2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

# Optimal Parameter Selection for Kernel Dynamic Mode Decomposition

Wataru Kurebayashi†, Sho Shirasaka‡ and Hiroya Nakao††

†Dept. of Information Engineering, Aomori University
2–3–1 Kohbata, Aomori, Aomori 030-0943, Japan
‡Graduate School of Information Science and Engineering, Tokyo Institute of Technology
††School of Engineering, Tokyo Institute of Technology
2–12–1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
Email: wkurebayashi@gmail.com

**Abstract**—The dynamic mode decomposition is a recently proposed algorithm for the mode decomposition of mixed-mode time series data based on the dynamical systems theory. The kernelization of this algorithm improves the estimation accuracy of dynamic modes, and therefore, facilitates wide-range applications of this method, while the optimal hyper-parameter selection for kernel dynamic mode decomposition still remains an open question. Here, we propose a formula for the hyper-patameter selection, and demonstrate the validity of our selection method.

## 1. Introduction

Dynamic mode decomposition (DMD) [1] is a recently proposed mode decomposition method for mixed-mode time series data. The advantage of this method is that it can directly take into account the latent dynamics of time series. In the field of fluid mechanics, for example, a conventional method called the proper orthogonal decomposition has been widely used, which performs the principal component analysis to reduce the dimensionality of time series, and does not take into account the latent dynamics. Since the DMD can capture the latent dynamics, it is useful for modeling the latent dynamics in terms of dynamical systems theory.

Although the computation of dynamic modes in the DMD is no more than a linear transformation, i.e., the DMD neglects the nonlinearity of time series, the theoretical basis of the DMD based on the Koopman operator can deal with nonlinear latent dynamics underlying time series, so the nonlinear extension of the DMD is expected to significantly enhance the potential of this method. Then, the extended DMD [2] and the kernel DMD [3] have been proposed. The extended DMD [2] is a modified version of the DMD that can approximate the Koopman operator more precisely by nonlinear regression. In addition, the kernel DMD [3] improves the computational efficiency of the extended DMD by what we call "kernel trick". Although the original DMD needs significantly high-dimensional time series data, e.g., simulation data in fluid mechanics, the extended and kernel DMDs can be applied to a wider class of time series. Therefore, these extensions are expected to significantly widen the applicability of the DMD.

However, the selection of optimal hyper-parameters of the extended and kernel DMDs still remains an open question. Since the selection of hyper-parameters deeply affects the estimation accuracy of the eigenvalues and eigenfunctions of the Koopman operator, it is important to establish the selection method for hyper-parameters. In this paper, we propose a selection method [4] for hyper-parameters of the kernel DMD, and demonstrate the validity of this method through numerical experiments.

## 2. Kernel dynamic mode decomposition

To begin with, we introduce a linear operator called the Koopman operator that gives a theoretical basis to the DMD. The DMD is an algorithm that computes the eigenvalues and eigenfunctions of the Koopman operator, which are used for the mode decomposition of time series.

We consider a discrete-time stationary dynamical system, described by

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

where $\boldsymbol{x}_t \in \Omega$ is a vector of state variables at time $t$ and $\boldsymbol{F}(\boldsymbol{x})$ is a map that represents the dyanamics of this system. Here, let $f(\boldsymbol{x})$ ($\boldsymbol{x} \in \Omega$) be an element of an appropriate functional space $\mathcal{F}$. The Koopman operator $\mathcal{K}$ corresponding to the system of Eq. (1) is defined as follows:

$$\mathcal{K}f(\boldsymbol{x}) = f(\boldsymbol{F}(\boldsymbol{x})), \tag{2}$$

that is, the Koopman operator $\mathcal{K}$ maps any function $f(\boldsymbol{x}) \in \mathcal{F}$ to a composite function $f(\boldsymbol{F}(\boldsymbol{x}))$. This operator can be interpreted as a time-shift operator that acts on an observable $f(\boldsymbol{x})$, because the following relation holds:

$$\mathcal{K}f(\boldsymbol{x})\big|_{\boldsymbol{x}=\boldsymbol{x}_t} = f(\boldsymbol{x}_{t+1}). \tag{3}$$

The difinition of the Koopman operator can be naturally extended to stochastic dynamical systems. Instead of Eq. (1), we here consider a stationary Malkov process:

$$\boldsymbol{x}_{t+1} \sim p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t) \tag{4}$$

In this case, the Koopman operator is defined as a conditional expectation as follows:

$$\mathcal{K}f(\boldsymbol{x}) = \int_{\Omega} f(\boldsymbol{x}')p(\boldsymbol{x}'|\boldsymbol{x})d\boldsymbol{x}' \tag{5}$$

Since the Koopman operator is a linear operator, we can consider the spectral decomposition of this operator. In the DMD, the Koopman operator is approximated only by discrete eigenpairs called point spectrum. Each discrete eigenvalue $\lambda_i$ ($i = 1, 2, 3, \ldots$) of the Koopman operator has a corresponding eigenfunction $\xi_i(\boldsymbol{x})$ such that

$$\mathcal{K}\xi_i(\boldsymbol{x}) = \lambda_i \xi_i(\boldsymbol{x}). \tag{6}$$

By using the eigenfunctions $\{\xi_i(\boldsymbol{x})\}$, the time series of state variables $\boldsymbol{x}_t$ can be decomposed as follows:

$$\boldsymbol{x}_t = \sum_{i=1}^{\infty} \boldsymbol{c}_i \xi_i(\boldsymbol{x}_t), \tag{7}$$

where $\{\boldsymbol{c}_i \in \Omega\}$ is appropriate coefficients such that $\boldsymbol{x} = \sum_i \boldsymbol{c}_i \xi_i(\boldsymbol{x})$. In addition, if the dynamics of $\boldsymbol{x}_t$ is deterministic, the time series $\boldsymbol{x}_t$ can be written as a sum of sinusoidal waveforms as follows:

$$\boldsymbol{x}_t = \sum_{i=1}^{\infty} \boldsymbol{c}_i \lambda_i^t \xi_i(\boldsymbol{x}_0). \tag{8}$$

Note that we neglected decaying amplitude components in Eq. (8) under the assumption of stationarity.

In the kernel DMD, the eigenvalues $\{\lambda_i\}$ and eigenfunctions $\{\xi_i(\boldsymbol{x})\}$ of the Koopman operator can be computed from time series data of state variables $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ by the following procedure. This algorithm has two hyper-parameters, the kernel parameter $\theta_k$ and the regularization parameter $\theta_r$. The kernel parameter represents the width of a kernel function. For example, we can use the Gaussian kernel with bandwidth parameter $\theta_k$, i.e., $k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\sum_i |x^{(i)} - x'^{(i)}|^2/\theta_k)$.

1. Computation of $(n-1)\times(n-1)$ matrices $K, L$ whose $(i,j)$-th element $K_{ij}, L_{ij}$ is given by

$$K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j), L_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_{j+1}). \tag{9}$$

2. Singular value decomposition of $K$:

$$K = U\Sigma V^{\top}, \tag{10}$$

where $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}], V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}]$ are orthogonal matrices, and $\Sigma$ is a diagonal matrix whose diagonal elements are $\sigma_1 \geq \ldots \geq \sigma_{n-1}$.

3. Let $n'$ be the number of singular values that covers $100 - \theta_r$ percents of the square sum of the all singular values of $K$. We approximate $K$ by the $n'$ largest singular values as

$$K \approx \tilde{U}\tilde{\Sigma}\tilde{V}^{\top} \tag{11}$$

where $\tilde{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n'}], \tilde{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n'}]$, and $\tilde{\Sigma}$ is a diagonal matrix whose diagonal elements are $\sigma_1 \geq \ldots \geq \sigma_{n'}$.

4. Computation of $\tilde{A}$ ($= \tilde{U}^{\top} L K^{-1} \tilde{U}$):

$$\tilde{A} = \tilde{U}^{\top} L \tilde{V} \tilde{\Sigma}^{-1}. \tag{12}$$

5. Eigendecomposition of $\tilde{A}^{\top}$. Let $\{\lambda_i\}$, $\{\boldsymbol{q}_i\}$ ($i = 1, \ldots, n'$) be the eigenvalues and eigenvectors of $\tilde{A}^{\top}$.

6. Computation of the eigenfunctions of the Koopman operator, $\{\xi_i(\boldsymbol{x})\}$, as follows:

$$\xi_i(\boldsymbol{x}) = \boldsymbol{q}_i^{\top} \tilde{U}^{\top} \boldsymbol{\phi}(\boldsymbol{x}), \tag{13}$$

where $\boldsymbol{\phi}(\boldsymbol{x}) = [k(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}, \boldsymbol{x}_{n-1})]^{\top}$.

Note that we introduced the regularization with hyper-parameter $\theta_r$ (steps 2–4), which is not included in Williams et al [3].

## 3. Criterion for parameter selection

In this section, we derive a criterion for selecting two hyper-parameters of the kernel DMD, the kernel parameter $\theta_k$ and the regularization parameter $\theta_r$. We evaluate the error between the exact and estimated Koopman operators, and select the two hyper-parameters that minimize the estimation error.

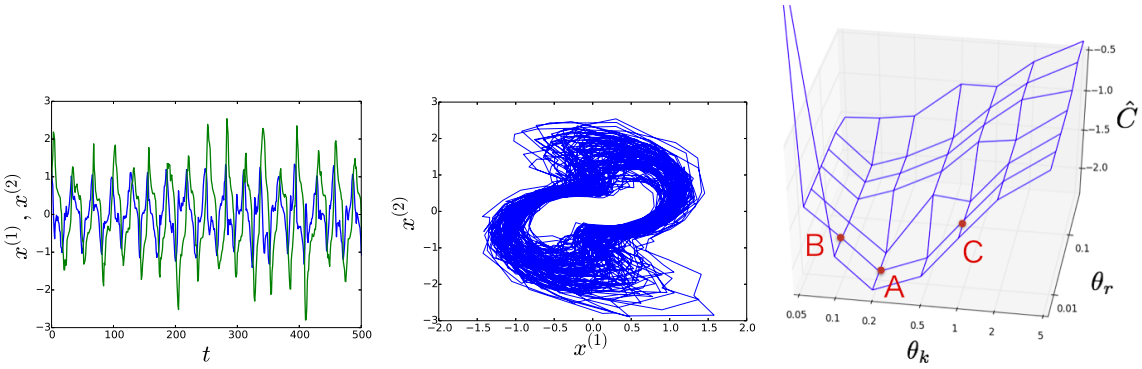
In order to evaluate the estimation error, we introduce the mean square error between the exact and estimated Koopman operators, $\mathcal{K}$ and $\hat{\mathcal{K}}$, as follows:

$$\mathrm{MSE} = \int_{\Omega}\int_{\Omega} \{p(\boldsymbol{x}|\boldsymbol{x}') - \hat{p}(\boldsymbol{x}|\boldsymbol{x}')\}^2 q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}', \tag{14}$$

where $\hat{p}(\boldsymbol{x}|\boldsymbol{x}')$ is the estimated conditional probability assumed implicitly in the algorithm of the kernel DMD, and $q(\boldsymbol{x})$ is the stationary probability density of the state $\boldsymbol{x}$. In Eq. (14), we calculate a square error ($\int_{\Omega}\{p(\boldsymbol{x}|\boldsymbol{x}') - \hat{p}(\boldsymbol{x}|\boldsymbol{x}')\}^2 d\boldsymbol{x}$) and its expectation value with respect to $\boldsymbol{x}'$, so we call Eq. (14) the mean square error. The advantage of this criterion is that it is plausible to approximately compute it only from data. For example, the Hilbert-Schmidt norm of $\mathcal{K} - \hat{\mathcal{K}}$ is a possible alternative of Eq. (14), but it is difficult to compute it only from data.

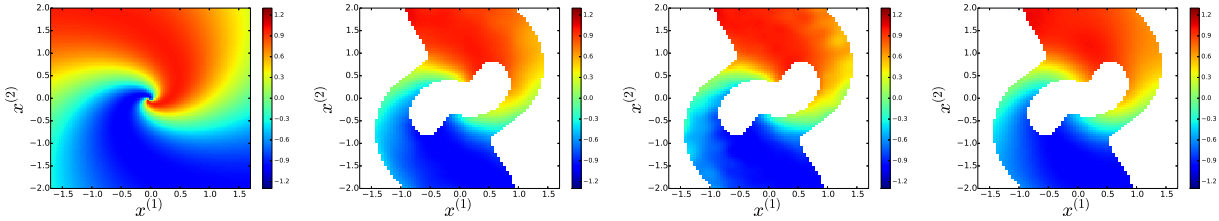
Equation (14) cannot be evaluated only by time series data, because it includes the unknown exact probability density $p(\boldsymbol{x}|\boldsymbol{x}')$. Then, from Eq. (14), we derive a cost function that can be evaluated only by time series data. We can rewrite Eq. (14) as follows:

$$\begin{aligned}\mathrm{MSE} = &\int_{\Omega}\int_{\Omega} \{p(\boldsymbol{x}|\boldsymbol{x}')\}^2 q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\ &- 2\int_{\Omega}\int_{\Omega} p(\boldsymbol{x}|\boldsymbol{x}')\hat{p}(\boldsymbol{x}|\boldsymbol{x}')q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\ &+ \int_{\Omega}\int_{\Omega} \{\hat{p}(\boldsymbol{x}|\boldsymbol{x}')\}^2 q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}'. \end{aligned} \tag{15}$$




(a) Time series. (b) Phase portrait. (c) Cost function.

Figure 1: (a,b) Time series data used in the numerical experiment, and (c) the cost function $\hat{C}$ for $0.05 \leq \theta_k \leq 5$ and $0.005 \leq \theta_r \leq 0.5$. The parameter sets A, B and C depicted in the figure are used as examples in the following.




(a) Exact. (b) KDMD (A: optimal kernel). (c) KDMD (B: narrower kernel). (d) KDMD (C: wider kernel).

Figure 2: Real part of the second Koopman eigenfunction ($\propto e^{i[\text{phase}]}$ ).

Since the first term of the right hand side of Eq. (15) does not depend on the hyper-parameters, we define a cost function $C$ as follows:

$$\begin{aligned} C &= \text{MSE} - \int_\Omega \int_\Omega \{p(\boldsymbol{x}|\boldsymbol{x}')\}^2 q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\ &= -2 \int_\Omega \int_\Omega p(\boldsymbol{x}|\boldsymbol{x}') \hat{p}(\boldsymbol{x}|\boldsymbol{x}') q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\ &\quad + \int_\Omega \int_\Omega \{\hat{p}(\boldsymbol{x}|\boldsymbol{x}')\}^2 q(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}'. \end{aligned} \tag{16}$$

The cost function $C$ can be approximately evaluated by time series data $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{\tilde{n}}\}$ that is not used for the estimation of the Koopman operator as follows:

$$\hat{C} = -\frac{2}{\tilde{n}-1} \sum_i^{\tilde{n}-1} \hat{p}(\tilde{\boldsymbol{x}}_{i+1}|\tilde{\boldsymbol{x}}_i) + \frac{1}{\tilde{n}} \sum_i^{\tilde{n}} \int_\Omega \{\hat{p}(\boldsymbol{x}|\tilde{\boldsymbol{x}}_i)\}^2 d\boldsymbol{x}. \tag{17}$$

Here, the estimated conditional probability $\hat{p}(\boldsymbol{x}|\boldsymbol{x}')$ assumed implicitly in the kernel DMD can be derived as follows (see Ref. [4] for derivation):

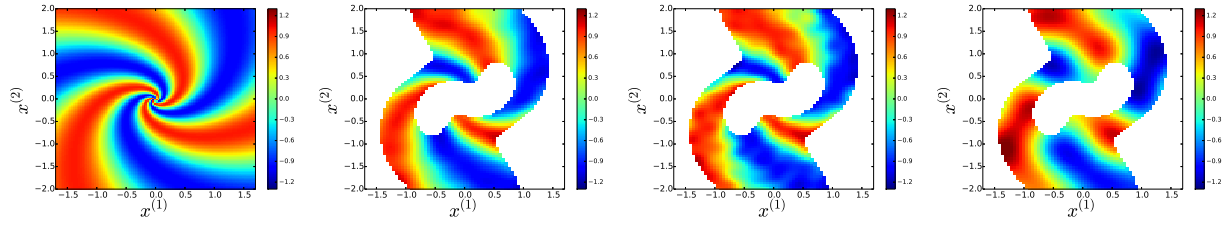
$$\hat{p}(\boldsymbol{x}|\boldsymbol{x}') = \boldsymbol{\phi}(\boldsymbol{x})^\top \tilde{U} (\tilde{U}^\top G \tilde{U})^{-1} \tilde{A} \tilde{U}^\top \boldsymbol{\phi}(\boldsymbol{x}'), \tag{18}$$

where $G$ is an $(n-1) \times (n-1)$ matrix whose $(i,j)$-th element is given by $\int_\Omega k(\boldsymbol{x}, \boldsymbol{x}_i) k(\boldsymbol{x}, \boldsymbol{x}_j) d\boldsymbol{x}$. By using Eq. (18), the cost function $\hat{C}$ can be written as

$$\begin{aligned} \hat{C} = &-\frac{2}{\tilde{n}-1} \sum_i^{\tilde{n}-1} \boldsymbol{\phi}(\tilde{\boldsymbol{x}}_{i+1})^\top \tilde{U} (\tilde{U}^\top G \tilde{U})^{-1} \tilde{A} \tilde{U}^\top \boldsymbol{\phi}(\tilde{\boldsymbol{x}}_i) \\ &+ \frac{1}{\tilde{n}} \sum_i^{\tilde{n}} \boldsymbol{\phi}(\tilde{\boldsymbol{x}}_i)^\top \tilde{U} \tilde{A}^\top (\tilde{U}^\top G \tilde{U})^{-1} \tilde{A} \tilde{U}^\top \boldsymbol{\phi}(\tilde{\boldsymbol{x}}_i). \end{aligned} \tag{19}$$

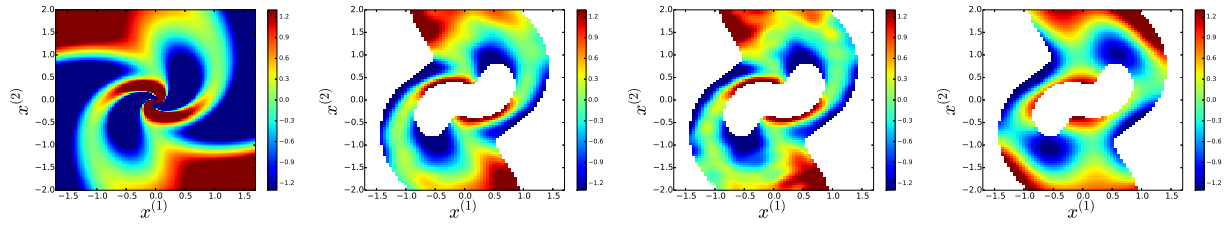
Thus, we can select the hyper-parameters, $\theta_k$ and $\theta_r$, by evaluating the const function $\hat{C}$ for each parameter set and selecting the parameter set $\theta_k, \theta_r$ that minimizes $\hat{C}$. The evaluation of $\hat{C}$ can be performed by the $m$-hold cross validation of Eq. (19) as follows:

1. Divide the time series $\{\boldsymbol{x}_t\}$ into $m$ subsets, denoted by $X_i$ $(i = 1, 2, \ldots, m)$.
2. For $i = 1, 2, \ldots, m$:
   (a) Apply the kernel DMD to all the time series except $X_i$ and compute $\tilde{A}$, $\tilde{U}$ and $G$.
   (b) Evaluate the cost function $\hat{C}$ of Eq. (19) by $\tilde{A}$, $\tilde{U}$, $G$ and $X_i$.
3. Average the values of $\hat{C}$ evaluated for $X_1, \ldots, X_m$.



(a) Exact. (b) KDMD (A: optimal kernel). (c) KDMD (B: narrower kernel). (d) KDMD (C: wider kernel).

Figure 3: Real part of the sixth Koopman eigenfunction ($\propto e^{i3[\mathrm{phase}]}$ ).



(a) Exact. (b) KDMD (A: optimal kernel). (c) KDMD (B: narrower kernel). (d) KDMD (C: wider kernel).

Figure 4: Real part of the fourteenth Koopman eigenfunction ($\propto$ [amplitude] $\times e^{i2[\mathrm{phase}]}$ ).

## 4. Numerical experiments

In order to confirm the validity of our method, we perform a numerical experiment, in which we apply our method to time series data genereted by a numerical simulation. As an example, we use two-dimensional time series data $\boldsymbol{x}_t = [x_t^{(1)}, x_t^{(2)}]^\top$ depicted in Fig. 1(a,b),. This time series is generated from a simple dynamical system described by $x_t^{(1)} = \exp(\rho_t + \sin 2\varphi_t)\cos(\varphi_t + \rho_t + \sin 2\varphi_t)$, $x_t^{(2)} = \exp(\rho_t + \sin 2\varphi_t)\sin(\varphi_t + \rho_t + \sin 2\varphi_t)$, $\varphi_{t+1} = \varphi_t + \pi/15 + 0.1\eta_t^{(1)}$, and $\rho_{t+1} = 0.9\rho_t + 0.1\eta_t^{(2)}$, where $\eta_t^{(1)}, \eta_t^{(1)}$ are independent Gaussian white noise of unit intensity. In this case, we can analytically derive the Koopman eigenfunctions as shown in Figs. 2–4(a).

We show the evaluation of the cost function $\hat{C}$ in Fig. 1(c), which is evaluated by 5-hold cross validation. The parameter sets A, B and C shown in Fig. 1(c) represent optimal, narrower and wider kernel parameters, which will be used as examples in the following discussion.

In Figs. 2–4, we show the eigenfunctions computed by the kernel DMD, which correspond to dynamical variables such as a phase or amplitude of oscillatory modes. In each figure, we compare (a) the exact eigenfunction with (b–d) the estimated eigenfunctions for the parameter sets (b) A, (c) B and (d) C. In each of Figs. 2–4, we see that the results for the parameter set A nicely fits to the exact eigenfunction, while the result for the parameter set B (narrower kernel) is noisy and overfitted, and the result for the parameter set C (wider kernel) is underfitted. Note that the region with no data point is shown in white, because the kernel DMD cannot estimate the eigenfunction in this region. From these results, we see that our method can find the optimal hyper-parameter for this time series data.

## 5. Summary

We proposed a method for selecting two hyper-parameters, $\theta_k, \theta_r$ [4], which deeply affect the estimation accuracy of Koopman eigenfunctions. In a numerical experiment, we demonstrated the validity of our method. By using the optimal hyper-parameters, we could estimate eigenfunctions of the Koopman operator in high precision. This result implies the importance of the optimal selection of hyper-parameters of the kernel DMD.


### Acknowledgment

W.K. is financially supported by JSPS KAKENHI JP16K16125 (Grant-in-Aid for Young Scientists (B)). S.S. is financially supported by JSPS KAKENHI JP15J12045. H.N. is financially supported by JSPS KAKENHI JP16K13847, JP16H01538.


### References


[1] P. J. Schmid, J. Fluid Mech. 656, 5–28 (2010).

[2] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, J. Nonlin. Sci. 25, 6, 1307–1346 (2015).

[3] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, arXiv:1411.2260 (2014).

[4] W. Kurebayashi, S. Shirasaka, and H. Nakao, in preparation (2016).