,1} \ z_{\chi,j,2} \ \cdots \ z_{\chi,j,n} \ z_{\psi,j,n}]^\top \in \{0, 1\}^{n+1}$  and  $\gamma'_j \in \mathbb{R}^{n+1}$ . Under the above conditions, new covariance  $\Sigma'_j$  becomes a Toeplitz matrix, whose inversion can be carried out with  $\mathcal{O}(m \log m)$  operations [19]. We assume these conditions unless otherwise stated. Note that the same benefit is achieved even for datasets generated from multiple time-series episodes so long as each episode is evenly spaced because in this case, the covariance is a block-diagonal matrix with Toeplitz components.

#### IV. INFERENCE AND LEARNING OF FSDMD

The model parameters in FSDMD can be learned using an approximative expectation-maximization (EM) algorithm [20], wherein the E-step and the M-step are iterated until convergence. We compute the approximate posterior of the latent variables in the E-step and update the model parameters in the M-step.

##### A. E-step

In the E-step, the posterior distribution of the latent variables should be inferred with a fixed  $\theta$  (and  $\mu$ ) obtained in the last M-step. Now, the posterior distribution can be expressed as follows:

$$\begin{aligned} & p(\chi_{1:n}, \psi_{1:n}, \varphi_{1:n}, \mathbf{z}'_{1:r}, \gamma'_{1:r} | \mathbf{x}_{1:n}, \mathbf{y}_{1:n}) \\ &= \frac{1}{Z} \left( \prod_{i=1}^n p(\mathbf{x}_i | \chi_i) p(\mathbf{y}_i | \psi_i) \right) \\ & \cdot \left( \prod_{i=1}^n \prod_{j=1}^r p(\chi_{j,i} | \varphi_{j,i}, z'_{j,i}) p(\psi_{j,i} | \varphi_{j,i}, z'_{j,i+1}) p(\varphi_{j,i}) \right) \\ & \cdot \left( \prod_{j=1}^r \prod_{i=1}^{n+1} p(z'_{j,i} | \gamma'_{j,i}) \right) \left( \prod_{j=1}^r p(\gamma'_j) \right), \end{aligned} \quad (12)$$

where  $Z$  is the partition function. Since an analytic computation of this is intractable, we resort to an approximation. We use the expectation propagation (EP) algorithm [21] because of the potential superiority of EP for spike-and-slab priors, as reported in literature [22], [23], [18]. EP leverages the factorization structure of the posterior such as the one shown in Eq. (12) by approximating each factor using a site approximation distribution. Below we describe the concept of EP and the actual inference procedures for FSDMD.

In EP, we approximate the posterior by another distribution  $q$ , which is defined according to the structure in Eq. (12), i.e.,

$$\begin{aligned} q(\chi_{1:n}, \psi_{1:n}, \varphi_{1:n}, \mathbf{z}'_{1:r}, \gamma'_{1:r}) &= \frac{1}{Z_{\text{EP}}} \left( \prod_{i=1}^n \tilde{h}_i^{(1)}(\chi_i) \tilde{h}_i^{(2)}(\psi_i) \right) \\ & \cdot \left( \prod_{i=1}^n \prod_{j=1}^r \tilde{h}_{j,i}^{(3)}(\chi_{j,i}, \varphi_{j,i}, z'_{j,i}) \tilde{h}_{j,i}^{(4)}(\psi_{j,i}, \varphi_{j,i}, z'_{j,i+1}) p(\varphi_{j,i}) \right) \\ & \cdot \left( \prod_{j=1}^r \prod_{i=1}^{n+1} \tilde{h}_{j,i}^{(5)}(z'_{j,i}, \gamma'_{j,i}) \right) \left( \prod_{j=1}^r p(\gamma'_j) \right). \end{aligned} \quad (13)$$

Here, terms  $\tilde{h}^{(1)}, \dots, \tilde{h}^{(5)}$  are defined using the *site parameters* of EP, namely,  $\mathbf{m}$ ,  $\mathbf{V}$ ,  $m$ ,  $v$ ,  $c$ ,  $u$ ,  $\eta$ ,  $\xi$ , and  $b$  as follows:

$$\begin{aligned} \tilde{h}_i^{(1)} &= \mathcal{CN}_{\chi_i}(\mathbf{m}_i^{(1)}, \mathbf{V}_i^{(1)}), \quad \tilde{h}_i^{(2)} = \mathcal{CN}_{\psi_i}(\mathbf{m}_i^{(2)}, \mathbf{V}_i^{(2)}), \\ \tilde{h}_{j,i}^{(3)} &= \mathcal{CN}_{\chi_{j,i}}(m_{j,i}^{(3)}, v_{j,i}^{(3)}) \mathcal{CN}_{\varphi_{j,i}}(c_{j,i}^{(3)}, u_{j,i}^{(3)}) \mathcal{B}_{z'_{j,i}}(\Phi(b_{j,i}^{(3)})), \\ \tilde{h}_{j,i}^{(4)} &= \mathcal{CN}_{\chi_{j,i}}(m_{j,i}^{(4)}, v_{j,i}^{(4)}) \mathcal{CN}_{\varphi_{j,i}}(c_{j,i}^{(4)}, u_{j,i}^{(4)}) \mathcal{B}_{z'_{j,i}}(\Phi(b_{j,i}^{(4)})), \\ \tilde{h}_{j,i}^{(5)} &= \mathcal{B}_{z'_{j,i}}(\Phi(b_{j,i}^{(5)})) \mathcal{N}_{\gamma'_{j,i}}(\eta_{j,i}^{(6)}, \xi_{j,i}^{(6)}). \end{aligned}$$

Note that  $q$  in Eq. (13) can be expressed in another form,

$$\begin{aligned} & q(\chi_{1:n}, \psi_{1:n}, \varphi_{1:n}, \mathbf{z}'_{1:r}, \gamma'_{1:r}) \\ &= \left( \prod_{i=1}^n \mathcal{CN}_{\chi_i}(\bar{\mathbf{m}}_i^{\chi}, \bar{\mathbf{V}}_i^{\chi}) \mathcal{CN}_{\psi_i}(\bar{\mathbf{m}}_i^{\psi}, \bar{\mathbf{V}}_i^{\psi}) \mathcal{CN}_{\varphi_i}(\bar{c}_i, \bar{U}_i) \right) \\ & \cdot \left( \prod_{j=1}^r \mathcal{N}_{\gamma_j}(\bar{\eta}_j, \bar{\Xi}_j) \prod_{i=1}^{n+1} \mathcal{B}_{z'_{j,i}}(\Phi(\bar{b}_{j,i})) \right), \end{aligned} \quad (14)$$

using the *global parameters* of EP, namely,  $\bar{\mathbf{m}}$ ,  $\bar{\mathbf{V}}$ ,  $\bar{c}$ ,  $\bar{U}$ ,  $\bar{\eta}$ ,  $\bar{\Xi}$ , and  $\bar{b}$ , which can be easily computed from the site parameters of EP (e.g.,  $\bar{\mathbf{V}}_i^{\chi} = \mathbf{V}_i^{(1)} + \mathbf{V}_i^{(3)}$ ) because each component of the distributions is now in the exponential family. These global parameters are utilized in the EP algorithm and also output as the results of posterior approximation.

Each of the site parameters is updated in turn so that  $q$  approximates the posterior; the updating procedures are shown in Algorithm 1. We omit the derivation of Algorithm 1 due to limitation of space. Readers can consult [18], wherein the EP for a model similar to FSDMD is presented in detail.

**Algorithm 1** (EP for FSDMD). Given data  $\mathbf{x}_{1:n}, \mathbf{y}_{1:n}$  and model parameters  $\mathbf{W}, \Lambda, \sigma^2, \mu$  obtained at the last M-step,

- 1) Update the site parameters of  $\tilde{h}^{(1)}$  and  $\tilde{h}^{(2)}$  by

$$\begin{aligned} (\mathbf{V}_i^{(1)})^{-1} &= \sigma^{-2} \mathbf{W}^H \mathbf{W}, \\ (\mathbf{V}_i^{(1)})^{-1} \mathbf{m}_i^{(1)} &= \sigma^{-2} \mathbf{W}^H \mathbf{x}_i, \\ (\mathbf{V}_i^{(2)})^{-1} &= \sigma^{-2} \Lambda^* \mathbf{W}^H \mathbf{W} \Lambda, \\ (\mathbf{V}_i^{(2)})^{-1} \mathbf{m}_i^{(2)} &= \sigma^{-2} \Lambda^* \mathbf{W}^H \mathbf{y}_i, \end{aligned}$$

where  $\Lambda^*$  denotes the complex conjugate of  $\Lambda$ .

- 2) Update global parameters  $\bar{\mathbf{m}}^{\chi}$ ,  $\bar{\mathbf{V}}^{\chi}$ ,  $\bar{\mathbf{m}}^{\psi}$ , and  $\bar{\mathbf{V}}^{\psi}$ .
- 3) Update the site parameters of  $\tilde{h}^{(3)}$  and  $\tilde{h}^{(4)}$  by the procedures presented in Appendix A.
- 4) Update global parameters  $\bar{\mathbf{m}}^{\chi}$ ,  $\bar{\mathbf{V}}^{\chi}$ ,  $\bar{\mathbf{m}}^{\psi}$ ,  $\bar{\mathbf{V}}^{\psi}$ ,  $\bar{c}$ ,  $\bar{U}$ , and  $\bar{b}$ .
- 5) Update the site parameters of  $\tilde{h}^{(5)}$  by the procedures presented in Appendix B.
- 6) Update global parameters  $\bar{\eta}$  and  $\bar{\Xi}$ .
- 7) Repeat from Step 2 to Step 6 until convergence.

##### B. M-step

In the M-step, given the approximate posterior,  $q$ , obtained from the last E-step, the set of model parameters,  $\theta = \{\mathbf{W}, \Lambda, \sigma^2\}$ , is updated by maximizing

$$\mathcal{Q} = \mathbb{E}_q [p(\mathbf{x}_{1:n}, \mathbf{y}_{1:n}, \chi_{1:n}, \psi_{1:n}, \varphi_{1:n}, \mathbf{z}'_{1:r}, \gamma'_{1:r})], \quad (15)$$

where  $\mathbb{E}_q[\cdot]$  is the expectation with regard to  $q$ . Note that, in the M-step, point estimates (rather than posterior distributions) of the parameters are obtained. By taking derivatives

of  $\mathcal{Q}$  with respect to the model parameters and equating them to zero, we obtain the update rules as follows. First, compute the following quantities:

$$\begin{aligned} \mathbf{S}_x &= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^H, & \mathbf{S}_y &= \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^H, \\ \mathbf{E}_1 &= \sum_{i=1}^n \mathbf{x}_i (\bar{\mathbf{m}}_i^x)^H, & \mathbf{E}_2 &= \bar{\mathbf{V}}^x + \sum_{i=1}^n \bar{\mathbf{m}}_i^x (\bar{\mathbf{m}}_i^x)^H, \\ \mathbf{E}_3 &= \sum_{i=1}^n \mathbf{y}_i (\bar{\mathbf{m}}_i^y)^H, & \mathbf{E}_4 &= \bar{\mathbf{V}}^y + \sum_{i=1}^n \bar{\mathbf{m}}_i^y (\bar{\mathbf{m}}_i^y)^H. \end{aligned}$$

Then,  $\mathbf{W}$  and  $\lambda$  are updated by

$$\mathbf{W} \leftarrow \mathbf{E}_{12} = \mathbf{E}_1 \mathbf{E}_2^{-1} \quad \text{and} \quad (16)$$

$$\lambda \leftarrow (\mathbf{E}_{12}^H \mathbf{E}_{12} \circ \mathbf{E}_4^T)^{-1} (\mathbf{E}_{12}^H \circ \mathbf{E}_3^T) \mathbf{1}, \quad (17)$$

where  $\circ$  denotes the element-wise product of matrices. Using the new values of  $\mathbf{W}$  and  $\lambda$ ,  $\sigma^2$  is updated by

$$\begin{aligned} \sigma^2 &\leftarrow \frac{1}{2nm} \text{tr} \left( \mathbf{S}_x - 2\text{Re} [\mathbf{E}_1 \mathbf{W}^H] + \mathbf{W} \mathbf{E}_2 \mathbf{W}^H \right. \\ &\quad \left. + \mathbf{S}_y - 2\text{Re} [\mathbf{E}_3 (\mathbf{W} \lambda)^H] + \mathbf{W} \lambda \mathbf{E}_4 (\mathbf{W} \lambda)^H \right). \end{aligned} \quad (18)$$

If necessary, a hyperparameter of GP,  $\mu_j$ , can also be updated by

$$\mu_j \leftarrow \text{esum}(\Sigma_j^{-1} \bar{\eta}_j) / (\text{esum}(\Sigma_j^{-1}) + 1), \quad (19)$$

where  $\text{esum}(\mathbf{A})$  is the sum of all the elements of  $\mathbf{A}$ . This is helpful when one would like to determine the total number of dynamic modes automatically from the data.

### C. Implementation Tips

Note that in general, the approximative EM algorithm can only find local solutions. We empirically found that using the outputs of the standard DMD as the initial value of the EM was helpful for avoiding meaningless local solutions. Moreover, note that there is no convergence guarantee for EP. However, by damping the updates with some forgetting rate (see, e.g., [20]), good convergence can often be achieved.

## V. RELATED WORK

A variant of DMD that is strongly related to the proposed method is the multi-resolution DMD (mrDMD) [24], in which DMD computation is recursively performed from the global signal to the local patches of the signal. We can extract temporally local modes using mrDMD, but such modes must be detected within the local patches of the signal. Because mrDMD segments the signal deterministically, the local modes may not be sufficiently detectable within each patch, which possibly poses a challenge in some applications. In contrast, the proposed method can adaptively determine the “on-off” of local modes.

The Koopman analysis or DMD for non-autonomous dynamical systems has been investigated by several other researchers. Proctor *et al.* [25] proposed DMD with control, considering external input signals when computing dynamic modes. Mezić and Surana [26] considered the analysis based on the Koopman operator for periodically varying systems. Maćešić *et al.* [27] presented a theory on applying DMD

to non-autonomous systems and proposed algorithms that computes dynamic modes using a local stencil of data. Note that our approach differs from these previous studies, especially in terms of the manner of expressing the non-autonomous nature of dynamics; FSDMD models a time-variant dynamical system by setting its dynamic modes “on” or “off” at each time.

Let us introduce other related studies from the viewpoint of probabilistic modeling of time-series data, which are mostly from signal processing or machine learning communities. One line of studies deal with switching linear dynamical systems (recent studies include [28], [29]; see references therein), in which states evolve according to one of multiple transition matrices at each time. Moreover, researchers have proposed factorial hidden Markov models such as [30], [31], [32], in which discrete and/or continuous hidden states are factorized as products of multiple primitive states. Infinite factorial dynamical models [33] are similar to our model in the sense that the “on-off” states of hidden dynamics are considered, but the manner of modeling and the use of these states are different. These studies will be helpful for developing more efficient modeling and inference methods for DMD on time-variant systems.

## VI. NUMERICAL EXAMPLES

In this section, we introduce two numerical examples of FSDMD application. In both examples, we used the Gaussian kernel  $k(i_1, i_2) = s \exp(-(i_2 - i_1)^2 / \ell^2)$  whose parameters,  $s$  and  $\ell$ , were set empirically without any intensive search.

### A. Decomposition of Local Waves

The first example is the decomposition of superposed traveling waves. In this example, one of the waves is temporally local, i.e., it rises suddenly and falls suddenly. Such phenomena should be considered when data may contain unobserved external effects or rapidly-decaying disturbances. The dataset used in this example was generated as a superposition of two decaying traveling waves:

$$\begin{aligned} y_1 &= \sin(3x - 2t) e^{-0.1t} \quad \text{and} \\ y_2 &= \begin{cases} \sin(2x - 5t) e^{-0.2t}, & 1/3\pi \leq t \leq 2/3\pi, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (20)$$

for  $x = 2\pi(j-1)/120$ ,  $j = 1, \dots, 120$ ,  $t = 2\pi(i-1)/120$ , and  $i = 1, \dots, 120$ . The generated traveling waves are plotted in Figure 2a. The task was to reconstruct these two component waves from their superposition by estimating the dynamic modes. We applied the standard DMD and FSDMD with  $r = 4$  and  $s = \ell = 1$ .

The waves reconstructed by DMD and FSDMD are shown in Figures 2b and 2c, respectively. Reconstruction by the standard DMD is “bumpy” because of the sudden rise and fall of  $y_2$ , which is not expected by the standard DMD. Meanwhile, reconstruction by the proposed model, FSDMD, is more accurate and the “off” state of  $y_2$  is successfully captured. Moreover, we show the true and estimated eigenvalues in Figure 3; this figure confirms that FSDMD successfully identified the eigenvalues of both traveling waves.

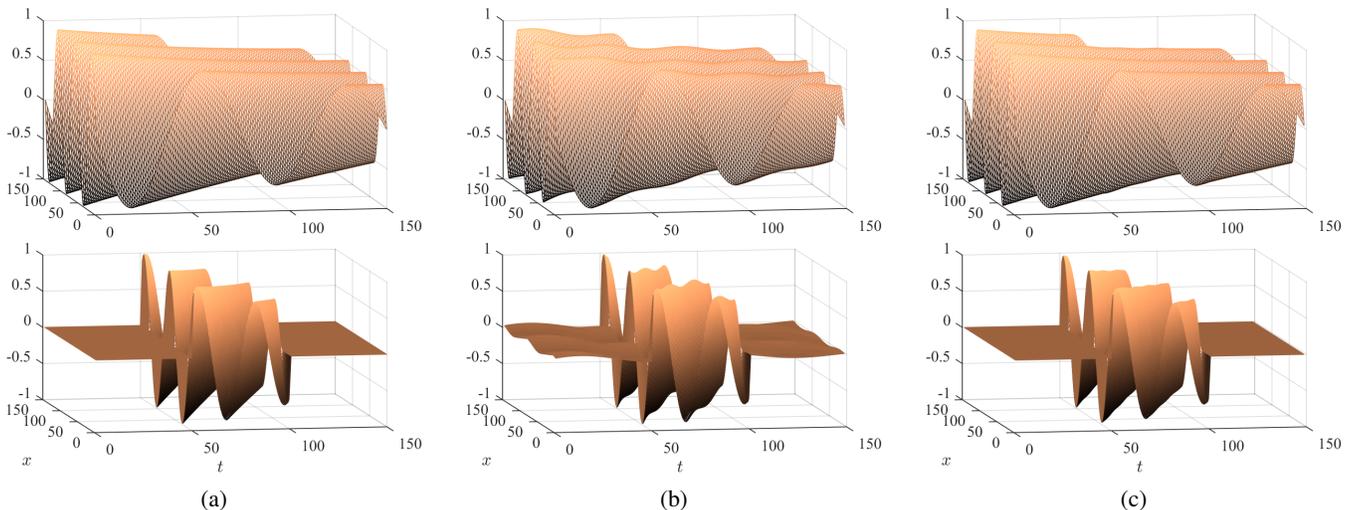


Fig. 2: Traveling waves generated by Eq. (20) or reconstructed by DMDs. The upper plots correspond to the global wave,  $y_1$ , and the lower plots correspond to the local wave,  $y_2$ . (a) True waves used to generate the dataset. (b) Reconstruction by the standard DMD. (c) Reconstruction by the proposed method, FSDMD.

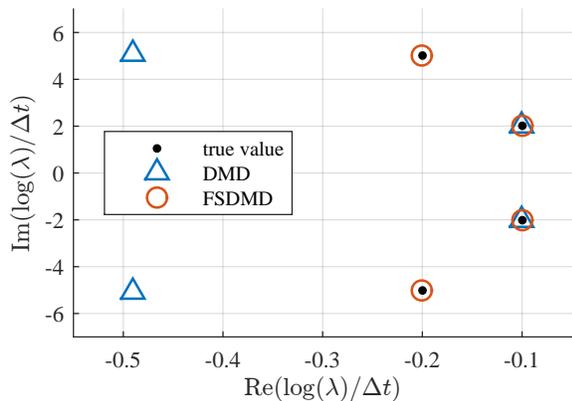


Fig. 3: True and estimated eigenvalues of the traveling waves. The standard DMD fails to estimate the eigenvalues of the local wave, whereas FSDMD successfully identifies all the eigenvalues.

### B. Analysis of Transient Fluid Flow

The second example is the analysis of the fluid flow behind a cylinder. It is well known that a repeating pattern of vortices is formed behind a cylinder in flow, and this pattern is called a Kármán’s vortex street. It is a limit cycle, and thus, a flow starting at some state approaches it as time advances. We generated a dataset by computing the vorticity of the flow field by a simulation<sup>1</sup>, which starts at unstable equilibrium (bottom of Figure 4a) and finally reaches the limit cycle (top of Figure 4a) via the transient regime between them (middle of Figure 4a). The size of flow field is  $440 \times 200$ , and we generated 400 snapshots. We analyzed the data using FSDMD with  $r = 21$ ,  $s = 1$ , and  $\ell = 80$ . The number of dynamic modes,  $r$ , was chosen empirically for presentation clarity; similar results were obtained with different  $r$ .

<sup>1</sup>We used COMSOL Multiphysics<sup>®</sup> for the simulation.

The estimated eigenvalues are shown in Figure 4b (wherein complex conjugates are omitted), and the on-off states of the corresponding dynamic modes are shown in the upper plot of Figure 4c. The lower plot of Figure 4c shows the magnitude of the lift coefficient around the cylinder over time, from which we can see that the flow is transient around time = 200 and approaches the limit cycle after time = 300. In Figure 4d, three of the estimated dynamic modes are shown.

We can notice two interesting points in the results, for example. First, dynamic mode #1 is activated only around time = 1. This is intuitive because the mode decays rapidly (i.e.,  $|\lambda_1| < 1$ ), and its spatial distribution looks corresponding to the unstable equilibrium. Second, some of the dynamic modes with non-zero frequencies, such as #8, #9, #10 #11, and #12, are activated in the middle of the transient regime. In particular, dynamic modes with the higher frequencies are turned on when the flow is the closer to the limit cycle.

## VII. CONCLUSION

In this paper, we have proposed an approach to computing dynamic mode decomposition for time-variant dynamical systems. The proposed method is an extension of the probabilistic DMD [9], in which each dynamic mode is associated with a binary variable that dictates “on-off” of the dynamic modes. Moreover, the temporal dependency of such binary variables is modeled using Gaussian process priors. The inference and learning of the proposed model can be realized with an approximative EM algorithm, where the posterior is approximated using the expectation propagation technique.

Several points should be explored in the future. For example, the automatic determination of the total number of dynamic modes can be performed more elegantly using the techniques of Bayesian nonparametrics. Furthermore, a more efficient inference and learning schemes, including online ones, should be developed.

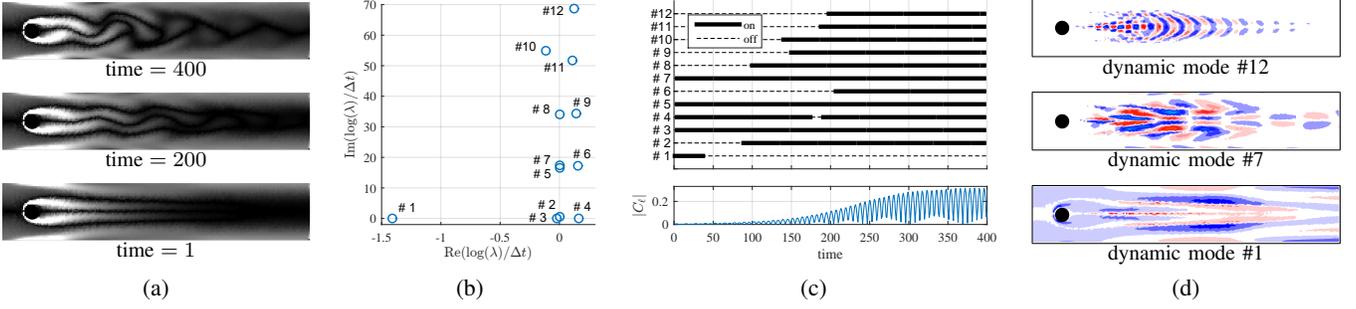


Fig. 4: (a) Three snapshots from the cylinder wake dataset; the flow is at unstable equilibrium at time = 1, the wake is occurring around time = 200, and finally a Kármán's vortex street is clearly observed at time = 400. (b) Eigenvalues estimated by FSDMD, numbered from 1 to 12. Conjugate eigenvalues are omitted for simplicity. (c) The upper plot shows the estimated on-off states of the dynamic modes from #1 to #12. The lower plot shows the magnitude of the lift coefficient around the cylinder; from this plot, we can see that the flow is transient around time = 200 and that the flow approaches the limit cycle after time = 300. The closer the flow approaches the limit cycle, dynamic modes with the higher frequency appear. (d) Contour plots of three dynamic modes.

## APPENDIX

### A. Update of Site Approximations (3) and (4)

In this appendix, the procedures for updating the site parameters of  $\tilde{h}_{j,i}^{(3)}$  are introduced briefly (the site parameters of  $\tilde{h}^{(4)}$  can be updated analogously). See [21] for the general procedures of EP.

First, compute a cavity distribution:

$$\hat{q}_{j,i}^{(3)} \propto \frac{q(\chi_{j,i}, \varphi_{j,i}, z'_{j,i})}{\tilde{h}^{(3)}(\chi_{j,i}, \varphi_{j,i}, z'_{j,i})} = \mathcal{CN}_{\chi_{j,i}}(\hat{m}_{j,i}^{(3)}, \hat{v}_{j,i}^{(3)}) \cdot \mathcal{CN}_{\varphi_{j,i}}(\hat{c}_{j,i}^{(3)}, \hat{u}_{j,i}^{(3)}) \cdot \mathcal{B}_{z'_{j,i}}(\Phi(\hat{b}_{j,i}^{(3)})),$$

whose parameters,  $\hat{m}^{(3)}$ ,  $\hat{v}^{(3)}$ ,  $\hat{c}^{(3)}$ ,  $\hat{u}^{(3)}$ , and  $\hat{b}^{(3)}$  can be computed easily because all the distributions are in the exponential family. Next, obtain  $q^*$  such that

$$q^* = \arg \min_{q'} D_{\text{KL}}(\hat{q}_{j,i}^{(3)} p(\chi_{j,i} | \varphi_{j,i}, z'_{j,i}) \parallel q'),$$

by matching the moments up to the second order. The moments of  $\hat{q}_{j,i}^{(3)} p(\chi_{j,i} | \varphi_{j,i}, z'_{j,i})$  (and thus  $q^*$ ) are

$$I_1^z = \Phi(\hat{b}_{j,i}^{(3)}) \mathcal{CN}_{\hat{m}_{j,i}^{(3)}}(\hat{c}_{j,i}^{(3)}, \hat{v}_{j,i}^{(3)} + \hat{u}_{j,i}^{(3)}),$$

$$I_0 = I_1^z + (1 - \Phi(\hat{b}_{j,i}^{(3)})) \mathcal{CN}_0(\hat{m}_{j,i}^{(3)}, \hat{v}_{j,i}^{(3)}),$$

$$I_1^X = I_1^z \left\{ \frac{\hat{m}_{j,i}^{(3)}}{\hat{v}_{j,i}^{(3)}} + \frac{\hat{c}_{j,i}^{(3)}}{\hat{u}_{j,i}^{(3)}} \right\} \left\{ \frac{1}{\hat{v}_{j,i}^{(3)}} + \frac{1}{\hat{u}_{j,i}^{(3)}} \right\}^{-1},$$

$$I_2^X = I_1^z \left( \left\{ \frac{1}{\hat{u}_{j,i}^{(3)}} + \frac{1}{\hat{v}_{j,i}^{(3)}} \right\}^{-1} + \frac{I_1^X \bar{I}_1^X}{(I_1^z)^2} \right),$$

$$I_1^\varphi = I_1^X + \hat{c}_{j,i}^{(3)}(I_0 - I_1^z), \quad \text{and}$$

$$I_2^\varphi = I_2^X + (\hat{u}_{j,i}^{(3)} + \hat{c}_{j,i}^{(3)} \hat{c}_{j,i}^{(3)})(I_0 - I_1^z),$$

where  $I_0$  is the zeroth moment,  $I_1^z$  is the first moment with regard to  $z'$ , and the other quantities are analogously defined.

Finally, update  $\tilde{h}_{j,i}^{(3)}$  by

$$\left( \tilde{h}_{j,i}^{(3)} \right)^{\text{new}} \propto \frac{q^*}{q_{j,i}^{(3)}},$$

whose parameters are again easily computed because the distributions are in the exponential family.

### B. Update of Site Approximation (5)

The outline of procedures for updating the site parameters of  $\tilde{h}_{j,i}^{(5)}$  is the same with the one for  $\tilde{h}^{(3)}$  and  $\tilde{h}^{(4)}$ , which is presented in Appendix A. That is, 1) we compute a cavity distribution, 2) conduct the moment matching, and 3) update the site parameters using the computed moments. Below we show a summary of computations. For the derivation, readers are recommended to see Section 4.5 of Andersen *et al.* [18].

First, let  $\hat{q}_{j,i}^{(5)}(z'_{j,i}, \gamma'_{j,i})$  be the cavity distribution, i.e.,

$$\begin{aligned} \hat{q}_{j,i}^{(5)}(z'_{j,i}, \gamma'_{j,i}) &\propto \frac{q(z'_{j,i}, \gamma'_{j,i})}{\tilde{h}_{j,i}^{(5)}(z'_{j,i}, \gamma'_{j,i})} \\ &= \mathcal{B}_{z'_{j,i}}(\Phi(\hat{b}_{j,i}^{(5)})) \cdot \mathcal{N}_{\gamma'_{j,i}}(\hat{\eta}_{j,i}^{(5)}, \hat{\xi}_{j,i}^{(5)}) \end{aligned}$$

Then, compute the moments of  $q^*$  by

$$I_1^z = \Phi(\hat{b}_{j,i}^{(5)}) \Phi(\hat{\alpha}_{j,i}^{(5)}),$$

$$I_0 = I_1^z + (1 - \Phi(\hat{b}_{j,i}^{(5)})) (1 - \Phi(\hat{\alpha}_{j,i}^{(5)})),$$

$$I_1^\gamma = I_0 \hat{\eta}_{j,i}^{(5)} + (2I_1^z - \Phi(\hat{\alpha}_{j,i}^{(5)})) \hat{\beta}_{j,i}^{(5)}, \quad \text{and}$$

$$\begin{aligned} I_2^\gamma &= (1 - \Phi(\hat{b}_{j,i}^{(5)})) \left\{ (\hat{\eta}_{j,i}^{(5)})^2 + \hat{\xi}_{j,i}^{(5)} \right. \\ &\quad \left. - \Phi(\hat{\alpha}_{j,i}^{(5)}) \hat{\rho}_{j,i}^{(5)} \right\} + I_1^z \hat{\rho}_{j,i}^{(5)}, \end{aligned}$$

where

$$\begin{aligned}\hat{\alpha}_{j,i}^{(5)} &= \frac{\hat{\eta}_{j,i}^{(5)}}{\sqrt{1 + \hat{\xi}_{j,i}^{(5)}}}, \\ \hat{\beta}_{j,i}^{(5)} &= \frac{\hat{\xi}_{j,i}^{(5)} \cdot \mathcal{N}_{\hat{\alpha}_{j,i}^{(5)}}(0, 1)}{\Phi(\hat{\alpha}_{j,i}^{(5)})\sqrt{1 + \hat{\xi}_{j,i}^{(5)}}}, \\ \hat{\kappa}_{j,i}^{(5)} &= \frac{(\hat{\xi}_{j,i}^{(5)})^2 \cdot \hat{\alpha}_{j,i}^{(5)} \cdot \mathcal{N}_{\hat{\alpha}_{j,i}^{(5)}}(0, 1)}{\Phi(\hat{\alpha}_{j,i}^{(5)})(1 + \hat{\xi}_{j,i}^{(5)})}, \\ \hat{\rho}_{j,i}^{(5)} &= 2\hat{\eta}_{j,i}^{(5)}(\hat{\eta}_{j,i}^{(5)} + \hat{\beta}_{j,i}^{(5)}) + (\hat{\xi}_{j,i}^{(5)} - (\hat{\eta}_{j,i}^{(5)})^2) - \hat{\kappa}_{j,i}^{(5)}.\end{aligned}$$

Finally, update  $\tilde{h}_{j,i}^{(5)}$  by

$$\left(\tilde{h}_{j,i}^{(5)}\right)^{\text{new}} \propto \frac{q^*}{q_{j,i}^{(5)}}.$$

### ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP18H03287.

### REFERENCES

- [1] M. Budišić, R. Mohr, and I. Mezić, “Applied Koopmanism,” *Chaos*, vol. 22, p. 047510, 2012.
- [2] I. Mezić, “On applications of the spectral theory of the Koopman operator in dynamical systems and control theory,” in *Proceedings of the 54th IEEE Conference on Decision and Control*, 2015, pp. 7034–7041.
- [3] A. Mauroy, F. Forni, and R. Sepulchre, “An operator-theoretic approach to differential positivity,” in *Proceedings of the 54th IEEE Conference on Decision and Control*, 2015, pp. 7028–7033.
- [4] A. Surana, “Koopman operator based observer synthesis for control-affine nonlinear systems,” in *Proceedings of the 55th IEEE Conference on Decision and Control*, 2016, pp. 6492–6499.
- [5] A. Mauroy and J. Goncalves, “Linear identification of nonlinear systems: A lifting technique based on the Koopman operator,” in *Proceedings of the 55th IEEE Conference on Decision and Control*, 2016, pp. 6500–6505.
- [6] A. Surana, M. O. Williams, M. Morari, and A. Banaszuk, “Koopman operator framework for constrained state estimation,” in *Proceedings of the 56th IEEE Conference on Decision and Control*, 2017, pp. 94–101.
- [7] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, “Spectral analysis of nonlinear flows,” *Journal of Fluid Mechanics*, vol. 641, pp. 115–127, 2009.
- [8] P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010.
- [9] N. Takeishi, Y. Kawahara, Y. Tabei, and T. Yairi, “Bayesian dynamic mode decomposition,” in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017, pp. 2814–2821.
- [10] I. Mezić, “Spectral properties of dynamical systems, model reduction and decompositions,” *Nonlinear Dynamics*, vol. 41, no. 1-3, pp. 309–325, 2005.
- [11] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, “On dynamic mode decomposition: Theory and applications,” *Journal of Computational Dynamics*, vol. 1, no. 2, pp. 391–421, 2014.
- [12] H. Arbabi and I. Mezić, “Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator,” *SIAM Journal on Applied Dynamical Systems*, vol. 16, no. 4, pp. 2096–2126, 2017.
- [13] N. Takeishi, Y. Kawahara, and T. Yairi, “Subspace dynamic mode decomposition for stochastic Koopman analysis,” *Physical Review E*, vol. 96, no. 3, p. 033310, 2017.
- [14] M. Korda and I. Mezić, “On convergence of extended dynamic mode decomposition to the koopman operator,” *Journal of Nonlinear Science*, vol. 28, no. 2, pp. 687–710, 2018.

- [15] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, “A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition,” *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307–1346, 2015.
- [16] Y. Kawahara, “Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis,” in *Advances in Neural Information Processing Systems*, vol. 29, 2016, pp. 911–919.
- [17] M. R. Andersen, O. Winther, and L. K. Hansen, “Bayesian inference for structured spike and slab priors,” in *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 1745–1753.
- [18] M. R. Andersen, A. Vehtari, O. Winther, and L. K. Hansen, “Bayesian inference for spatio-temporal spike-and-slab priors,” *Journal of Machine Learning Research*, vol. 18, no. 139, pp. 1–58, 2017.
- [19] J. P. Cunningham, K. V. Shenoy, and M. Sahani, “Fast Gaussian process methods for point process intensity estimation,” in *Proceedings of the 25th International Conference on Machine Learning*, 2008, pp. 192–199.
- [20] T. Minka and J. Lafferty, “Expectation-propagation for the generative aspect model,” in *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002, pp. 352–359.
- [21] T. Minka, “Expectation propagation for approximate Bayesian inference,” in *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, 2001, pp. 362–369.
- [22] D. Hernández-Lobato, J. M. Hernández-Lobato, and P. Dupont, “Generalized spike-and-slab priors for Bayesian group feature selection using expectation propagation,” *Journal of Machine Learning Research*, vol. 14, pp. 1891–1945, 2013.
- [23] T. Peltola, P. Jylänki, and A. Vehtari, “Expectation propagation for likelihoods depending on an inner product of two multivariate random variables,” in *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, 2014, pp. 769–777.
- [24] J. N. Kutz, X. Fu, and S. L. Brunton, “Multiresolution dynamic mode decomposition,” *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 2, pp. 713–735, 2016.
- [25] J. L. Proctor, S. L. Brunton, and J. N. Kutz, “Dynamic mode decomposition with control,” *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.
- [26] I. Mezić and A. Surana, “Koopman mode decomposition for periodic/quasi-periodic time dependence,” in *IFAC-PapersOnLine*, vol. 49, no. 18, 2016, pp. 690–697.
- [27] S. Maćešić, N. Črnjarić-Žic, and I. Mezić, “Koopman operator family spectrum for nonautonomous systems – part 1,” 2017, arXiv:1703.07324.
- [28] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, “Bayesian nonparametric inference of switching dynamic linear models,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1569–1585, 2011.
- [29] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski, “Bayesian learning and inference in recurrent switching linear dynamical systems,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017, pp. 914–922.
- [30] Z. Ghahramani and M. I. Jordan, “Factorial hidden Markov models,” *Machine Learning*, vol. 29, no. 2-3, pp. 245–273, 1997.
- [31] S. Chiappa and D. Barber, “Bayesian factorial linear gaussian state-space models for biosignal decomposition,” *IEEE Signal Processing Letters*, vol. 14, no. 4, pp. 267–270, 2007.
- [32] J. A. Quinn, C. K. I. Williams, and N. McIntosh, “Factorial switching linear dynamical systems applied to physiological condition monitoring,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1537–1551, 2009.
- [33] I. Valera, F. J. R. Ruiz, L. Svensson, and F. Perez-Cruz, “Infinite factorial dynamical model,” in *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 1666–1674.