Available online at www.sciencedirect.com
ScienceDirect
Physica D 227 (2007) 78–99
PHYSICA D
www.elsevier.com/locate/physd

Modeling and nonlinear parameter estimation with Kronecker product representation for coupled oscillators and spatiotemporal systems

Chen Yao*, Erik M. Bollt

Department of Math and Computer Science, Clarkson University, Potsdam, NY 13699-5815, USA

Received 8 September 2005; received in revised form 8 November 2006; accepted 5 December 2006
Available online 21 December 2006
Communicated by J.E. Pearson

Abstract

We focus here on the question of global modeling, that is, building an ordinary differential equation (ODE) of minimal dimensions which models a given multivariate time dependent data-set. We give proof of convergence of the given method. We will furthermore introduce a new method which adapts the least-squares best approximation by a Kronecker product representation [C.F. Van Loan, N. Pitsianis, Approximation with Kronecker products. www.math.tkk.fi/ mhuhtane/kronecker.ps], to analyze underlying structure when it exists as a system of coupled oscillators. Several examples are presented including diffusively coupled chaotic systems. Some interesting prospective problems will be briefly discussed in the conclusions.

Keywords: Modeling; Nonlinear parameter estimation; Kronecker product; Coupled oscillators; Spatiotemporal systems

1. Introduction

In this paper, we address the question of global modeling which we define to be the following: given multivariate time-series, can we produce a set of ordinary differential equations of minimal dimension which closely reproduces the data? Consider for example a data-set which comes from a digital movie of a spatiotemporal, and perhaps chaotic, process such as a reaction diffusion system in the plane. We concern ourselves here with producing a minimal set of ODEs to model such data during the duration of the movie. While it is known that any the solution of any dissipative system can be arbitrarily well approximated by the solution of a three-dimensional system [1], we will take a constructive point of view. A common situation in multivariate dynamical data-sets is for there to exist an underlying structure, by which we mean that there is a repetition apparent in the underlying equations, with coupling between the repeated elements. Of course, the fitted version of the equations with inherent modeling errors can make it difficult to find underlying structure by naked-eye inspection. We will introduce here a new method to uncover underlying structure, when it exists, in terms of an optimal Kronecker product representation, by adapting the work of Van Loan and Pitsianis [19] to the dynamical systems setting.

Modeling time-series and multivariate time-series obtained from measurements of a dynamical system has long been a fundamental problem presenting itself to the experimental scientists. If we have a reliable model of such a system, then further work such as short-time prediction and control can proceed. Predicting the future evolution of dynamical systems has been a main goal of scientific modeling for centuries. The classic approach has been to build a global model, based on fundamental laws, yielding a differential equation which describes the motion of states. "This requires strong assumptions. A good fit of the data to the model validates the assumptions", [21]. Weigenbend and Gershenfeld make a distinction between weak modeling

* Corresponding author.
E-mail address: ycdecember@yahoo.com.cn (C. Yao).

0167-2789/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2006.12.006

(data rich and theory poor) and strong modeling (data poor and theory rich). This is related to, "...the distinction between memorization and generalization...". In this framework, the current work should be cast as producing a global but weak model.

Given a time-series from a "chaotic" dynamical system, data-only local analysis, using the methods of embedding and attractor reconstruction has become routine [22,21,23,7] based on the Taken's embedding theorem, [24]. Local methods to construct the model of systems have been developed by many authors, including, [7,8,6,14,4]. These methods are called local because a model is fitted for each neighborhood of the phase space, making an atlas of models. They are furthermore data-driven or weak models.

On the other hand, many researchers have developed techniques to (globally) model continuous time systems with ODEs [6, 5,9,11] by treating each data point as an initial value problem for the ODE. An important question associated with many of the modeling techniques used by previous researchers is how good the modeling method is. Some researchers consider the question of truncating the model [6,16]. They assume that the vector field of the dynamical system can be modeled by a series expansion in some basis,

$$F(X, P) = \lim_{N\to\infty} \sum_{k=0}^{N} p^k \phi^k(X). \tag{1}$$

For practical computational reasons, the series must be truncated. Interestingly, Brown, et al. [6] use a synchronization criterion as a nontrivial method to validate their models, by requiring that the model should be able to synchronize identically to the experimental data.

These works, and the problem we study here, fall into a class of problems called parameter estimation, which is related to, but not the same as, the problem of filtering whereby data is incorporated recursively into an evolving model. The now classic Kalman filtering method [20] is the best known filtering method for optimally estimating linear models, and the Kalman method can be readily extended to filtering models consisting of linear combinations of a general basis [25]. There have also been works to analyze the convergence of the extended Kalman filter, including a paper by Krener [12]. In that paper, the author demonstrates that if the initial estimation error of the filter is not too large, the error will tend to go to zero exponentially as time goes to infinity.

In this paper we will consider modeling unknown dynamical systems that are generally assumed to be governed by global autonomous ordinary differential equations (ODEs),

$$\mathrm{d}X/\mathrm{d}t = F(X, P). \tag{2}$$

We will assume the vector fields $F(X, P)$ described in terms of a series expansion in a well chosen basis set of functions,

$$F(X, P) = \sum_{k=0}^{\infty} p_k \phi_k(X). \tag{3}$$

We choose here the Taylor polynomials as the basis functions. We thoroughly analyze convergence properties of the explicit Euler method for our integration method, but we also discuss and produce numerical evidence for higher-ordered schemes. We prove that when the forward Euler integration is applied, the error between the modeling parameters and the exact parameters is in the order of the sampling rate. Supporting this statement, we prove here several useful lemmas followed by the convergence theorem of the parameters with respect to the time step size.

Once we have obtained our best model, or specifically the parameters of the assumed general model form, we furthermore show here a new method to uncover underlying structure of the equations. How can we know if there are coupled oscillators in the unknown system by investigating the modeling parameters? Based on a recent matrix approximation theory by Van Loan [19], we introduce an optimal estimation within the classes of systems which can be written in terms of the Kronecker product [18]; our algorithm uncovers subsystem structure in a large system consisting of many coupled simpler elements.

Finally, we will discuss the important issue of model quality. How should we evaluate the error between the original unknown system and the modeling system? We ask, how well does the model reproduce the data? In this regard, we will discuss how an exact system is expected to compare to a slightly perturbed system. The perturbation parameters can be added to the ODEs of the exact system and the initial conditions. After adding the perturbation parameter, we compare the time-series of two systems to see how the error between those two systems is bounded. Since we know that even if we had estimated the model exceedingly close to the original system, that sensitive dependence on initial conditions would cause exponential divergence of a prediction from a true solution. We incorporate this notion into our evaluation of quality.

The contents of this paper are outlined as follows. In Section 2, the modeling technique is given in an introductory example, and the convergence theorems are presented. In Section 3, the definition and properties of Kronecker products are given, and the approach to check the existence of a good Kronecker product representation is discussed. In Section 4, we evaluate the error between the original system and the modeling system by discussing the perturbed system. In Section 5, we give an experimental example to

show how to verify the existence of oscillators in an unknown system by investigating the time-series data from this system, and some prospective problems are discussed briefly. Conclusions are given in Section 6.

2. Nonlinear parameter estimation

In this section, we review the main technique we used to obtain approximations to the system parameters P. Suppose the experimental multi-variate time-series data $U(t)$ is obtained from an unknown dynamical system. Assuming no precedent information about the exact ODEs of the system, we write

$$\frac{\mathrm{d}\overrightarrow{X}}{\mathrm{d}t} = F(\overrightarrow{X}, P), \tag{4}$$

where $\overrightarrow{X}(t) \in R^N$, $N \gg 1$ is the phase-space dimension, and P denotes a set of system parameters. We model time-series data with expected ODEs using an explicit Euler integration approach, but we discuss higher-ordered methods in subsequent sections. This technique models the time-series data via

$$y^{(n+1)} = y^{(n)} + h_t F(y^{(n)}, P), \tag{5}$$

where $y^{(n)} = y(nh_t)$. Given this equation, the modeling procedure is reduced to finding a vector field $F(X, P)$, which can be written as a series expansion in terms of appropriate basis functions:

$$F(X, P) = \sum_{k=0}^{\infty} p_k \phi_k(X), \tag{6}$$

where $\{\phi_k(X)\}$ is a set of basis functions and the p_k are parameters that will be determined by the modeling procedure. In this paper, the basis functions used in the modeling procedure are polynomials. In the first part of this section, we give a simple introductory example, the system of N coupled Rossler oscillators, to present this technique. In the second part, we give the proof of convergence of the approximation parameters to analyze the quality of the approximation.

2.1. The modeling procedure by example

We consider the system of N coupled Rossler oscillators [2] with periodic boundary conditions:

$$\begin{aligned} \mathrm{d}x_i/\mathrm{d}t &= -\omega_i y_i - z_i + \epsilon(x_{i+1} + x_{i-1} - 2x_i) \\ \mathrm{d}y_i/\mathrm{d}t &= \omega_i x_i + a y_i \\ \mathrm{d}z_i/\mathrm{d}t &= b + z_i(x_i - c), \end{aligned} \tag{7}$$

where ω_i $(i = 1, \ldots, N)$ is the mean frequency of the ith oscillator, ϵ is the coupling parameter, and a, b, and c are the parameters of the individual Rossler oscillator. We choose $a = 0.165, b = 0.2, c = 10$ and $\omega_i = 1$ so that each oscillator exhibits a chaotic attractor when initial conditions are chosen from the basin of attraction [2]. Rossler equations are particularly simple and have only one nonlinear component, therefore providing us with a better opportunity to show the basic steps. In this part, for the purpose of illustration, we choose the relatively small array, $N = 3$. We let

$$\overrightarrow{X} = [x_1, y_1, z_1, \ldots, x_3, y_3, z_3]^{\mathrm{T}}. \tag{8}$$

We produce synthetic test data, by integration of the system by a fourth-order Runge–Kutta algorithm. We obtain a numerical multivariate time-series $\{\overrightarrow{X}^{(i)}\}$ that represents the 3 coupled system. With an assumption that the numerical time-series is all the information we have about the motion of the system, this time-series is expected to reproduce the original system Eq. (4) or at least systems approximating to it. Before we do the modeling, we investigate the original system to get some *a priori* expectation about the modeling method results.

Rewriting these 3 coupled Rossler oscillators, we express the equations in the Eq. (1) form:

$$\dot{\overrightarrow{X}} = P_1 \overrightarrow{X} + P_2 \begin{bmatrix} x_1 z_1 \\ x_2 z_2 \\ x_3 z_3 \end{bmatrix} + P_3, \tag{9}$$

where

$$P_1 = \begin{bmatrix} -2\epsilon & -1 & -1 & \epsilon & 0 & 0 & \epsilon & 0 & 0 \\ 1 & a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c & 0 & 0 & 0 & 0 & 0 & 0 \\ \epsilon & 0 & 0 & -2\epsilon & -1 & -1 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -c & 0 & 0 & 0 \\ \epsilon & 0 & 0 & \epsilon & 0 & 0 & -2\epsilon & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 \\ 0 \\ b \\ 0 \\ 0 \\ b \\ 0 \\ 0 \\ b \end{bmatrix}. \tag{10}$$

Hence, the basis functions of Eq. (9) are constant function, linear function and quadratic function. The parameters of this system are P_1, P_2 and P_3. Therefore for this nonlinear system, the polynomials are suitable to be used as basis functions. Combining the basis functions into one vector, we have the parameter field P in Eq. (11).

$$\dot{\vec{X}} = P \begin{bmatrix} \phi_1(\vec{X}) \\ \phi_2(\vec{X}) \\ \phi_3(\vec{X}) \end{bmatrix}, \tag{11}$$

where ϕ_1, ϕ_2 and ϕ_3 are constant, linear and quadratic functions of $\vec{X}$, respectively. The parameter matrix P is the unknown we try to solve in the modeling procedure.

Now suppose we are given a multi-variate time-series data-set $U(t)$ from the system of Rossler oscillators and do not know the exact functions of this system. In the modeling procedure, we assume and use the polynomials as the basis functions. We will discuss the role of the Taylor's series assumption, and minimal basis choice in subsequent sections.

The modeling procedure gives us an approximation A to the exact parameter field P via some integration schemes. From Eq. (5), we know for each time t, the $\dot{\vec{u}}$ can be approximated by

$$\dot{\vec{u}} = \frac{\vec{u}^{(t+h)} - \vec{u}^{(t)}}{h}, \tag{12}$$

where h is a fixed time step and $\vec{u}^{(t)}$ is the data value at time t. The modeling equation for each time t is

$$\vec{u}^{(t+h)} - \vec{u}^{(t)} = h \cdot A\, \Phi(\vec{u}^{(t)}), \tag{13}$$

where $\Phi(\vec{u})$ is a polynomial function of $\vec{u}$, which contains all possible combinations of linear terms and quadratic terms of $\vec{u}$ and constant 1 for the constant terms. The system parameters P are now coded in the notation within the matrix A as it incorporates the various linear combinations of the assumed basis functions. We consider the system as a continuous dynamical system. Therefore, the equations for all time and space are:

$$\left[\vec{u}^{(1)} - \vec{u}^{(0)}, \vec{u}^{(2)} - \vec{u}^{(1)}, \ldots, \vec{u}^{(n)} - \vec{u}^{(n-1)}\right] = h \cdot A \left[\Phi(\vec{u}^{(0)}), \Phi(\vec{u}^{(1)}), \ldots, \Phi(\vec{u}^{(n-1)})\right], \tag{14}$$

here $\vec{u}^{(n)}$ is u at time $t = nh$ and A is the expected approximating parameter matrix.

Now we give the technique of fitting the given data-set into a desired equation form. According to the system of Rossler oscillators, fitting by quadratic basis functions can produce a result good enough. Generally, the assumption is that we can obtain as accurate a parameter matrix for P as desired by increasing the order of the polynomial. However, for some cases, it is not useful to have a polynomial with a higher order in the fitting procedure. In a series of papers, the minimum description length principle for truncating a model is given and applied [6,16].

2.2. Convergence of the parameter matrix

Now we focus on the obtained parameter matrix A. It is well known that the truncated error between $\dot{\vec{u}}$ and $\frac{\vec{u}^{(t+h)} - \vec{u}^{(t)}}{h}$ originating from the Euler integration approach is linearly dependent on the time step size h. In this part, we discuss the relations among the parameter matrices obtained by different time step sizes.

This part is divided into several subparts. In the first subpart, the review for Euler integration is offered. In the second subpart, some preliminary definitions and lemmas leading to a convergence theorem are given. In the third subpart, the subsequent convergence of the parameter matrix by forward Euler is discussed. In the fourth subpart, some numerical results of the convergence of the parameters by using higher-order integrations are produced.

2.2.1. Review for forward Euler integration

From Taylor's theorem [13], for any function $\overrightarrow{u}(t)$, which is infinitely many times differentiable on $(t_0 - R, t_0 + R)$, we can find its Taylor's series around the given center t_0,

$$\overrightarrow{u}(t) = \overrightarrow{u}(t_0) + \overrightarrow{u}'(t_0)(t - t_0) + \frac{\overrightarrow{u}''(t_0)}{2!}(t - t_0)^2 + \frac{\overrightarrow{u}'''(t_0)}{3!}(t - t_0)^3 + \cdots + \frac{\overrightarrow{u}^{(n)}(t_0)}{n!}(t - t_0)^n + \cdots \tag{15}$$

here R is the radius of convergence of the series. Denoting h as the difference $t - t_0$, we can have an expression of $\overrightarrow{u}'(t_0)$:

$$\overrightarrow{u}'(t) = \frac{\overrightarrow{u}(t) - \overrightarrow{u}(t_0)}{h} - \frac{h}{2!}\overrightarrow{u}''(t_0) - \frac{h^2}{3!}\overrightarrow{u}'''(t_0) - \cdots - \frac{h^{n-1}}{n!}\overrightarrow{u}^{(n)}(t_0) + \cdots. \tag{16}$$

From this equation we can see the error by the forward Euler's method is:

$$\left|\overrightarrow{u}'(t) - \frac{\overrightarrow{u}(t) - \overrightarrow{u}(t_0)}{h}\right| = \frac{h}{2!}\overrightarrow{u}''(t_0) + \cdots + \frac{h^{n-1}}{n!}\overrightarrow{u}^{(n)}(t_k) + \cdots. \tag{17}$$

A better approximation can be obtained if we decrease the h, the sampling rate. And the error is in the order of h. We are thus motivated to discuss the error between parameter matrices obtained by two different time step sizes.

2.2.2. Preliminary definitions and lemmas

In this subpart, we give some preliminary definitions and lemmas, which will be used in the later proofs of the convergence theorem illustrating the nonlinear parameter estimation technique converges with respect to the sampling rate h. First, we define two new matrices as the following according to Eq. (14),

Definition. Denote

$$X_n = \left[\overrightarrow{u}^{(1)} - \overrightarrow{u}^{(0)}, \overrightarrow{u}^{(2)} - \overrightarrow{u}^{(1)}, \ldots, \overrightarrow{u}^{(n)} - \overrightarrow{u}^{(n-1)}\right] \tag{18}$$

and

$$Y_n = \left[\Phi(\overrightarrow{u}^{(0)}), \Phi(\overrightarrow{u}^{(1)}), \ldots, \Phi(\overrightarrow{u}^{(n-1)})\right]. \tag{19}$$

Therefore we have an equation for the fixed data sampling rate h_n:

$$X_n = h_n A_n Y_n. \tag{20}$$

For the unknown parameters A_n, that is equivalent to

$$A_n = \frac{1}{h_n} X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1}. \tag{21}$$

Notice that this is a least square interpretation of Eq. (20).

Now we have the definition of the different A_n w.r.t. different data sampling rates, h_n. In the modeling procedure, we choose different amounts of the data from the time-series data-set U w.r.t. different data sampling rates, h_n, to do the fitting steps. If the chosen sampling rate h_n is small, more information will be applied in the fitting procedure. It means the sizes of the matrices X_n and Y_n are large and the parameters resulting are more accurate than if obtained by a larger sampling rate $h_{n'}$. Therefore, to compare different parameter matrices, we need to investigate the data matrices X_n and Y_n with different n. However, it is impossible to compare two matrices of different sizes directly. To meet this challenge, we can enlarge the size of one of them so that the enlarged matrix is in the same size as the other one.

In the following lemmas, we discuss two parameters A_n and $A_{n'}$ according to two sampling rates h_n and $h_{n'}$, where $h_n = k \cdot h_{n'}$. The numbers of columns of the matrices $X_{n'}$ and $Y_{n'}$ are k times of those of columns of X_n and Y_n. We can build two new matrices $X_{n,k}$ and $Y_{n,k}$ by repeating each column of X_n and Y_n k times so that $X_{n,k}$ and $Y_{n,k}$ are the same size as $X_{n'}$ and $Y_{n'}$ respectively. Before we compare $X_{n,k}$, $Y_{n,k}$ with $X_{n'}$, $Y_{n'}$, we give a lemma to show the parameters $A_{n,k}$ obtained by column-repeated matrices $X_{n,k}$ and $Y_{n,k}$ are equal to the parameters A_n.

Lemma 1. *The parameters* $A_n = \frac{1}{h_n} X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1}$ *are equivalent to the parameters* $A_{n,k} = \frac{1}{h_n} X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}$.

Proof. The column-repeated matrices $X_{n,k}$ and $Y_{n,k}$ are given below:

$$\begin{aligned} X_{n,k} &= [X_n(:,1), \ldots, X_n(:,1), X_n(:,2), \ldots, X_n(:,2), \ldots, X_n(:,n), \ldots, X_n(:,n)] \\ &= \Big[\overrightarrow{u}^{(1)} - \overrightarrow{u}^{(0)}, \ldots, \overrightarrow{u}^{(1)} - \overrightarrow{u}^{(0)}, \overrightarrow{u}^{(2)} - \overrightarrow{u}^{(1)}, \ldots, \overrightarrow{u}^{(2)} - \overrightarrow{u}^{(1)}, \ldots, \overrightarrow{u}^{(n)} \\ &\quad - \overrightarrow{u}^{(n-1)}, \ldots, \overrightarrow{u}^{(n)} - \overrightarrow{u}^{(n-1)}\Big]. \end{aligned} \tag{22}$$

$$\begin{aligned} Y_{n,k} &= [Y_n(:,1),\ldots,Y_n(:,1),Y_n(:,2),\ldots,Y_n(:,2),\ldots,Y_n(:,n),\ldots,Y_n(:,n)] \\ &= \left[\Phi(\overrightarrow{u}^{(0)}),\ldots,\Phi(\overrightarrow{u}^{(0)}),\Phi(\overrightarrow{u}^{(1)}),\ldots,\Phi(\overrightarrow{u}^{(1)}),\ldots,\Phi(\overrightarrow{u}^{(n-1)}),\ldots,\Phi(\overrightarrow{u}^{(n-1)})\right]. \end{aligned} \tag{23}$$

Choosing the jth columns of $X_{n,k}$ and $Y_{n,k}$ arbitrarily, we know of the existence of the corresponding columns of X_n and Y_n with the same values of the chosen columns, and the sampling rates used to find parameters A_n and $A_{n,k}$ are both h_n. Therefore, the values of parameters A_n and $A_{n,k}$ given by the following equations are the same:

$$\begin{aligned} X_n &= h_n A_n Y_n \\ X_{n,k} &= h_n A_{n,k} Y_{n,k}. \quad \square \end{aligned} \tag{24}$$

This lemma tells us that it is suitable to find the relations between A_n and $A_{n'}$ by comparing $A_{n,k}$ with $A_{n'}$, which is given by the equation $A_{n'} = \frac{1}{h_{n'}} X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1}$, and the properties of A_n are saved in $A_{n,k}$. Now we investigate the errors between A_n and $A_{n'}$. From the discussion above, the error is given by:

$$E = \|A_{n'} - A_n\| = \|A_{n'} - A_{n,k} + A_{n,k} - A_n\|,$$

here we obtain this equation by adding and subtracting $A_{n,k}$ so that we can evaluate the error by splitting the right hand side function into two parts by the triangle inequality:

$$E \le \|A_{n'} - A_{n,k}\| + \|A_{n,k} - A_n\|.$$

From Lemma 1 we know the second term is zero. Express $A_{n'}$ and $A_{n,k}$ by Eqs. (21) and (24):

$$E \le \left\| \frac{1}{h_{n'}} X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - \frac{1}{h_n} X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} \right\|,$$

factor the constant $\frac{1}{k h_{n'}}$ out:

$$E \le \left| \frac{1}{k h_{n'}} \right| \cdot \|k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\|,$$

add and subtract a term $k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}$:

$$E \le \left| \frac{1}{k h_{n'}} \right| \cdot \|k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} + k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} - X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\|,$$

combine the common terms and split into two parts by the triangle inequality:

$$E \le \left| \frac{1}{k h_{n'}} \right| \cdot [|k| \cdot \|X_{n'} Y_{n'}^{\mathrm{T}}\| \cdot \|(Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\| + \|k X_{n'} Y_{n'}^{\mathrm{T}} - X_{n,k} Y_{n,k}^{\mathrm{T}}\| \cdot \|(Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\|],$$

add and subtract a term $k X_{n'} Y_{n,k}^{\mathrm{T}}$ to the second term:

$$\begin{aligned} E \le & \left| \frac{1}{h_{n'}} \right| \|X_{n'} Y_{n'}^{\mathrm{T}}\| \cdot \|(Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\| \\ & + \left| \frac{1}{k h_{n'}} \right| \cdot \|k X_{n'} Y_{n'}^{\mathrm{T}} - k X_{n'} Y_{n,k}^{\mathrm{T}} + k X_{n'} Y_{n,k}^{\mathrm{T}} - X_{n,k} Y_{n,k}^{\mathrm{T}}\| \cdot \|(Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\|, \end{aligned}$$

split the second term and simplify the whole function:

$$\begin{aligned} E \le & \left| \frac{1}{h_{n'}} \right| \cdot \|X_{n'}\| \cdot \|Y_{n'}^{\mathrm{T}}\| \cdot \|(Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\| + \left| \frac{1}{h_{n'}} \right| \cdot \|X_{n'}\| \cdot \|Y_{n'}^{\mathrm{T}} - Y_{n,k}^{\mathrm{T}}\| \cdot \|(Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\| \\ & + \left| \frac{1}{k h_{n'}} \right| \cdot \|k X_{n'} - X_{n,k}\| \cdot \|Y_{n,k}^{\mathrm{T}}\| \cdot \|(Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\|. \end{aligned} \tag{25}$$

In order to estimate the error, terms in the Eq. (25) are discussed in the following lemmas. To show those lemmas, we use Taylor's series, singular value decomposition and spectral perturbation theory of matrices, [10].

Lemma 2. *$\|X_n\|$ is in the order of h_n.*

Proof. The matrix X_n is given in Eq. (18), and from the Taylor's series in Eq. (16), we know that the norm of each column in X_n is in the order of h_n. Therefore we can draw a conclusion that:

$$\|X_n\| = C_1 \cdot h_n + \mathcal{O}(h_n^2), \tag{26}$$

where C_1 is a constant. Similarly, $\|X_{n'}\|$ converges with $h_{n'}$. $\square$

Lemma 3. *The error* $\|E_y\| = \|Y_{n'} - Y_{n,k}\|$ *is in the order of* $h_{n'}$.

Proof. The matrices $Y_{n,k}$ and $Y_{n'}$ are given below:

$$Y_{n'} = \left[\Phi(\overrightarrow{u}^{(0)}), \Phi(\overrightarrow{u}^{(h_{n'})}), \ldots, \Phi(\overrightarrow{u}^{(N-h_{n'})})\right]. \tag{27}$$

$$Y_{n,k} = \left[\Phi(\overrightarrow{u}^{(0)}), \ldots, \Phi(\overrightarrow{u}^{(0)}), \Phi(\overrightarrow{u}^{(h_n)}), \ldots, \Phi(\overrightarrow{u}^{(h_n)}), \ldots, \Phi(\overrightarrow{u}^{(N-h_n)}), \ldots, \Phi(\overrightarrow{u}^{(N-h_n)})\right], \tag{28}$$

here we are talking about the time space $[0, N]$.

Consider the jth column in $Y_{n'}$ and $Y_{n,k}$:

$$Y_{n'}^{(j)} = \Phi(\overrightarrow{u}^{(j-1)\cdot h_{n'}}). \tag{29}$$

For any integer k there are k cases of jth column in $Y_{n,k}$.

(i) $j = km$ (m is any positive integer):

$$Y_{n,k}^{(j)} = \Phi(\overrightarrow{u}^{\frac{j-k}{k}\cdot h_n}), \tag{30}$$

therefore, from the Taylor' series of $Y_{n'}^{(j)}$ at the center $\frac{j-k}{k}\cdot h_n$, the error

$$\|Y_{n'}^{(j)} - Y_{n,k}^{(j)}\| = \|(k-1)h_{n'}\Phi'(\overrightarrow{u}^{(j-k)h_{n'}}) + \mathcal{O}(h_{n'}^2)\|. \tag{31}$$

(ii) $j = km + 1$:

$$Y_{n,k}^{(j)} = \Phi(\overrightarrow{u}^{\frac{j-1}{k}\cdot h_n}), \tag{32}$$

therefore

$$\|Y_{n'}^{(j)} - Y_{n,k}^{(j)}\| = 0. \tag{33}$$

Generally, when $j = km + a$ (m and a are positive integers and $a = 1, \ldots, k$)

$$Y_{n,k}^{(j)} = \Phi(\overrightarrow{u}^{\frac{j-a}{k}\cdot h_n}), \tag{34}$$

therefore the error between the jth columns of $Y_{n'}$ and $Y_{n,k}$ is:

$$\|Y_{n'}^{(j)} - Y_{n,k}^{(j)}\| = \|(a-1)h_{n'}\Phi'(\overrightarrow{u}^{(j-a)h_{n'}}) + \mathcal{O}(h_{n'}^2)\|. \tag{35}$$

Hence, the error between matrices $\|E_y\| = C_2 \cdot h_{n'} + \mathcal{O}(h_{n'}^2)$. □

Lemma 4. *The error* $\|E_x\| = \|k \cdot X_{n'} - X_{n,k}\|$ *is in the order of* $h_{n'}^2$.

Proof. Matrices $X_{n'}$ and $X_{n,k}$ are given below:

$$X_{n'} = \left(\overrightarrow{u}^{(h_{n'})} - \overrightarrow{u}^{(0)}, \overrightarrow{u}^{(2h_{n'})} - \overrightarrow{u}^{(h_{n'})}, \cdots, \overrightarrow{u}^{(N)} - \overrightarrow{u}^{(N-h_{n'})}\right). \tag{36}$$

$$X_{n,k} = \left(\overrightarrow{u}^{(h_n)} - \overrightarrow{u}^{(0)}, \ldots \overrightarrow{u}^{(h_n)} - \overrightarrow{u}^{(0)}, \overrightarrow{u}^{(2h_n)} - \overrightarrow{u}^{(h_n)}, \ldots, \overrightarrow{u}^{(N)} - \overrightarrow{u}^{(N-h_n)}\right). \tag{37}$$

Consider the jth columns in $X_{n'}$ and $X_{n,k}$:

$$X_{n'}^{j} = \left(\overrightarrow{u}^{(jh_{n'})} - \overrightarrow{u}^{(j-1)h_{n'}}\right). \tag{38}$$

For any integer k, $j = km + a$ (m and a are positive integers and $a = 1, \ldots, k$)

$$X_{n,k}^{j} = \left(\overrightarrow{u}^{((\frac{j-a}{k}+1)h_n)} - \overrightarrow{u}^{((\frac{j-a}{k})h_n)}\right). \tag{39}$$

From Taylor's expansion we know

$$\|X_{n'}^{j}\| = \|h_{n'}\overrightarrow{u}'^{(j-1)h_{n'}} + \mathcal{O}(h_{n'}^2)\|. \tag{40}$$

$$\|X_{n,k}^{j}\| = \|h_n\overrightarrow{u}'^{(j-a)h_{n'}} + \mathcal{O}(h_n^2)\|. \tag{41}$$

Therefore,

$$\begin{aligned}\|k \cdot X^j_{n'} - X^j_{n,k}\| &= \|k \cdot h_{n'} \overrightarrow{u}'^{(j-1)h_{n'}} - h_n \overrightarrow{u}'^{(j-a)h_{n'}} + \mathcal{O}(h^2_{n'})\| \\ &\le |k \cdot h_{n'}| \cdot \|\overrightarrow{u}'^{(j-1)h_{n'}} - \overrightarrow{u}'^{(j-a)h_{n'}}\| + \mathcal{O}(h^2_{n'}) \\ &\le |k \cdot h_{n'}| \cdot \|(a-1)h_{n'} \overrightarrow{u}''^{(j-a)h_{n'}} + \mathcal{O}(h^2_{n'})\| + \mathcal{O}(h^2_{n'}) \\ &= \mathcal{O}(h^2_{n'}). \end{aligned} \tag{42}$$

To get the last second step, we extend $\overrightarrow{u}'^{(j-1)h_{n'}}$ as a Taylor's series around the center $\overrightarrow{u}'^{(j-a)h_{n'}}$. □

Lemma 5. *The error* $\|E_{yy}\| = \|(Y_{n'}Y^{\mathrm{T}}_{n'})^{-1} - (Y_{n,k}Y^{\mathrm{T}}_{n,k})^{-1}\|$ *is in the order of* $h_{n'}$.

Proof. Since every matrix has an SVD, let us consider the SVDs of $Y_{n'} = U_{n'}W_{n'}V^{\mathrm{T}}_{n'}$ and $Y_{n,k} = U_{n,k}W_{n,k}V^{\mathrm{T}}_{n,k}$. Therefore:

$$\|(Y_{n'}Y^{\mathrm{T}}_{n'})^{-1} - (Y_{n,k}Y^{\mathrm{T}}_{n,k})^{-1}\| = \left\| U_{n'}\frac{1}{W^2_{n'}}U^{\mathrm{T}}_{n'} - U_{n,k}\frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\|,$$

add and subtract a term $U_{n'}\frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k}$, and split into two parts by the triangle inequality:

$$\begin{aligned} &\le \left\| U_{n'}\frac{1}{W^2_{n'}}U^{\mathrm{T}}_{n'} - U_{n'}\frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\| + \left\| U_{n'}\frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} - U_{n,k}\frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\| \\ &\le \|U_{n'}\| \cdot \left\| \frac{1}{W^2_{n'}}U^{\mathrm{T}}_{n'} - \frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\| + \|U_{n'} - U_{n,k}\| \cdot \left\| \frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\|, \end{aligned}$$

the norm of any unitary matrix is 1:

$$= \left\| \frac{1}{W^2_{n'}}U^{\mathrm{T}}_{n'} - \frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\| + \|U_{n'} - U_{n,k}\| \cdot \left\| \frac{1}{W^2_{n,k}} \right\|,$$

add and subtract a term $\frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n'}$ to the first term and split into two parts by the triangle inequality:

$$\begin{aligned} &\le \left\| \frac{1}{W^2_{n'}}U^{\mathrm{T}}_{n'} - \frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n'} \right\| + \left\| \frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n'} - \frac{1}{W^2_{n,k}}U^{\mathrm{T}}_{n,k} \right\| + \|U_{n'} - U_{n,k}\| \cdot \left\| \frac{1}{W^2_{n,k}} \right\| \\ &= \left\| \frac{1}{W^2_{n'}} - \frac{1}{W^2_{n,k}} \right\| + 2\left\| \frac{1}{W^2_{n,k}} \right\| \cdot \|U_{n'} - U_{n,k}\| \\ &= \left\| \frac{(W_{n,k} + W_{n'})(W_{n,k} - W_{n'})}{W^2_{n'} \cdot W^2_{n,k}} \right\| + 2\left\| \frac{1}{W^2_{n,k}} \right\| \cdot \|U_{n'} - U_{n,k}\|. \end{aligned} \tag{43}$$

From perturbation theory [10], we know:

$$|W^i_{n,k} - W^i_{n'}| \le \sigma(E_y) = \|E_y\|_2. \tag{44}$$

Let

$$A = \begin{pmatrix} 0 & Y^{\mathrm{T}}_{n,k} \\ Y_{n,k} & 0 \end{pmatrix}, \tag{45}$$

and

$$A + h_{n'}B = \begin{pmatrix} 0 & Y^{\mathrm{T}}_{n'} \\ Y_{n'} & 0 \end{pmatrix}, \tag{46}$$

then $\begin{pmatrix} 0 \\ U^i_{n,k} \end{pmatrix}$ and $\begin{pmatrix} V^i_{n,k} \\ 0 \end{pmatrix}$ are eigenvectors of A. Similarly $\begin{pmatrix} 0 \\ U^i_{n'} \end{pmatrix}$ and $\begin{pmatrix} V^i_{n'} \\ 0 \end{pmatrix}$ are eigenvectors of $A + h_{n'}B$. Let $L_1 = \begin{pmatrix} 0 \\ U^1_{n,k} \end{pmatrix}$ and $L'_1 = \begin{pmatrix} 0 \\ U^1_{n'} \end{pmatrix}$, then we know the error between L_1 and L'_1 is bounded [10].

$$\begin{aligned} \|L_1 - L'_1\| &= \|h_{n'}Z_1 + h^2_{n'}Z_2 + \cdots\| \\ &= \left\| h_{n'}\sum_{j=1}^{n} t_{j1}L_j + h^2_{n'}\sum_{j=1}^{n} t_{j2}L_j + \cdots \right\| \end{aligned}$$

$$= |h_{n'}| \cdot \left\| \sum_{j=1}^{n} t_{j1} L_j + h_{n'} \sum_{j=1}^{n} t_{j2} L_j + \cdots \right\|. \tag{47}$$

Therefore $\|U_{n'} - U_{n,k}\|$ is bounded and of the order of $h_{n'}$. And from Eq. (44) and Lemma 3, the first term in Eq. (43) is also of the order of $h_{n'}$. Hence $\|E_{yy}\|$ is in the order of $h_{n'}$. □

2.2.3. Convergence theorems by using forward Euler integration

In this part, we give the convergence theorem, which shows the parameter estimation technique converges with respect to the sampling rate $h_{n'}$.

Convergence Theorem 1. *The error $\|E_A\| = \|A_{n'} - A_n\|$ is in the order of $h_{n'}$ by forward Euler integration.*

Proof. As in the discussion in the previous part, the error between $A_{n'}$ and A_n is given below by using the triangle inequality:

$$\|E_A\| = \|A_{n'} - A_n\|, \tag{48}$$

adding and subtracting a term $A_{n,k}$ so that $A_{n'}$ and $A_{n,k}$ are of the same size:

$$= \|A_{n'} - A_{n,k} + A_{n,k} - A_n\|, \tag{49}$$

by the triangle inequality and Eqs. (21) and (24):

$$\begin{aligned} &\leq \|A_{n'} - A_{n,k}\| + \|A_{n,k} - A_n\| \\ &= \left\| \frac{1}{h_{n'}} X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - \frac{1}{h_n} X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} \right\| + \left\| \frac{1}{h_n} X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} - \frac{1}{h_n} X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} \right\|, \end{aligned} \tag{50}$$

the second term is zero by Lemma 1:

$$\begin{aligned} &= \left\| \frac{1}{h_{n'}} X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - \frac{1}{k h_{n'}} X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} \right\| \\ &= \left| \frac{1}{k h_{n'}} \right| \left\| k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} \right\|, \end{aligned} \tag{51}$$

adding and subtracting a term $k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}$ to have two parts:

$$= \left| \frac{1}{k h_{n'}} \right| \| k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} + k X_{n'} Y_{n'}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} - X_{n,k} Y_{n,k}^{\mathrm{T}} (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1} \|, \tag{52}$$

by the triangle inequality:

$$\leq \left| \frac{1}{k h_{n'}} \right| [|k| \|X_{n'} Y_{n'}^{\mathrm{T}}\| \cdot \|(Y_{n'} Y_{n'}^{\mathrm{T}})^{-1} - (Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\| + \|k X_{n'} Y_{n'}^{\mathrm{T}} - X_{n,k} Y_{n,k}^{\mathrm{T}}\| \cdot \|(Y_{n,k} Y_{n,k}^{\mathrm{T}})^{-1}\|], \tag{53}$$

from singular value decomposition:

$$\begin{aligned} &= \left| \frac{1}{h_{n'}} \right| \|X_{n'} Y_{n'}^{\mathrm{T}}\| \cdot \left\| U_{n'} \frac{1}{W_{n'}^2} U_{n'}^{\mathrm{T}} - U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}} \right\| \\ &\quad + \left| \frac{1}{k h_{n'}} \right| \cdot \|k X_{n'} Y_{n'}^{\mathrm{T}} - k X_{n'} Y_{n,k}^{\mathrm{T}} + k X_{n'} Y_{n,k}^{\mathrm{T}} - X_{n,k} Y_{n,k}^{\mathrm{T}}\| \cdot \left\| U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}} \right\| \\ &= \left| \frac{1}{h_{n'}} \right| \|X_{n'}\| \cdot \|Y_{n'}^{\mathrm{T}}\| \cdot \left\| U_{n'} \frac{1}{W_{n'}^2} U_{n'}^{\mathrm{T}} - U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}} \right\| + \left| \frac{1}{h_{n'}} \right| \cdot \|X_{n'}\| \cdot \|Y_{n'}^{\mathrm{T}} - Y_{n,k}^{\mathrm{T}}\| \cdot \left\| U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}} \right\| \\ &\quad + \left| \frac{1}{k h_{n'}} \right| \cdot \|k X_{n'} - X_{n,k}\| \|Y_{n,k}^{\mathrm{T}}\| \cdot \left\| U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}} \right\|. \end{aligned} \tag{54}$$

From Lemmas 2 and 5, $\|X_{n'}\|$ and $\|U_{n'} \frac{1}{W_{n'}^2} U_{n'}^{\mathrm{T}} - U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}}\|$ are both in the order of $h_{n'}$ so that the first term in Eq. (54) is in the order of $h_{n'}$. From Lemma 3, $\|Y_{n'}^{\mathrm{T}} - Y_{n,k}^{\mathrm{T}}\|$ is in the order of $h_{n'}$ too and $\|U_{n,k} \frac{1}{W_{n,k}^2} U_{n,k}^{\mathrm{T}}\|$ equals $\|\frac{1}{W_{n,k}^2}\|$ so that the second term is in the order of $h_{n'}$. From Lemma 4, $\|k X_{n'} - X_{n,k}\|$ is in the order of $h_{n'}^2$ so that the third term is in the order of $h_{n'}$. Therefore, the error $\|E_A\|$ is in the order of $h_{n'}$, meaning the parameter matrix converges with $h_{n'}$. □

From the proof above, we know that not only the error of the parameter matrix A is decreasing with h, but also the size of the data matrices used to calculate A is increasing, eventually beyond the possibility of direct computation on the computer. If the given time-series data is huge, can we skip some columns of the original time-series to obtain a good matrix A by using the smaller data-set?

Let A_{ns} be the smaller data-set obtained by choosing the multiplicity of sth columns of A_n and extending this smaller data-set into the original size by repeating those columns s times. That is, A_{ns} choose every sth column from A_n, and repeat it $s-1$ times. Now we discuss the error between these two matrices A_{ns} and A_n, since from Lemma 1, we know the parameter obtained by skip-only data matrices is equal to A_{ns} obtained by skip-repeated matrices.

Theorem. *Let* $A_{ns} = [A_n(:,1), \ldots, A_n(:,1), A_n(:,s+1), \ldots, A_n(:,s+1), \ldots, A_n(:,(n-1)s+1)]$, *then* $\|A_{ns} - A_n\| = \mathcal{O}(h)$.

Proof. The proof is similar to the proof of the convergence theorem in the previous part:

$$\|A_{ns} - A_n\|,$$

by definitions in Eq. (21):

$$= \left\| \frac{1}{h_n} X_{ns} Y_{ns}^{\mathrm{T}} (Y_{ns} Y_{ns}^{\mathrm{T}})^{-1} - \frac{1}{h_n} X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} \right\|$$
$$= \left| \frac{1}{h_n} \right| \cdot \| X_{ns} Y_{ns}^{\mathrm{T}} (Y_{ns} Y_{ns}^{\mathrm{T}})^{-1} - X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} \|,$$

adding and subtracting a term $X_{ns} Y_{ns}^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1}$:

$$= \left| \frac{1}{h_n} \right| \cdot \| X_{ns} Y_{ns}^{\mathrm{T}} (Y_{ns} Y_{ns}^{\mathrm{T}})^{-1} - X_{ns} Y_{ns}^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} + X_{ns} Y_{ns}^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} - X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} \|,$$

by the triangle inequality:

$$\leq \left| \frac{1}{h_n} \right| \cdot [\| X_{ns} Y_{ns}^{\mathrm{T}} (Y_{ns} Y_{ns}^{\mathrm{T}})^{-1} - X_{ns} Y_{ns}^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} \| + \| X_{ns} Y_{ns}^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} - X_n Y_n^{\mathrm{T}} (Y_n Y_n^{\mathrm{T}})^{-1} \|]$$
$$= \left| \frac{1}{h_n} \right| [\| X_{ns} Y_{ns}^{\mathrm{T}} \| \cdot \| (Y_{ns} Y_{ns}^{\mathrm{T}})^{-1} - (Y_n Y_n^{\mathrm{T}})^{-1} \| + \| X_{ns} Y_{ns}^{\mathrm{T}} - X_n Y_n^{\mathrm{T}} \| \cdot \| (Y_n Y_n^{\mathrm{T}})^{-1} \|],$$

by singular value decomposition:

$$= \left| \frac{1}{h_n} \right| \left[\| X_{ns} Y_{ns}^{\mathrm{T}} \| \cdot \left\| U_{ns} \frac{1}{W_{ns}^2} U_{ns}^{\mathrm{T}} - U_n \frac{1}{W_n^2} U_n^{\mathrm{T}} \right\| + \| X_{ns} Y_{ns}^{\mathrm{T}} - X_{ns} Y_n^{\mathrm{T}} + X_{ns} Y_n^{\mathrm{T}} - X_n Y_n^{\mathrm{T}} \| \cdot \| (Y_n Y_n^{\mathrm{T}})^{-1} \| \right]$$
$$\leq \left| \frac{1}{h_n} \right| \left[\| X_{ns} Y_{ns}^{\mathrm{T}} \| \cdot \left\| U_{ns} \frac{1}{W_{ns}^2} U_{ns}^{\mathrm{T}} - U_n \frac{1}{W_n^2} U_n^{\mathrm{T}} \right\| + (\| X_{ns} \| \cdot \| Y_{ns}^{\mathrm{T}} - Y_n^{\mathrm{T}} \| + \| X_{ns} - X_n \| \cdot \| Y_n^{\mathrm{T}} \|) \cdot \| (Y_n Y_n^{\mathrm{T}})^{-1} \| \right]. \tag{55}$$

Since the terms in this proof are similar to those in the previous theorem, we have shortened the explanation. □

Now we present a graph to show and validate the convergence with $h_{n'}$. Fig. 1 shows the error between the modeling parameter and the exact parameter with h. The red line (.-) is log(error) vs log(h). The blue (*-) and green (o-) lines are log(h) vs $log(h)$ and 2 log(h) vs log(h), respectively. From the graph, we see that the red line is almost parallel to the blue line, the trend of which means the error is in the order of h, not h^2.

2.2.4. Higher order convergence theorems

In fact, the order of convergence in Theorem 1 depends on the fact that the scheme used in Eq. (13) is $\mathcal{O}(h)$. If instead, we use higher order schemes to approximate $\dot{X}$ in Eq. (4) then a comparable theorem can be proved that convergence of A_n is in higher order of h.

We use a method based on Taylor's series to generate the higher order schemes. Since all data from the unknown system is obtained, the approximation to $\dot{X}(t)$ can be found by the values of X at other times. Here we use a five-point stencil to approximate $\dot{X}$ at time t. The involved points used to find an approximation to the value of $\dot{X}(t)$ are $X(t-2)$, $X(t-1)$, $X(t)$, $X(t+1)$ and $X(t+2)$. The linear combination of these five points is:

$$L(X(t))_5 = a_1 X(t+2) + a_2 X(t+1) + a_3 X(t) + a_4 X(t-1) + a_5 X(t-2), \tag{56}$$

here a_i $(i = 1, 2, \ldots, 5)$ are coefficients of this linear combination. By the Taylor's series of $X(t+i)$, the linear combination can be extended as:

$$L(X(t))_5 = a_1 \left(X(t) + 2hX'(t) + \frac{(2h)^2}{2!} X''(t) + \frac{(2h)^3}{3!} X'''(t) + \frac{(2h)^4}{4!} X^{(4)}(t) + \cdots \right)$$

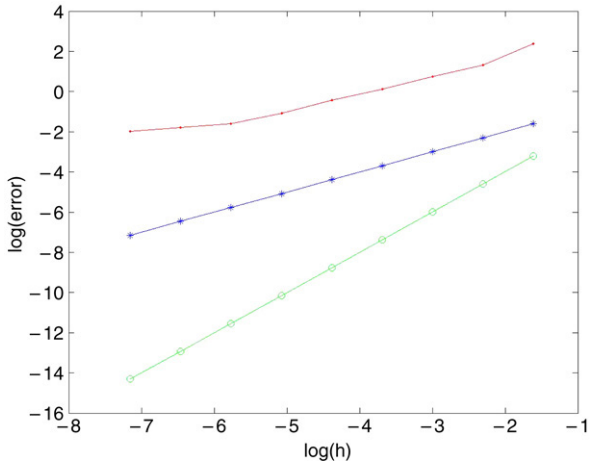

Fig. 1. The error between the modeling parameters of the order of h. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$+a_2\left(X(t)+hX'(t)+\frac{h^2}{2!}X''(t)+\frac{h^3}{3!}X'''(t)+\frac{h^4}{4!}X^{(4)}(t)+\cdots\right)+a_3X(t)$$
$$+a_4\left(X(t)-hX'(t)+\frac{h^2}{2!}X''(t)-\frac{h^3}{3!}X'''(t)+\frac{h^4}{4!}X^{(4)}(t)+\cdots\right)$$
$$+a_5\left(X(t)-2hX'(t)+\frac{(2h)^2}{2!}X''(t)-\frac{(2h)^3}{3!}X'''(t)+\frac{(2h)^4}{4!}X^{(4)}(t)+\cdots\right). \tag{57}$$

In this section, we will use second order, third order and fourth order schemes to compare with the Euler integration. The coefficients in Eq. (57) for the higher cases are given below.

1. First order scheme (forward Euler integration): $a_1 = a_4 = a_5 = 0$, $a_2 = 1$, and $a_3 = -1$.

   $$X(t+1) - X(t) = hX'(t) + \mathcal{O}(h^2),$$

   therefore,

   $$X'(t) = \frac{X(t+1) - X(t)}{h} + \mathcal{O}(h). \tag{58}$$

2. Second order scheme: $a_1 = a_3 = a_5 = 0$, $a_2 = 1$, and $a_4 = -1$.

   $$X(t+1) - X(t-1) = 2hX'(t) + \mathcal{O}(h^3),$$

   therefore,

   $$X'(t) = \frac{X(t+1) - X(t-1)}{2h} + \mathcal{O}(h^2). \tag{59}$$

3. Third order scheme: $a_1 = 1$, $a_2 = -2$, $a_3 = 9$, $a_4 = -10$, and $a_5 = 2$.

   $$X(t+2) - 2X(t+1) + 9X(t) - 10X(t-1) + 2X(t-2) = 6hX'(t) + \mathcal{O}(h^4),$$

   therefore,

   $$X'(t) = \frac{X(t+2) - 2X(t+1) + 9X(t) - 10X(t-1) + 2X(t-2)}{6h} + \mathcal{O}(h^3). \tag{60}$$

4. Fourth order scheme: $a_1 = 1$, $a_2 = -8$, $a_3 = 0$, $a_4 = 8$, and $a_5 = -1$.

   $$X(t+2) - 8X(t+1) + 8X(t-1) - X(t-2) = -12hX'(t) + \mathcal{O}(h^5),$$

   therefore,

   $$X'(t) = \frac{X(t+2) - 8X(t+1) + 8X(t-1) - X(t-2)}{-12h} + \mathcal{O}(h^4). \tag{61}$$

Fig. 2 shows the convergence of the parameters by using higher order schemes. The slopes of the red, green, purple and blue lines approximate 1, 2, 3 and 4, respectively. We see that the slopes of these four lines depend on the orders of the used schemes. Therefore, from these numerical results, we strongly suspect that the convergence rate of the parameters depends on the order of the scheme used, although our method of rigorous proof is only allowed for the explicit Euler scheme.

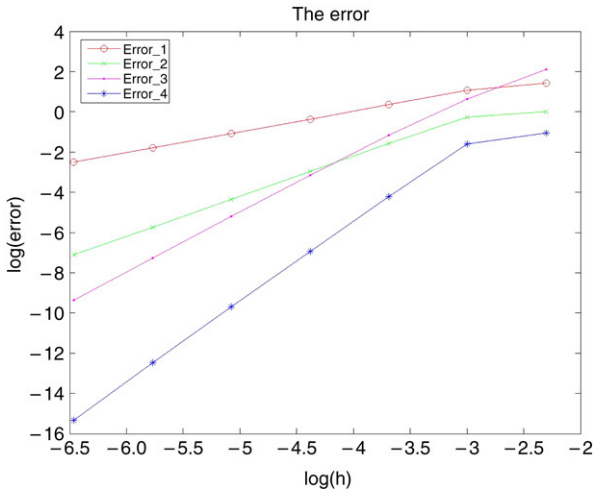

Fig. 2. The errors between the modeling parameters depend on the schemes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In this section, we introduced a technique to model the time-series data from the unknown dynamical system, and also the convergence theorems of the parameters are discussed. In the next section, we will introduce a useful definition, the Kronecker product, to help us investigate the structure of the unknown system by looking at the parameters given in the modeling procedure.

3. Kronecker product representation

In the previous section, we gave a technique to model the time-series data from an unknown dynamical system. If this unknown dynamical system consists of many smaller coupled oscillators, can we discover the property of the system by investigating the parameters given in the modeling procedure? In this section, first, we will review the definition of Kronecker product [18] and some useful properties. Second, we will analyze the structure of the parameters from the system of N coupled Rossler oscillators to show why Kronecker products can be applied. Finally, we will discuss some cases on finding the Kronecker product representations of the parameters.

3.1. Kronecker product and its properties

Definition. Let A be an $n \times p$ matrix and B an $m \times q$ matrix. The $mn \times pq$ matrix

$$A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,p}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,p}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1}B & a_{n,2}B & \cdots & a_{n,p}B \end{bmatrix} \tag{62}$$

is called the *Kronecker product* of A and B. It is also called the direct product or the tensor product [18].

Some properties of the Kronecker product:

1. $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.
2. $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$, $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$.
3. For scalar a, $a \otimes A = A \otimes a = aA$.
4. For conforming matrices, $(A \otimes B)(C \otimes D) = AC \otimes BD$.
5. For partitioned matrices, $[A_1, A_2] \otimes B = [A_1 \otimes B, A_2 \otimes B]$.
6. $(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}}$, $(A \otimes B)^H = A^H \otimes B^H$.

The question we are interested in is for a matrix $A \in R^{m \times n}$ with $m = m_1 m_2$ and $n = n_1 n_2$, there exists a decomposition of A such that

$$\|A - B \otimes C\|_F^2, \tag{63}$$

can be minimized, where $B \in R^{m_1 \times n_1}$, $C \in R^{m_2 \times n_2}$, and "$\otimes$" denotes the Kronecker product. This problem is investigated intensively in [19], in which the problem of finding proper B and C is solved by computing the largest singular value and associated singular vectors of a permuted version of A, and the properties of B and C, which minimize the error given in Eq. (63), are discussed according to the properties of A.

3.2. Kronecker product representation

From Eq. (62), we know Kronecker products copy the matrix C and multiply it by a constant from the matrix B to produce a new matrix A. For one common class of systems with coupled oscillators, the repeated blocks can be found in the parameter field corresponding to a proper order of the ODEs. This fact gives us the motivation to consider the Kronecker product representations of the parameter fields of dynamical systems. On one hand, the original parameters can be decomposed into two small size matrices so that the storage room is saved. On the other hand, for an unknown system, a good Kronecker product representation of the modeling parameters gives us some information about the system, and especially in the aspect that this system is suspected to have coupled chaotic oscillators.

3.2.1. Kronecker product representation of Rossler oscillators

In this section, we analyze the structure of the parameters of the system of N coupled Rossler oscillators to show how Kronecker product representation works for the system of coupled chaotic oscillators. Here we discuss the parameter matrix of the linear term of the ODEs given in Eq. (9). For the parameters of the higher terms, we can use a similar method of analysis.

$$P_1 = \begin{bmatrix} -2\epsilon & -1 & -1 & \epsilon & 0 & 0 & \epsilon & 0 & 0 \\ 1 & a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c & 0 & 0 & 0 & 0 & 0 & 0 \\ \epsilon & 0 & 0 & -2\epsilon & -1 & -1 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -c & 0 & 0 & 0 \\ \epsilon & 0 & 0 & \epsilon & 0 & 0 & -2\epsilon & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c \end{bmatrix}. \tag{64}$$

Here the parameter P_1 is obtained by sorting the ODEs by the method of Eq. (8). Let

$$R = \begin{bmatrix} -2\epsilon & -1 & -1 \\ 1 & a & 0 \\ 0 & 0 & -c \end{bmatrix}. \tag{65}$$

We notice there are repeating Rs in the parameter matrix P_1. And when ϵ is zero, a good Kronecker product decomposition of P_1 can be identified, according to the definition of the Kronecker product.

$$P_1 = \begin{bmatrix} R & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & R \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes R = B \otimes R. \tag{66}$$

Generally, for a small nonzero value of ϵ, the left side matrix B in the Eq. (66) is not the identity, but almost diagonal with constants close to 1 in the diagonal position and the error $\|P_1 - B \otimes R\|_F^2$ is not zero, but very small.

3.2.2. Investigation of permutations

Here, we discuss how the order of the elements in the vector $\overrightarrow{X}$ affects the error $\|P_1 - B \otimes R\|_F^2$. In addition, we will give a general way to find a good order of the ODEs so that a good Kronecker product representation can be identified, when it exists.

In the case discussed in the previous part, the P_1 is generated when we set up $\overrightarrow{X} = [x_1, y_1, z_1, \ldots, x_3, y_3, z_3]^{\mathrm{T}}$. If we change the order of the elements in $\overrightarrow{X}$ in the following way:

$$X = [x_1, x_2, x_3, y_1, \ldots, z_2, z_3]^{\mathrm{T}}, \tag{67}$$

then the system can be written as:

$$\dot{X} = \begin{bmatrix} -2\epsilon & \epsilon & \epsilon & -1 & 0 & 0 & -1 & 0 & 0 \\ \epsilon & -2\epsilon & \epsilon & 0 & -1 & 0 & 0 & -1 & 0 \\ \epsilon & \epsilon & -2\epsilon & 0 & 0 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & a & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & a & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -c & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c \end{bmatrix} X + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 z_1 \\ x_2 z_2 \\ x_3 z_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ b \\ b \\ b \end{bmatrix}. \tag{68}$$

We notice there is no repeating block in the diagonal position of the linear parameter matrix P_1'. Thus, a good Kronecker product representation of P_1' cannot be obtained, in this form and in this order. Thus, we know the error $\|P_1 - B \otimes R\|_F^2$ relies on the order of the elements in the vector $\overrightarrow{X}$. Therefore how to find a good order of the ODEs becomes important.

Assume we have the exact parameter matrix A and a permuted parameter matrix A'. Let M be the permutation matrix such that

$$A' = MAM^{\mathrm{T}} \tag{69}$$

Suppose A, A' and M have the Kronecker product representations, each,

$$A = B \otimes C, \qquad A' = B' \otimes C', \qquad M = M_1 \otimes M_2. \tag{70}$$

Then follows,

$$\begin{aligned} &B' \otimes C' = A' = MAM^{\mathrm{T}} = M(B \otimes C)M^{\mathrm{T}} \\ &= (M_1 \otimes M_2)(B \otimes C)(M_1 \otimes M_2)^{\mathrm{T}} \\ &\text{By the property 6 of Kronecker products} \\ &= (M_1 \otimes M_2)(B \otimes C)(M_1^{\mathrm{T}} \otimes M_2^{\mathrm{T}}). \\ &\text{By property 4 of Kronecker products} \end{aligned} \tag{71}$$

$$= (M_1 B \otimes M_2 C)(M_1^{\mathrm{T}} \otimes M_2^{\mathrm{T}}) = (M_1 B M_1^{\mathrm{T}}) \otimes (M_2 C M_2^{\mathrm{T}}). \tag{72}$$

If the sizes of B' and C' are the same as those of $M_1 B M_1^{\mathrm{T}}$ and $M_2 C M_2^{\mathrm{T}}$, B' and C' are the same as $M_1 B M_1^{\mathrm{T}}$ and $M_2 C M_2^{\mathrm{T}}$ because of the uniqueness of the Kronecker product decompositions of the same size. Therefore for a parameter matrix from an unknown system, which has coupled oscillators, a good Kronecker product representation of the parameter can be found by permuting the order of the fitting vector. Thus we summarize with the following proposition.

Proposition 1. *If A and A' are similar by a permutation matrix M, as indicated by Eq. (69), and M itself has a Kronecker decomposition $M = (M_1 \otimes M_2)$ of appropriate dimension as indicated by Eq. (71), then $B' = M_1 B M_1^{\mathrm{T}}$ and $C' = M_2 C M_2^{\mathrm{T}}$ may also be permuted in the Kronecker decomposition as shown in Eq. (72).*

Furthermore, other permutations of A where M does not have an appropriately sized Kronecker decomposition are not expected to give a zero error, or perhaps even a small residual, Kronecker product approximation.

Now we will explicitly check the relationship between the errors and the permutation, choosing a family of permutations with a prior expected Kronecker structure. Use M as a permutation operator and assume we have a data-set X, which can be split into two parts, XL and XR, denoting the left part and the right part of X respectively. Applying M to X, the new data-set is generated in the following way:

$$M(X)(:, k : k+1) = [XL(:, i); XR(:, j)], \tag{73}$$

here we form $M(X)$ by inserting XR into XL so that the columns of XL become into the odd number columns and the columns of XR become into the even number columns in $M(X)$.

To find a best order of X, we can apply the operator M several times and compare the errors from all permutations of X. In Fig. 3, we show the relationship between the number of times the permutation is applied, and the resulting errors. We notice that the errors from iterations of the permutation are periodic since, after several permutations, the data-set X goes back to the original one. Therefore, we can develop more schemes for permutations to find a better order.

3.2.3. Existence of Kronecker product representation

In the previous part, we know the permutations can be applied to find a good Kronecker product representation. However, trying all possible permutations is computationally expensive. Can we know if there exists a good Kronecker product representation without trying to fit all possible permutations? Let us go back to the properties of the Kronecker product to answer this question. The following theorem about eigenvalues and eigenvectors [18] is very helpful to offer an insight into this problem.

Theorem of eigenvalues and eigenvectors [18]

The Kronecker product of two vectors $u \in R^p$ and $v \in R^r$ is a vector $u \otimes v \in R^{pr}$ given by $u \otimes v = [u^{\mathrm{T}}v_1, \ldots, u^{\mathrm{T}}v_r]^{\mathrm{T}}$. Suppose that $B \in R^{r,r}$ and $C \in R^{s,s}$ and

$$Au_i = \lambda_i u_i, \quad i = 1, \ldots, r, \qquad Bv_j = \mu_j v_j, \quad j = 1, \ldots, s, \tag{74}$$

then for $i = 1, \ldots, r$, $j = 1, \ldots, s$

$$(A \otimes B)(u_i \otimes v_j) = \lambda_i \mu_j (u_i \otimes v_j). \tag{75}$$

$$(A \otimes I_s + I_r \otimes B)(u_i \otimes v_j) = (\lambda_i + \mu_j)(u_i \otimes v_j). \tag{76}$$

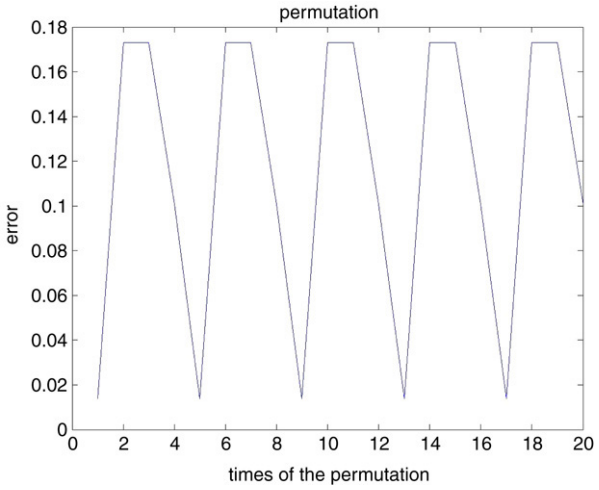

Fig. 3. Permutations vs. errors.

Thus the eigenvalues of a Kronecker product (sum) are the products (sums) of the eigenvalues of the factors. The eigenvectors of a Kronecker product (sum) are the products of the eigenvectors of the factors.

Consider the following typical situation, which helps to interpret the implications of the above theorem to indicate existence of a structured Kronecker decomposition.

If one of A and B is the identity, supposing it is $A = I$ without loss of generality, then there are repeating blocks in the diagonal position of $A \otimes B$. Furthermore, the eigenvalues of $A \otimes B$ are structured in the following way. The eigenvalues of $A \otimes B$ consist of each of the r eigenvalues of B appearing with multiplicity such that it is repeated p-times. Therefore, whenever we look at the spectrum of a matrix C, and we see repeated eigenvalues, we should suspect that there exists a structured Kronecker decomposition whose size is indicated by the multiplicity of the repetition. In general, noise, fitting numerics and other corruptions would prevent the observed matrix C from exactly having repeating elements in its spectrum. Continuity theorems of the spectrum of a matrix with respect to small perturbations of the matrix allow us to state the following principle.

Remark. If C has approximately repeating elements in its spectrum, then a Kronecker decomposition of structured size indicated by the repetition length is indicated.

In Eq. (64), we notice there are repeating blocks R in the diagonal position, and remaining elements in the matrix P_1 are almost zero. If the value of ϵ is zero, the matrix P_1 can be decomposed as the Kronecker product of the identity and the repeating block R. Hence, from the remark above, the spectrum of the matrix P_1 is three groups of the repeated eigenvalues of the matrix R. However, the value of ϵ is generally nonzero, the fact of which means the Kronecker product representation matrix B in Eq. (66) is almost the identity. Moreover, the eigenvalues of the matrix B are not the same constants, but some values close by. Therefore, the spectrum of the matrix P_1 is three groups of the close by values.

Fig. 4 shows the spectrum of the fitted parameter of the 3 coupled Rossler system. In Fig. 4, three groups of the values of the spectrum can be seen. From this fact, we can suspect that there exists a good Kronecker decomposition of the matrix P_1. Therefore, without knowing the exact parameter of the system, we are still able to uncover the structure of the system.

4. On evaluation of error

Once a parameter matrix is found, we can approximately reproduce the data, which is expected to approximate to the original data set. We are interested in evaluating the quality of our fit. There are at least two reasonable ways to describe the quality of the fit.

1. Residual error of the data produced by the fitted ODE system as compared to the given data.
2. How close is the right hand side of the fitted ODE to the original ODE?

The first type of error is reasonable in the sense in that the fitting criterion was based on the least squares minimum residual. The second criterion is what we really want to reduce, however, it is not possible to directly compute since we do not have the function which produced the data. The theorems from the previous section concern convergence of the fitted ODE with respect to decreasing sampling. But we cannot directly compare F, the true ODE and $\tilde{F}$, the fitted ODE.

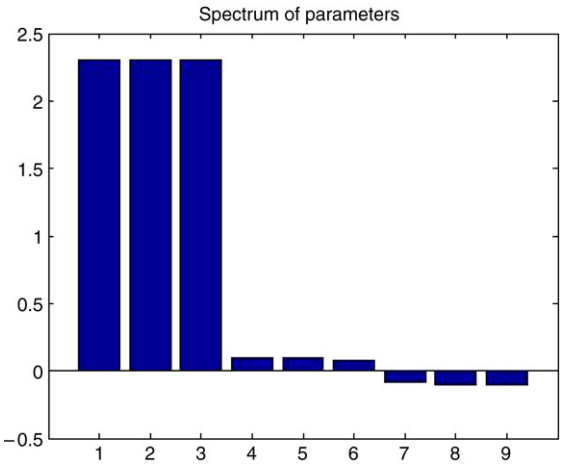

Fig. 4. Spectrum of the parameter of the 3 coupled Rossler system.

4.1. Even good fits can produce large errors

In this section we wish to point out that even if we have done an excellent job of fitting $\tilde{F}$ to be very close to the true F which produced that data, if the ODE is for example chaotic, that sensitive dependence can cause the outputs of $\dot{x} = F$ and $\dot{x} = \tilde{F}$ to produce very different solutions.

Consider for example, the Lorenz system [2],

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= -xz + \gamma x - y \\ \dot{z} &= xy - bz. \end{aligned} \tag{77}$$

Now suppose we have a data set from the Lorenz system, and with that data set we have done an exceedingly good job of fitting the equations of motion; for the sake of argument, suppose we get,

$$\begin{aligned} \dot{x} &= (\sigma + \epsilon)(y - x) \\ \dot{y} &= -xz + \gamma x - y \\ \dot{z} &= xy - bz. \end{aligned} \tag{78}$$

By any standard, this should be considered to be an excellent result, but now we point out that even so, the residual error might point out a large error.

Consider the relationship between the perturbations and the errors that occur in the original and reproduced systems, which we model as a non-autonomous ODE of the form

$$\dot{x} = f(x) + \epsilon g(x, t, \epsilon), \quad x \in R^n \tag{79}$$

where f and g are of class C^r, $r \geq 1$. ϵ is the perturbation parameter. Denoting $x(x_0, t)$ as the solution of the unperturbed autonomous system $\dot{x} = f(x)$ with the initial condition x_0 and $x_\epsilon(x_0', t)$ as the solution of the perturbed system (Eq. (79)) with the initial condition x_0', we will be mainly interested in the following question: In comparison with the unperturbed system, how do the perturbed initial condition and perturbation parameter ϵ affect the solution of the perturbed system?

From the theorem of dependence on initial condition [15], based on Gromwall's inequality, we know that,

$$\|x(x_0, t) - x(x_0', t)\| \leq \|x_0 - x_0'\| e^{Kt}, \tag{80}$$

where K is the Lipschitz constant of $f(x)$ in R^n. From Corollary 3.1.7 of [3], we know that if $\|g\|$ is uniformly bounded by M and the initial conditions of the two systems are the same, then,

$$\|x_\epsilon(x_0, t) - x(x_0, t)\| \leq \frac{\epsilon M}{K}[e^{Kt} - 1], \tag{81}$$

which is another form of Gronwall's inequality. Therefore, the reproduced data set can be considered as the data obtained by the perturbed system with a perturbed initial condition. From the two inequalities above, we can compare the data set of the perturbed system with the data set of the original system.

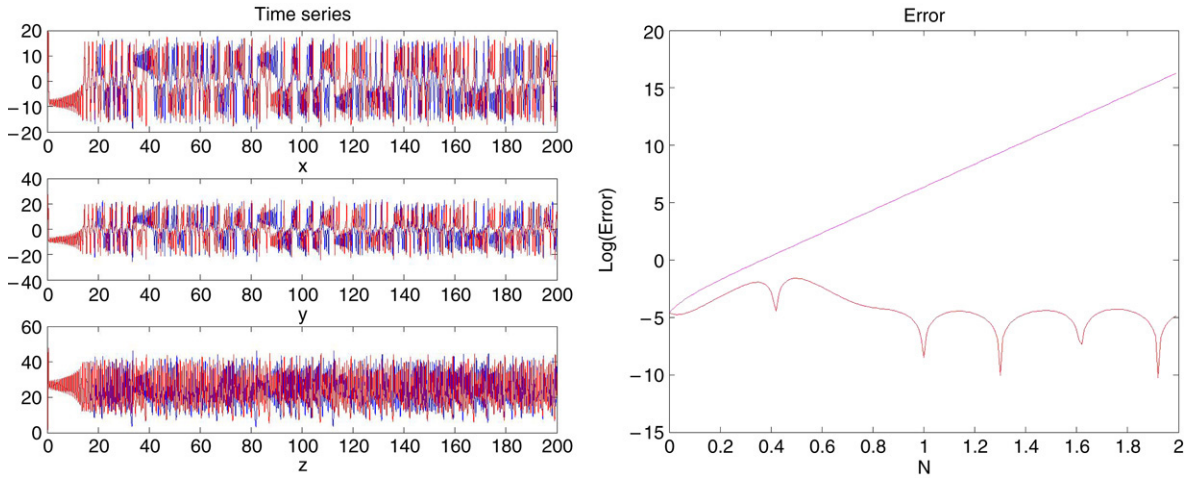

Fig. 5. Error between the unperturbed Lorenz and perturbed Lorenz systems.

$$\|x_\epsilon(x_0', t) - x(x_0, t)\| \le \|x_\epsilon(x_0', t) - x(x_0', t)\| + \|x(x_0', t) - x(x_0, t)\| \le \frac{\epsilon M}{K}[e^{Kt} - 1] + \|x_0 - x_0'\| e^{Kt}. \tag{82}$$

Consider the Lorenz system [2], and the "very-well-fitted" version of that system as above in Eqs. (77) and (78). We choose the values of the constants σ, γ, b as $\sigma = 10, b = \frac{8}{3}, \gamma = 28$ and ϵ is the perturbation parameter for the x oscillator. When ϵ is zero, the system is unperturbed and is the exact Lorenz system. We choose $\epsilon = 0.01$, and the difference between the initial conditions $x_0' - x_0 = 0.01$. The Lipschitz constant K for the x oscillator is easily found analytically to be,

$$K = |\sigma|. \tag{83}$$

The graph in Fig. 5 shows the error between the two systems and the upper bound found by Eq. (82).

In Fig. 5 (Left), we see time-series produced by the two systems Eqs. (77) and (78) respectively. The blue curve is the time-series data from the unperturbed system and the red curve is from the perturbed system. From the figure, we can see there is an overlap of each oscillator for early times, meaning the error between the two systems is initially very small. Eventually, the two curves diverge from each other. Fig. 5 (right) shows a log–log plot of the error between x oscillators of two systems and the upper bound.

The point here is that even with such a well fitted equation Eq. (77) to estimate Eq. (78), residual error can grow exponentially quickly. In fact, the Gromwall's inequality based analysis above serves well to bound these errors analytically, but a Lyapunov exponent analysis would accurately describe the growth rate of the errors. What would be more useful would be a way to consider the positive question of whether a good model has been found.

4.2. The synchronization criterion

In the work of Brown, et al. [6], they suggest a useful synchronization criterion as a nontrivial method to evaluate, whether the fitted ODE is likely to be close to the true ODE which produced the data. Their idea is as follows. If the fitted ODE $\dot{y} = \tilde{F}(y, A)$ is exactly the same as $\dot{x} = F(x, P)$, that is $\|F - \tilde{F}\| = 0$, then we should hope that there exists a coupling scheme,

$$\dot{x} = F(x, P) \tag{84}$$

$$\dot{y} = \tilde{F}(y, A) + \epsilon E(y - x), \tag{85}$$

such that there will be identical synchronization between the two systems, $y(t) \to x(t)$ as $t \to \infty$. Here ϵ is a coupling parameter and E is a coupling matrix. The idea then is that if $\tilde{F}$ is not necessarily *exactly* the same as F, but only close to F, that the same coupling scheme will still produce some form of synchronization, probably nearly identical synchronization. That is the generalized synchronization manifold will likely be of the form,

$$y = \rho(x), \tag{86}$$

and ρ is nearly the identity function. This criterion offers a positive test to validate a good fit, however, it is obviously not always obvious how to find a good ϵ and coupling matrix E. Even if such pairs exist, it is not obvious how much $\tilde{F}$ and F may differ and still produce a positive result. Nonetheless, when it does work, and it does tend to work well, this is a useful way to validate that the fitted model $\tilde{F}$ is likely to be close to F without actually knowing F, since only the data from Eq. (84) is required.

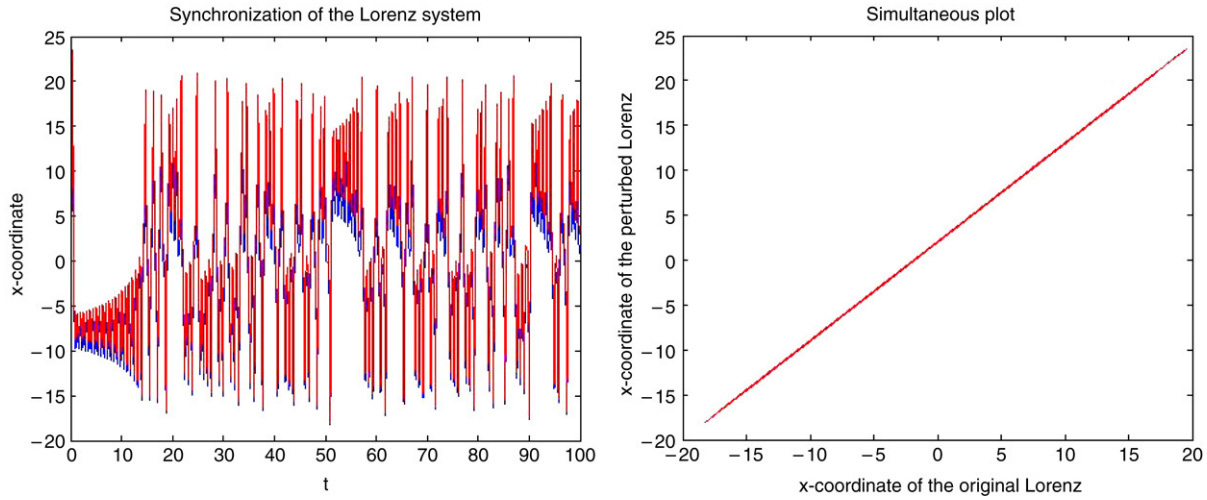

Fig. 6. Synchronization of the Lorenz system.

To validate the Brown method with our previous example, let F be as Eq. (77) and suppose the fitted $\tilde{F}$ is as in Eq. (78). Fig. 6 shows an application of the Brown method for the Lorenz system. We choose $\epsilon = 1$ and the perturbation of the initial conditions as 2.

Fig. 6 (Left) shows the x-coordinates plotted from Eqs. (77) and (78). The difference of the initial conditions between two x-coordinates is 1. The blue trajectory is from the unperturbed Lorenz system, and the red trajectory is from the perturbed Lorenz system. The synchronization is observed. Fig. 6 (right) is a simultaneous plot of one trajectory from the left plot versus the other. We can see that the plot lines up along the diagonal since the two trajectories are synchronized.

5. PDE examples and some prospective work

While the above theory of modeling is developed for ODEs which reproduce a given data-set, we discuss here how evolving spatiotemporal patterns, which should in principle require a PDE, can be approximately reproduced by estimating a set of ODEs which describe time evolving Fourier coefficients with a well chosen set of basis functions. First we give an example data-set from a simple PDE.

5.1. Numerical experiments: A reaction diffusion equation

Consider the reaction diffusion equations of two species and one spatial variable, $u(x,t), v(x,t) \in \Re, x \in [0, L]$ [17]:

$$\begin{aligned} \frac{\partial u}{\partial t} &= D\frac{\partial^2 u}{\partial x^2} + \frac{1}{\epsilon}[v - f(u)] \\ \frac{\partial v}{\partial t} &= D\frac{\partial^2 v}{\partial x^2} - u + \alpha, \end{aligned} \tag{87}$$

with initial conditions $[-2, -4]$ and boundary conditions,

$$u(0,t) = -2, \qquad u(L,t) = -2, \quad \text{and,} \quad v(0,t) = -4, \qquad v(L,t) = -4. \tag{88}$$

Here, the parameters ϵ and α are assumed to be positive and fixed. We choose the nonlinear term $f(u)$ to be given by,

$$f(u) = u^2 + u^3. \tag{89}$$

We construct an empirical data-set $(u(x_i, t_j), v(x_i, t_j))$ by a finite element solver [26].

Assuming this multivariate time-series data-set is the only information we know about this unknown system, we estimate the modeling parameters by using a system of nonlinear equations,

$$\begin{bmatrix} \dot{u} \\ v \end{bmatrix} = A \begin{bmatrix} u \\ v \\ Q(u,v) \\ C(u,v) \end{bmatrix}, \tag{90}$$

where we assume $Q(u, v)$ and $C(u, v)$ are all possible combinations of quadratic and cubic terms. See Fig. 7 which shows the u and v data-sets vs. the reproduced data-sets.

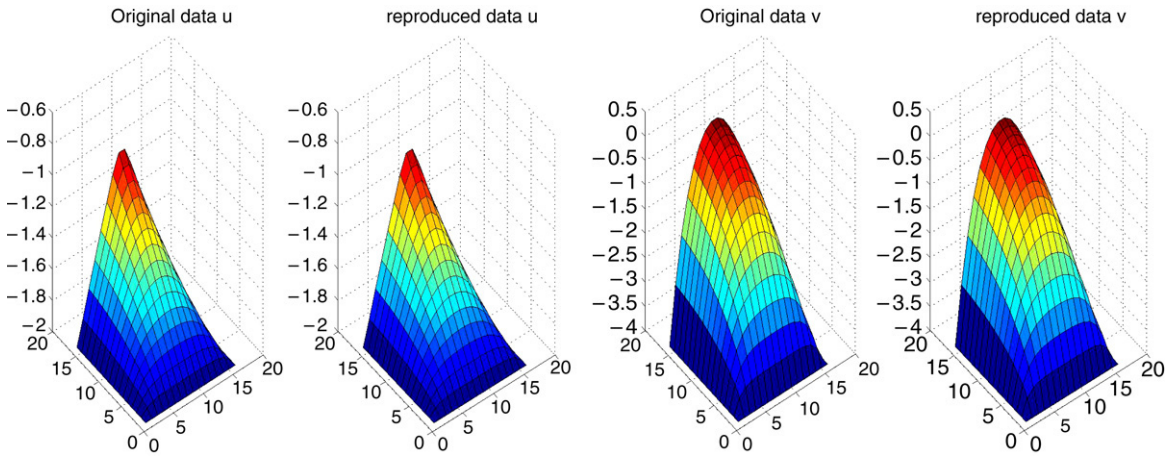

Fig. 7. The original data from the reaction diffusion equation (87) and the reproduced data from the best cubic estimate Eq. (90).

The left graph is the original data of u, v and the right graph is the reproduced data. From the figures we know that the data fitting procedure works well and the error $\|u - u'\|$ is 4.3121e-12. Therefore we know that parameter estimation is achieved for this nonlinear reaction diffusion equation. Moreover, if we use u data alone to do the data fitting, we still get a nice graph.

Fig. 8 shows the original data u and reproduced u obtained by original u only. Although the graph looks good, the error $\|u - u'\| = 4.1836$ is much larger. An explanation of this surprisingly good fit using only $u(x, t)$ data is that the v equation is linear and changes much slower than the u^2 and u^3 terms in the u equation. Therefore, ignoring the v can still give us a good fit in this case.

Now let us consider whether a good Kronecker product representation of the fitted parameter A exists or not. We can use the equation of the system to examine the structure of the parameter matrix A. Using h_t and h_x to denote time step size and space step size, respectively, we have an equation for each time jh_t when $N = 3$ for convenience.

$$\frac{1}{h_t} \cdot \begin{bmatrix} u(x_1, t_{j+1}) - u(x_1, t_j) \\ v(x_1, t_{j+1}) - v(x_1, t_j) \\ u(x_2, t_{j+1}) - u(x_2, t_j) \\ v(x_2, t_{j+1}) - v(x_2, t_j) \\ u(x_3, t_{j+1}) - u(x_3, t_j) \\ v(x_3, t_{j+1}) - v(x_3, t_j) \end{bmatrix} = A \cdot \begin{bmatrix} u(x_1, t_j) \\ v(x_1, t_j) \\ u(x_2, t_j) \\ v(x_2, t_j) \\ u(x_3, t_j) \\ v(x_3, t_j) \\ Q(u, v) \\ C(u, v) \end{bmatrix}, \tag{91}$$

here A is:

$$A = \begin{bmatrix} -2a & 1/\epsilon & a & & a & & \\ -1 & -2a & & a & & a & \\ a & & -2a & 1/\epsilon & a & & \\ & a & -1 & -2a & & a & \\ a & & a & & -2a & 1/\epsilon & \\ & a & & a & -1 & -2a & \\ & & & & & & A_1' \end{bmatrix}, \tag{92}$$

where,

$$a = D/h_x^2, \tag{93}$$

since we use the centered difference approximation,

$$\frac{\partial^2 u(x_i, t)}{\partial x^2} \approx \frac{u(x_{i-1}, t) - 2u(x_i, t) + u(x_{i+1,t})}{h_x^2}. \tag{94}$$

Here we only discuss the parameter of the linear term, which is given explicitly in Eq. (92). Two repeating blocks in the matrix A can be noticed:

$$B_1 = \begin{bmatrix} -2a & 1/\epsilon \\ -1 & -2a \end{bmatrix}, \qquad B_2 = \begin{bmatrix} a & \\ & a \end{bmatrix}. \tag{95}$$

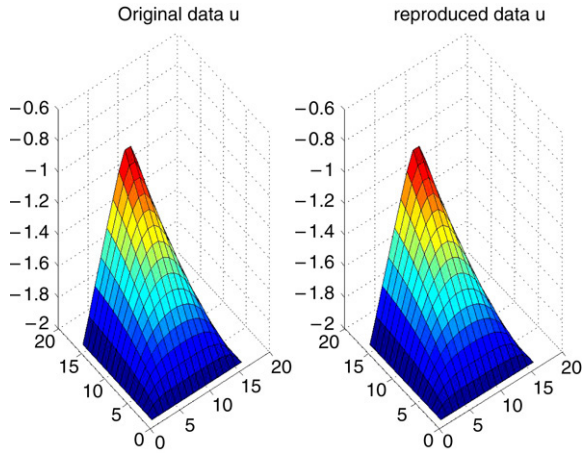

Fig. 8. The original data and the reproduced data of the reaction diffusion equation (87) by using only the $u(x,t)$ data.

When a is very small, a fact which means the D in Eq. (93) is much smaller than h_x, we notice there are repeating blocks B_1 in the diagonal position, and remaining elements are all approaching zero, so that a good representation of the linear parameter can be found. However, generally, D is always larger than h_x, meaning we can not ignore either of the repeating blocks, so that there exist more than one repeating block in the parameter matrix. Therefore, we can not find a good Kronecker product decomposition, meaning there is no obvious oscillator in the reaction diffusion system as expected for a large value of D.

If we re-sort the data, A', another form of A, could be found,

$$\frac{1}{h_t}\cdot\begin{bmatrix} u(x_1,t_{j+1})-u(x_1,t_j)\\ u(x_2,t_{j+1})-u(x_2,t_j)\\ \vdots\\ u(x_m,t_{j+1})-u(x_m,t_j)\\ v(x_1,t_{j+1})-v(x_1,t_j)\\ \vdots\\ v(x_m,t_{j+1})-v(x_m,t_j)\end{bmatrix} = A'\cdot\begin{bmatrix} u(x_1,t_j)\\ u(x_2,t_j)\\ \vdots\\ u(x_m,t_j)\\ v(x_1,t_j)\\ \vdots\\ v(x_m,t_j)\\ Q(u,v)\\ C(u,v)\end{bmatrix}, \tag{96}$$

here A can be written as:

$$A' = \begin{bmatrix} -2a & a & & & a & 1/\epsilon & & & & & \\ a & -2a & a & & & & 1/\epsilon & & & & \\ \vdots & & & & & & \vdots & \vdots & & & \\ a & & & a & -2a & & & & 1/\epsilon & & \\ -1 & & & & & -2a & a & & a & & A_1 \\ & -1 & & & & a & -2a & a & & & \\ \vdots & & & \vdots & & & & & & & \\ & & & & -1 & a & & & a & -2a & \end{bmatrix}. \tag{97}$$

Investigating the A' given in Eq. (97), we could not find repeating blocks in A'. We predict that there is no good Kronecker product representation for A' when we sort the data of u and v as Eq. (96).

5.2. Time delay embedding and alternate coordinate

Consider fitting data from the Rossler system. Proceeding naively to fit the best general quadratic model to data shown in Fig. 9, produced by the following equations [2],

$$\begin{aligned} \mathrm{d}x/\mathrm{d}t &= -y - z\\ \mathrm{d}y/\mathrm{d}t &= x + ay\\ \mathrm{d}z/\mathrm{d}t &= b + z(x - c). \end{aligned} \tag{98}$$

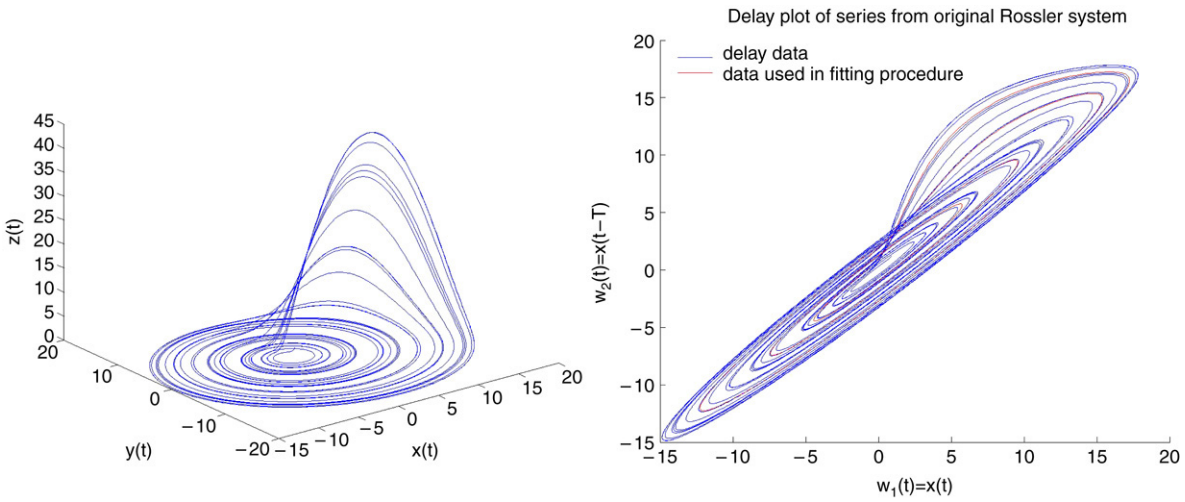

Fig. 9. The original data and the time delay plot of Rossler system Eq. (98).

For the Rossler system, $\mathrm{d}X/\mathrm{d}t = F(X, P)$ in Eq. (98), we can make the delay plot by the data of delay coordinates [21–23]. Here the data from the Rossler system, $(x(t), y(t), z(t)) \in \mathcal{X}$ and the delay embedding data can be obtained by the following equation:

$$w(t) = [x(t), x(t - T), x(t - 2T)], \tag{99}$$

for each time t and $w(t) \in \mathcal{W}$.

Fig. 9 shows the original time-series of Rossler system in an exact x, y, z (Fig. 9 (left)) vs. delay plot in delay variable $(x(t), x(t - T), x(t - 2T))$ (Fig. 9 (right)). Because of the embedding theory, the right plot has shape similar to the other. We guess there is a homeomorphism H between the original data and the delay data. We denote the original system as $\mathcal{X}$ and delay system as $\mathcal{W}$.

$$\begin{array}{ccccc} \Phi & : & \mathcal{X} & \to & \mathcal{X} \\ \mathcal{H} & \downarrow & & & \downarrow \\ \phi & : & \mathcal{W} & \to & \mathcal{W}. \end{array} \tag{100}$$

From Eq. (2), we know $\dot{X} = F(X, P)$, which generates a flow on $\mathcal{X}$. Now we ask if in general we should expect the original equations have a particular (say quadratic) form, and the same simple form should carry over to governing equations of the conjugate variables. Assuming there is an equation for the W variable such that,

$$\dot{W} = G(W, P'). \tag{101}$$

Suppose there is a homeomorphism H between X and W and also F generates a flow on X. Then,

$$\dot{W} = \dot{H}(X) = \frac{\mathrm{d}H}{\mathrm{d}X} \cdot \dot{X} = \frac{\mathrm{d}H}{\mathrm{d}X}(X) \cdot F(X, P) = \frac{\mathrm{d}H}{\mathrm{d}X}(H^{-1}(W)) \cdot F(H^{-1}(W), P). \tag{102}$$

Therefore the equation obtained by delay data is a composition function of F and H^{-1}. If H is linear, the equation obtained by delay data may be in the same degree as F. In general, H may be a very complicated function, the fact of which cannot guarantee there is a new function G, which is similar to F, obtained by the data fitting procedure.

Nonetheless, in the case of the simple Rossler example, we do find excellent fitted models. Fig. 10 shows the reproduced data of the delay plot in Fig. 9 (right) by using a quadratic fitting procedure. In Fig. 10 (left), the red curve is the data of the delay plot used in the fitting procedure. The blue curve (–) is reproduced data by the fourth order scheme and the green one by the forward Euler integration. Fig. 10 (right) is the time-series plots of each coordinate.

6. Conclusion

Our purpose in this paper is to develop a technique to generate a parameter field for a dynamical system. The technique we used to find the parameter field involves modeling the system by forward Euler integration. In contrast to the paper by Brown et al. [6], we use high ordered schemes to fit the derivatives, rather than integrals of the vector field, such as the Adams–Moulton methods. Furthermore, we provide rigorous proof that the method leads to linear convergence of system parameters with respect to sampling time of the data. In order to generate more accurate results, other basis functions instead of polynomials can be used. We also show that the Kronecker product is a useful approach to examine underlying structure of the unknown system if a good permutation has been developed.

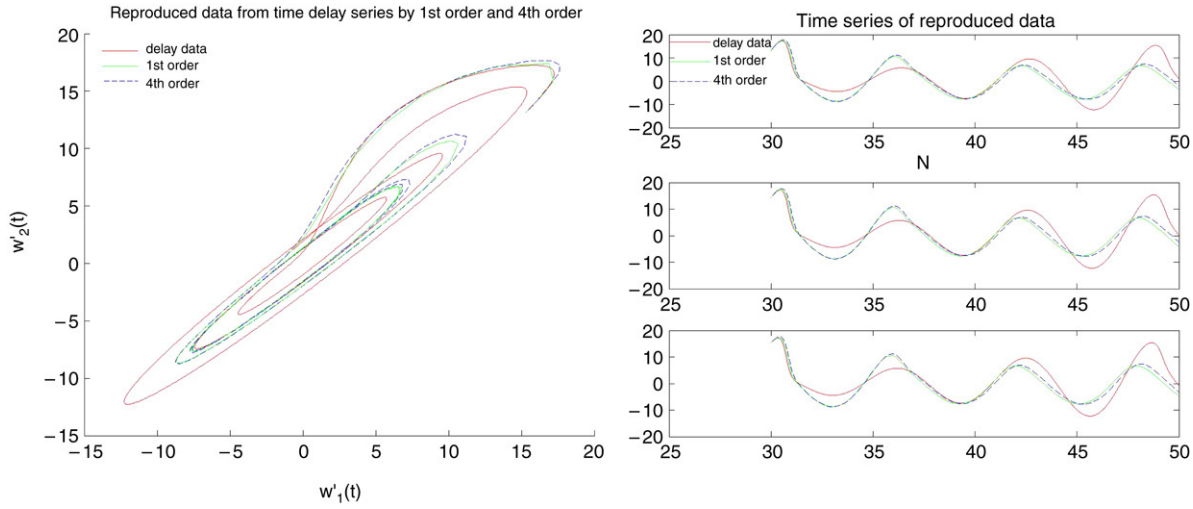

Fig. 10. The delay data and the reproduced data from the delay plot of the Rossler system Eq. (98). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Acknowledgment

CY and EMB have been supported by the National Science Foundation under DMS0404778.

References

[1] J.C. Robinson, All possible chaotic dynamics can be approximated in three dimensions, Nonlinearity 11 (1998) 529–545.
[2] K. Alligood, T.D. Sauer, J.A. Yorke, Chaos, An Introduction to Dynamical Systems, Springer, New York, 1996.
[3] N. Berglund, Perturbation Theory of Dynamical Systems, 2001.
[4] E. Bollt, Model selection, confidence, and scaling in predicting chaotic time-series, Internat. J. Bifur. Chaos in Appl. Sci. Eng. 10 (6) (2000) 1407–1422.
[5] J.L. Breeden, A. Hubler, Reconstructing equations of motion from experimental data with unobserved variables, Phys. Rev. A 42 (1990) 5817.
[6] R. Brown, N. Rulkov, E. Tracy, Modeling and synchronizing chaotic systems from time-series data, Phys. Rev. E 49 (1994) 5.
[7] J.D. Farmer, J.J. Sidorowich, Predicting chaotic time-series, Phys. Rev. Lett. 59 (1987) 845.
[8] J. Farmer, J. Sidorowich, Exploiting chaos to predict the future and reduce noise, in: Y.C Lee (Ed.), Evolution, Learning, and Cognition, World Scientific Press, 1988, p. 277.
[9] M. Giona, Functional reconstruction and local prediction of chaotic time-series, Phys. Rev. A 44 (1991) 3496.
[10] G.H. Golub, C.F. Van Loan, Matrix, Computations, third ed., The Johns Hopkins University Press, 1996.
[11] G. Gouesbet, Reconstruction of the vector fields of continuous dynamical systems from numerical scalar time-series, Phys. Rev. A 43 (1994) 5321.
[12] J. Krener, The Local Convergence of the Extended Kalman Filter, 2004, http://arxiv.org/abs/math.OC/0212255.
[13] W. Kosmala, A Friendly Introduction to Analysis; Single and Multivariable, second ed., Pearson Education, NJ, 2004.
[14] E.J. Kostelich, J.A. Yorke, Noise reduction in dynamical systems, Phys. Rev. A 38 (1988) 1649.
[15] L. Perko, Differential Equations and Dynamical Systems, Third ed., Springer-Verlag, New York, 2001.
[16] J. Rissanen, Ann. Stat. 11 (1983) 416; Automatica 14 (1978) 465; Stochastic Complexity in Statistical Inquiry, World scientific, NJ, 1989.
[17] I. Triandaf, I.B. Schwartz, Karhunen–Loeve mode control of chaos in a reaction–diffusion process, Phys. Rev. E 56 (1997) 204.
[18] C.F. Van Loan, Computational Frameworks for the Fast Fourier Transform, SIAM Publications, Philadelphia, PA, 1992.
[19] C.F. Van Loan, N. Pitsianis, Approximation with Kronecker products. www.math.tkk.fi/ mhuhtane/kronecker.ps.
[20] G. Welch, G. Bishop, An Introduction to the Kalman Filter, 2004.
[21] A.S. Weigenbend, N.A. Gershenfeld, Times Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1993.
[22] H. Abarbanel, Analysis of Observed Chaotic Data, Springer, New York, 1996.
[23] H. Kantz, T. Schrieber, Nonlinear Time-Series Analysis, Cambridge U. Press, New York, 1997.
[24] F. Takens, in: D. Rand, L.-S. Young (Eds.), Dynamical Systems and Turbulence, in: Lecture Notes in Mathematics, vol. 898, Springer, Berlin, 1980, p. 366.
[25] W.J. Rugh, Linear Systems Theory, second ed., Prentice Hall, Upper Saddle River, NJ, 1996.
[26] Matlab 7, The MathWorks, Natick, MA, 2005.