2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Modeling Nonlinear Dynamic System in RKHS through the Koopman Operator

Satomi Sugaya*, Yoshihiko Susuki†, Atsushi Ishigame†, Andrea Mammoli‡, Manel Martínez-Ramón‡

* School of Math, Science & Engineering, Central New Mexico Community College, USA
† Department of Electrical and Information Systems, Osaka Prefecture University, Japan
‡ Electrical and Computer Engineering Department, The University of New Mexico, USA
Email: satomisu@gmail.com, susuki@eis.osakafu-u.ac.jp, manel@unm.edu

Abstract—Koopman Operator is a linear but infinite-dimensional operator defined for a nonlinear dynamical system and captures full information of the system. We present a formulation in Reproduced Kernel Hirbert Space (RKHS) for modeling a nonlinear dynamic system in order to develop relevant linear estimators. The KO is represented as a linear estimator in RKHS, and its parameters are determined using the well-known Gaussian process models. This leads to structures useable in modeling and nowcasting that account for the nonlinear behavior of the system.

1. Introduction

Koopman Operator (KO) methods allow us to work in the linear formulation on analysis of systems whose underlying dynamics is nonlinear through the knowledge of observations [1, 2]. Although KO becomes infinite-dimensional (acts on a space of functions defined on the state space of a system, called observables), it provides full information on the underlying nonlinear system. While finding the complete solution to a nonlinear system is in general impossible, we have abundance of data on observables arising from the dynamics and can use them effectively in terms of spectra of the KO. Various applications of this new view are reported, for example, in fluid mechanics [2] and power grids [3].

Reproducing Kernel Hilbert Spaces (RKHS) have been extensively used by the Machine Learning community due to the fact that by the virtue of the Generalized Representer Theorem [4, 5] a linear estimator (to any nonlinear observable-dynamics) can be expressed and optimized in such spaces, which in turn provides an estimator with nonlinear properties. Several approaches have been presented which combine the KO method with RKHS [6].

In this short paper, we present a formulation in RHKS for modeling a nonlinear dynamic system in order to develop linear estimators. The essence of this work is to derive a relevant representation of KO from a finite number of samples of the underlying nonlinear dynamics. The so-called extended dynamic mode decomposition [7] and its kernel version [8] are reported for this line of research. Following [8, 6], we utilize the kernel idea to generate a finite number of observables that are defined by the sampled dynamics. Thus, the RKHS formalism is utilized in the current nonlinear modeling and linear estimators. Since the structural complexity of such estimators is in general arbitrary, we use Tikhonov type regularization based on the well known Gaussian Process (GP) models [9]. By making use of the Representer theorem, this leads to structures useable in modeling and nowcasting that account for the nonlinear behavior of the system.

2. Koopman Spectral Analysis

Here we provide a brief introduction of KO for nonlinear dynamic systems and its spectral analysis [10, 11]. Consider a discrete-time dynamic system evolving on a finite n dimensional manifold $\mathbb{X}$: for $\boldsymbol{x}_k \in \mathbb{X}$ and $k \in \mathbb{Z}$,

$$\boldsymbol{x}_{k+1} = \boldsymbol{T}(\boldsymbol{x}_k) \tag{1}$$

where $\boldsymbol{T}$ maps $\mathbb{X}$ to itself and is considered to be nonlinear in general. Here, we do not address the problem of finding the state evolution $\boldsymbol{x}_k$, but that of the scalar-valued observable, which is a function of the state $\boldsymbol{x}$, $g : \mathbb{X} \to \mathbb{C}$. In the following, we will use $\mathcal{H}$ to represent a given space of observables, *i.e.*, $g \in \mathcal{H}$. Thus, the evolution of the observable g itself under an iteration of $\boldsymbol{T}$ is represented by the so-called Koopman Operator (KO) $\mathbf{U} : \mathcal{H} \to \mathcal{H}$ such that

$$\mathbf{U}g = g \circ \boldsymbol{T}. \tag{2}$$

Note that $\mathbf{U}$ is a linear operator that acts on the function space $\mathcal{H}$ (by choice, it can become the Hilbert space), and hence it is infinite-dimensional. The decomposition of an observable in this space in terms of eigen-functions of $\mathbf{U}$ is called the Koopman mode decomposition [10, 11].

Let $\phi_i \in \mathcal{H} \backslash \{0\}$ be the i-th eigen-function of $\mathbf{U}$ with eigen-value λ_i: $\mathbf{U}\phi_i = \lambda_i \phi_i$. If g lies in the subspace spanned by the eigen-functions, then g can be written as

$$g(\boldsymbol{x}_k) = \sum_{i=1}^{\infty} \phi_i(\boldsymbol{x}_k) v_i \tag{3}$$

where $v_i = \langle \phi_i, g \rangle \in \mathbb{C}$ (inner-product defined on $\mathcal{H}$). In the same manner, for a vector-valued observable $\boldsymbol{g} = (g_1, \ldots, g_m)^\top$ where $g_j \in \mathcal{H}$ and $\top$ the transpose operation, we have

$$\boldsymbol{g}(\boldsymbol{x}_k) = \sum_{i=1}^{\infty} \phi_i(\boldsymbol{x}_k) \boldsymbol{v}_i \tag{4}$$

where $\boldsymbol{v}_i = (\langle \phi_i, g_1 \rangle, \ldots, \langle \phi_i, g_m \rangle)^\top \in \mathbb{C}^m$. From (1) and the definition of eigen-functions, it is appropriate to write

$$\phi_i(\boldsymbol{x}_k) = (\mathbf{U}^k \phi_i)(\boldsymbol{x}_0) = \lambda_i^k \phi_i(\boldsymbol{x}_0). \tag{5}$$

Thus, we have

$$g(\boldsymbol{x}_k) = \sum_{i=1}^{\infty} \lambda_i^k \phi_i(\boldsymbol{x}_0) v_i, \tag{6}$$

and

$$\boldsymbol{g}(\boldsymbol{x}_k) = \sum_{i=1}^{\infty} \lambda_i^k \phi_i(\boldsymbol{x}_0)\boldsymbol{v}_i. \tag{7}$$

The vector $\boldsymbol{v}_i$ is called the Koopman mode [11] of the dynamic system (1), corresponding to the given observable g.

3. Review of Linear System Modeling

Consider a linear dynamical system on $\mathbb{X} = \mathbb{R}^M$ (finite M) given by

$$\boldsymbol{x}_{k+1} = \mathsf{A}\boldsymbol{x}_k, \tag{8}$$

where $\boldsymbol{x}_k \in \mathbb{R}^M$ and $\mathsf{A} \in \mathbb{R}^{M\times M}$. Because we know the dynamics on $\mathbb{X}$ completely in this case, we do not need to go through the observable space. In [11] the eigen-values and eigen-vectors of A are related to eigen-values and eigen-functions of the associated Koopman operator $\mathbf{U}$.

The eigen-values and vectors of A can be found from the Krylov method (*e.g.* [12]), which is extended in the Koopman mode decomposition of nonlinear dataset [11]. Now consider a finite number of samples of the linear state dynamics $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ and construct the so-called Krylov subspace. spanned by the $m-1$ samples $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m-1}\}$.

For simplicity of the review, first we introduce the special case where $\boldsymbol{x}_m$ is given by a linear combination of all the past data, *i.e.*,

$$\boldsymbol{x}_m = c_0\boldsymbol{x}_0 + \cdots + c_{m-1}\boldsymbol{x}_{m-1} = \sum_{i=0}^{m-1} c_i\boldsymbol{x}_i, \tag{9}$$

where $c_0, \ldots, c_{m-1}$ are the coefficients. By defining K and $\boldsymbol{c}$ as

$$\mathsf{K} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{m-1}], \tag{10}$$

$$\boldsymbol{c} = [c_0, c_1, \ldots, c_{m-1}]^\top, \tag{11}$$

we have

$$\boldsymbol{x}_m = \mathsf{K}\boldsymbol{c}. \tag{12}$$

Here, the data $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ is also represented as

$$[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m] = \mathsf{AK}. \tag{13}$$

Thus, we derive

$$\mathsf{AK} = \mathsf{KC}, \tag{14}$$

where C is the m-dimensional companion matrix defined as

$$\mathsf{C} = \begin{bmatrix} 0 & 0 & \ldots & 0 & c_0 \\ 1 & 0 & & 0 & c_1 \\ 0 & 1 & & 0 & c_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & 1 & c_{m-1} \end{bmatrix}. \tag{15}$$

If $\boldsymbol{u}_i$ and λ_i are an eigen-vector and an eigen-value of C, then multiplying (14) by $\boldsymbol{u}_i$ from the right gives

$$\mathsf{AK}\boldsymbol{u}_i = \mathsf{KC}\boldsymbol{u}_i = \mathsf{K}\lambda\boldsymbol{u}_i = \lambda\mathsf{K}\boldsymbol{u}_i, \tag{16}$$

resulting in $\boldsymbol{v}_i \equiv \mathsf{K}\boldsymbol{u}_i$ being the eigenvector of A with the same eigenvalue. Thus, by computing $\boldsymbol{c}$ from the data, it is possible to compute the eigen-values and eigen-vectors of A, namely, obtain the linear dynamic model.

For a more general case, when $\boldsymbol{x}_m \neq \mathsf{K}\boldsymbol{c}$, then the approximate eigenvalue and vector to A is given by minimizing the error, *i.e.*, the difference between the actual $\boldsymbol{x}_m$ and $\mathsf{K}\boldsymbol{c}$:

$$\boldsymbol{r} = \boldsymbol{x}_m - \mathsf{K}\boldsymbol{c}. \tag{17}$$

This is an optimization problem where $\boldsymbol{c}$ is chosen so that $\boldsymbol{r} \perp \text{span}\,\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m-1}\}$. Thus, we have

$$\mathsf{AK} = \mathsf{KC} - \boldsymbol{r}\boldsymbol{e}^\top, \tag{18}$$

where $\boldsymbol{e} = [0, \ldots, 0, 1]^\top \in \mathbb{R}^M$. The approximate eigenvalues and vectors of A are given by that of C, and are called the *Ritz* vales and vectors [11]. If K is of full-column rank, then the vector $\boldsymbol{c}$ is found from (17) as

$$\boldsymbol{c} = \left(\mathsf{K}^\top\mathsf{K}\right)^{-1}\mathsf{K}^\top\boldsymbol{x}_m. \tag{19}$$

In the Machine learning context, the solution for $\boldsymbol{c}$ given in (19) is nothing but the result of a linear least-square-error regression. Below, we present a method commonly used in Machine learning community to find such $\boldsymbol{c}$ form nonlinear data set.

4. Nonlinear System Modeling using Gaussian Processes

4.1. RKHS representation of the Koopman operator

Consider a nonlinear transformation $\varphi(\cdot)$ into an RKHS provided by a kernel dot product $\varphi_i^\top\varphi_j = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and where $\varphi(\boldsymbol{x}_i) = \varphi_i$. The existence of such a dot product is proved as Mercer's theorem [13]. Here, we denote the data matrix inside the RKHS as

$$\mathcal{K}_{m-1} = [\varphi_0, \varphi_1 \ldots, \varphi_{m-1}], \tag{20}$$

and the kernel matrices of dot products between data as

$$\mathsf{H}_{m-1} = \mathcal{K}_{m-1}^\top\mathcal{K}_{m-1}, \tag{21}$$

$$\boldsymbol{h}_{m-1,m} = \mathcal{K}_{m-1}^\top\varphi_m. \tag{22}$$

Assume now an estimator

$$\varphi_m = f(\mathcal{K}_{m-1}) + r_m \tag{23}$$

where

$$f(\mathcal{K}_{m-1}) = \mathcal{K}_{m-1}\boldsymbol{c} \tag{24}$$

is the Koopman operator represented through the data matrix inside the RKHS.

A Gaussian Process (GP) approach uses a Bayesian perspective to adjust the weight vector $\boldsymbol{c}$ by applyimg a criterion based on maximum likelihood rather than a minimum mean square error [9]. In general, this perspective provides with a solution expressed in terms of a probability distribution rather than a deterministic value.

4.2. Model inference

The GP assumes that r and $\boldsymbol{c}$ are both latent random variables for which a probabilistic model should be provided. For the $\boldsymbol{c}$, a Gaussian prior distribution is proposed with the form

$$p(\boldsymbol{c}) = \mathcal{N}(\boldsymbol{c}|\mathbf{0}, \Sigma_{\mathsf{c}}). \tag{25}$$

The model also assumes that r_m is a Gaussian random variable whose components are zero mean, independent and identically distributed (i.i.d.) with variance σ_n^2. Thus, the corresponding likelihood is a Gaussian distribution expressed as

$$p(r_m|\boldsymbol{c}, \mathcal{K}_{m-1}) = \mathcal{N}(r_m|0, \sigma_n^2 \boldsymbol{I}). \tag{26}$$

Therefore, φ_m is also a random variable whose mean is precisely $f(\mathcal{K}_{m-1})$, this is

$$p(\varphi_m|\boldsymbol{c}, \mathcal{K}_{m-1}) = \mathcal{N}(r_m|\mathcal{K}_{m-1}\boldsymbol{c}, \sigma_n^2 \boldsymbol{I}). \tag{27}$$

The goal of the GP method is to maximize this likelihood marginalized with respct to the posterior probability distribution of $\boldsymbol{c}$ with respect to the observations. This posterior can be found by virtue of the Bayes' rule as

$$p(\boldsymbol{c}|\mathcal{K}_{m-1}) \propto p(\boldsymbol{c}) p(\varphi_m|\boldsymbol{c}, \mathcal{K}_{m-1}). \tag{28}$$

Since this product is a product of Gaussians, the result is a square exponential, which needs to be further normalize to have probability distribution properties. The solution of this expression, given (25) and (27), is

$$p(\boldsymbol{c}|\mathcal{K}_{m-1}) = \mathcal{N}(\mathbf{c}|\bar{\boldsymbol{c}}, \mathsf{A}^{-1}) \tag{29}$$

where

$$\begin{aligned} \bar{\boldsymbol{c}} &= \sigma_n^{-2} \mathsf{A}^{-1} \mathcal{K}_{m-1}^\top \varphi_m, \\ \mathsf{A} &= \sigma_n^{-2} \mathcal{K}_{m-1}^\top \mathcal{K}_{m-1} + \Sigma_c^{-1}. \end{aligned} \tag{30}$$

Therefore, we can express $\bar{\boldsymbol{c}}$ and A in terms of the kernel matrix (21) and vector (22) as

$$\begin{aligned} \bar{\boldsymbol{c}} &= \left(\mathsf{H}_{m-1} + \sigma_n^2 \Sigma_c^{-1}\right)^{-1} \boldsymbol{h}_{m-1,m}, \\ \mathsf{A} &= \sigma_n^{-2} \mathsf{H}_{m-1} + \Sigma_c^{-1}. \end{aligned} \tag{31}$$

Since the distribution is a Gaussian, the mean is also the maximum a posteriori (MAP) of the distribution, and the covariance expresses, in particular, a confidence interval on the estimation of the parameters. The expression is computed as a function of kernel dot products, thus it still exists when the RKHS is of infinite dimension.

Assume now that the kernel matrix H_{m-1} in (31) is constructed using a kernel function $k(\cdot,\cdot)$ that depends on a set of parameters. The matrix A will thus depend on these parameters, plus the variance parameter σ_n^2. If the whole set of parameters is denoted as $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_J]^\top$, the optimization of the estimator can be performed by maximization of the posterior of $\boldsymbol{c}$ with respect to $\boldsymbol{\theta}$, this is

$$\boldsymbol{\theta}_{\text{MAP}} = \arg\max_{\boldsymbol{\theta}_{\text{MAP}}} \log p(\boldsymbol{c}|\mathcal{K}_{m-1}, \boldsymbol{\theta})|_{\boldsymbol{c}=\bar{\boldsymbol{c}}}. \tag{32}$$

By taking derivatives with respect to the parameters and equaling them to zero, we find

$$\left.\frac{d \log p(\boldsymbol{c}|\mathcal{K}_{m-1}, \boldsymbol{\theta})}{d\theta_j}\right|_{\boldsymbol{c}=\bar{\boldsymbol{c}}} = -\frac{1}{2}\text{tr}\left(\mathsf{A}\frac{d\mathsf{A}^{-1}}{d\theta_j}\right), \tag{33}$$

which can be used to maximize the log posterior with respect the parameters using a simple gradient descent.

Assuming that a new set of observations $\boldsymbol{x}_j^*$, $0 \leq j \leq m-1$, is available, then a prediction φ_m^* can be computed. The predictive distribution for this sample can be computed by marginalizing the distribution of the prediction $f^* = f(\varphi_m)$ with respect to the parameter vector $\boldsymbol{c}$. This marginalization produces a Gaussian distribution with mean $\bar{f}^*$ and covariance $\mathcal{C}_{f^*}$:

$$\begin{aligned} \bar{f}^* &= \mathcal{K}_{m-1}^* \boldsymbol{c}, \\ \mathcal{C}_{f^*} &= \mathcal{K}_{m-1}^* \mathsf{A}^{-1} \mathcal{K}_{m-1}^{*\top}. \end{aligned} \tag{34}$$

These two elements are a vector and a matrix that live in the RKHS, thus they are, in general, of infinite dimension. Nevertheless, they are expressed in a subspace of dimension m. The projection of the prediction in these space will have a covariance matrix that can be expressed as

$$\mathsf{C}_{f^*} = \mathsf{H}^* \mathsf{A}^{-1} \mathsf{H}^{*\top} \tag{35}$$

where $\mathsf{H}^* = \mathcal{K}_{m-1}^{*\top} \mathcal{K}_{m-1}$ is a matrix containing all dot products between the test data $\mathcal{K}_{m-1}^*$ and training data $\mathcal{K}_{m-1}$

4.3. A GP model for the prediction preimage

Still, the vector $\bar{f}^*$ lives into the RKHS. In order to recover a prediction into the original space, one can apply a preimage procedure, by projecting the prediction into the input space using a projection matrix. This is, for an arbitrary prediction f^*, a projection $\boldsymbol{x}$ into the input space can be written as

$$\boldsymbol{x}_m = \mathcal{W}^\top \bar{f}^* + \boldsymbol{e} \tag{36}$$

where $\boldsymbol{e}$ is the projection error. This error can be considered a Gaussian vector whose components have zero mean and variance σ_m^2, $1 \leq m \leq M$. Again, the matrix $\mathcal{W}$ can be considered a latent variable with a Gaussian prior, and then a posterior for this variable can be maximized. Moreover, it can be assumed that the RKHS has infinite dimension, and using the Generalized Representer Theorem [4, 5], this vector has a dual representation in terms of a linear combination of a training data, this is

$$\mathcal{W} = \mathcal{K}_{m-1} \mathsf{B} \tag{37}$$

where B is a matrix of coefficients of finite dimension, and $\mathcal{K}_{m-1}$ a set of (training) data into the RKHS. Thus, we have

$$\boldsymbol{x} = \mathsf{B}^\top \mathcal{K}_{m-1}^\top \bar{f}^* + \boldsymbol{e}. \tag{38}$$

The projection then becomes

$$\begin{aligned} \boldsymbol{x} &= \mathsf{B}^\top \mathcal{K}_{m-1}^\top \mathcal{K}_{m-1}^* \boldsymbol{c} + \boldsymbol{e} \\ &= \mathsf{B}^\top \mathsf{H}^{*\top} \boldsymbol{c} + \boldsymbol{e}. \end{aligned} \tag{39}$$

A multiple input multiple output Gaussian process scheme is presented in [14] that can be applied here directly. While in that work the approach is nonlinear, in this case, we only need to apply a particularization, where the projection is a linear one, since $\mathsf{H}^* \boldsymbol{c}$ is a finite dimension vector. The benefits of using this approach instead of a standard least squares preimage solution are based on the fact that the GP approach applies a prior over the parameters that regularizes the solution. The optimum regularization includes the estimation of the noise variances σ_m^2, which is performed by simply maximizing the log likelihood of the data with respect to these parameters.

5. Conclusion

Nonlinear system modeling using the Koopman operator can be straightforwardly applied using the so called "rather old kernel trick", which consists on a nonlinear transformation into a Hilbert space endowed with a kernel dot product. Such procedure provides the operator with nonlinear properties, and an arbitrary expressive capacity if the RKHS is of infinite dimension. This is the case when the kernel dot product is, for example, of the exponential family, for example, if a square exponential or a Matérn function is used (see e.g. [15]). In these cases, a regularization of the operator is preceptive. Indeed, if the space has infinite dimension, then the complexity of the estimator becomes arbitrary, and it is guaranteed to overfit [16]. In this short paper, we briefly present a nonlinear version of the operator, and then we propose the use of the well known Gaussian Processes (GP) for the parameter estimation. The GP method proposes the use of a Bayesian approach to the estimation of the parameters by using a prior distribution for them that, in turn, acts as a regularizer that minimizes the overfitting of the model.

The inference of the linear linear parameters is given as in the standard GP presented in [9]. The inference to find optimal values for the kernel and noise parameters can be done by maximizing the log posterior of the linear ones.

The prediction model is formulated into the RKHS, so a preimage technique is needed if the prediction needs to be expressed into the input space. In general, this problem does not have an exact solution. It can be ill posed easily lead to overfitting. Therefore, a regularized solution is needed.

We use a straightforward application of the Generalized represented theorem, we show that this task is performed by a simple linear transformation matrix, that can be, in turn, optimized using any linear algorithm. While procedures as the Ridge Regression [17] are straightforward and there are regularized, we propose the use of a MIMO or multitask GP model for the estimation, due to its optimal regularization capabilities, through the use of Bayesian inference to optimize the transformation. Besides, this approach produces a probabilistic model of the prediction that can be useful to estimate the confidence intervals of the prediction.

Acknowledgements

This work is supported in part by JSPS KAKENHI Grant #15H03964, The Specific Support Project of Osaka Prefecture University and grants TEC2014-52289-R and P2013ICE-2933 (Spanish Goverment).

References

[1] M. Budišić, R. Mohr, and I. Mezić, "Applied Koopmanism," *CHAOS*, vol. 22, p. 047510, December 2012.

[2] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," *Annu. Rev. Fluid Mech.*, vol. 45, pp. 357–378, January 2013.

[3] Y. Susuki, I. Mezić, F. Raak, and T. Hikihara, "Applied Koopman operator theory for power systems technology," *NOLTA, IEICE*, vol. 7, pp. 430–459, October 2016.

[4] G. Kimeldorf and G. Wahba, "Some results on tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, no. 1, pp. 82–95, 1971.

[5] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Computational learning theory*, pp. 416–426, Springer, 2001.

[6] Y. Kawahara, "Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis," in *Advances in Neural Information Processing Systems*, vol. 29, pp. 911–919, 2016.

[7] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition," *J. Nonlinear Sci.*, vol. 25, pp. 1307–1346, June 2015.

[8] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, "A kernel-based method for data-driven Koopman spectral analysis," *J. Comput. Dyn.*, vol. 2, pp. 247–265, December 2015.

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[10] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dynamics*, vol. 41, no. 1, pp. 309–325, 2005.

[11] 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, Dec 2009.

[12] A. Ruhe, "Rational Krylov sequence methods for eigenvalue computation," *Linear Algebra and Its Application*, vol. 58, pp. 391–405, 1984.

[13] M. A. Aizerman, E. M. Braverman, and L. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and remote Control*, vol. 25, pp. 821–837, 1964.

[14] E. V. Bonilla, K. M. Chai, and C. Williams, "Multitask gaussian process prediction," in *Advances in Neural Information Processing Systems 20* (J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, eds.), pp. 153–160, 2008.

[15] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, June 2004.

[16] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[17] C. Saunders, A. Gammerman, and V. Vovk, "Ridge regression learning algorithm in dual variables," in *Proceedings of the 15th International Conference on Machine Learning (ICML)*, (Madison, WI, USA), pp. 515–521, July 1998.