

# Data-Driven Bifurcation Analysis via Learning of Homeomorphism

Wentao Tang

WENTAO\_TANG@NCSTU.EDU

Department of Chemical and Biomolecular Engineering, North Carolina State University,  
Raleigh, NC 27695-7905, USA

**Editors:** A. Abate, K. Margellos, A. Papachristodoulou

## Abstract

This work proposes a data-driven approach for bifurcation analysis in nonlinear systems when the governing differential equations are not available. Specifically, *regularized regression with barrier terms* is used to learn a *homeomorphism* that transforms the underlying system to a reference linear dynamics – either an explicit reference model with desired qualitative behavior, or Koopman eigenfunctions that are identified from some system data under a reference parameter value. When such a homeomorphism fails to be constructed with low error, bifurcation phenomenon is detected. A case study is performed on a planar numerical example where a pitchfork bifurcation exists.

**Keywords:** Bifurcation, Nonlinear Dynamics, Conjugacy, Koopman Eigenfunctions

## 1. Introduction

Bifurcation, namely the dependence of qualitative properties on parameters and the loss of structural stability at critical parameter values, is a characteristic behavior of nonlinear dynamical systems (Strogatz, 2015), which often has important engineering consequences (e.g., runaway reactors) (Kevrekidis, 1987; Balakotaiah et al., 1999). This work focuses on *local bifurcations* near an equilibrium point, i.e., the change of local convergence behavior, instead of global/non-local features such as homoclinic or heteroclinic orbits and limit cycles. If the system model is available as ordinary differential equations (ODEs), then by constructing a homeomorphism (bijective and continuous mapping with continuous inverse) between the system model and a normal form, bifurcations can be classified into, e.g., fold, Andronov-Hopf, cusp, and Bogdanov-Takens types (Guckenheimer and Holmes, 2013; Kuznetsov, 2023). Numerical algorithms can be used for the continuation of ODE solutions to generate accurate bifurcation maps (Kuznetsov and Meijer, 2019).

However, an accurate first-principles model can be difficult to obtain in systems with detailed underlying mechanisms or too complex to be amenable to analysis. In such cases, a *data-driven bifurcation analysis* approach is needed to discover the occurrence of bifurcation when the parameter value varies, thus assisting the system design and control. In this work, we assume that accurate state measurements can be obtained (from noiseless simulations), and propose a *regression* approach to detect bifurcation under the following two use cases.

- In the first case, a reference model is supposed to be available a priori. Such a model has the desired local qualitative behavior of the system and is known to be topologically equivalent (conjugate) to the underlying system under a reference parameter value ( $\alpha_0$ ). Thus, by determining a homeomorphism that transforms the data to conform to this reference model, conclusion can be made on whether under the system bifurcates between the reference and the given parameter value ( $\alpha_1$ ).

- In the second case, a reference ODE model is unavailable. Instead, it is assumed that two datasets are available under  $\alpha_0$  and  $\alpha_1$ . Without the model, empirical Koopman eigenfunctions are assumed to have been identified under  $\alpha_0$ , which can be achieved via extended mode decomposition (Williams et al., 2015), deep learning (Lusch et al., 2018), or sparse regression (Kaiser et al., 2021). Hence, a homeomorphism is sought to transform the data under  $\alpha_1$  to conform to the linear dynamics of these Koopman eigenfunctions.

The homeomorphism learning problem is formulated as a convex optimization one, with regularization on the coefficient magnitude in the regression problem, as well as barrier terms to enforce the non-degeneracy of the learned homeomorphic mapping. When the resulting loss (error) metric is not sufficiently low, local bifurcation is detected. A threshold value for such a bifurcation detection is theoretically obtained in this paper; however, practically, the bifurcation should be considered to occur when the error starts to significantly increase.

The proposed work is related to the recent advances in using sparse regression for finding governing equations (Brunton et al., 2016), deep learning for solving dynamical systems (Karniadakis et al., 2021), as well as on Koopman operators (Mauroy et al., 2020; Otto and Rowley, 2021; Colbrook and Townsend, 2024) and their applications in control (Korda and Mezić, 2018b; Huang and Vaidya, 2022; Narasingam et al., 2023). We refer to Mezić (2020) for a profound investigation of the spectrum structure of Koopman operators, where it is noted that point spectrum information (eigenfunctions) are often sufficient<sup>1</sup>. This paper is in particular motivated by the work of Bollt et al. (2018) that focused on “matching” dynamics, namely the determination of a homeomorphism between two conjugate systems based on their Koopman eigenfunctions, as well as its extension in Redman et al. (2022) to determine the equivalence of numerical algorithms as dynamical systems.

## 2. Preliminaries and Problem Settings

Here we consider a parameterized continuous-time dynamical system in the form of

$$\dot{x}(t) = f(x(t); \alpha) =: f_\alpha(x(t)), \quad (1)$$

where  $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$  is the state vector and  $\alpha \in \mathcal{A} \subseteq \mathbb{R}^q$  represents the parameters. We assume that  $f$  is smooth to guarantee the existence and uniqueness of solution under any parameter value. Denote the flow of (1) by a single-parameter group  $S_{f_\alpha}^t$ , i.e.,  $x(t_0 + t) = S_{f_\alpha}^t(x(t_0))$ . To study the bifurcation behavior, we consider the conjugacy between nonlinear systems.

**Definition 1** *Two dynamical systems  $\dot{x} = f(x)$  and  $\dot{y} = g(y)$ , on  $\mathcal{X} \subseteq \mathbb{R}^n$  and  $\mathcal{Y} \subseteq \mathbb{R}^n$ , respectively, are said to be conjugate if there exists a homeomorphism  $h : \mathcal{X} \rightarrow \mathcal{Y}$ , such that the flows of these two systems, denoted as  $S_f^t$  and  $S_g^t$ , respectively, satisfy*

$$S_g^t \circ h = h \circ S_f^t. \quad (2)$$

In other words, two conjugate systems are identical via an invertible change of coordinates. Conjugacy is an equivalent relation between  $n$ -dimensional dynamical systems, i.e., if  $\dot{x} = f(x)$  is known to be conjugate to  $\dot{y} = g(y)$ , which is further conjugate to another  $\dot{z} = k(z)$ , then the first system must be conjugate to the last.

---

1. Specifically, it was shown that if a global stabilizing equilibrium exists, then Koopman eigenfunctions can be extended from a small local neighborhood of the equilibrium to the entire attraction basin, and that for limit cycle systems, any function that is  $L^2$  in the cyclic coordinate and analytical in the remaining coordinates can be expanded as a series of countably many Koopman eigenfunctions.

**Definition 2** *The parametric system (1) is said to bifurcate between parameter values  $\alpha_0$  and  $\alpha_1 \in \mathcal{A}$  if the system at  $\alpha_1$  is not conjugate to the system at  $\alpha_0$ .*

If the system (1) at a “reference parameter”  $\alpha_0$  is conjugate to an a priori “standard system”  $\dot{y} = g(y)$ , then to judge whether the system at  $\alpha_1$  is conjugate to the system at  $\alpha_0$ , it suffices to construct a homeomorphism from the system at  $\alpha_1$  and the “standard system”. In this work, we focus on the detection of local bifurcation. Without loss of generality, assume that the equilibrium around which we are interested in is the origin (namely  $f_\alpha(x) = 0$  for all  $\alpha \in \mathcal{A}$ ). Under mild conditions, the standard system  $\dot{y} = g(y)$  can be chosen as the linearization of  $\dot{x} = f_{\alpha_0}(x)$  at  $x = 0$ .

**Assumption 1** *Let  $A = \partial f_{\alpha_0}(0)/\partial x$  (the Jacobian matrix). We assume that the eigenvalues of  $A$ ,  $(\lambda_1, \dots, \lambda_n)$ , are of type  $(C, \nu)$  for some  $C, \nu > 0$ . That is, for any  $\lambda_k$  and  $(m_1, \dots, m_n) \in \mathbb{N}^n$  with  $\sum_{r=1}^n m_r \geq 2$ , the following inequality holds:  $|\lambda_k - \sum_{r=1}^n m_r \lambda_r| \geq C(\sum_{r=1}^n m_r)^{-\nu}$ .*

**Fact 1** (cf. [Arnold \(1988\)](#), §24) *Under the above assumption, the system (1) at  $\alpha_0$  is conjugate to the linearized system  $\dot{y} = Ay$ .*

Now our first problem in consideration is stated as follows.

**Problem 1** *Assume that the system  $\dot{y} = g(y) = Ay$  on  $\mathbb{R}^n$  is given. Suppose that at  $\alpha = \alpha_1$ , a dataset comprising of (sufficiently many) orbits, sampled at a frequency of  $1/\tau$ :*

$$\mathcal{D} = \left\{ \left( x_0^{(i)}, x_1^{(i)}, \dots, x_P^{(i)} \right) \mid i = 1, \dots, M \right\} \quad (3)$$

*is available, where  $x_k^{(i)} = S_{f_\alpha}^\tau(x_{k-1}^{(i)})$  for  $i = 1, \dots, M$  and  $k = 1, \dots, P$ . Learn a homeomorphism  $h : \mathcal{X} \rightarrow \mathcal{Y}$  such that the following equations approximately hold:*

$$h \left( x_k^{(i)} \right) = e^{A\tau} h \left( x_{k-1}^{(i)} \right), \quad i = 1, \dots, M, \quad k = 1, \dots, P. \quad (4)$$

However, it is not always realistic to assume that the model at  $\alpha_0$  is known so that the Jacobian  $A$  can be easily obtained. We consider a second setting where  $n$  eigenfunctions of the Koopman operators are known, or can be first identified from data, for the system at  $\alpha_0$ .

**Definition 3** *The Koopman operators, denoted as  $U_{f_\alpha}^t$ , for the parametric system (1), refer to the family of linear operators on the space of continuous functions on  $\mathcal{X}$ , such that*

$$U_{f_\alpha}^t \circ \psi = \psi \circ S_{f_\alpha}^t, \quad \forall \psi \in \mathcal{C}(\mathcal{X}). \quad (5)$$

*Here any  $\psi$  is called an observable. We say that  $\psi$  is an eigenfunction of the Koopman operators associated with (complex-valued) eigenvalue  $\lambda$ , if*

$$(U_{f_\alpha}^t \circ \psi)(x) = e^{\lambda t} \psi(x), \quad \forall x \in \mathcal{X}, \quad (6)$$

*and a generalized eigenfunction, if for some integer  $k \geq 1$ ,  $(d/dt - \lambda)^k \psi(x) = 0$ .*

In other words, the eigenfunction  $\psi$  is an observable whose value gains a factor of  $e^{\lambda\tau}$  as time proceeds by  $\tau$ . It is natural that if two dynamical systems  $\dot{x} = f(x)$  and  $\dot{y} = g(y)$  are conjugate through a homeomorphism  $h$  (i.e.,  $y = h(x)$ ), then from any eigenfunction  $\psi$  of the former system,  $\psi \circ h^{-1}$  specifies an eigenfunction of the latter system associated with the same eigenvalue. For the parametric system (1) at  $\alpha_0$ , if it is conjugate to a linear system  $\dot{y} = Ay$ , then the dynamics can be fully described by the eigenfunctions of the latter. For this linear system, the eigenfunctions are related to the Jordan canonical form of the matrix  $A$ .

**Fact 2 (cf. Mezić (2020), §3)** *If a set of functions  $\psi = (\psi_1, \dots, \psi_n)$  satisfy  $\dot{\psi}(x) = J\psi(x)$ , where  $J$  is the Jordan canonical form of  $A$ , then they are generalized Koopman eigenfunctions of the linear system  $\dot{y} = Ay$ . Moreover, these generalized eigenfunctions come from an invertible transform of  $y$ ; specifically  $\psi_i = w_i^\top y$  ( $i = 1, \dots, n$ ), with  $\{w_1, \dots, w_n\}$  being a basis of  $\mathbb{R}^n$ .*

Without knowing the model at  $\alpha_0$ , the eigenfunctions can be identified from data instead. A common approach to this end is extended dynamic mode decomposition (EDMD) (Williams et al., 2015), where a dictionary of basis functions on  $x$  is used to lift the states  $x$  into a high-dimensional space, and then a linear dynamics  $A$  is found via regression. Alternatively, instead of obtaining  $A$ , a sparse regression on dictionary functions can be performed under any given  $\lambda$  to determine a Koopman eigenfunction (Kaiser et al., 2021). It has been proved that for the Koopman operator estimated on a finite number  $N$  of basis functions and sample size  $M$ , as  $M \rightarrow \infty$ , it will converge to the optimal approximation, (Klus et al., 2016), and as  $N \rightarrow \infty$ , the finite-rank approximation will converge to the true Koopman operator in strong operator topology (Korda and Mezić, 2018a).

Hence, finding the Koopman eigenfunctions at  $\alpha = \alpha_1$  that evolve according to the same linear dynamics as the data-identified eigenfunctions at  $\alpha = \alpha_0$  serves as an approximation for detecting bifurcation. The problem we are interested in is follows.

**Problem 2** *Assume that the eigenfunctions  $\psi = (\psi_1, \dots, \psi_n)$  are known to satisfy  $\dot{\psi} = A\psi$  for a given  $A$  for the system (1) at  $\alpha = \alpha_0$ . Suppose that at  $\alpha_1$ , a (sufficiently large) dataset in the form of (3) is available. Learn a homeomorphism  $h : \mathcal{X} \rightarrow \mathcal{Y}$  such that (4) holds.*

Based on above discussions, detecting local bifurcation around a given equilibrium based on a known reference model  $\dot{x} = f_{\alpha_0}(x)$  and based on Koopman eigenfunctions  $(\psi_1, \dots, \psi_n)$  identified from data can have the same formulation. With Assumption 1, which allows the reference system to be conjugate to the linearized system, the only difference between the two problems is the assignment of matrix  $A$ . Now, focusing on (4), we aim at determining the homeomorphism  $h$  from data through a regression approach for both the *model-aware* Problem 1 and *model-free* Problem 2.

### 3. Proposed Approach

#### 3.1. Least Squares Formulation

Now we choose a basis  $\{\varphi_1, \dots, \varphi_N\}$  for the components of  $h$ , i.e., to approximate  $h$  by

$$\begin{bmatrix} h_1(x) & \cdots & h_n(x) \end{bmatrix} = \begin{bmatrix} \varphi_1(x) & \cdots & \varphi_N(x) \end{bmatrix} \begin{bmatrix} \theta_{11} & \cdots & \theta_{1n}(x) \\ \vdots & \ddots & \vdots \\ \theta_{N1} & \cdots & \theta_{Nn}(x) \end{bmatrix}, \quad (7)$$

namely  $h_j(x) = \sum_{k=1}^N \theta_{kj} \varphi_k(x)$  and  $h(x) = h_\Theta(x) := \Theta^\top \varphi(x)$ , where  $\Theta \in \mathbb{R}^{N \times n}$  is the coefficient matrix to be learned and  $\varphi = [\varphi_1, \dots, \varphi_N]^\top$ . Choosing a polynomial basis as  $\varphi$  is desirable, due to the approximability of any continuous function by polynomials in  $\mathcal{C}^\infty$  (Stone-Weierstrass theorem) as well as in Sobolev space (Meyers-Serrin theorem). The latter property is needed since we need  $h$  to be an invertible mapping, which boils down to constraining the Jacobian of  $h$  to be always non-singular (assuming that  $h$  is differentiable). With this parameterization, the error on the  $i$ -th sampled orbit  $(x_0^{(i)}, \dots, x_P^{(i)})$  resulted from the parameters  $\Theta$  is thus the collection of vectors

$$e_k^{(i)}(\Theta) = \Theta^\top \varphi \left( x_k^{(i)} \right) - e^{A\tau} \Theta^\top \varphi \left( x_{k-1}^{(i)} \right), \quad i = 1, \dots, M, \quad k = 1, \dots, P. \quad (8)$$

Denoting

$$\Phi_k = \begin{bmatrix} \vdots \\ \varphi(x_k^{(i)})^\top \\ \vdots \end{bmatrix} \in \mathbb{R}^{M \times N}, \quad E_k(\Theta) = \begin{bmatrix} \vdots \\ e_k^{(i)}(\Theta)^\top \\ \vdots \end{bmatrix} \in \mathbb{R}^{M \times n}, \quad (9)$$

and  $\exp(A\tau) = \bar{A}$ , then we have  $E_k(\Theta) = \Phi_k\Theta - \Phi_{k-1}\Theta\bar{A}^\top$  for  $k = 1, \dots, P$ . Hence, we can have a least square formulation of minimizing the total squared Frobenius norm of  $E_1, \dots, E_P$ , scaled by the length of prediction horizon  $P$  and sample size  $M$ :

$$\min_{\Theta} \frac{1}{2MP} \sum_{k=1}^P \|E_k(\Theta)\|_F^2 = \frac{1}{2MP} \sum_{k=1}^P \|\Phi_k\Theta - \Phi_{k-1}\Theta\bar{A}^\top\|_F^2. \quad (10)$$

### 3.2. Regularization and Barrier Terms

Here we use a Tikhonov regularization on the coefficients, with  $\beta$  being the regularization hyperparameter, in order to smooth the contours of  $h$  to mitigate generalization loss. (The  $\ell_2$ -regularization can be replaced by a LASSO term, which seems to be unnecessary as sparsity of non-zero coefficients in  $h$  does not have a particular meaning.) To enforce the invertibility of  $h_\Theta := \Theta^\top \varphi$ , we assume that all basis functions are differentiable (which is obviously true for polynomials) and require that the Jacobian matrix of  $h_\Theta$  evaluated at each  $x_0^{(i)}$  must have full rank. That is, by letting  $C^{(i)} = \partial\varphi(x_0^{(i)})/\partial x$ , we require that  $\Theta^\top C^{(i)}$  must not have zero singular values for all  $i = 1, \dots, M$ , i.e.,  $\Theta^\top C^{(i)} C^{(i)\top} \Theta \succ 0$ . To impose this positive definiteness constraint, we adopt a log-barrier term with barrier hyperparameter  $\mu > 0$ . This results in additional objective terms:  $-\mu \log \det(\Theta^\top C^{(i)} C^{(i)\top} \Theta)$  defined on all sample points  $i = 1, \dots, M$ .

However, the log-barrier term above, despite being convex, can be hardly handled by solvers (e.g., `cvxpy`) due to the quadratic dependence on  $\Theta$  in the log-determinant. For a practical treatment, in view of the relation  $\Theta^\top C^{(i)} C^{(i)\top} \Theta \succeq \Theta^\top C^{(i)} + C^{(i)\top} \Theta - I$ , we instead define the barrier terms:  $-\mu \log \det(\Theta^\top C^{(i)} + C^{(i)\top} \Theta - I)$ . It should also be noted that such barrier terms have the side effect of rewarding  $h$  to have a Jacobian with large determinant. To mitigate this effect, we force penalization term to be nonnegative. Now, the optimization problem to be solved is as follows.

$$\begin{aligned} \min_{\Theta \in \mathbb{R}^{N \times n}} J_\mu(\Theta) = & \frac{1}{2MP} \sum_{k=1}^P \|\Phi_k\Theta - \Phi_{k-1}\Theta\bar{A}^\top\|_F^2 + \frac{1}{2}\beta\|\Theta\|_F^2 \\ & + \frac{\mu}{M} \sum_{i=1}^M \max \left\{ 0, -\log \det \left( \Theta^\top C^{(i)} + C^{(i)\top} \Theta - I \right) \right\}. \end{aligned} \quad (11)$$

In this convex optimization formulation (11),  $\mu$  is considered as a tunable hyperparameter that gives a tradeoff between the mean squared error (MSE) plus the Tikhonov regularization cost versus the Jacobian magnitude. To select an optimal value  $\mu = \mu_*$ , we consider a *scaled MSE* after the optimization problem above on  $J_\mu(\Theta)$  is solved:

$$L_\mu := \frac{\left[ \frac{1}{MP} \sum_{k=1}^P \|\Phi_k\Theta - \Phi_{k-1}\Theta\bar{A}^\top\|_F^2 \right]^{1/2}}{\frac{1}{M} \sum_{i=1}^M |\det(\Theta^\top C^{(i)})|}, \quad (12)$$

in which the denominator is interpreted as a stochastic approximation of the measure of the mapped state space. Specifically, suppose that the sample of initial states  $\{x_0^{(i)} | i = 1, \dots, M\}$  are obtained from a probability distribution  $\rho$  on  $\mathcal{X}$ . Then the denominator is in fact

$$\frac{1}{M} \sum_{i=1}^M \left| \frac{\partial h(x_0^{(i)})}{\partial x} \right| \approx \int_{\mathcal{X}} \left| \frac{\partial h(x)}{\partial x} \right| d\rho(x) = \rho_*(\mathcal{Y}), \quad (13)$$

in which  $|\cdot|$  for a Jacobian matrix is used to denote the absolute value of its determinant and  $\rho_*$  is the push-forward of  $\rho$  onto  $\mathcal{Y}$  under  $h$ . In particular, if the sampling is uniform on a compact  $\mathcal{X}$ , then  $\rho_*(\mathcal{Y}) = \text{mes}(\mathcal{Y})/\text{mes}(\mathcal{X})$ , where  $\text{mes}$  is the Lebesgue measure. Hence, the optimal barrier hyperparameter  $\mu_*$  is chosen *a posteriori* according to the minimal  $L_\mu$  in (12).

Using the scaled MSE  $L_\mu$  as the posterior loss metric, the regularization parameter  $\lambda$  is also tunable. Specifically,  $K$ -fold cross-validation is used to evaluate the effect of regularization. That is, the entire dataset is first partitioned into  $K$  portions. Then, in each of the  $K$  parallel training experiments, one portion is used for the validation following the optimization on the remaining data. The geometric mean of the  $K$  generalization losses is taken as the validated  $L_\mu$  under the specific  $\lambda$  value. Furthermore, the richness of polynomial basis function, specified by the maximum allowed degree  $d$ , is also tuned according to the training loss and validation loss. The sample size  $M$  should be moderately large, but not excessive so as to reduce the computational time for optimization. The same principle applies to the selection of prediction horizon  $P$  and sampling frequency  $1/\tau$ .

### 3.3. Generalization Loss

In this subsection, we aim to establish theoretical probabilistic upper bounds on the generalized MSE and cumulative prediction error, supposing that bifurcation does not occur, i.e., the actual dynamics can be matched to  $A$ . Now let us denote by  $L_*$  the posterior minimized  $L_\mu$  value at  $\mu_*$ ,  $\Theta_*$  the optimized parameter value under  $\mu_*$ , and  $h_* = \Theta_*^\top \varphi$  the corresponding homeomorphic mapping learned. According to (12), we have

$$\frac{1}{MP} \sum_{k=1}^P \|\Phi_k \Theta - \Phi_{k-1} \Theta A^\top\|_F^2 \leq L_*^2 \left[ \frac{1}{M} \sum_{i=1}^M \left| \det(\Theta^\top C^{(i)}) \right| \right]^2, \quad (14)$$

which boils down to

$$\frac{1}{MP} \sum_{i=1}^M \sum_{k=1}^P \|h(x_k^{(i)}) - \bar{A}h(x_{k-1}^{(i)})\|^2 \leq L_*^2 \left[ \frac{1}{M} \sum_{i=1}^M \left| \frac{\partial h(x_0^{(i)})}{\partial x} \right| \right]^2. \quad (15)$$

**Assumption 2** The initial states  $\{x_0^{(i)}\}_{i=1}^M$  are independently sampled from a bounded  $\mathcal{X}$ .

Since  $\mathcal{X}$  is bounded, according to Hoeffding's inequality, with confidence  $1 - \delta_1$ , we have

$$\mathbb{E} \left[ \frac{1}{P} \sum_{k=1}^P \left\| (h_* \circ S_{f_\alpha}^{k\tau})(x) - \bar{A} (h_* \circ S_{f_\alpha}^{(k-1)\tau})(x) \right\|^2 \right] \leq \text{L.H.S. of (15)} + F_* \sqrt{\frac{\log(1/\delta_1)}{2M}}, \quad (16)$$

and with confidence  $1 - \delta_2$ , we have

$$\frac{1}{M} \sum_{i=1}^M \left| \frac{\partial h_*(x^{(i)})}{\partial x} \right| \leq \mathbb{E} \left[ \left| \frac{\partial h_*(x)}{\partial x} \right| \right] + H_* \sqrt{\frac{\log(1/\delta_2)}{2M}}, \quad (17)$$

where the expectation is on the population  $x \sim \rho$  and

$$F_* := \sup_{x \in \mathcal{X}} \frac{1}{P} \sum_{k=1}^P \left\| \left( h_* \circ S_{f_\alpha}^{k\tau} \right) (x) - \bar{A} \left( h_* \circ S_{f_\alpha}^{(k-1)\tau} \right) (x) \right\|^2, \quad H_* := \sup_{x \in \mathcal{X}} \left| \frac{\partial h_*(x)}{\partial x} \right|.$$

Thus, with confidence  $1 - \delta$  where  $\delta = \delta_1 + \delta_2$ , we have for the system at  $\alpha$  the following estimation of the generalization error on matching the target linear dynamics  $A$ :

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{P} \sum_{k=1}^P \left\| \left( h_* \circ S_{f_\alpha}^{k\tau} \right) (x) - \bar{A} \left( h_* \circ S_{f_\alpha}^{(k-1)\tau} \right) (x) \right\|^2 \right] \leq \\ L_*^2 \left\{ \mathbb{E} \left[ \left| \frac{\partial h_*(x)}{\partial x} \right| \right] + H_* \sqrt{\frac{\log(1/\delta_2)}{2M}} \right\}^2 + F_* \sqrt{\frac{\log(1/\delta_1)}{2M}}. \end{aligned} \quad (18)$$

Therefore, we have the following conclusion.

**Theorem 4** *Under the previous assumptions, if the system (1) does not bifurcate between  $\alpha_0$  and  $\alpha$ , then when  $M$  is sufficiently large, with confidence  $1 - \delta$ ,*

$$\mathbb{E} \left[ \frac{1}{P} \sum_{k=1}^P \left\| \left( h_* \circ S_{f_\alpha}^{k\tau} - \bar{A} h_* \circ S_{f_\alpha}^{(k-1)\tau} \right) (x) \right\|^2 \right] \leq L_*^2 \mathbb{E} \left[ \left| \frac{\partial h_*(x)}{\partial x} \right| \right]^2 + \mathcal{O} \left( \sqrt{\frac{\log(1/\delta)}{M}} \right). \quad (19)$$

In particular, we consider the case when  $A$  is Hurwitz, so that  $\|\bar{A}\| \leq \alpha$  for some  $\alpha \in (0, 1)$ . In this case, the expectation term on the left-hand side in the above theorem can be converted into the *cumulative squared prediction error* starting from the initial state  $x$ , with the coefficient  $C_\alpha$  being the  $\ell_2$ -gain of the discrete-time linear system  $\bar{A}$ , which depends on  $\alpha$ . It is possible to further bound the expected continuous-time cumulative prediction error, which is not detailed here.

**Corollary 5** *If  $\|\bar{A}\| \leq \alpha < 1$ , then under the previous assumptions, if the system (1) does not bifurcate between  $\alpha_0$  and  $\alpha$ , then at large  $M$ , with confidence  $1 - \delta$ ,*

$$\mathbb{E} \left[ \frac{1}{P} \sum_{k=1}^P \left\| \left( h_* \circ S_{f_\alpha}^{k\tau} - \bar{A}^k h_* \right) (x) \right\|^2 \right] \leq C_\alpha L_*^2 \mathbb{E} \left[ \left| \frac{\partial h_*(x)}{\partial x} \right| \right]^2 + \mathcal{O} \left( \sqrt{\frac{\ln(1/\delta)}{M}} \right). \quad (20)$$

Clearly, the generalization loss mainly depends on the optimized loss  $L_*$ , which is a scaled MSE. The magnitude of this quantity is therefore of interest. Indeed, if bifurcation does not occur and  $h_*$  can be obtained from a dense infinite-dimensional subspace of all continuous functions on  $\mathcal{X}$ , then  $L_*$  can approach 0 with a ground-truth solution. However, as the components of  $h$  are restricted to polynomials of degree not exceeding  $d$ , there exists an optimal parameter choice  $\Theta_*$  and a corresponding  $h_* = h_{\Theta_*}$  to achieve the lowest possible error

$$\epsilon_d^2 := \min_{\Theta} \mathbb{E} \left[ \frac{1}{P} \sum_{k=1}^P \left\| \left( h \circ S_{f_\alpha}^{k\tau} - \bar{A} h \circ S_{f_\alpha}^{(k-1)\tau} \right) (x) \right\|^2 \right] \quad (21)$$

among all such degree-constrained  $h$  that makes  $|\partial h_*(x)/\partial x| \geq 1$  almost everywhere. Hence for this specific feasible solution,  $h_*$ , the barrier term in the objective function (11) is zero and thus



independent of  $\mu$ . Denote this feasible objective as  $J_o$ . Hence, at any given  $\mu$ , after solving (11) to the optimum, the objective  $J_\mu$  and the corresponding solution  $\Theta_\mu$  must be such that

$$\frac{1}{2MP} \sum_{k=1}^P \|\Phi_k \Theta_\mu - \Phi_{k-1} \Theta_\mu \bar{A}^\top\|_F^2 + \frac{\mu}{M} \sum_{i=1}^M \max \left\{ 0, -\ln \det \left( \Theta_\mu^\top C^{(i)} + C^{(i)\top} \Theta_\mu - I \right) \right\} \leq J_o. \quad (22)$$

It follows, by relaxing the first summation term, that

$$\frac{1}{M} \sum_{i=1}^M \left[ 1 - |\det(\Theta_\mu^\top C^{(i)})| \right] \leq \frac{1}{2\mu} J_o \quad \stackrel{(12)}{\Rightarrow} \quad L(\mu) \leq \frac{\sqrt{2J_o}}{1 - \frac{1}{2\mu} J_o}, \quad (23)$$

and thus, by taking the infimum over  $\mu > 0$ ,  $L_*^2 \leq 2J_o$ . Similar to the previous discussions, we give a probabilistic upper bound on the right-hand side,  $2J_o$ , which is the MSE reached by  $\Theta_o$ . That is, with confidence  $1 - \delta$ ,

$$L_*^2 \leq \epsilon_d^2 + F_d \sqrt{\frac{\ln(1/\delta)}{2M}} + \beta \|\Theta_o\|_F^2, \quad (24)$$

where

$$F_d := \sup_{x \in \mathcal{X}} \frac{1}{P} \sum_{k=1}^P \left\| \left( h_o \circ S_{f_\alpha}^{k\tau} - \bar{A} h_o \circ S_{f_\alpha}^{(k-1)\tau} \right) (x) \right\|^2.$$

**Theorem 6** *By solving the barrier optimization problem (11) and tuning the barrier constant  $\mu$  according to the criterion (12), the learned mapping  $h$  has the following generalization loss on the linear dynamics with confidence  $1 - \delta$ :*

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{P} \sum_{k=1}^P \left\| \left( h_* \circ S_{f_\alpha}^{k\tau} - \bar{A} h_* \circ S_{f_\alpha}^{(k-1)\tau} \right) (x) \right\|^2 \right] \leq \\ \left[ \epsilon_d^2 + F_d \sqrt{\frac{\ln(1/\delta)}{2M}} + \beta \|\Theta_o\|_F^2 \right] \mathbb{E} \left[ \left\| \frac{\partial h_*(x)}{\partial x} \right\| \right] + \mathcal{O} \left( \sqrt{\frac{\ln(1/\delta)}{M}} \right). \end{aligned} \quad (25)$$

In the sample size limit ( $M \rightarrow \infty$ ), the linearization error is caused by  $\epsilon_d$  and  $F_d$  (which should converge to 0 as  $d \rightarrow \infty$ ) as well as  $\beta$ , the regularization parameter. Similar to the Corollary 5, if  $A$  is stable, then the left-hand side above can be replaced by the expected cumulative prediction error.

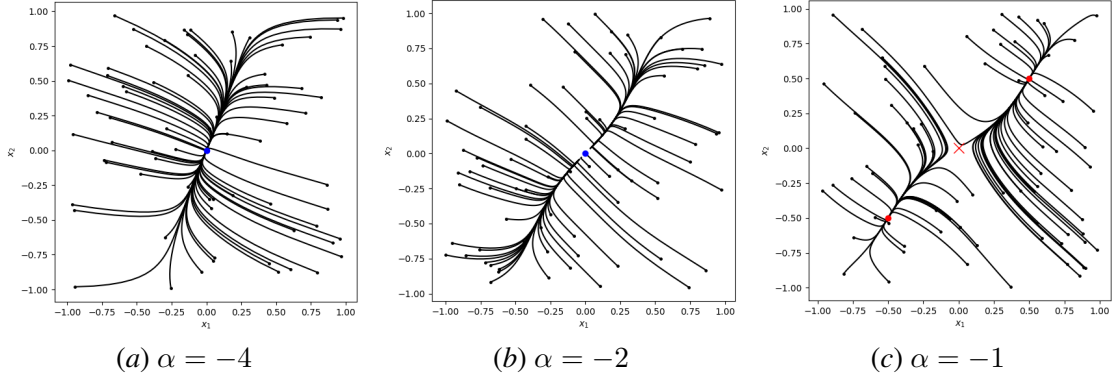
#### 4. Numerical Case Study

Now we consider the following planar system example (Example 8.1.3, [Strogatz \(2015\)](#)):

$$\dot{x}_1 = \alpha x_1 + x_2 + (1/2\pi) \sin(2\pi x_1), \quad \dot{x}_2 = x_1 - x_2. \quad (26)$$

When  $\alpha < -2$ , the origin is a stable equilibrium; when  $\alpha > -2$ , the origin becomes a saddle and two other stable equilibria appear. This is known as a supercritical pitchfork bifurcation. At  $\alpha_0 = -4$ , the linearized system at the origin gives two negative eigenvalues:  $-2 \pm \sqrt{3}$ . We will examine the existence of local bifurcation as  $\alpha_1$  sweeps from  $-4$  to  $-1$ . The phase diagram for  $\alpha = -4, -2$ , and  $-1$  are shown in Fig. 1. In the model-aware setting, bifurcation detection aims at constructing  $h$  to match the linearized dynamics with  $A = \text{diag}(-2 - \sqrt{3}, -2 + \sqrt{3})$ .




 Figure 1: Phase diagram at different  $\alpha$  values.

Assuming that the model is unavailable, through a data-driven approach (Kaiser et al., 2021), we may seek eigenfunctions at any given eigenvalue  $\lambda$  (using polynomials of degree up to 5). Such eigenvalues should be chosen at the values where the linear dynamics show a minimal RMSE (root mean square error). From Fig. 2(a), where  $\lambda$  ranges from  $-5$  to  $-0.1$ , it is seen that the RMSE has four local minima: at  $\lambda_1 = -3.4995$ ,  $\lambda_2 = -0.5970$ , and two other points in between. We choose the eigenvalues at the two ends that best avoid dependence between  $\psi_1$  and  $\psi_2$ , as illustrated in Figs. 2(b) and 2(c). The convergence of  $(\psi_1, \psi_2)$  to zero explains the attraction of states to the origin in two dimensions as shown in Fig. 1(a). In the model-free setting, bifurcation detection needs to match the Koopman-linearized dynamics with  $A = \text{diag}(-3.4995, -0.5970)$ .

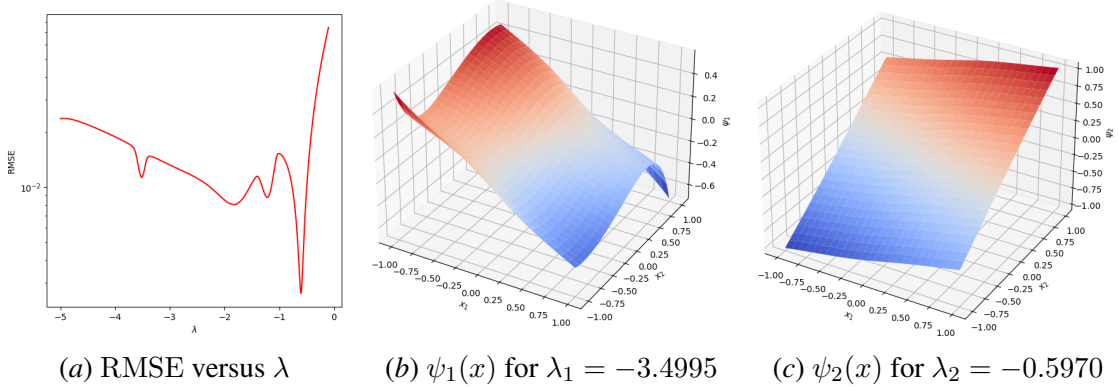


Figure 2: Koopman eigenfunctions constructed from data by varying eigenvalue.

The barrier constant  $\mu$ , regularization constant  $\beta$ , and prediction horizon  $P$  for bifurcation detection are tuned according to the MSE term (red circles), regularization term (red triangles) and cross-validated  $L_*^2$  value (blue squares), as shown in Fig. 3. The tuning is performed under  $\alpha = \alpha_0 = -4$ . The following hyperparameters are chosen:  $\mu = 10^{-3}$  (when  $L_*$  has almost settled on a minimal value) and  $\beta = 10^{-4}$  (when the regularization term is slightly less than the MSE term). For the choice of  $P$ , since  $\tau = 1/10$ , it is noted that when  $P = 12$  (i.e., with a prediction horizon of 1.2 time units), the MSE term and regularization term become close. When  $P > 20$ , the effect of  $P$  becomes small – while the accuracy is improved, longer computational time is needed for high  $P$ . Hence,  $P = 5, 10, 15, 20$ , and  $25$  will be used for bifurcation detection.

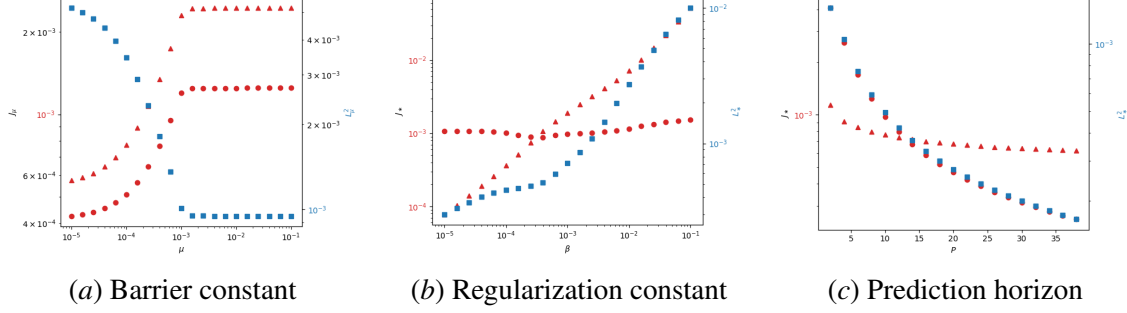
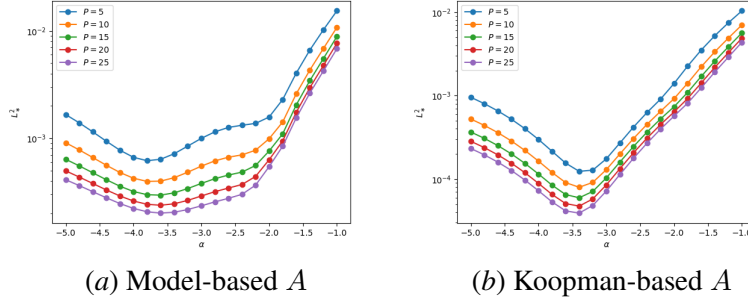


Figure 3: Hyperparameter tuning.

Using the proposed method, we sample  $M = 100$  initial points in  $\mathcal{X} = [-1, 1] \times [-1, 1]$  and perform the homeomorphism learning under varying values of  $\alpha_1$  from  $-5$  to  $-1$  with an increment of  $0.2$ . Both model-aware and model-free (Koopman eigenfunction-based) settings are considered. The dependence of loss  $L_*^2$  on system parameter  $\alpha$  is shown in Fig. 4(a) and 4(b), respectively, for these two cases. When  $\alpha > -2$ ,  $L_*^2$  increases exponentially with  $\alpha$ , reflecting the occurrence of bifurcation. However, in the model-free setting, such an exponential growth is triggered before the bifurcation actually takes place, and the value of  $L_*^2$  is sensitive to  $\alpha$  even without bifurcation. In contrast, in the model-based setting, the behavior before and after bifurcation are highly different, demonstrating that for local bifurcation detection, if the local linearization  $A = \partial f_{\alpha_0}(0)/\partial x$  is available, this  $A$  matrix is a better linear dynamics to be matched. On the other hand, the lower value of  $L_*^2$  in the model-free case explains that when using Koopman eigenfunctions, one can obtain better linearization in the provided region  $\mathcal{X}$  under the reference parameter value  $\alpha_0 = -4$ . But such a Koopman linearization appears to be less robust to parameter variations.

Figure 4: Bifurcation detection as the system parameter  $\alpha$  varies.

## 5. Conclusion

In this paper, a data-driven method of determining local bifurcation is proposed, where it is assumed that a linear model is available at the reference system parameter value, either by taking the Jacobian of the model or by identifying the Koopman eigenfunctions. Specifically, in the absence of bifurcation, a homeomorphism (linearly parameterized) that makes the target system conjugate to the reference system is learned through solving a regularized regression problem with convex barrier terms that enforces invertibility. Thus, bifurcation is detected when the homeomorphism can not be learned with sufficiently low error. Further research will investigate the approaches for non-local bifurcations related to limit cycles, and homoclinic or heteroclinic orbits.

## Acknowledgments

This work is supported by the faculty startup fund from NC State University. The author would also like to thank Prof. Joseph Sang-II Kwon (Texas A&M University) and Prof. Xunyuanyun Yin (Nanyang Technological University) for very beneficial discussions. During the 2022 AIChE Annual Meeting in Phoenix, Arizona, Prof. Ioannis G. Kevrekidis at Johns Hopkins University raised a question to Xunyuanyun on EDMD, which stimulated me in the audience to develop the idea in this paper.

## References

- Vladimir Igorevich Arnold. *Geometrical methods in the theory of ordinary differential equations*. Springer, 2<sup>nd</sup> edition, 1988.
- Vemuri Balakotaiah, Sandra MS Dommeti, and Nikunj Gupta. Bifurcation analysis of chemical reactors and reacting flows. *Chaos*, 9(1):13–35, 1999.
- Erik M Bollt, Qianxiao Li, Felix Dietrich, and Ioannis Kevrekidis. On matching, and even rectifying, dynamical systems through koopman operator eigenfunctions. *SIAM Journal on Applied Dynamical Systems*, 17(2):1925–1960, 2018.
- Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy Sciences*, 113(15):3932–3937, 2016.
- Matthew J Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *Communications on Pure and Applied Mathematics*, 77(1):221–283, 2024.
- John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer, 2013.
- Bowen Huang and Umesh Vaidya. A convex approach to data-driven optimal control via Perron-Frobenius and Koopman operators. *IEEE Transactions on Automatic Control*, 67(9):4778–4785, 2022.
- Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of Koopman eigenfunctions for control. *Machine Learning: Science and Technology*, 2(3):035023, 2021.
- George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- Ioannis G Kevrekidis. A numerical study of global bifurcations in chemical dynamics. *AIChE Journal*, 33(11):1850–1864, 1987.
- Stefan Klus, Peter Koltai, and Christof Schütte. On the numerical approximation of the Perron-Frobenius and Koopman operator. *Journal of Computational Dynamics*, 3(1):51–77, 2016.
- Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the Koopman operator. *Journal of Nonlinear Science*, 28:687–710, 2018a.

- Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018b.
- Yuri A Kuznetsov. *Elements of applied bifurcation theory*. Springer, 4<sup>th</sup> edition, 2023.
- Yuri A Kuznetsov and Hil G E Meijer. *Numerical bifurcation analysis of maps*. Cambridge University Press, 2019.
- Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.
- Alexandre Mauroy, Y Susuki, and Igor Mezić. *Koopman operator in systems and control*. Springer, 2020.
- Igor Mezić. Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry. *Journal of Nonlinear Science*, 30(5):2091–2145, 2020.
- Abhinav Narasingam, Sang Hwan Son, and Joseph Sang-II Kwon. Data-driven feedback stabilisation of nonlinear systems: Koopman-based model predictive control. *International Journal of Control*, 96(3):770–781, 2023.
- Samuel E Otto and Clarence W Rowley. Koopman operators for estimation and control of dynamical systems. *Annual Review on Control, Robotics, and Autonomous Systems*, 4:59–87, 2021.
- William T Redman, Maria Fonoberova, Ryan Mohr, Ioannis G Kevrekidis, and Igor Mezić. Algorithmic (semi-) conjugacy via Koopman operator theory. In *2022 IEEE 61<sup>st</sup> Conference on Decision and Control (CDC)*, pages 6006–6011. IEEE, 2022.
- Steven H Strogatz. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. CRC Press, 2<sup>nd</sup> edition, 2015.
- Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.