Clarkson University

Modeling Low-Dimensional Submanifolds In Dynamical Systems

A Dissertation

by

Chen Yao

Department of Mathematics and Computer Science

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

(Mathematics)

2007

Accepted by the Graduate School

_____

Date Dean

Abstract

Dimensionality reduction, which can be applied to produce a low-dimensional dynamical system that replicates either approximately, or topologically, the output of a high-dimensional dynamical system, plays an important role in the study of dynamical systems. In many cases, there exists a low-dimensional submanifold in a high-dimensional system, even in infinite dimensional systems in the case of a PDE with a low-dimensional attractor.

In this dissertation, we present a new technique to model a low-dimensional dynamical system embedded in a high-dimensional dynamical system. The technique involves a method of discovering a low-dimensional nonlinear submanifold, known only empirically through the data set from a high-dimensional system, and a method of globally modeling the low-dimensional manifold to obtain an analytical form that reproduces a data set approximating the one on the low-dimensional manifold. When a reliable analytical form of the low-dimensional system is obtained, further work such as short-time prediction, structure investigation, and control may be possible to accomplish.

In an example of the Lorenz attractor, we focus on how to perform the dimensionality reduction on a computationally generated data set that is very large, and has an irregular distribution. In examples of fast-slow systems, we apply the technique of modeling systems embedded onto low-dimensional submanifolds on fast-slow systems constructed by simple systems and slow manifolds. In these examples, we empirically show that the

results of analytical forms are as good as expected. In examples of very high-dimensional systems, image spaces and video sequences, we consider each image as a data point to obtain major underlying subspace by the modeling technique.

The technique of modeling low-dimensional systems developed here does not require analytical forms of the original systems, and also, the embedding low-dimensional submanifolds can be highly nonlinear. Moreover, in some specified systems, the analytical forms generated by the modeling technique can give us some insight into the structure of the system restricted on the low-dimensional submanifolds.

Acknowledgments

I avail myself of the opportunity to express my gratitude to my advisor, Erik Bollt, through whose endeavor, I am able to make my accomplishments. He has dedicated his time and efforts to walk me through the journey of the doctoral program. I am grateful he has been constantly providing leadership and encouragement in time of my hardship. His prudent working attitude, optimism and confidence not only have helped me reach my goal, but also will continue to have influence on my future career. I would like to thank him for his enlightenment and wisdom, with which he has brought into the field of research that through years of undertaking, eventually will see the daylight.

I would like to thank the Clarkson University Mathematics Department faculty for their supports and hospitalities. I would also like to extend my thanks to my colleagues, Leya Tesmenitsky, and Xunyang Shen, who have helped me edit this dissertation.

Last but not the least, I especially would like to thank my parents and my husband for their generous supports and endless love.

Contents

List of Tables

List of Illustrations

Chapter 1
Introduction

In this dissertation, we will present a technique to explore a problem of how to model a high-dimensional process by a low-dimensional process, when such a low-dimensional process exists. The main concern of the technique will be to develop a method to produce an analytical form to approximate the equations on a low-dimensional invariant submanifold that is embedded in some high-dimensional dynamical system. Therefore, there are two major aspects to focus on:

1. Discover a low-dimensional manifold in a given high-dimensional process. When a dynamical system has a stable invariant submanifold, onto which a set of empirical data is attracted, we will apply a method to discover the low-dimensional submanifold. In this sense, the method to be applied can be described as dimensionality reduction.

2. Produce an analytical form that is able to reproduce a data set approximating the low-dimensional data set. When a low-dimensional submanifold is obtained empirically by some dimensionality reduction methods, we will perform a global modeling technique to attempt to uncover the underlying form of the data set attracted onto the submanifold.

1.1 Discovery of low-dimensional manifolds

In many cases, such as dissipative dynamical systems and singularly perturbed systems [12, 16] of the form

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{F}(\mathbf{x}, \mathbf{y}) \\ \epsilon\dot{\mathbf{y}} &= \mathbf{G}(\mathbf{x}, \mathbf{y}), \end{aligned} \tag{1.1}$$

where $\mathbf{x} \in \Re^m$, $\mathbf{y} \in \Re^n$, $\mathbf{F} : \Re^m \times \Re^n \to \Re^m$, and $\mathbf{G} : \Re^m \times \Re^n \to \Re^n$, in some domain $D \subset \Re^m \times \Re^n$, ϵ is a small parameter and $0 < \epsilon << 1$, and variable $\mathbf{x}$ is called slow variable and $\mathbf{y}$ is called fast variable since $\frac{\dot{\mathbf{x}}}{\dot{\mathbf{y}}}$ can be of order ϵ, there exist low-dimensional invariant manifolds or attractors. Many techniques have been developed in applied dynamical systems in order to discover a low-dimensional invariant manifold by reducing the order or dimension of input empirical data, or even an analytical model, which is given by a high-dimensional process.

Several methods, including singular perturbation theory [12, 16], the inertial manifold theory [41], and the method of multiple scales [25], attempt to find a low-dimensional equation, which is restricted to some stable invariant manifold, and dynamics of which is same (conjugate) as the long term behavior of the original system. These methods usually require an analytical form of the system to reduce the system onto a stable invariant manifold analytically, however, they are not always appropriate to apply when only a data set is known empirically.

There are also data-driven and empirical techniques developed to reduce the order or dimension of a system when only measurement (data) of the system is available. Galerkin's method and finite element methods [18], which project

dynamical systems onto a linear subspace spanned by some chosen basis functions, are well developed, but these methods are essentially linear. Therefore, they have to either make severe errors when the invariant manifold is intrinsically nonlinear, or retain many basis functions in order to capture the original system. Moreover, they also require an analytical form of the system. Amongst the empirically driven linear methods is the very popular proper orthogonal decomposition (POD), or Karhunen-Loeve analysis [18, 24, 31]. However, either of them are also fundamentally a linear analysis. Therefore, this linear view may not be appropriate for systems in which a nonlinear manifold is embedded.

The ISOMAP (isometric mapping), is a new nonlinear algorithm which was developed recently by Tenenbaum [4, 47, 52] in the machine learning community for the purpose of image processing. This is a method of modeling an embedding manifold as a graph, known only empirically through a data set. On the assumption that a given data set is a good representation of an unknown manifold, the ISOMAP approximates the manifold by an undirected graph (generated by the given data set) that preserves the geodesics[1] of the true manifold. Generally, this method discovers intrinsically low-dimensional structures embedded in high-dimensional data sets. One of the essential advantages of this method is that it attempts to give a more faithful representation of the global structure of data, especially when the structure is nonlinear.

---

[1] A geodesic is the shortest path between points along the manifold in the space.

In this dissertation, we will focus on the nonlinear analysis, ISOMAP method, to show how to discover a low-dimensional manifold when systems are known only empirically through a data set.

1.2 Global modeling

Modeling time-series and multivariate time-series obtained from measurements of a dynamical system has long been a fundamental problem presenting itself to experimental scientists. If a reliable model of such a system is obtained, further work such as short-time prediction and control may be possible. 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," [57]. Weigenbend and Gershenfeld make a distinction between weak modeling (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.

With a time-series from a “chaotic" dynamical system, data-only local analysis, using the methods of embedding and attractor reconstruction has become routine [1, 57, 23, 14] based on the Taken’s embedding theorem [51]. Local methods

of constructing the model of systems have been developed by many authors, including [14, 15, 10, 29, 6]. 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 model dynamical systems that are continuous in time and are naturally modeled by global autonomous ordinary differential equations (ODEs), $\frac{dX}{dt} = F(X)$, where $X \in \Re^n$[10, 9, 19, 21]. Their goal is to build vector fields F. They assume that the vector field $F(x)$ of the dynamical system can be modeled by a series expansion in some basis,

$$F(X) = \lim_{N\to\infty} \sum_{k=0}^{N} p^k \phi^k(X). \tag{1.2}$$

where $\{\phi^k(X)\}$ is a set of basis functions and p^ks are parameters whose values must be determined by the modeling procedure. 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 [10, 39]. Interestingly, Brown, *et. .al.* [10] use a synchronization criterion as a nontrivial method to validate their models, requiring that the model should be able to synchronize identically to the experimental data.

These works, and the problem we study, 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 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 [45]. There have also been works to analyze the convergence of the extended Kalman filter, including in a paper by Krener [27]. 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 dissertation, we will present a global modeling technique, which is based on explicit Euler integration and higher order schemes, to model a given data set, or multivariate time series obtained from measurements of dynamical systems, in order to obtain a modeling system in form of a set of ODEs, which reproduces orbits whose evolution in time mimics the behavior of multivariate time series, or data set properly. Moreover, the convergence theorems of the parameters determined by the modeling procedure are proved for the purpose of comparing the modeling parameters with the exact form of the original system in view of the quality of the modeling technique.

We will furthermore introduce a new method, which is an application of Kronecker product by adapting the work of Van Loan and Pitsianis [55], to uncover underlying structure of the system by investigating the modeling parameters for some specific systems, such as dynamical systems with coupled oscillators. In this dissertation, an example of N-coupled Rössler oscillators is introduced to show a

good structure that can be discovered by the Kronecker product representation of the modeling parameters.

Once a modeling system is produced, the reproduced data is able to be generated for the purpose of error evaluations. In this dissertation, the quality of the modeling system is evaluated based on the theorem of dependence on initial condition [37], which is based on Gromwall's inequality to give us a bound of the error, and Lyapunov analysis [2, 37, 50], which is a global approach to investigating the stability of a fixed point of the error system. Moreover, there is a useful synchronization criterion [10] that can be applied as a nontrivial method to evaluate error.

The technique of modeling a low-dimensional submanifold, which is to be introduced in this dissertation, is based on two aspects aforementioned. Given a data set empirically, which is assumed to be attracted onto a low-dimensional invariant submanifold, we are able to perform the global modeling technique on a low-dimensional data set, which can be produced by applying methods of dimensionality reductions on the given data set, to obtain an analytical form of the modeling system, which reproduces a data set approximating the given one.

The organization of this dissertation is outlined below.

Chapter 2 gives a review of one nonlinear dimensionality reduction method, ISOMAP, for the purpose of obtaining a low-dimensional data set if high-dimensional empirical data is given by dynamical systems.

In Chapter 3, a global modeling technique is introduced by an example of the Rössler oscillator. Proofs of convergence theorems of the parameters, which are produced by modeling technique based on forward Euler integration and higher order schemes, are given in Chapter 3.3.

In Chapter 4.1, an introduction of Kronecker product is given. An example of N-coupled Rössler oscillators given in Chapter 4.2 reveals a good structure, which is able to be discovered by Kronecker product representation of the parameter matrix. In Chapter 4.3, the existence of Kronecker product representation is discussed. Chapter 4 closes with an example of structure investigation of a system of reaction diffusion equations.

In Chapter 5.1, errors between evolutions of the orbits generated by the exact system and the modeling system are evaluated based on the theorem of dependence on initial condition and Lyapunov analysis. Synchronization criterion is shown in Chapter 5.2. Chapter 5 closes with an example on time delay embedding of the Rössler system.

Chapter 6.1 shows an application of the modeling technique of the Lorenz system. In this example, we focus on a problem of how to obtain a data set with a good density for the purpose of performing the ISOMAP method. Chapter 6.2 shows three examples of fast-slow systems with low-dimensional embedding slow-manifolds, the existence of which allow us to produce analytical forms of the underlying structure of the systems by the modeling technique. Chapter 6 closes

with high-dimensional examples of image spaces and video sequences.

Chapter 2
Nonlinear Dimensionality Reduction: ISOMAP Method

The dimensionality reduction of multivariate data is a fundamental problem in dynamical systems. Its goal is to produce a low-dimensional dynamical system, which roughly stated, maintains properties of the high-dimensional system either approximately, or topologically [46]. In many dissipative cases, there exists a low-dimensional dynamical system for some high-dimensional system due to a slow manifold, inertial manifold, or center manifold, which can be found by several methods, including singular perturbation theory [12, 16], the inertial manifold theory [41], and the method of multiple scales [25]. However, those methods focus on the analytical forms of the original systems, which sometimes cannot be obtained.

Data-driven and empirical techniques are also applied to reduce the dimension. These methods are developed to operate when there is a low-dimensional manifold embedded linearly, or almost linearly in the given ambient space. Among those methods is a very popular method, proper orthogonal decomposition (POD), which generates an optimal basis for the model reduction of a set of functions, or a given empirical data set [17]. POD, which is also called principal component analysis (PCA), or KL (Karhun-Loeve) analysis [17, 31, 33, 48] in terms of time averaged optimal basis, is fundamentally a linear analysis dealing with the time-averaged covariance matrix.

Therefore, these methods are fundamentally flawed when the embedding manifold being modeled is highly nonlinear.

In this chapter, a review of a nonlinear method, ISOMAP, will be given for the purpose of discovering a nonlinear low-dimensional manifold.

2.1 ISOMAP Method

Recently, the ISOMAP method has been applied to model a low-dimensional nonlinear manifold in dynamical systems [7]. We will adapt this algorithm here for the purposes of modeling nonlinear manifolds in the context of dynamical systems. This is a natural way to obtain a nonlinear invariant manifold, which is the goal of the usual method, POD, or KL analysis, which however is an entirely linear analysis, hence inherently limited.

2.1.1 ISOMAP Algorithm

For a description of the ISOMAP method, we assume a data set X consisting of N points in n-dimensional Euclidean space $X = \{x_i\}_{i=1}^N \subset \Re^n$, which is assumed to lie on a submanifold $M^d \subset \Re^d$ of lower dimension $d << n$. The algorithm attempts to obtain a corresponding data set $Y = \{y_i\}_{i=1}^N \subset \Re^d$ appropriately embedded in a lower dimensional manifold. In the following, we will describe X as represented in the variables of the ambient space[2] $\Re^n$, and Y as represented in the intrinsic

[2] By ambient space, we mean an embedding sapce $\Re^n$. That is, any point x_i in the data set X is in $\Re^n$.

variables of the manifold M^d.

ISOMAP is a manifold learning algorithm that extends the MDS (multi-dimensional scaling) by using approximations of geodesic distances with the shortest paths on a discrete graph, which approximates the manifold, instead of directly applying MDS to a matrix of Euclidean distances. Three main steps are involved to develop the ISOMAP algorithm, which we will briefly review hereunder.

1. **Build a neighborhood graph G to approximate the manifold**. Graph $G = (X, E)$ consists of each point x_i in $X = \{x_i\}_{i=1}^N$, which can also be considered as the set of vertices of the graph G, and edges $E = \{x_i, x_j\}$ which are ordered pairs of edges connecting pairs of points x_i, x_j in X. There are two ways to construct neighbors, meaning connecting points, in G, ϵ-neighborhoods and neighborhoods of k-nearest neighbors. To build a ϵ-neighborhoods graph, we construct the graph G containing edges $\{x_i, x_j\}$ iff the Euclidean distance between points x_i and x_j, $|| x_i - x_j|| < \epsilon$ for some ϵ. To build a k-nearest neighbors graph, we construct the graph G containing edges $\{x_i, x_j\}$ iff x_i is one of the k nearest neighbors of x_j (and vice versa), with respect to the Euclidean distance. The discrete graph G is constructed to approximate the intrinsic manifold M^d.

2. **Compute geodesic distances (shortest paths) and construct the matrix D_G.** Let D_G be a matrix of graph distances corresponding to the graph G and D_M be

a matrix of geodesic distances[3] for all pairs of data points in X. D_G is going to be defined to approximate the geodesic distance D_M. For each pair of points x_i, and x_j in X, if there is a direct connection, the existence of which means that the edge $\{x_i, x_j\}$ is in the graph G, we use the graph distance

$$D_G(i,j) = ||x_i - x_j||,$$

with respect to the Euclidean distance, to approximate the geodesic distance $D_M(i,j)$. If the edge $\{x_i, x_j\}$ is not in the graph G, the distance between two points x_i and x_j is defined by the following function,

$$d_G(i,j) = \min_P(||x_i - x_{i1}|| + \cdots + ||x_{in} - x_j||)$$

where $P = (x_{i1}, \ldots, x_{in})$ varies over all paths along the edges of G connecting x_i and x_j. Thus the graph distance matrix D_G is obtained for all pairs of data points in X to approximate the true geodesic distances D_M of the manifold. This approximation can be improved by increasing the number of the points in X, or modifying the value of ϵ, or k. If ϵ or k is chosen too large, or the density of the data set X is too low, some neighbors could be in separate parts of the manifold, and the approximation is poor. If ϵ or k is chosen too small, the graph G might not have enough information about the manifold. There are popular methods to compute shortest paths of the graph, including Floyd's algorithm for small to medium sized data set or Dijtsra's algorithm for small to large data set.

[3] The length of the shortest path between points along the manifold in the space.

3. **Apply MDS (multi-dimensional scaling)** [13] **to the resulting graph distance matrix D_G.** MDS requires only the graph distance matrix D_G as input, which was generated in the step 2, to form intrinsic variables Y. A review of MDS will be given below.

2.1.2 Review of MDS (Multi-dimensional scaling)

Multi-dimensional scaling (MDS) [13] is a method of searching for a low-dimensional space by using a distance matrix of points. The distances between points are also called the original dissimilarities. The identifying aspect of MDS is that it uses only the dissimilarity matrix, which usually represents the Euclidean distances between pair of points, whereas by contrast POD works directly with the original set of points.

Let X be a set of N points in an n-dimensional Euclidean space $X = \{x_i\}_{i=1}^{N} \subset \Re^n$, where $x_i = (x_{i1}, x_{i2}, ..., x_{in})^T$. Assume that the only information we have about the data set X is the Euclidean distance matrix D, each element of which $[D]_{ij}$ is given by the equation involving the ith and jth points in X,

$$[D]_{ij} = d_{ij}^2 = (x_i - x_j)^T (x_i - x_j).$$

In the process of MDS, the original Euclidean coordinates can be obtained from the given derived Euclidean distance matrix D, and a matrix of projected d-dimensional data set Y will also be formed. The description of MDS is given as following.

1. **Find matrix** $A = -\frac{1}{2}D$**.**

2. **Find matrix** $B = HAH$, where H is the centering matrix,

$$H = I - n^{-1}1 \cdot 1^T,$$

where I is the $N \times N$ identity, and 1 is a $N \times 1$ vector of ones, $1 = (1, 1, ..., 1)^T_{1\times N}$. The proof that the obtained matrix B is a product of the matrix of the original data set X can be found in [62, 13].

$$B = XX^T, \tag{2.1}$$

where $X = [x_1, x_2, ..., x_N]^T$ is the $N \times n$ matrix of the original coordinates. Hence the method of recovering of the original coordinates is given, however, to obtain an explicit form of matrix X, the spectral decomposition of matrix B has to be applied.

3. **Find the spectral decomposition of matrix** B**.** B is a $N \times N$ symmetric, positive semi-definite matrix of rank n, and hence has n non-negative eigenvalues and $N - n$ zero eigenvalues. Therefore matrix B can be written in terms of its spectral decomposition,

$$B = V\Lambda V^T,$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_N)$, the diagonal matrix of eigenvalues $\{\lambda_i\}$ of B, and $V = [v_1, v_2, ..., v_N]$, the matrix of corresponding normalized eigenvectors, as result $v_i^T v_i = 1$. For convenience the eigenvalues of B are labelled in the

way that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$. Due to the $N - n$ zero eigenvalues, B can also be rewritten as

$$B = V_n \Lambda_n V_n^T,$$

where

$$\Lambda_n = diag(\lambda_1, \lambda_2, ..., \lambda_n),\ V_n = [v_1, v_2, ..., v_n].$$

Therefore as $B = XX^T$ discussed in the Eq. (2.1), the explicit form of the coordinates of the original points has been recovered by matrix X,

$$X = V_n \Lambda_n^{\frac{1}{2}}, \tag{2.2}$$

where $\Lambda_n^{\frac{1}{2}} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda n})$, and V_n is the matrix of the corresponding eigenvectors.

4. **Choose an appropriate number $d < n$ as the dimension of the intrinsic variables Y.** A possible value of d can be found by considering the ratio of $\frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$. The matrix of intrinsic variables Y is given by

$$Y = V_d \Lambda_d^{\frac{1}{2}}, \tag{2.3}$$

where V_d and $\Lambda_d^{\frac{1}{2}}$ use the top d (significant) eigenvalues and eigenvectors of V and Λ, respectively. These first d significant coordinates form intrinsic variables Y, in which the dimension of each point is $d < n$. Moreover, MDS method

minimizes the residual error,

$$\begin{aligned} E &= trace(XX^T - YY^T)^2 \\ &= \sum_{k=1}^{n} (\lambda_k - \lambda_k^*)^2, \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} \lambda_k^* &= \max(\lambda_k, 0) \qquad k = 1, ..., d \\ &= 0 \qquad\qquad\quad k = d+1, ..., n. \end{aligned}$$

When we apply the MDS to the ISOMAP algorithm, a nonlinear analysis occurs since the dissimilarity used in the application of MDS in the ISOMAP is the graph distance matrix D_G instead of the Euclidean matrix. To form a lower dimensional intrinsic variables Y, we follow the general steps of the MDS. However, the matrix X constructed in Eq. (2.2) is not the original coordinates of the data set used in ISOMAP, in spite of the fact of which, the intrinsic variables Y discussed in the ISOMAP method can still be formed by the MDS.

As discussed in step 2 of the ISOMAP algorithm, the choice of ϵ or k is very important to get a good graph G. To choose a proper ϵ or k, we need to consider the data set X. X is a finite sampling subset of the true manifold M, therefore whether the data set X represents the manifold properly will affect the results of the ISOMAP. Tenenbaum discussed some of the theoretical claims for ISOMAP, and a full proof of the asymptotic convergence theorem was given in [4]. The goal is that the graph distance D_G converges asymptotically to the geodesic distance D_M, the

fact of which can be guaranteed under the following conditions.

Asymptotic Convergence Theorem [4]**:** Given λ_1, λ_2, $\mu > 0$, then for α sufficiently large the inequalities

$$1 - \lambda_1 \leq \frac{\text{graph distance}(D_G)}{\text{geodesic distance}(D_M)} \leq 1 + \lambda_2$$

hold with probability at least $1 - \mu$. Here α is the sampling condition of X, which says that for every point m in the true manifold M there is a data point $x_i \in X$ for which $d_M(m, x_i) = \alpha$. This condition is very important to guarantee that a good approximation can be obtained by using the sampling data set X with a proper sampling condition α. However, it is hard to check that if the sampling data set X satisfies the sampling condition because we can not always have information about the true manifold.

In order to operate the ISOMAP method to discover a low-dimensional manifold, a good data set is required. Generally, we assume that data set is generated experimentally while the knowledge of systems, such as the analytical form of the model, is unknown. However, it is hard to verify whether the collected data set is a good sample of the embedding manifold. If the given data set does not meet the sampling condition required by the asymptotic convergence theorem, a proper low-dimensional structure may not be obtained by the ISOMAP method. This is when we ask: "For a given dynamical system, can we expect to produce a good data set computationally?"

In practice, the data set of a dynamical system generated experimentally

by using an arbitrary initial condition is able to be used to perform the ISOMAP method for the purpose of discovering a lower dimensional process. However, there still exist some “bad" initial conditions in the specific dynamical systems, such as the fixed points, the periodic points, the eventually periodic points, and the asymptotically periodic points, because the data sets generated by those initial points cannot be a good representation of the system, the ISOMAP method will fail to discover a good representation of the low-dimensional submanifold. In this dissertation, we will give examples of some dynamical systems, the empirical data of which is appropriate for the ISOMAP method, to reveal that the data set can be generated to satisfy the sampling condition as expected, and the ISOMAP method is capable to be performed in dynamical systems.

2.2 Conclusions

In this chapter, we have given a review of the ISOMAP method in order to discover a nonlinear low-dimensional manifold. For the purpose of applying the ISOMAP method to dynamical systems, a data set that satisfying the asymptotic convergence theorem [4] is expected. In the Chapter 6, we will give an example of the Lorenz system to present a resampling technique to obtain a data set that is appropriate to apply the ISOMAP method. For the future work, a problem of relationships between the initial distributions of ensemble of initial conditions of dynamical systems, and the invariant measure on the invariant set, becomes interesting.

Chapter 3
Global Modeling and Parameter Estimation

3.1 Introduction

In this chapter, we will address the question of global modeling, the definition of which we will take is given as the following: Given multivariate time series, can we produce a set of ordinary differential equations of minimal dimension, which is able to reproduce a data set approximating the given time series? 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 focus here on producing a minimal set of ODEs to "optimally" model such data during the duration of the movie. While it is known that the solution of any dissipative system can be arbitrarily well approximated by the solution of a three-dimensional system [40], we will take a constructive point of view.

We consider modeling unknown dynamical systems that are generally assumed to be governed by global autonomous ordinary differential equations (ODEs),

$$dX/dt = F(X, P), \tag{3.1}$$

where $X \in \Re^n$, $F : \Re^n \to \Re^n$ is a vector field, and P is the parameter field. We assume the vector field $F(X, P)$ is described in terms of a series expansion in a well chosen basis

set of functions,

$$F(X,P) = \sum_{k=0}^{\infty} P_k \phi_k(X), \tag{3.2}$$

where $\{\phi_k(X)\}$ is a set of basis functions and P_ks are parameters whose values must be determined by the global modeling technique.

3.2 Global Modeling Technique

Suppose multivariate time series, such as the solution of an evolution equation

$$u_t = f(u),$$

sampled on a grid in x, and in time t,

$$\{u^n\}_{n=0:M} = \{u(x,t_n)\}_{n=0:M} \tag{3.3}$$

is given. Assuming the system where the time series come from has an exact forms of ODEs,

$$\frac{dX}{dt} = F(X,P), \tag{3.4}$$

where $X \in \Re^n$, $F : \Re^n \rightarrow \Re^n$, $n >> 1$ is the phase-space dimension, and P denotes a set of system parameters. We model time series data with an expected form of ODEs by using an explicit Euler integration approach, but we will also discuss higher-ordered schemes in subsequent sections. The technique based on an explicit Euler integration models the time-series data via

$$X^{(n+1)} \approx X^{(n)} + hF(X^{(n)},P), \tag{3.5}$$

where h is a step size of time t, and $X^{(n)}$ is the value of X at time $t = nh$. Given this

equation, the modeling procedure can be described as a procedure to find a vector field $F(X, P)$, which can be written as a series expansion in terms of appropriate basis functions,

$$F(X, P) = \sum_{k=0}^{\infty} P_k \phi_k(X), \tag{3.6}$$

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 chapter, the basis functions used in modeling procedure are polynomials. In order to present this technique explicitly, we give a simple introductory example of the Rössler system.

Consider the system of Rössler oscillators [2] with periodic boundary conditions:

$$\begin{aligned} dx/dt &= -y - z \\ dy/dt &= x + ay \\ dz/dt &= b + z(x - c). \end{aligned} \tag{3.7}$$

where a, b, and c are the parameters of the Rössler oscillator. We choose $a = 0.165, b = 0.2$, and $c = 10$ so that the oscillator exhibits a chaotic attractor [2]. Rössler equations are particularly simple and have only one nonlinear component, therefore providing us with a better opportunity to show the basic steps. Letting

$$X = [x, y, z]^T, \tag{3.8}$$

we can obtain synthetic test data, a numerical multivariate time series $\{X^{(i)}\}_{i=1:M}$, which represents the Rössler oscillator, by integration of the system by a fourth-order Runge-Kutta algorithm built in Matlab [34]. With an assumption that the numerical time-series is all the information we have about the motion of the system, this time series is to be modeled to find an expected form of ODE, which approximates

the original system, in the form of Eq. (3.4). Before we perform the modeling technique, we investigate the original system to get some *a prior* expectation about the results of modeling procedure.

Rewriting the Rössler oscillator, we express the equations in the form of Eq. (1.2),

$$\dot{X} = P_0 + P_1 X + P_2 X^2, \tag{3.9}$$

where $\dot{X}$ is the derivative of X w.r.t. time t, and

$$P_0 = \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix},\ P_1 = \begin{bmatrix} 0 & -1 & -1 \\ 1 & a & 0 \\ 0 & 0 & -c \end{bmatrix},\ P_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},\ X^2 = \begin{bmatrix} x^2 \\ xy \\ xz \\ y^2 \\ yz \\ z^2 \end{bmatrix}. \tag{3.10}$$

Hence, the basis functions of Eq. (3.9) are constant function, linear function, and quadratic function.

$$\phi_1(X) = 1, \phi_2(X) = X, \text{ and } \phi_3(X) = X^2. \tag{3.11}$$

The parameters of this system are P_0, P_1 and P_2. Combining the basis functions into one vector, we have the parameter field P in Eq. (3.12).

$$\dot{X} = P \begin{bmatrix} \phi_1(X) \\ \phi_2(X) \\ \phi_3(X) \end{bmatrix},\ P = [P_0, P_1, P_2]. \tag{3.12}$$

This is the exact system given in the view of the matrices. The parameter matrix P is the unknown we try to determine in the modeling procedure.

Now suppose we are given a multivariate time series data set $\{u^{(n)}\}_{n=0:M}$ in the form of Eq. (3.3) from the system of Rössler oscillator and do not know the

exact functions of this system. A modeling system, which is assumed in the form of $\dot{u} = F(u, A)$, is able to be obtained when an approximating parameter A is determined by the modeling procedure with a proper choice of the basis functions.

Modeling procedure gives us an approximation A to the exact parameter field P via some integration schemes. From Eq. (3.5), we know at time t, the value of $\dot{u}$ can be approximated by

$$\dot{u} \approx \frac{u(t+h) - u(t)}{h}, \tag{3.13}$$

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

$$u(t+h) - u(t) \approx h \cdot F(u(t), A), \tag{3.14}$$

where the vector field F of this system can also be expressed as following

$$F(u, A) = A_0\phi_1(u) + A_1\phi_2(u) + A_2\phi_3(u) = A\Phi(u),$$

where $A = [A_0, A_1, A_2]$, and $\Phi(u) = [\phi_1(u), \phi_2(u), \phi_3(u)]^T$ is a vector of polynomial functions of u given in Eq. (3.11), which contains constant 1 for the constant terms, and all possible combinations of linear terms and quadratic terms of u. The parameters P of the original system are now coded in the notation within the matrix A that 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 space are:

$$\left[u^{(1)} - u^{(0)}, \cdots, u^{(M)} - u^{(M-1)}\right] \approx h \cdot A \left[\Phi(u^{(0)}), \cdots, \Phi(u^{(M-1)})\right], \tag{3.15}$$

here $u^{(n)} = u(nh)$ is the value of u at time $t = nh$, and A is the expected approximating parameter matrix. By the given data set $\{u^n\}_{n=0:M}$, the left matrix in Eq. (3.15), and the matrix of polynomial functions $\Phi(u)$ are able to be constructed, such that the parameter matrix A is able to be obtained in the modeling procedure. In the following, the parameter matrix A is called modeling parameter.

We gave the technique of modeling the given data set into a desired equation form in the above. For the system of Rössler oscillator, modeling 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 modeling procedure. In a series of papers, a method of choosing a proper set of basis functions, which is called as the minimum description length principle, is introduced and applied [10, 39] for the purpose of truncating the modeling system.

3.3 Parameter Estimation

In this section, we will not only thoroughly analyze convergence properties of the explicit Euler method for our integration method, but also discuss and produce numerical evidence for higher-ordered schemes. We will prove that when the forward Euler integration is applied, the error between the modeling parameters is in order of the sampling rate. Supporting this statement, we will prove here several

useful lemmas followed by the convergence theorem of the parameters with respect to the time step size.

Here we will focus on the modeling parameter A, which was introduced in the previous section in Eq. (3.15). The use of the explicit Euler integration in the introductory example gives us some insight into the properties of the parameter A. It is well known that the truncated error between $\dot{u}$ and $\frac{u(t+h)-u(t)}{h}$ originating from the Euler integration approach is linearly dependent on the time step size h. In this section, we will discuss the relationship among the modeling parameters obtained by different time step sizes.

From Taylor's theorem [28], for any function $u(t)$, 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,

$$u(t) = u(t_0) + u'(t_0)(t - t_0) + \frac{u''(t_0)}{2!}(t - t_0)^2 + \cdots + \frac{u^{(n)}(t_0)}{n!}(t - t_0)^n + \cdots . \quad (3.16)$$

Here we assume time t is in the convergence interval $I = (t_0 - R, t_0 + R)$, where R is the radius of convergence of the series. Denoting h as the difference $t - t_0$, we can have an expression of $u'(t_0)$:

$$u'(t_0) = \frac{u(t) - u(t_0)}{h} - \frac{h}{2!}u''(t_0) - \cdots - \frac{h^{n-1}}{n!}u^{(n)}(t_0) + \cdots . \quad (3.17)$$

From this equation we can see the error by the forward Euler integration is :

$$\left| u'(t_0) - \frac{u(t) - u(t_0)}{h} \right| = \left| \frac{h}{2!}u''(t_0) + \cdots + \frac{h^{n-1}}{n!}u^{(n)}(t_0) + \cdots \right| . \quad (3.18)$$

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.

3.3.1 Lemmas

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, two new matrices are defined as following according to Eq. (3.15) for the rest of this chapter,

Definition: Denote

$$X_n = \left[u^{(1)} - u^{(0)}, u^{(2)} - u^{(1)}, \cdots, u^{(n)} - u^{(n-1)}\right] \tag{3.19}$$

and

$$Y_n = \left[\Phi(\overrightarrow{u}^{(0)}), \Phi(\overrightarrow{u}^{(1)}), \cdots, \Phi(\overrightarrow{u}^{(n-1)})\right]. \tag{3.20}$$

here n denotes the number of time grids, $u^{(i)}$ is the value of function u at time $t = ih_n$, where h_n is the sampling rate w.r.t. n, and Φ is a vector of basis functions.

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

$$X_n \approx h_n A_n Y_n. \tag{3.21}$$

In order to obtain the unknown parameters A_n, we use the following equation to find an approximation of A_n,

$$A_n = \frac{1}{h_n} X_n Y_n^T (Y_n Y_n^T)^{-1}. \tag{3.22}$$

Notice that this is a least square interpretation of the Eq. (3.21).

We notice we can form different A_n w.r.t. different data sampling rate, h_n. In the modeling procedure, we choose different amount of the data from the time series data set $\{u^n\}$ w.r.t. different data sampling rate, h_n, to do the modeling steps. If the chosen sampling rate h_n is small, more information will be applied in the modeling 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 will discuss two parameters A_n and $A_{n'}$, which are given by the Eq. (3.22) according to two sampling rates h_n and $h_{n'}$, where $h_n = k \cdot h_{n'}$, $k > 1$. 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 that 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 $A_n = \frac{1}{h_n} X_n Y_n^T (Y_n Y_n^T)^{-1}$ are equivalent to the parameters $A_{n,k} = \frac{1}{h_n} X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}$.

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

$$X_{n,k} = \left[\underbrace{X_n(:,1), \cdots, X_n(:,1)}_{k-\text{columns}}, \underbrace{X_n(:,2), \cdots, X_n(:,2)}_{k-\text{columns}}, \ \cdots, \ \underbrace{X_n(:,n), \cdots, X_n(:,n)}_{k-\text{columns}} \right]$$
$$= \left[u^{(1)} - u^{(0)}, \ \cdots, \ u^{(2)} - u^{(1)}, \ \cdots, \ u^{(2)} - u^{(1)}, \ \cdots, \ u^{(n)} - u^{(n-1)} \right]. \tag{3.23}$$

$$Y_{n,k} = \left[\underbrace{Y_n(:,1), \cdots, Y_n(:,1)}_{k-\text{columns}}, \underbrace{Y_n(:,2), \cdots, Y_n(:,2)}_{k-\text{columns}}, \ \cdots, \ \underbrace{Y_n(:,n), \cdots, Y_n(:,n)}_{k-\text{columns}} \right]$$
$$= \left[\Phi(u^{(0)}), \ \cdots, \ \Phi(u^{(1)}), \ \cdots, \ \Phi(u^{(1)}), \ \cdots, \ \Phi(u^{(n-1)}), \ \cdots, \ \Phi(u^{(n-1)}) \right]. \tag{3.24}$$

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:

$$\begin{aligned} A_n &= \tfrac{1}{h_n} X_n Y_n^T (Y_n Y_n^T)^{-1} \\ A_{n,k} &= \tfrac{1}{h_n} X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}. \end{aligned} \tag{3.25}$$

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 $A_{n'} = \frac{1}{h_{n'}} X_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1}$, and the properties of A_n are saved in $A_{n,k}$.

Now we investigate the errors between A_n and $A_{n'}$ w.r.t. the p-norm when $p = 1$ [55]. From the discussion above, the error is given by:

$$E = ||A_{n'} - A_n||_1 = ||A_{n'} - A_{n,k} + A_{n,k} - A_n||_1,$$

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 triangle inequality:

$$E \leq ||A_{n'} - A_{n,k}||_1 + ||A_{n,k} - A_n||_1.$$

From lemma 1 we know the second term is zero. Expressing $A_{n'}$ and $A_{n,k}$ by Eq. (3.22), (3.25):

$$E \leq ||\frac{1}{h_{n'}} X_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - \frac{1}{h_n} X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1,$$

and factoring the constant $\frac{1}{kh_{n'}}$ out, we have,

$$E \leq |\frac{1}{kh_{n'}}| \cdot ||kX_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1,$$

subtracting and adding a term $kX_{n'} Y_{n'}^T (Y_{n,k} Y_{n,k}^T)^{-1}$, we have an equivalent form,

$$\begin{aligned} E \leq |\tfrac{1}{kh_{n'}}| \cdot ||&kX_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - kX_{n'} Y_{n'}^T (Y_{n,k} Y_{n,k}^T)^{-1} \\ &+ kX_{n'} Y_{n'}^T (Y_{n,k} Y_{n,k}^T)^{-1} - X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1, \end{aligned}$$

combining the common terms and splitting into two parts by triangle inequality, we have,

$$\begin{aligned} E \leq &|\tfrac{1}{kh_{n'}}| \cdot |k| \cdot ||X_{n'} Y_{n'}^T||_1 \cdot ||(Y_{n'} Y_{n'}^T)^{-1} - (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ &+ |\tfrac{1}{kh_{n'}}| \cdot ||kX_{n'} Y_{n'}^T - X_{n,k} Y_{n,k}^T||_1 \cdot ||(Y_{n,k} Y_{n,k}^T)^{-1}||_1 \end{aligned}$$

adding and subtracting a term $kX_{n'} Y_{n,k}^T$ to the second term, we have,

$$\begin{aligned} E \leq &|\tfrac{1}{h_{n'}}| ||X_{n'} Y_{n'}^T||_1 \cdot ||(Y_{n'} Y_{n'}^T)^{-1} - (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ &+ |\tfrac{1}{kh_{n'}}| \cdot ||kX_{n'} Y_{n'}^T - kX_{n'} Y_{n,k}^T + kX_{n'} Y_{n,k}^T - X_{n,k} Y_{n,k}^T||_1 \cdot ||(Y_{n,k} Y_{n,k}^T)^{-1}||_1, \end{aligned}$$

splitting the second term and simplifying the whole function, we have,

$$\begin{aligned} E \;\leq\; & |\tfrac{1}{h_{n'}}| \cdot ||X_{n'}||_1 \cdot ||Y_{n'}^T||_1 \cdot ||(Y_{n'} Y_{n'}^T)^{-1} - (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ + \;& |\tfrac{1}{h_{n'}}| \cdot ||X_{n'}||_1 \cdot ||Y_{n'}^T - Y_{n,k}^T||_1 \cdot ||(Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ + \;& |\tfrac{1}{kh_{n'}}| \cdot ||kX_{n'} - X_{n,k}||_1 \cdot ||Y_{n,k}^T||_1 \cdot ||(Y_{n,k} Y_{n,k}^T)^{-1}||_1. \end{aligned} \tag{3.26}$$

In order to estimate the error in the norm based on the columns in the matrices, terms in the Eq. (3.26) are discussed in the following lemmas. To show those lemmas, the

theories of Taylor's series, singular value decomposition, and spectral perturbation theory of matrices [20] are applied.

Lemma 2: $||X_n||_1 = \mathcal{O}(h_n)$**.**

Proof: The matrix X_n is given in Eq. (3.19), and from the Taylor's series in Eq. (3.17), 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:

$$||X_n||_1 = C_1 \cdot h_n + \mathcal{O}(h_n^2), \tag{3.27}$$

where C_1 is a constant. Similarly, $||X_{n'}||_1$ converges with $h_{n'}$.

Lemma 3: $||E_y||_1 = ||Y_{n'} - Y_{n,k}||_1 = \mathcal{O}(h_{n'})$**.**

Proof: According to $Y_n = \left[\Phi(\overrightarrow{u}^{(0)}), \Phi(\overrightarrow{u}^{(1)}), \cdots, \Phi(\overrightarrow{u}^{(n-1)})\right]$ defined in Eq. (3.20),The matrices $Y_{n,k}$ and $Y_{n'}$ are given below:

$$\begin{aligned} Y_{n'} &= \left[\Phi(\overrightarrow{u}^{(0)}), \Phi(\overrightarrow{u}^{(1)}), \cdots, \Phi(\overrightarrow{u}^{(n'-1)})\right] && (3.28) \\ &= \left[\Phi(u(0)), \Phi(u(h_{n'})), \cdots, \Phi(u(N - h_{n'}))\right]. && (3.29) \end{aligned}$$

$$\begin{aligned} Y_{n,k} &= \left[\Phi(u^{(0)}),\ \cdots,\ \Phi(u^{(1)}),\ \cdots,\ \Phi(u^{(1)}),\ \cdots,\ \Phi(u^{(n-1)}),\ \cdots,\ \Phi(u^{(n-1)})\right] && (3.30) \\ &= \left[\Phi(u(0)), \cdots, \Phi(u(h_n)), \cdots, \Phi(u(h_n)), \cdots, \Phi(u(N - h_n)), \cdots, \Phi(u(N - h_n))\right] && (3.31) \end{aligned}$$

here $u^{(i)} = u(ih)$, and we are talking about the time space $[0, N]$.

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

$$Y_{n'}^{(j)} = \Phi(u(j-1) \cdot h_{n'}). \tag{3.32}$$

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

1. $j = km$, (m is any positive integer):

$$Y_{n,k}^{(j)} = \Phi(u(\frac{j-k}{k} \cdot h_n)), \tag{3.33}$$

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

$$||Y_{n'}^{(j)} - Y_{n,k}^{(j)}||_1 = ||(k-1)h_{n'}\Phi'(u(j-k)h_{n'})) + \mathcal{O}(h_{n'}^2)||_1. \tag{3.34}$$

2. $j = km + 1$:

$$Y_{n,k}^{(j)} = \Phi(u(\frac{j-1}{k} \cdot h_n)), \tag{3.35}$$

therefore

$$||Y_{n'}^{(j)} - Y_{n,k}^{(j)}||_1 = 0. \tag{3.36}$$

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

$$Y_{n,k}^{(j)} = \Phi(u(\frac{j-a}{k} \cdot h_n)), \tag{3.37}$$

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

$$||Y_{n'}^{(j)} - Y_{n,k}^{(j)}|| = ||(a-1)h_{n'}\Phi'(u((j-a)h_{n'})) + \mathcal{O}(h_{n'}^2)||_1. \tag{3.38}$$

Hence, the error between matrices $||E_y||_1 = C_2 \cdot h_{n'} + \mathcal{O}(h_{n'}^2)$.

Lemma 4: $||E_x||_1 = ||k \cdot X_{n'} - X_{n,k}||_1 = \mathcal{O}(h_{n'}^2)$**.**

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

$$X_{n'} = \left[\; u(h_{n'}) - u(0), \quad u(2h_{n'}) - u(h_{n'}), \quad \cdots, \quad u(N) - u(N - h_{n'}) \; \right] . \tag{3.39}$$

$$X_{n,k} = \left[\; u(h_n) - u(0), \quad \cdots \quad , \quad u(2h_n) - u(h_n), \quad \cdots , \quad u(N) - u(N - h_n) \; \right] . \tag{3.40}$$

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

$$X^j_{n'} = \left(u(jh_{n'}) - u((j-1)h_{n'}) \right). \tag{3.41}$$

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

$$X^j_{n,k} = \left(u((\frac{j-a}{k} + 1)h_n) - u((\frac{j-a}{k})h_n) \right). \tag{3.42}$$

From Taylor's expansion we know

$$||X^j_{n'}||_1 = ||h_{n'}u'((j-1)h_{n'}) + \mathcal{O}(h^2_{n'})||_1. \tag{3.43}$$

$$||X^j_{n,k}||_1 = ||h_n u'((j-a)h_{n'}) + \mathcal{O}(h^2_n)||_1. \tag{3.44}$$

Therefore,

$$\begin{aligned} &||k \cdot X^j_{n'} - X^j_{n,k}||_1 \\ &= ||k \cdot h_{n'}u'((j-1)h_{n'}) - h_n u'((j-a)h_{n'}) + \mathcal{O}(h^2_{n'})||_1 \\ &\leq |k \cdot h_{n'}| \cdot ||u'((j-1)h_{n'}) - u'((j-a)h_{n'}||_1 + \mathcal{O}(h^2_{n'}) \\ &\leq |k \cdot h_{n'}| \cdot ||(a-1)h_{n'}u''((j-a)h_{n'}) + \mathcal{O}(h^2_{n'})||_1 + \mathcal{O}(h^2_{n'}) \\ &= \mathcal{O}(h^2_{n'}). \end{aligned} \tag{3.45}$$

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

Lemma 5: $||E_{yy}||_1 = ||(Y_{n'}Y^T_{n'})^{-1} - (Y_{n,k}Y^T_{n,k})^{-1}||_1 = \mathcal{O}(h_{n'})$**.**

Proof: Since every matrix has a singular value decomposition (SVD), let us

consider the SVDs of $m \times n'$ matrices $Y_{n'}$ and $Y_{n,k}$.

$$Y_{n'} = U_{n'} W_{n'} V_{n'}^T, Y_{n,k} = U_{n,k} W_{n,k} V_{n,k}^T$$

Therefore:

$$\begin{aligned} &||(Y_{n'} Y_{n'}^T)^{-1} - (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ &= ||U_{n'} \tfrac{1}{W_{n'}^2} U_{n'}^T - U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1, \end{aligned}$$

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

$$\begin{aligned} &\leq ||U_{n'} \tfrac{1}{W_{n'}^2} U_{n'}^T - U_{n'} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 + ||U_{n'} \tfrac{1}{W_{n,k}^2} U_{n,k}^T - U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 \\ &\leq ||U_{n'}||_1 \cdot ||\tfrac{1}{W_{n'}^2} U_{n'}^T - \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 + ||U_{n'} - U_{n,k}||_1 \cdot ||\tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1, \end{aligned}$$

the 2-norm of any unitary matrix is 1, $||U_{n'}||_2 = ||U_{n,k}^T||_2 = 1$, and $||U_{n'}||_1 \leq$

$\sqrt{m}||U_{n'}||_2 = \sqrt{m}$ and $||U_{n,k}^T||_1 \leq \sqrt{m}||U_{n,k}^T|||_2 = \sqrt{m}$, therefore,:

$$||(Y_{n'} Y_{n'}^T)^{-1} - (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \leq \sqrt{m} \cdot [||\frac{1}{W_{n'}^2} U_{n'}^T - \frac{1}{W_{n,k}^2} U_{n,k}^T||_1 + ||U_{n'} - U_{n,k}||_1 \cdot ||\frac{1}{W_{n,k}^2}||_1],$$

add and subtract a term $\sqrt{m} \frac{1}{W_{n,k}^2} U_{n'}^T$ to the first term and split into two parts by

triangle inequality:

$$\begin{aligned} &\leq \sqrt{m}||\tfrac{1}{W_{n'}^2} U_{n'}^T - \tfrac{1}{W_{n,k}^2} U_{n'}^T||_1 + \sqrt{m}||\tfrac{1}{W_{n,k}^2} U_{n'}^T - \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 + \sqrt{m}||U_{n'} - U_{n,k}||_1 \cdot ||\tfrac{1}{W_{n,k}^2}||_1 \\ &= m||\tfrac{1}{W_{n'}^2} - \tfrac{1}{W_{n,k}^2}||_1 + 2\sqrt{m}||\tfrac{1}{W_{n,k}^2}||_1 \cdot ||U_{n'} - U_{n,k}||_1 \\ &= m||\tfrac{(W_{n,k} + W_{n'})(W_{n,k} - W_{n'})}{W_{n'}^2 \cdot W_{n,k}^2}||_1 + 2\sqrt{m}||\tfrac{1}{W_{n,k}^2}||_1 \cdot ||U_{n'} - U_{n,k}||_1. \end{aligned} \tag{3.46}$$

From perturbation theory [20], we know:

$$|W_{n,k}^i - W_{n'}^i| \leq \sigma(E_y) = ||E_y||_2 \leq \sqrt{n}||E_y||_1. \tag{3.47}$$

Let

$$M = \begin{pmatrix} 0 & Y_{n,k}^T \\ Y_{n,k} & 0 \end{pmatrix}, \tag{3.48}$$

and

$$M + h_{n'}B = \begin{pmatrix} 0 & Y_{n'}^T \\ Y_{n'} & 0 \end{pmatrix}, \tag{3.49}$$

then $\begin{pmatrix} 0 \\ U_{n,k}^i \end{pmatrix}$ and $\begin{pmatrix} V_{n,k}^i \\ 0 \end{pmatrix}$ are eigenvectors of M. Similarly $\begin{pmatrix} 0 \\ U_{n'}^i \end{pmatrix}$ and $\begin{pmatrix} V_{n'}^i \\ 0 \end{pmatrix}$ are eigenvectors of $M + h_{n'}B$. Let $L_1 = \begin{pmatrix} 0 \\ U_{n,k}^1 \end{pmatrix}$ and $L_1' = \begin{pmatrix} 0 \\ U_{n'}^1 \end{pmatrix}$, then we know the error between L_1 and L_1' is bounded [20].

$$\begin{aligned} &||L_1 - L_1'||_1 \\ &= ||h_{n'}Z_1 + h_{n'}^2 Z_2 + \cdots||_1 \\ &= ||h_{n'}\Sigma_{j=1}^n t_{j1}L_j + h_{n'}^2\Sigma_{j=1}^n t_{j2}L_j + \cdots||_1 \\ &= |h_{n'}| \cdot ||\Sigma_{j=1}^n t_{j1}L_j + h_{n'}\Sigma_{j=1}^n t_{j2}L_j + \cdots||_1. \end{aligned} \tag{3.50}$$

Therefore $||U_{n'} - U_{n,k}||_1$ is bounded and in order of $h_{n'}$. And from Eq. (3.47) and lemma 3, the first term in Eq. (3.46) is also in the order of $h_{n'}$. Hence $||E_{yy}||_1$ is in the order of $h_{n'}$.

3.3.2 Convergence Theorems by Using Forward Euler Integration

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

Convergence Theorem 1: The error $||E_A||_1 = ||A_{n'} - A_n||_1 = O(\ h_{n'})$ by forward Euler integration.

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

$$||E_A||_1 = ||A_{n'} - A_n||_1, \tag{3.51}$$

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

$$= ||A_{n'} - A_{n,k} + A_{n,k} - A_n||_1, \tag{3.52}$$

by triangle inequality and Eq. (3.22, 3.25):

$$\begin{aligned} &\leq ||A_{n'} - A_{n,k}||_1 + ||A_{n,k} - A_n||_1 \\ &= ||\tfrac{1}{h_{n'}} X_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - \tfrac{1}{h_n} X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ &+ ||\tfrac{1}{h_n} X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1} - \tfrac{1}{h_n} X_n Y_n^T (Y_n Y_n^T)^{-1}||_1, \end{aligned} \tag{3.53}$$

the second term is zero by lemma 1:

$$\begin{aligned} &= ||\tfrac{1}{h_{n'}} X_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - \tfrac{1}{k h_{n'}} X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ &= |\tfrac{1}{k h_{n'}}|||k X_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1, \end{aligned} \tag{3.54}$$

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

$$\begin{aligned} = |\tfrac{1}{k h_{n'}}|||&k X_{n'} Y_{n'}^T (Y_{n'} Y_{n'}^T)^{-1} - k X_{n'} Y_{n'}^T (Y_{n,k} Y_{n,k}^T)^{-1} \\ &+ k X_{n'} Y_{n'}^T (Y_{n,k} Y_{n,k}^T)^{-1} - X_{n,k} Y_{n,k}^T (Y_{n,k} Y_{n,k}^T)^{-1}||_1, \end{aligned}$$

by triangle inequality:

$$\begin{aligned} \leq |\tfrac{1}{k h_{n'}}|[&|k|||X_{n'} Y_{n'}^T||_1 \cdot ||(Y_{n'} Y_{n'}^T)^{-1} - (Y_{n,k} Y_{n,k}^T)^{-1}||_1 \\ &+ ||k X_{n'} Y_{n'}^T - X_{n,k} Y_{n,k}^T||_1 \cdot ||(Y_{n,k} Y_{n,k}^T)^{-1}||_1], \end{aligned}$$

from singular value decomposition:

$$\begin{aligned} &= |\tfrac{1}{h_{n'}}|||X_{n'} Y_{n'}^T||_1 \cdot ||U_{n'} \tfrac{1}{W_{n'}^2} U_{n'}^T - U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 \\ &+ |\tfrac{1}{k h_{n'}}| \cdot ||k X_{n'} Y_{n'}^T - k X_{n'} Y_{n,k}^T + k X_{n'} Y_{n,k}^T - X_{n,k} Y_{n,k}^T||_1 \cdot ||U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 \\ &= |\tfrac{1}{h_{n'}}|||X_{n'}||_1 \cdot ||Y_{n'}^T||_1 \cdot ||U_{n'} \tfrac{1}{W_{n'}^2} U_{n'}^T - U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 \\ &+ |\tfrac{1}{h_{n'}}| \cdot ||X_{n'}||_1 \cdot ||Y_{n'}^T - Y_{n,k}^T||_1 \cdot ||U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1 \\ &+ |\tfrac{1}{k h_{n'}}| \cdot ||k X_{n'} - X_{n,k}||_1 ||Y_{n,k}^T|| \cdot ||U_{n,k} \tfrac{1}{W_{n,k}^2} U_{n,k}^T||_1. \end{aligned} \tag{3.55}$$

From Lemma 2 and 5, $||X_{n'}||_1$ and $||(Y_{n'}Y_{n'}^T)^{-1} - (Y_{n,k}Y_{n,k}^T)^{-1}||_1 = ||U_{n'}\frac{1}{W_{n'}^2}U_{n'}^T - U_{n,k}\frac{1}{W_{n,k}^2}U_{n,k}^T||_1$ are both in the order of $h_{n'}$ so that the first term in Eq. (3.55) is in the order of $h_{n'}$. From lemma 3 , $||Y_{n'}^T - Y_{n,k}^T||_1$ is in the order of $h_{n'}$ too so that the second term is in the order of $h_{n'}$. From lemma 4, $||kX_{n'} - X_{n,k}||_1$ is in the order of $h_{n'}^2$ so that the third term is in the order of $h_{n'}$. Therefore, the error $||E_A||_1$ is in the order of $h_{n'}$, meaning the parameter matrix converges with $h_{n'}$.

Now we give a numerical example on the convergence of the modeling

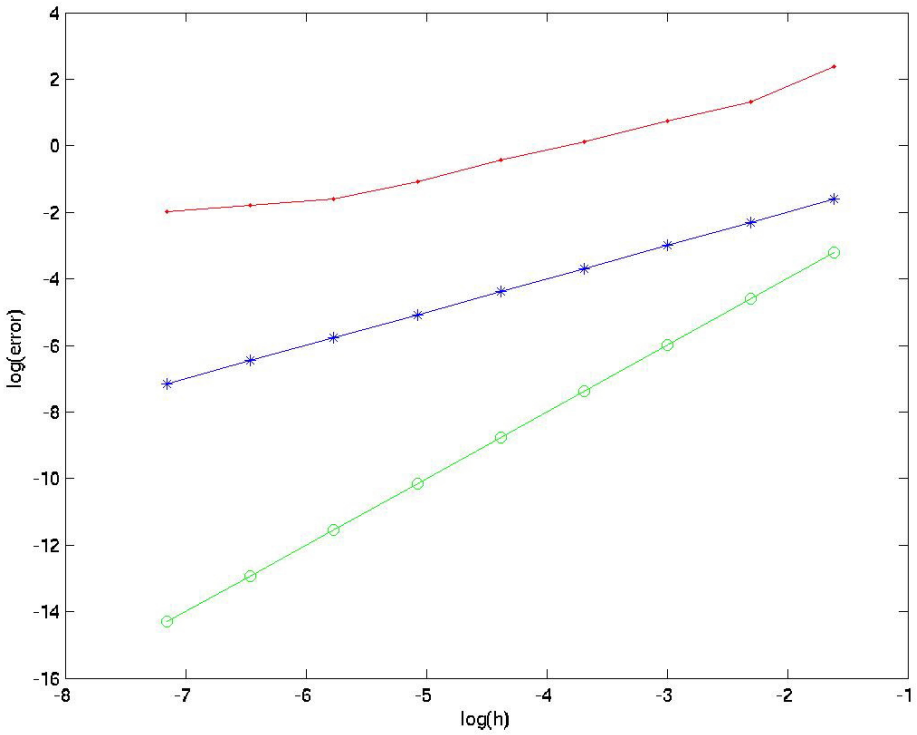

Figure 3.1: Example on Rössler System: The Errors between the Modeling Parameters A_n Are in Order of h_n. This is a loglog plot of the errors $E = ||A_n - A_{n'}||_1$ vs. different sampling rate h_n. The slope of the curve(.-) is almost 1, the fact of which implies the error is in order of h_n.

parameter A_n w.r.t. different sampling rate h_n by applying a multivariate time series data set $\{u^{(n)}\}_{n=0:M}$ generated from the system of Rössler oscillator [2] by Matlab. We present a graph to show and validate the convergence with $h_{n'}$. Figure (3.1) shows the error between the modeling parameters A_n with h_n. The red line(.-) is $log(error)$ vs $log(h_n)$. The lines (*-) and (o-) are $log(h_n)$ vs $log(h_n)$ and $2log(h_n)$ vs $log(h_n)$, respectively. From the graph, we see the red line is almost parallel to the blue line, the trend of which means the error is in order of h_n, not h_n^2.

From the proofs above, we know that not only the error of the parameter matrix A_n is decreasing with h_n, but also the size of the data matrices used to calculate A_n 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_n 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 that the parameter obtained by skip-only data matrices is equal to A_{ns} obtained by skip-repeated matrices.

Theorem: Let $A_{ns} = [A_n(:,1), \cdots, A_n(:,1), A_n(:,s+1), \cdots, A_n(:, (n-1)s+1)]$**, then** $||A_{ns} - A_n||_1 = \mathcal{O}(h)$.

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

part:

$$||A_{ns} - A_n||_1,$$

by definitions in Eq. (3.22):

$$\begin{aligned} &= ||\tfrac{1}{h_n} X_{ns} Y_{ns}^T (Y_{ns} Y_{ns}^T)^{-1} - \tfrac{1}{h_n} X_n Y_n^T (Y_n Y_n^T)^{-1}||_1 \\ &= |\tfrac{1}{h_n}| \cdot ||X_{ns} Y_{ns}^T (Y_{ns} Y_{ns}^T)^{-1} - X_n Y_n^T (Y_n Y_n^T)^{-1}||_1, \end{aligned}$$

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

$$\begin{aligned} &= |\tfrac{1}{h_n}| \cdot ||X_{ns} Y_{ns}^T (Y_{ns} Y_{ns}^T)^{-1} - X_{ns} Y_{ns}^T (Y_n Y_n^T)^{-1} \\ &\quad + X_{ns} Y_{ns}^T (Y_n Y_n^T)^{-1} - X_n Y_n^T (Y_n Y_n^T)^{-1}||_1, \end{aligned}$$

by triangle inequality:

$$\begin{aligned} &\leq |\tfrac{1}{h_n}| \cdot [||X_{ns} Y_{ns}^T (Y_{ns} Y_{ns}^T)^{-1} - X_{ns} Y_{ns}^T (Y_n Y_n^T)^{-1}||_1 \\ &\quad + ||X_{ns} Y_{ns}^T (Y_n Y_n^T)^{-1} - X_n Y_n^T (Y_n Y_n^T)^{-1}||_1] \\ &= |\tfrac{1}{h_n}| [||X_{ns} Y_{ns}^T||_1 \cdot ||(Y_{ns} Y_{ns}^T)^{-1} - (Y_n Y_n^T)^{-1}||_1 \\ &\quad + ||X_{ns} Y_{ns}^T - X_n Y_n^T||_1 \cdot ||(Y_n Y_n^T)^{-1}||_1], \end{aligned}$$

by singular value decomposition:

$$\begin{aligned} &= |\tfrac{1}{h_n}| [||X_{ns} Y_{ns}^T||_1 \cdot ||U_{ns} \tfrac{1}{W_{ns}^2} U_{ns}^T - U_n \tfrac{1}{W_n^2} U_n^T||_1 \\ &\quad + ||X_{ns} Y_{ns}^T - X_{ns} Y_n^T + X_{ns} Y_n^T - X_n Y_n^T||_1 \cdot ||(Y_n Y_n^T)^{-1}||_1] \\ &\leq |\tfrac{1}{h_n}| [||X_{ns} Y_{ns}^T||_1 \cdot ||U_{ns} \tfrac{1}{W_{ns}^2} U_{ns}^T - U_n \tfrac{1}{W_n^2} U_n^T||_1 \\ &\quad + (||X_{ns}||_1 \cdot ||Y_{ns}^T - Y_n^T||_1 + ||X_{ns} - X_n||_1 \cdot ||Y_n^T||_1) \cdot ||(Y_n Y_n^T)^{-1}||_1]. \end{aligned}$$

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

3.3.3 Higher Order Convergence Theorem

In fact, the order of convergence in convergence theorem 1 depends on the fact that the scheme used in Eq. (3.14) is $\mathcal{O}(h)$. If instead, we use higher order schemes to approximate $\dot{X}$ in Eq. (3.4) 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 time. 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 are:

$$L(X(t))_5 = a_1X(t+2)+a_2X(t+1)+a_3X(t)+a_4X(t-1)+a_5X(t-2), \quad (3.56)$$

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:

$$\begin{aligned} L(X(t))_5 &= a_1(X(t) + 2hX'(t) + \tfrac{(2h)^2}{2!}X''(t) + \tfrac{(2h)^3}{3!}X'''(t) + \ldots) \\ &+ a_2(X(t) + hX'(t) + \tfrac{h^2}{2!}X''(t) + \tfrac{h^3}{3!}X'''(t) + \ldots) \\ &+ a_3X(t) \\ &+ a_4(X(t) - hX'(t) + \tfrac{h^2}{2!}X''(t) - \tfrac{h^3}{3!}X'''(t) + \ldots) \\ &+ a_5(X(t) - 2hX'(t) + \tfrac{(2h)^2}{2!}X''(t) - \tfrac{(2h)^3}{3!}X'''(t) + \ldots). \end{aligned} \quad (3.57)$$

In this section, we will use 2nd order, 3rd order and 4th order schemes for the purpose of comparison with the Euler integration. The coefficients in Eq. (3.57) 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,

   $$X'(t) = \frac{X(t+1) - X(t)}{h} + \mathcal{O}(h). \quad (3.58)$$

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

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

therefore,

$$X'(t) = \frac{X(t+1) - X(t-1)}{2h} + \mathcal{O}(h^2). \tag{3.59}$$

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,

$$X'(t) = \frac{X(t+2) - 2X(t+1) + 9X(t) - 10X(t-1) + 2X(t-2)}{6h} + \mathcal{O}(h^3). \tag{3.60}$$

4. Forth 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,

$$X'(t) = \frac{X(t+2) - 8X(t+1) + 8X(t-1) - X(t-2)}{-12h} + \mathcal{O}(h^4). \tag{3.61}$$

These higher order schemes are not unique w.r.t. the coefficients a_1, a_2, a_3, a_4, and a_5. From the Eqs. (3.58,3.59,3.60,3.61), we know that the error between $X'(t)$ and the approximations are in higher order of h.

Figure (3.2) shows the convergence of the parameters by applying higher order schemes to the time series data set $\{u^{(n)}\}_{n=0:M}$ generated from the system of

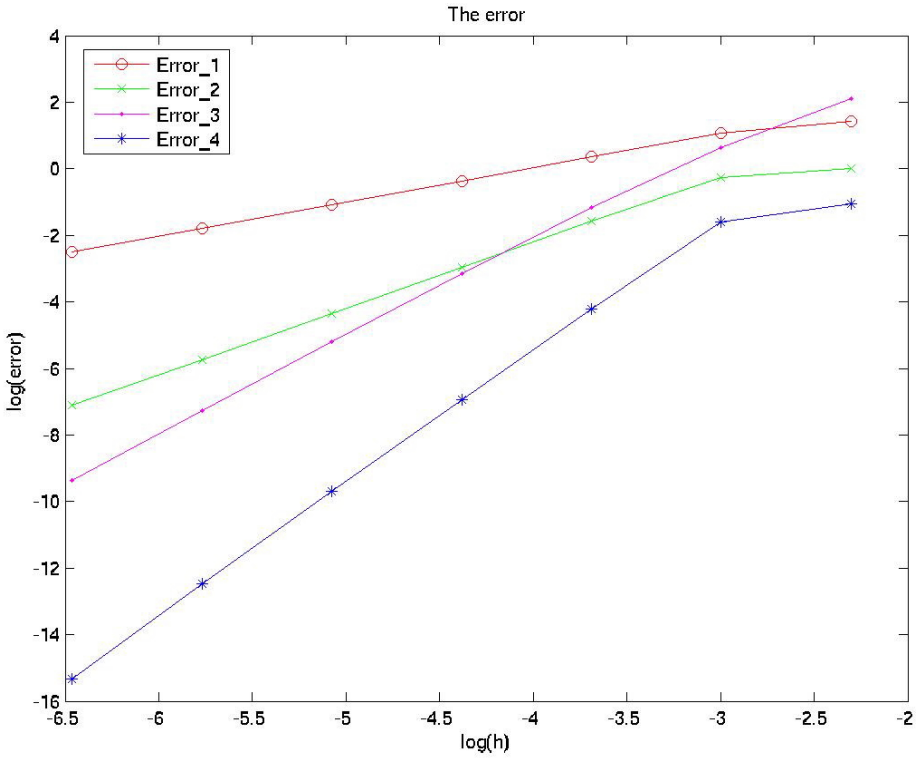

Figure 3.2: Example on Rössler System: The Errors between the Modeling Parameters A_n by Higher Order Schemes. This is a loglog plot of the errors $E = ||A_n - A_{n'}||_1$ vs. different sampling rate h_n by different schemes. The slopes of curves (o-), ($\times$-), (.-), and (*-) are approximating 1, 2, 3 and 4, respectively, the fact of which implies the convergence rates of the parameters depend on the schemes used in the modeling procedure.

Rössler oscillator [2] by Matlab. The slopes of the lines (o-), (×-), (.-), and (*-) are approximating 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 rates of the parameters depend on the order of the scheme used, although our method of rigorous proof is only allowed for the explicit Euler scheme.

3.4 Conclusions

In this chapter, we have introduced a global modeling technique, which is based on forward Euler integration and high order schemes, and investigated the errors among the parameter matrices given by the modeling technique, and also given the convergence theorems of the parameters. To improve this global modeling technique, we can try to use some other functions, such as trigonometric functions, as the basis functions for a better fitting of the systems that have some periodic properties. Once a good model, or specifically the parameters of the assumed general model form, is obtained, we will furthermore show a new method to uncover underlying structure of the system in next chapter.

Chapter 4
Structure Investigation

A common situation in multivariate dynamical data sets is that there exists an underlying structure, by which we mean that there is a repetition apparent in the underlying equations, with coupling between the repeated elements. Evidently, 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 [55] to the dynamical systems setting.

4.1 Kronecker Product and Its Properties

The definition of the Kronecker product and its properties will be given as the following,

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

$$C = A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,p}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,p}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1}B & a_{n,2}B & \cdots & a_{n,p}B \end{bmatrix}. \tag{4.1}$$

is called the **Kronecker product** of A and B. It is also called the direct product or the tensor product. [55]

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. $(A \otimes B)^T = A^T \otimes B^T, (A \otimes B)^H = A^H \otimes B^H$.

The property we are interested in is that for any 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

$$||A - B \otimes C||_F^2, \tag{4.2}$$

can be minimized, where $|| \cdot ||$ denotes the Frobenius norm, $B \in R^{m_1 \times n_1}$, $C \in R^{m_2 \times n_2}$. This problem is investigated intensively in [56], 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. (4.2).

4.2 N-Coupled Rössler Oscillators

To give a good example on discovery of the underlying structure of the unknown system by investigating the parameters from the modeling procedure, we consider

the system of N coupled Rössler oscillators [2] with periodic boundary conditions:

$$\begin{aligned} \dot{x_i} &= -\omega_i y_i - z_i + \epsilon(x_{i+1} + x_{i-1} - 2x_i) \\ \dot{y_i} &= \omega_i x_i + a y_i \\ \dot{z_i} &= b + z_i(x_i - c), \end{aligned} \tag{4.3}$$

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 Rössler 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]. In this part, for the purpose of illustration, we choose the relatively small array, $N = 3$. We construct a vector X as the following,

$$X = [x_1, y_1, z_1, \ldots, x_3, y_3, z_3]^T. \tag{4.4}$$

Since the global modeling technique gives us an approximation parameter to the exact parameter, we investigate the original system to get some *a prior* expectation about the underlying structure of the system of coupled Rössler oscillators.

Rewriting this system of 3-coupled Rössler oscillators as the following,

$$\dot{X} = P_0 + P_1 X + P_2 \begin{bmatrix} x_1 z_1 \\ x_2 z_2 \\ x_3 z_3 \end{bmatrix}, \tag{4.5}$$

where

$$P_0 = \begin{bmatrix} 0 \\ 0 \\ b \\ 0 \\ 0 \\ b \\ 0 \\ 0 \\ b \end{bmatrix}, \; P_1 = \begin{bmatrix} -2\epsilon & -1 & -1 & \epsilon & 0 & 0 & \epsilon & 0 & 0 \\ 1 & a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c & 0 & 0 & 0 & 0 & 0 & 0 \\ \epsilon & 0 & 0 & -2\epsilon & -1 & -1 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -c & 0 & 0 & 0 \\ \epsilon & 0 & 0 & \epsilon & 0 & 0 & -2\epsilon & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c \end{bmatrix}, \; P_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.6}$$

Hence, the basis functions of Eq. (4.5) are constant function, linear function and quadratic function. The parameters of this system are P_0, P_1 and P_2. 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. (4.7).

$$\dot{X} = P \begin{bmatrix} \phi_1(X) \\ \phi_2(X) \\ \phi_3(X) \end{bmatrix}, \tag{4.7}$$

where ϕ_1, ϕ_2 and ϕ_3 are constant, linear and quadratic functions of X, respectively.

$$\phi_1(X) = 1, \phi_2(X) = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ \vdots \\ x_3 \\ y_3 \\ z_3 \end{bmatrix}, \phi_3(X) = \begin{bmatrix} x_1 z_1 \\ x_2 z_2 \\ x_3 z_3 \end{bmatrix}.$$

The parameter matrix $P = [P_0, P_1, P_2]$ is the unknown we try to determine in the modeling procedure.

Now suppose we are given a multivariate time series data set $\{u^{(n)}\}_{n=0:M}$ from the system of Rössler oscillators and do not know the exact functional form of the vector field. Modeling procedure gives us an approximation A to the exact parameter P via some integration schemes. According to the system of Rössler oscillators, quadratic polynomials are proper to be chosen as the basis functions. In the following sections we will focus on the exact parameter P instead, since we can assume that the error between the modeling parameter A and the system parameter P is small enough to ignore when discussing the underlying structure of the system.

4.3 Kronecker Product Representation

From Eq. (4.1), we know Kronecker products copy the matrix B and multiply it by a constant from the matrix A to produce a new matrix C. 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 particularly in the aspect that this system is suspected to have coupled copies of nearly identical chaotic oscillators.

In this section, we will analyze the structure of the parameters of the system of N-coupled Rössler 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. (4.5). For the parameters of the higher

terms, we can use a similar method of analysis.

$$P_1 = \begin{bmatrix} -2\epsilon & -1 & -1 & \epsilon & 0 & 0 & \epsilon & 0 & 0 \\ 1 & a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c & 0 & 0 & 0 & 0 & 0 & 0 \\ \epsilon & 0 & 0 & -2\epsilon & -1 & -1 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -c & 0 & 0 & 0 \\ \epsilon & 0 & 0 & \epsilon & 0 & 0 & -2\epsilon & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c \end{bmatrix}. \tag{4.8}$$

Here the parameter P_1 is obtained by sorting the ODEs in the way of Eq. (4.4). Let

$$R = \begin{bmatrix} -2\epsilon & -1 & -1 \\ 1 & a & 0 \\ 0 & 0 & -c \end{bmatrix}. \tag{4.9}$$

We notice there are repeating Rs in the parameter matrix P_1. And when $\epsilon = 0$, a good Kronecker product decomposition of P_1 can be identified, according to the definition of Kronecker product.

$$P_1|_{\epsilon=0} = \begin{bmatrix} R & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & R \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes R = B \otimes R. \tag{4.10}$$

Generally, for a small nonzero value of ϵ, the left side matrix B in the Eq. (4.10) 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.

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

As discussed in the previous part, the P_1 is generated when we set up $X = [x_1, y_1, z_1, \ldots, x_3, y_3, z_3]^T$. If we change the order of the elements in X in the

following way:

$$X = [x_1, x_2, x_3, y_1, \ldots, z_2, z_3]^T, \tag{4.11}$$

the system can be written as:

$$\dot{X} = P_0' + P_1' X + P_2' \begin{bmatrix} x_1 z_1 \\ x_2 z_2 \\ x_3 z_3 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ b \\ b \\ b \end{bmatrix} + \begin{bmatrix} -2\epsilon & \epsilon & \epsilon & -1 & 0 & 0 & -1 & 0 & 0 \\ \epsilon & -2\epsilon & \epsilon & 0 & -1 & 0 & 0 & -1 & 0 \\ \epsilon & \epsilon & -2\epsilon & 0 & 0 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & a & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & a & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -c & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -c \end{bmatrix} X + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 z_1 \\ x_2 z_2 \\ x_3 z_3 \end{bmatrix} \tag{4.12}$$

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 X. Therefore how to find a good order of the ODEs becomes important to investigate the underlying structure of the system.

4.3.1 Investigation of Permutation

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

$$A' = MAM^T. \tag{4.13}$$

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

$$A = B \otimes C,\ A' = B' \otimes C',\ M = M_1 \otimes M_2. \tag{4.14}$$

Then A' can be expressed by the Kronecker product representations of A and M,

$$\begin{aligned} &B' \otimes C' = A' = MAM^T = M(B \otimes C)M^T \\ &= (M_1 \otimes M_2)(B \otimes C)(M_1 \otimes M_2)^T, \end{aligned} \tag{4.15}$$

By the property 6 of Kronecker products, we have

$$= (M_1 \otimes M_2)(B \otimes C)(M_1^T \otimes M_2^T),$$

By the property 4 of Kronecker products, the Eq. (4.15) is equivalent to

$$B' \otimes C' = (M_1 B \otimes M_2 C)(M_1^T \otimes M_2^T) = (M_1 B M_1^T) \otimes (M_2 C M_2^T). \tag{4.16}$$

If the sizes of B' and C' are the same as those of $M_1 B M_1^T$ and $M_2 C M_2^T$, B' and C' are the same as $M_1 B M_1^T$ and $M_2 C M_2^T$ because of the uniqueness of the Kronecker product decomposition in 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 summarize with the following proposition.

Proposition 1: If A and A' are similar by a permutation matrix M, as indicated by Eq. (4.13), and A, A', and M have Kronecker decompositions $A = B \otimes C$, $A' = B' \otimes C'$, and $M = M_1 \otimes M_2$ of appropriate dimension as indicated by Eq. (4.14), then $B' = M_1 B M_1^T$ and $C' = M_2 C M_2^T$ may also be permuted in the Kronecker decomposition as shown in Eq. (4.16).

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 matrix $P_{m\times n}$, which can be split into two parts, P_L and P_R, denoting the left part and the right part of P respectively. Here we assume that the matrix P has even columns without loss of generality. Applying the permutation operator M to the matrix P, a new matrix is generated in the following way:

$$M(P)^{(2i-1)} = P_L^{(i)},\ M(P)^{(2i)} = P_R^{(i)},\ i = 1 : \frac{n}{2},\ n \text{ is even}, \tag{4.17}$$

where $M(P)^{(2i-1)}$ is the $(2i-1)$th column in the new matrix $M(P)$, $P_L^{(i)}$ is the ith column in P_L, and $P_R^{(i)}$ is the ith column in P_R. We form M to P by inserting the P_L into P_R so that the columns of P_L become the odd number columns in the new matrix $M(P)$ and the columns of P_R become the even number columns in $M(P)$.

To find a best order of P, we can apply the operator M several times and compare the errors from all permutations of P. In figure (4.1), we show the relationship between the number of times the permutation operator M applied to the matrix P_1 given in Eq. (4.6), and the resulting errors among the Kronecker product representations. We notice that the errors from iterations of the permutation are periodic since, after several permutations, the matrix P_1 goes back to the original one.

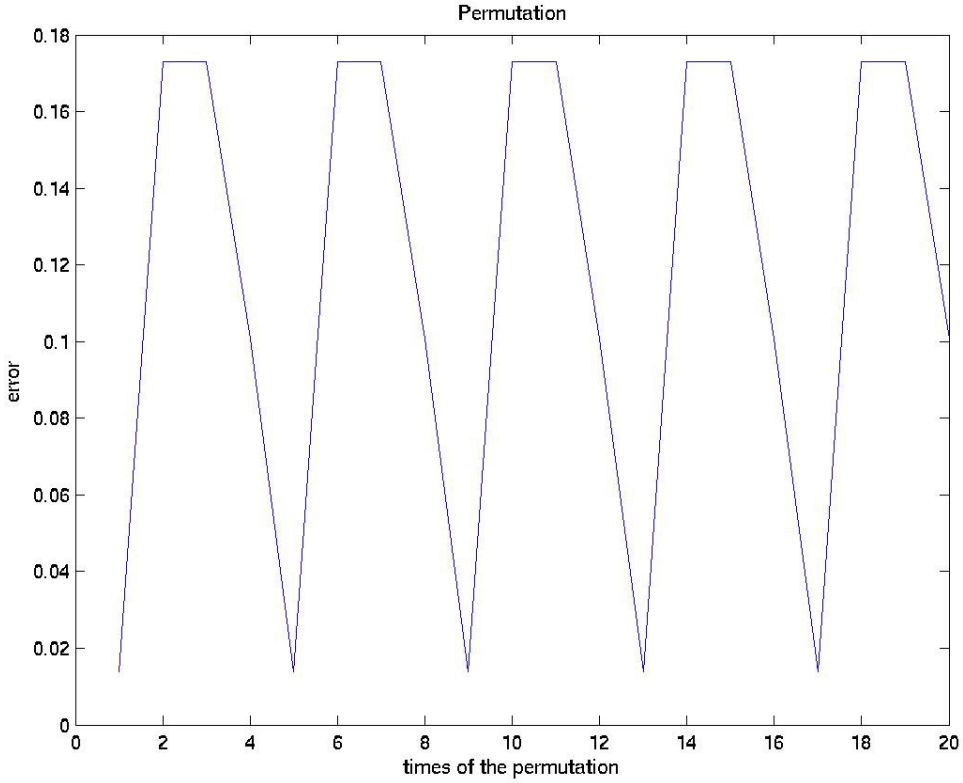

Figure 4.1: Permutations vs. Errors. This is plot of the errors between the $M^n(P_1)$ and its Kronecker product representations in the Frobenius norm. The x-axis is the number of the iterations n, and the y-axis is the values of errors for the different n.

4.3.2 Existence of Kronecker Product Representation

From the previous section, 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.

Definition: The Kronecker product of two vectors: 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 $u \otimes v = [u^T v_1, \ldots, u^T v_r]^T$.

The following theorem about eigenvalues and eigenvectors [55] is very helpful to offer an insight into this problem.

Theorem of eigenvalues and eigenvectors [55]**:** Suppose that $A \in R^{r,r}$ and $B \in R^{s,s}$ and

$$Au_i = \lambda_i u_i, i = 1, \ldots, r, Bv_j = \mu_j v_j, j = 1, \ldots, s, \tag{4.18}$$

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

$$(A \otimes B)(u_i \otimes v_j) = \lambda_i \mu_j (u_i \otimes v_j),$$
$$(A \otimes I_s + I_r \otimes B)(u_i \otimes v_j) = (\lambda_i + \mu_j)(u_i \otimes v_j).$$

Thus the eigenvalues of a Kronecker product (sum) are the products (sums) of the eigenvalues of the factors (decompositions). 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 $C = A \otimes B$. Furthermore, the eigenvalues of $C = A \otimes B$ are structured in the following way: The eigenvalues of C 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 (eigenvalues) 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. (4.8), 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

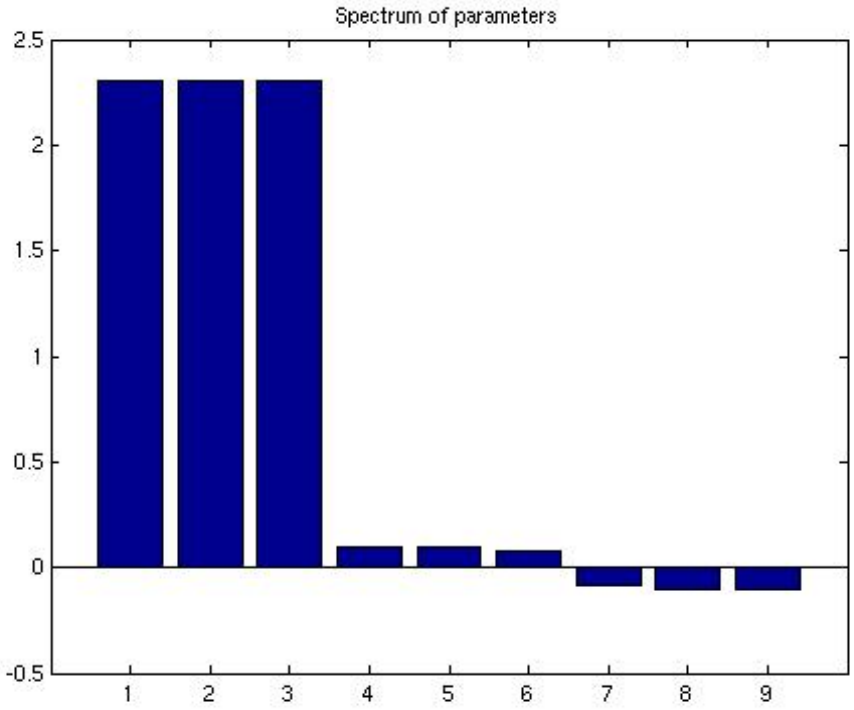

Figure 4.2: Spectrum of the Parameter Matrix P_1 of the 3-Coupled Rössler System. There are three groups of the close by values in this plot.

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. (4.10) 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.

Figure (4.2) shows the spectrum of the fitted parameter of the 3 coupled Rössler system. In Fig. (4.2), three groups of the values of the spectrum can be seen. From this fact, we can suspect that there exists a good Kronecker product

representation of the matrix P_1. Therefore, with modeling parameter of the system, we are able to uncover the underlying structure of the system.

4.4 Reaction Diffusion Equations

While the theory of modeling given in Chapter 3 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.

Consider the reaction diffusion equations of two species and one spatial variable, $u(x,t), v(x,t) \in \Re$, $x \in [0, L]$ [54]:

$$\begin{array}{l} \frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + \frac{1}{\epsilon}[v - f(u)] \\ \frac{\partial v}{\partial t} = D\frac{\partial^2 v}{\partial x^2} - u + \alpha \end{array}, \tag{4.19}$$

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

$$u(0,t) = -2,\ u(L,t) = -2,\ \text{and}\ v(0,t) = -4,\ v(L,t) = -4. \tag{4.20}$$

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

$$f(u) = u^2 + u^3, \tag{4.21}$$

and $D = 0.032249$, $\alpha = 0.01$, $\epsilon = 0.01$. We obtain an empirical data set $\{u(x_i, t_j), v(x_i, t_j)\}_{i=1:M, j=1:N}$ by a finite element solver [34].

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,

$$\begin{bmatrix} \dot{u} \\ v \end{bmatrix} = A \begin{bmatrix} u \\ v \\ Q(u,v) \\ C(u,v) \end{bmatrix}, \tag{4.22}$$

where we assume $Q(u,v)$ and $C(u,v)$ are all possible combinations of quadratic and cubic terms. Figure (4.3, 4.4) shows the u and v data sets vs. the reproduced data sets u' and v' generated by modeling system with the modeling parameter A in Eq. 4.22.

The left graphs are the original data of u, v and the right graphs are the reproduced data u', v'. From the figures we know, the fitting data procedure works well and the error $||u - u'||_1$ is 4.3121×10^{-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 will still get a nice graph.Figure. (4.5) shows the original data u and reproduced u' obtained by original u only. Although the graph looks good, the error $||u - u'||_1 = 4.1836$ 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

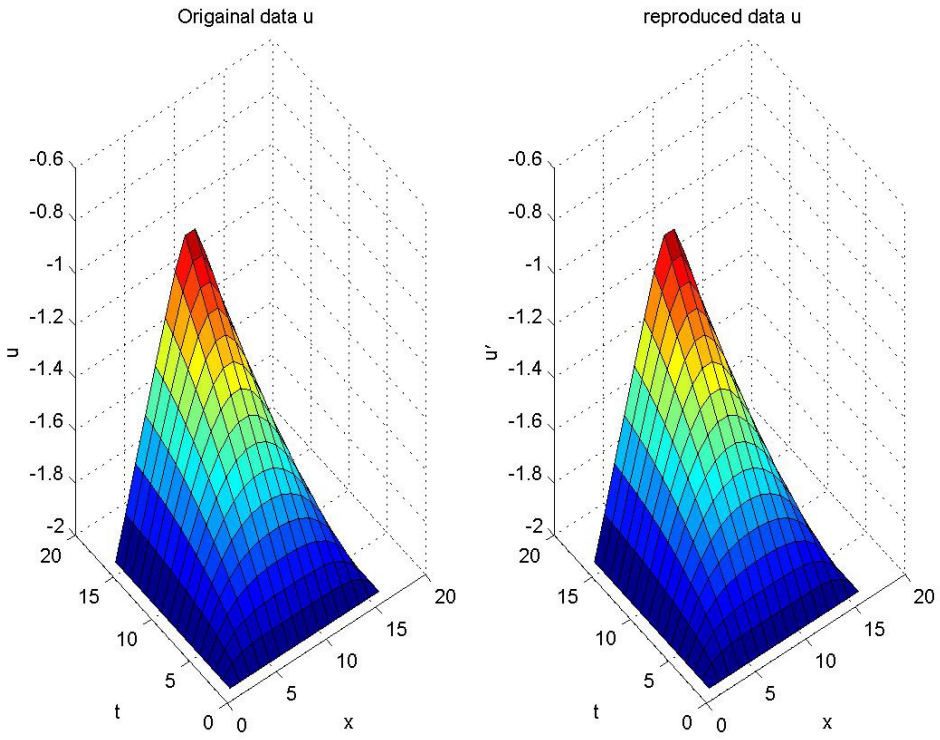

Figure 4.3: The Original Data $u(x,t)$ from the Reaction Equation and the Reproduced Data $u'(x,t)$ from the Modeling System of Best Cubic Fitting.

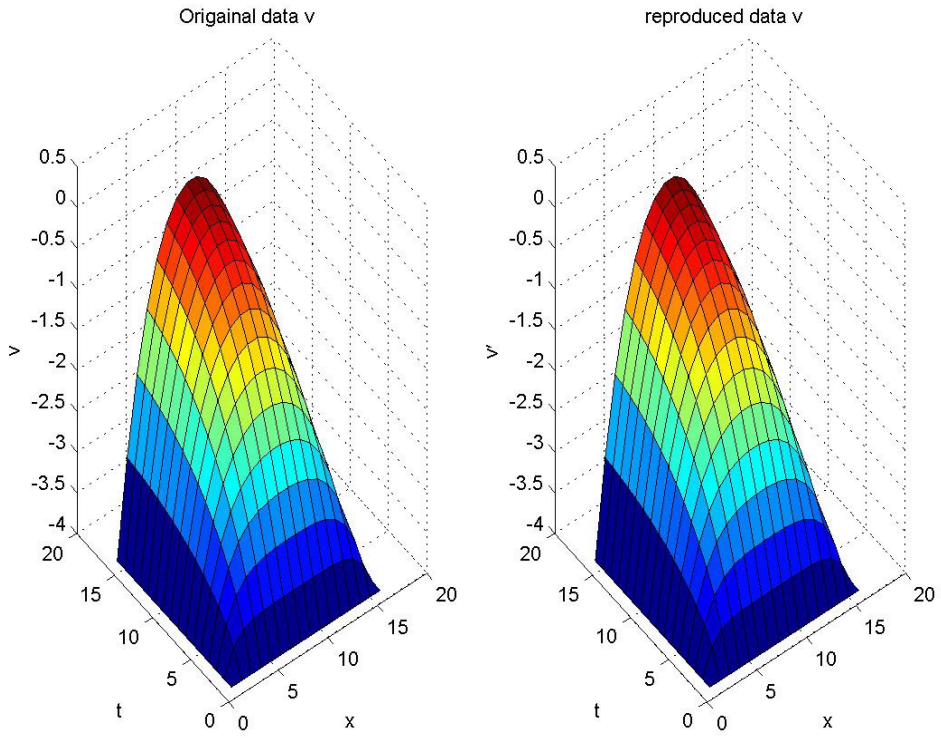

Figure 4.4: The Original Data $v(x,t)$ from the Reaction Equation and the Reproduced Data $v'(x,t)$ from the Modeling System of Best Cubic Fitting.

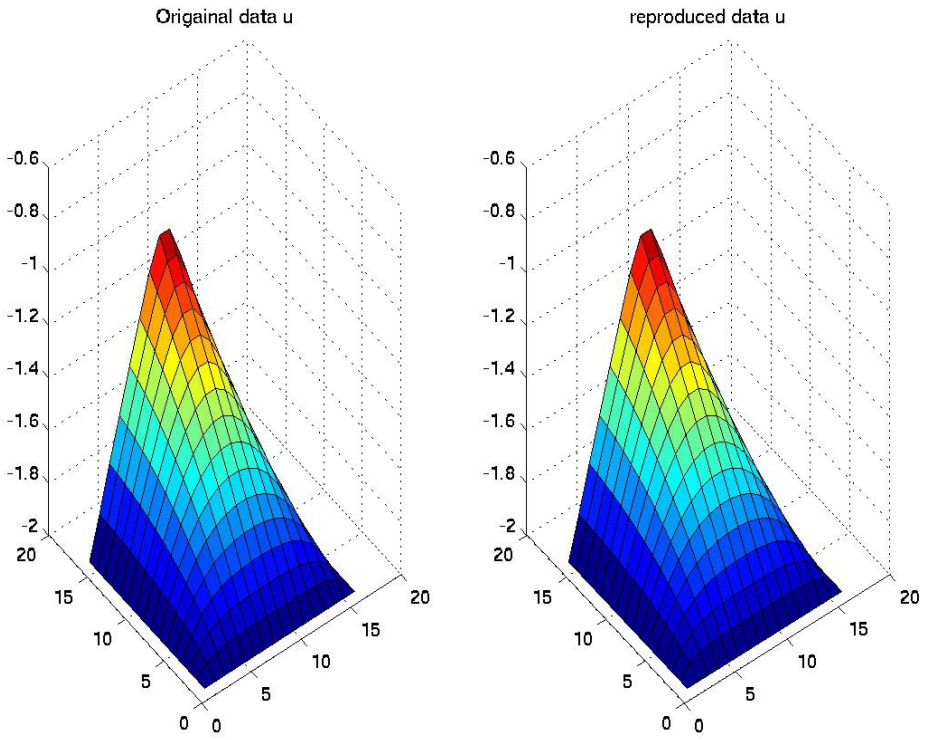

Figure 4.5: The Original Data $u(x,t)$ and the Reproduced Data of the Reaction Diffusion Equation by Fitting Only the $u(x,t)$ Data.

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.

$$\frac{1}{h_t} \cdot \begin{bmatrix} u(x_1, t_{j+1}) - u(x_1, t_j) \\ v(x_1, t_{j+1}) - v(x_1, t_j) \\ u(x_2, t_{j+1}) - u(x_2, t_j) \\ v(x_2, t_{j+1}) - v(x_2, t_j) \\ u(x_3, t_{j+1}) - u(x_3, t_j) \\ v(x_3, t_{j+1}) - v(x_3, t_j) \end{bmatrix} = A \cdot \begin{bmatrix} u(x_1, t_j) \\ v(x_1, t_j) \\ u(x_2, t_j) \\ v(x_2, t_j) \\ u(x_3, t_j) \\ v(x_3, t_j) \\ Q(u, v) \\ C(u, v) \end{bmatrix}, \tag{4.23}$$

here A is :

$$A = \begin{bmatrix} -2a & 1/\epsilon & a & & a & & \\ -1 & -2a & & a & & a & \\ a & & -2a & 1/\epsilon & a & & A_1' \\ & a & -1 & -2a & & a & \\ a & & a & & -2a & 1/\epsilon & \\ & a & & a & -1 & -2a & \\ & & & & & & \end{bmatrix}, \tag{4.24}$$

where A_1' is the parameter of the nonlinear terms $Q(u, v)$ and $C(u, v)$,

$$Q(u, v) = \begin{bmatrix} u(x_1, t_j) * u(x_1, t_j) \\ u(x_1, t_j) * u(x_2, t_j) \\ u(x_1, t_j) * u(x_3, t_j) \\ \vdots \\ v(x_2, t_j) * v(x_3, t_j) \\ v(x_3, t_j) * v(x_3, t_j) \end{bmatrix}, C(u, v) = \begin{bmatrix} u^3(x_1, t_j) \\ u^2(x_1, t_j)u(x_2, t_j) \\ u(x_1, t_j)u^2(x_2, t_j) \\ \vdots \\ v^2(x_2, t_j) * v(x_3, t_j) \\ v^3(x_3, t_j) \end{bmatrix}$$

and

$$a = D/h_x^2, \tag{4.25}$$

since we use the centered difference approximation to model the data set,

$$\frac{\partial^2 u(x_i, t)}{\partial x^2} \approx \frac{u(x_{i-1}, t) - 2u(x_i, t) + u(x_{i+1,t})}{h_x^2}. \tag{4.26}$$

Here we only discuss the parameter of the linear term, which is given explicitly in

Eq. (4.24). Two repeating blocks in the matrix A can be noticed :

$$B_1 = \begin{bmatrix} -2a & 1/\epsilon \\ -1 & -2a \end{bmatrix}, B_2 = \begin{bmatrix} a & \\ & a \end{bmatrix} \tag{4.27}$$

When a is very small, implying the D in Eq. (4.25) is much smaller than h_x^2, 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 resort the data, A', another form of A could be found according to a new order,

$$\frac{1}{h_t} \cdot \begin{bmatrix} u(x_1, t_{j+1}) - u(x_1, t_j) \\ u(x_2, t_{j+1}) - u(x_2, t_j) \\ \vdots \\ u(x_m, t_{j+1}) - u(x_m, t_j) \\ v(x_1, t_{j+1}) - v(x_1, t_j) \\ \vdots \\ v(x_m, t_{j+1}) - v(x_m, t_j) \end{bmatrix} = A' \cdot \begin{bmatrix} u(x_1, t_j) \\ u(x_2, t_j) \\ \vdots \\ u(x_m, t_j) \\ v(x_1, t_j) \\ \vdots \\ v(x_m, t_j) \\ Q(u, v) \\ C(u, v) \end{bmatrix}, \tag{4.28}$$

here A' can be written as :

$$A' = \begin{bmatrix} -2a & a & & & a & 1/\epsilon & & & & & \\ a & -2a & a & & & & 1/\epsilon & & & & \\ \vdots & & & & & & \vdots & \vdots & & & \\ a & & & a & -2a & & & & 1/\epsilon & & \\ -1 & & & & & -2a & a & & a & & A'_1 \\ & -1 & & & & a & -2a & a & & & \\ \vdots & & & \vdots & & & & & & & \\ & & & & -1 & a & & & a & -2a & \end{bmatrix}, \tag{4.29}$$

Investigating the A' given in Eq. (4.29), 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. (4.28).

4.5 Summary

In this chapter, we have given an example of the system of coupled Rössler oscillators to answer the question: how can we know if there are coupled oscillators in an unknown system by investigating the modeling parameters? Based on a recent matrix approximation theory by Van Loan [56], we have introduced an optimal estimation within the classes of systems, which can be written in terms of the Kronecker product [55], the representation of which uncovers subsystem structure in a large system consisting of many coupled simpler elements.

Chapter 5
Data Reproduction by Modeling Systems

In this chapter, we will discuss the important issue of model quality: How should we evaluate the error between the original unknown system and the modeling system? Consequently, the question of how well the model reproduces the data has to be addressed. 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 into 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 closely to the original system, that sensitive dependence to initial conditions will cause exponential divergence of a prediction from a true solution. Thus, we will incorporate this notion into evaluations of quality.

5.1 Error Evaluation

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

1. Residual error of the data reproduced by the modeling ODE system as compared to the given data.

2. How close is the right hand side of the modeling ODE to the original ODE?

The first type of error is reasonable in the sense that the fitting criteria are based on least squares minimum residual. The second type of error is what we really want to reduce, however, it is not possible to directly compute it since we do not have the function that produces the original data set. The theorems from the previous chapters concern convergence of the modeling ODE with respect to decreasing the sampling rate, but we cannot directly compare $\mathbf{f}$, the true ODE and $\widetilde{\mathbf{f}}$, the modeling ODE.

5.1.1 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 $\widetilde{\mathbf{f}}$ to be very close to the true $\mathbf{f}$ which produces the data set, if the ODE is for example autonomous and chaotic, the sensitive dependence can cause the outputs of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ and $\dot{\mathbf{x}} = \widetilde{\mathbf{f}}(\mathbf{x})$ to produce very different solutions.

Consider for example, the Lorenz system [2],

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= -xz + \gamma x - y \\ \dot{z} &= xy - bz. \end{aligned} \tag{5.1}$$

Now suppose we have a data set from the Lorenz system, and with that data set we have done an exceedingly good job of modeling the equations of motion. Suppose we get the modeling equations in the following form,

$$\begin{aligned} \dot{x} &= (\sigma + \epsilon)(y - x) \\ \dot{y} &= -xz + \gamma x - y \\ \dot{z} &= xy - bz. \end{aligned} \tag{5.2}$$

By any standard, this should be considered to be a good 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

$$\dot{x} = f(x) + \epsilon g(x, t, \epsilon), \qquad x \in R^n, \tag{5.3}$$

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. (5.3)) with the initial condition x_0', we will be mainly interested in the following question: In comparison with the unperturbed system, how the perturbed initial condition and perturbation parameter ϵ effect the solution of the perturbed system?

From the theorem of dependence on initial condition [37], based on Gromwall's inequality, we know that,

$$||x(x_0, t) - x(x_0', t)|| \leq ||x_0 - x'||^{Kt}, \tag{5.4}$$

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 two systems are the same, then,

$$||x_\epsilon(x_0, t) - x(x_0, t)|| \leq \frac{\epsilon M}{K}[e^{Kt} - 1], \tag{5.5}$$

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,

$$\begin{aligned} & ||x_\epsilon(x_0', t) - x(x_0, t)|| \qquad (5.6) \\ \leq \; & ||x_\epsilon(x_0', t) - x(x_0', t)|| + ||x(x_0', t) - x(x_0, t)|| \\ \leq \; & \frac{\epsilon M}{K}[e^{Kt} - 1] + ||x_0 - x_0'||^{Kt} \end{aligned}$$

Consider the Lorenz system [2], and the “very-well-fitted" version of that system as above in Eqs. (5.1, 5.2). 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 the $\epsilon = 0.01$, and the difference between the initial conditions $x_0' - x_0 = 0.01$. The Lipschitz constant K for the x oscillator is found analytically to be,

$$K = |\sigma|. \qquad (5.7)$$

Figure 5.1 shows the time series of two systems with the chosen perturbations. In figure 5.1, we see time-series produced by the two systems Eq. (5.1) and Eq. (5.2) respectively. The dashed curve is the time-series data from the unperturbed system and the solid 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 two systems is initially very small. Eventually, the two curves diverge from each other. Fig. 5.2 shows a loglog plot of the error between the x oscillators of two systems

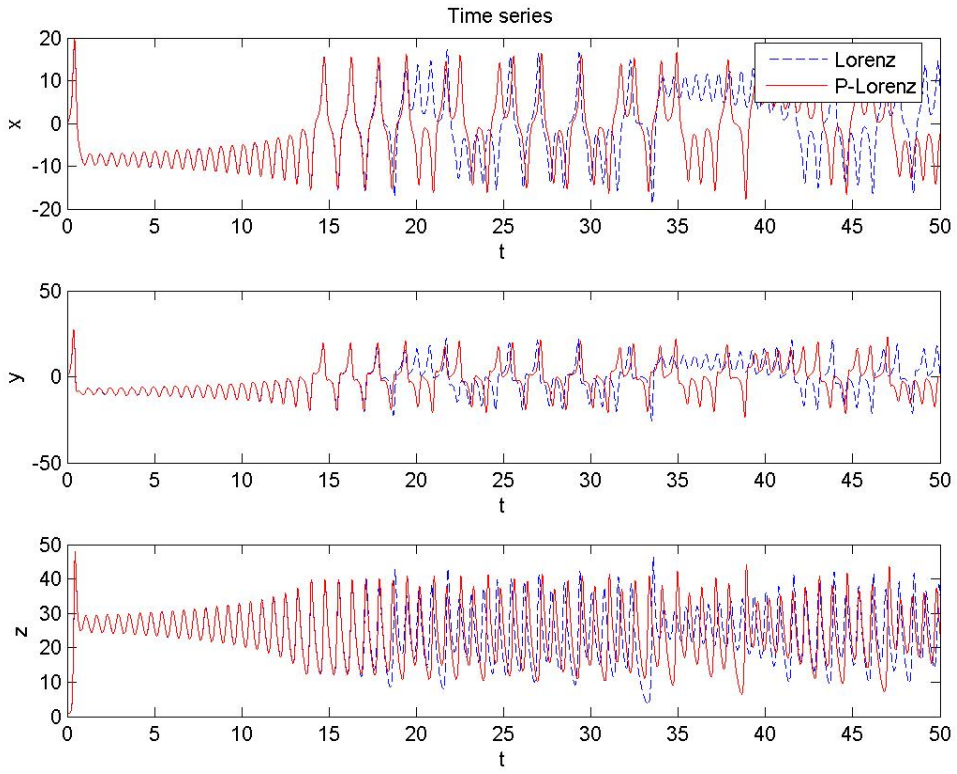

Figure 5.1: Time Series of the Lorenz System and the Perturbed Lorenz System.

and the upper bound given by Eq. (5.6).

The point here is that even with such a well fitted equation Eq. (5.2) to estimate Eq. (5.1), estimated residual error could grow exponentially. In fact, the Gromwall's inequality based analysis serves well to bound these errors analytically, but a Lyapunov exponent analysis [2, 50] would accurately describe the growth rate of the errors. Instead, it would be more useful to be a way to consider whether a good model has been found.

5.1.2 Lyapunov Analysis

The theory of Lyapunov functions, a generalization of potential energy functions, is a global approach to investigating whether a nonhyperbolic equilibrium [4] point is stable, asymptotically stable or unstable. As we know, local linear stability analysis tells us that in the neighborhood of an equilibrium, solution trajectories are attracted to the equilibrium if the eigenvalues of the linear part of the equation have negative real parts, while Lyapunov functions can tell us that the initial values from a large region converge to an equilibrium. Moreover, they can sometimes be used to determine stability of equilibria where the eigenvalues of $D\mathbf{f}$ have real parts zero.

Let $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_n(t))^T$ be a solution of the n-dimensional

[4] A point $x_0 \in \Re^n$ is called an equilibrium point of $\dot{x} = \mathbf{f}(x)$ if $\mathbf{f}(x_0) = 0$. An equilibrium point x_0 is called a hyperbolic equilibrium point of $\dot{x} = \mathbf{f}(x)$ if none of the eigenvalues of the matrix $D\mathbf{f}(x_0)$ have zero real part. An equilibrium is a nonhyperbolic equilibrium point if it is not a hyperbolic equilibrium point.

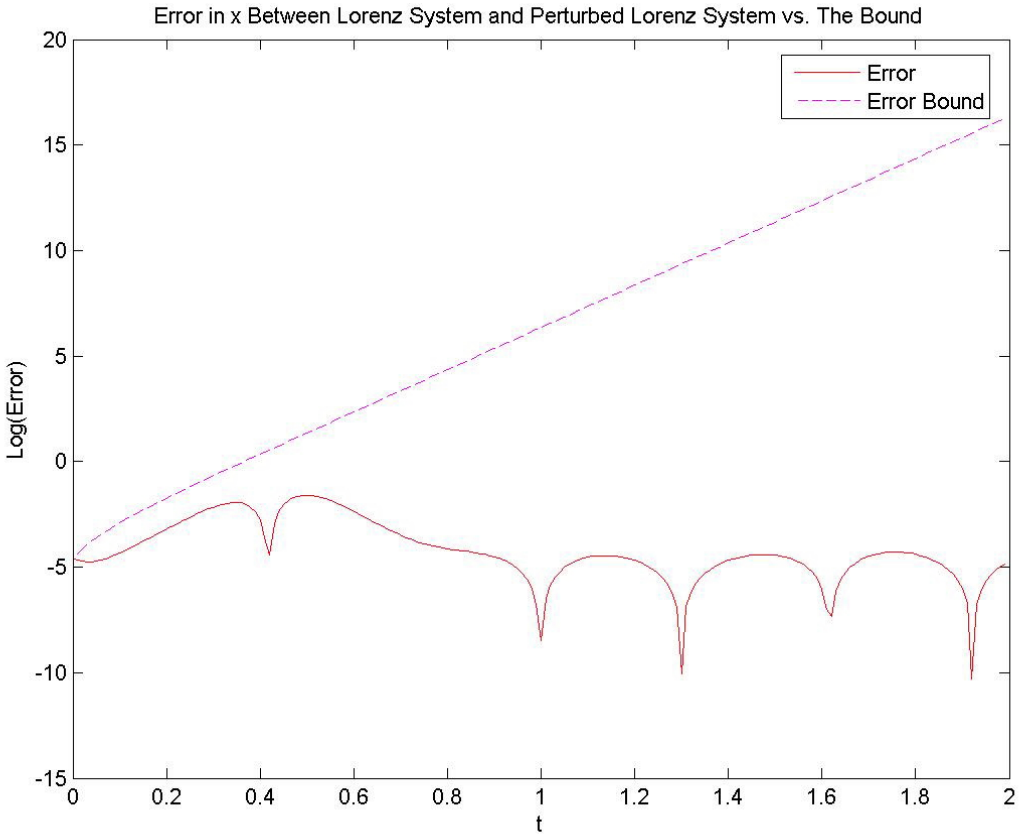

Figure 5.2: Error Between the Lorenz System and Perturbed Lorenz System In x. The solid curve is the error is x between two systems, and the dashed curve is the bound of the error given in Eq. (5.6)

system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{5.8}$$

of differential equations. Suppose we can construct a function $V(\mathbf{x})$, which can be considered as the energy function of the system given in Eq. (5.8). To measure the rate of change of $V(\mathbf{x})$ along a solution trajectory of the system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, we can obtain the derivative of V with respect to t. By the chain rule and the ODEs, the derivative of $V(\mathbf{x})$ along the solution $\mathbf{x}(\mathbf{t}) = (x_1(t), x_2(t), ..., x_n(t))^T$ is able to be found,

$$\begin{aligned} \dot{V}(\mathbf{x}) &= \dot{V}(x_1(t), x_2(t), ..., x_n(t)) = \frac{\partial V}{\partial x_1}\frac{dx_1}{dt} + ... + \frac{\partial V}{\partial x_n}\frac{dx_n}{dt} \\ &= \frac{\partial V}{\partial x_1} f_1(\mathbf{x}) + ... + \frac{\partial V}{\partial x_n} f_n(\mathbf{x}). \end{aligned}$$

The energy function $V(\mathbf{x})$ can be used to investigate the stability of equilibria in the sense of the theory of Lyapunov functions. A function V satisfying the following hypotheses is called a Lyapunov function [2, 37, 50].

Definition: Let $\mathbf{x}_0$ be an equilibrium of the system of ODEs given in Eq. (5.8). A function $V : \Re^n \longrightarrow \Re$ is called a Lyapunov function for $\mathbf{x}_0$ if for some neighborhood N of $\mathbf{x}_0$, the following conditions are satisfied:

1. $V(\mathbf{x}_0) = 0$, and $V(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{x}_0$ in N.

2. $\dot{V}(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ in N.

In addition, the proof of the theorem of Lyapunov functions given in [37] shows the stability of equilibria by using the derivative of Lyapunov functions. If

$\dot{V}(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ in N, $\mathbf{x}_0$ is stable. If $\dot{V}(\mathbf{x}) < 0$ for all $\mathbf{x}$ in $N \backslash \{\mathbf{x}_0\}$, $\mathbf{x}_0$ is asymptotically stable[5].

Consider the Lorenz system (Eq. (5.1)) as an example. We will show that every trajectory approaches the origin as $t \to \infty$ when $\gamma < 1$, meaning the origin is globally stable, by using the theory of Lyapunov functions. The origin is an equilibrium point for this system and

$$D\mathbf{f}(\mathbf{0}) = \begin{bmatrix} -\sigma & \sigma & 0 \\ \gamma & -1 & 0 \\ 0 & 0 & -b \end{bmatrix}.$$

Thus $D\mathbf{f}(\mathbf{0})$ has eigenvalues $\lambda_1 = -\frac{1}{2} - \frac{\sigma}{2} + \frac{\sqrt{1-2\sigma+\sigma^2+4\gamma\sigma}}{2}$, $\lambda_2 = -\frac{1}{2} - \frac{\sigma}{2} - \frac{\sqrt{1-2\sigma+\sigma^2+4\gamma\sigma}}{2}$, and $\lambda_3 = -b$; i.e., $\mathbf{x}_0 = \mathbf{0}$ is a nonhyperbolic equilibrium point. So we consider a Lyapunov function [50] of the form

$$V(x, y, z) = \frac{1}{\sigma}x^2 + y^2 + z^2. \tag{5.9}$$

The function $V(x, y, z)$ satisfies the two conditions:

1. $V(0, 0, 0) = 0$, and $V(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

2. $\dot{V} = \frac{2}{\sigma}x\dot{x} + 2y\dot{y} + 2z\dot{z} = \frac{2}{\sigma}x(\sigma y - \sigma x) + 2y(-xz + \gamma x - y) + 2z(xy - bz)$

   $= 2(xy - x^2) + 2(\gamma xy - xyz - y^2) + 2(xyz - bz^2)$

   $= 2(\gamma + 1)xy - 2(x^2 + y^2 + bz^2)$

   $= -2(x - \frac{\gamma+1}{2}y)^2 - 2[1 - (\frac{\gamma+1}{2})^2]y^2 - 2bz^2.$

Here $\dot{V}$ is strictly negative if $\gamma < 1$ and $(x, y, z) \neq (0, 0, 0)$. Moreover, $\dot{V} = 0$

[5] $\mathbf{x}_0$ is asympototically stable if it is stable and if there exists a $\delta > 0$ such that for all $\mathbf{x} \in N_\delta(\mathbf{x}_0)$, the $\lim_{t\to\infty} \phi_t(\mathbf{x}) = \mathbf{x}_0$; here $\phi_t(\mathbf{x})$ is the flow of the ODEs.

implies $(x, y, z) = (0, 0, 0)$. Hence the claim that the origin is globally stable for $\gamma < 1$ is established.

Now consider the original Lorenz system and the perturbed Lorenz system given in the following forms, respectively,

$$\begin{aligned} \dot{x}_1 &= \sigma(y_1 - x_1) \\ \dot{y}_1 &= -x_1 z_1 + \gamma x_1 - y_1 \\ \dot{z}_1 &= x_1 y_1 - b z_1, \end{aligned} \tag{5.10}$$

and

$$\begin{aligned} \dot{x}_2 &= (\sigma + \epsilon)(y_2 - x_2) \\ \dot{y}_2 &= -x_2 z_2 + \gamma x_2 - y_2 \\ \dot{z}_2 &= x_2 y_2 - b z_2. \end{aligned} \tag{5.11}$$

We can construct a system of error $\mathbf{E}$ between oscillators by taking the difference between those two systems. The system of error is given in form of

$$\begin{aligned} \dot{E}_x &= \sigma(y_1 - x_1) - (\sigma + \epsilon)(y_2 - x_2) = \sigma(E_y - E_x) - \epsilon(y_2 - x_2) \\ \dot{E}_y &= -x_1 z_1 + \gamma x_1 - y_1 - (-x_2 z_2 + \gamma x_2 - y_2) = \gamma E_x - E_y - x_1 z_1 + x_2 z_2 \\ \dot{E}_z &= x_1 y_1 - b z_1 - (x_2 y_2 - b z_2) = -b E_z + x_1 y_1 - x_2 y_2, \end{aligned} \tag{5.12}$$

where $E_x = x_1 - x_2$, $E_y = y_1 - y_2$, and $E_z = z_1 - z_2$. Considering the local stability of the origin of this system, we can obtain the linearization at $(E_x, E_y, E_z) = (0, 0, 0)$, which means when $x_1 = x_2$, $y_1 = y_2$, and $z_1 = z_2$,

$$\begin{aligned} \dot{E}_x &= \sigma(E_y - E_x) \\ \dot{E}_y &= \gamma E_x - E_y \\ \dot{E}_z &= -b E_z, \end{aligned} \tag{5.13}$$

by omitting the nonlinearities in Eq. (5.12). The differential equation for E_z is decoupled and shows that $E_z(t)$ approaches 0 exponentially fast. The other two

oscillators are governed by the system

$$\begin{bmatrix} \dot{E}_x \\ \dot{E}_y \end{bmatrix} = \begin{bmatrix} -\sigma & \sigma \\ \gamma & -1 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix},$$

with trace $\tau = -\sigma - 1 < 0$ and determinant $\Delta = \sigma(1 - \gamma)$. If $\gamma > 1$, the point $(E_x, E_y, E_z) = (0, 0, 0)$ is a saddle point [50] because $\Delta < 0$. Including the decaying E_z direction, the saddle point has one outgoing and two incoming directions. If $\gamma < 1$, all directions are incoming and the point $(E_x, E_y, E_z) = (0, 0, 0)$ is a sink. Specifically, since $\tau^2 - 4\Delta > 0$, the point $(0, 0, 0)$ is a stable node [50]. In order to investigate the global stability of the point $(E_x, E_y, E_z) = (0, 0, 0)$, a Lyapunov function of the system of error can be considered, however, the system of error given in the Eq. (5.12) is too complicated to give us a hint as to how to construct a proper Lyapunov function. Therefore the Lyapunov exponent [2, 50] of the system of error can be used to evaluate the change of the error $\mathbf{E}$ for a short period of time.

The local behavior of the system of error varies among the three directions. For a given initial point, which can be obtained by a small value of perturbation of the initial condition of the Lorenz system, we consider a sphere of the initial point of infinitesimal radius evolving into an ellipse as the map is iterated. The average growth rate of the longest orthogonal direction of the ellipse can be considered as the first Lyapunov number of the orbit, and its natural logarithm is called the Lyapunov exponent. In the numerical study of the error system given in the Eq. (5.12), we are trying to find how $\mathbf{E}$ grows. By plotting the natural logarithm of the error, $\ln ||\mathbf{E}(t)||$, versus t, we find curves that are close to straight lines with positive slopes, the fact

of which implies that the error $\mathbf{E}$ increases exponentially fast. Numerically, the growth rates of three oscillators E_x, E_y, and E_z are $\lambda_x = 0.2639$, $\lambda_y = 0.2539$, and $\lambda_z = 0.2676$. Therefore $\mathbf{E}(t)$ could be approximated by the following form,

$$||\mathbf{E}(t)|| \sim ||\mathbf{E}(t_0)||e^{\boldsymbol{\lambda} t}, \tag{5.14}$$

where $\boldsymbol{\lambda}$ is the largest growth rate of three oscillators of the error system, and $\boldsymbol{\lambda} = \lambda_z = 0.2676$. The number $\boldsymbol{\lambda}$ can be also considered as the Lyapunov exponent. By the Eq. (5.14), we can predict the growth of the error $\mathbf{E}(t)$, which generally shows the quality of the fitting ODEs in the sense of error evaluation.

Figure (5.3) shows the growth rate of the E_z-oscillator of the error system. The approximation of the growth rate $\lambda_z = 0.2676$ is obtained by the linear function fitting of the trajectory $E_z(t)$. The measurement is not perfect since there is always some error when the linear function fitting is applied and we should average over many different points on the same trajectory of E_z to get the true value of λ_z.

5.2 Synchronization Criterion

In the work of Brown, *et. al*. [10], they suggest a useful synchronization criterion that can be applied as a nontrivial method to evaluate whether the fitted ODE is likely to be close to the true ODE, which produces the data. Their idea is as following: If the fitted ODE $\dot{y} = \widetilde{\mathbf{f}}(y, A)$ are exactly the same as $\dot{x} = \mathbf{f}(x, P)$, that is $\|\mathbf{f} - \widetilde{\mathbf{f}}\| = 0$,

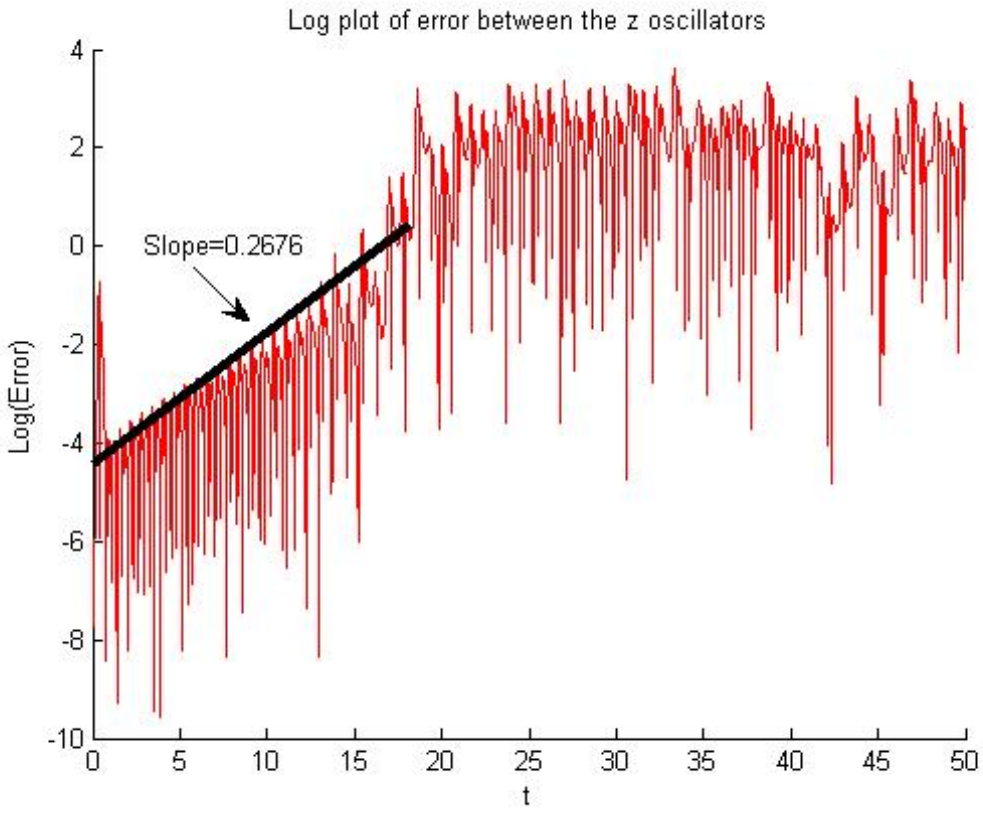

Figure 5.3: The Natural Logarithm of the Error between the z-Oscillators of the Lorenz System and Perturbed Lorenz System.

then we should hope that there exists a coupling scheme

$$\dot{x} = \mathbf{f}(x, P) \tag{5.15}$$

$$\dot{y} = \widetilde{\mathbf{f}}(y, A) + \epsilon C(x, y), \tag{5.16}$$

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 C is a coupling matrix. The idea then is that if $\widetilde{\mathbf{f}}$ is not necessarily *exactly* the same as $\mathbf{f}$, but only close to $\mathbf{f}$, then 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,

$$y = \rho(x), \tag{5.17}$$

where ρ is nearly the identity function. This criterion offers a positive test to validate a good fit, however, it is not always obvious how to find a good ϵ and coupling matrix C. Even if such pairs exist, it is not obvious how much $\widetilde{\mathbf{f}}$ and $\mathbf{f}$ 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 $\widetilde{\mathbf{f}}$ is likely to be close to $\mathbf{f}$ without actually knowing $\mathbf{f}$, since only the data from Eq. (5.15) is required.

To validate the Brown method with our previous example, let $\mathbf{f}$ be as Eq. (5.1) and suppose the fitted $\widetilde{\mathbf{f}}$ is as in Eq. (5.2). Figure (5.4) shows an application of the

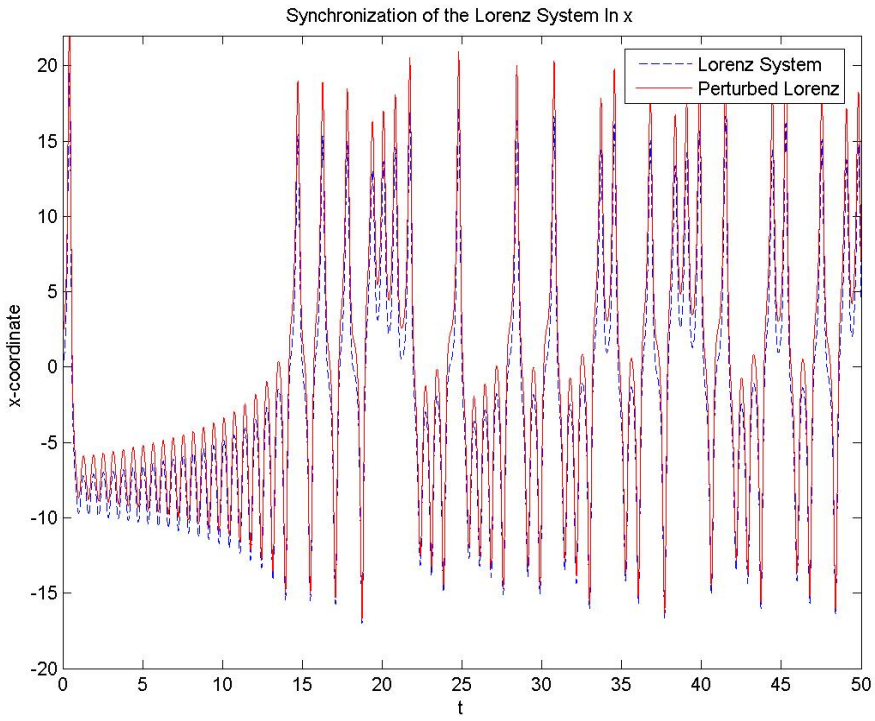

Figure 5.4: Time Series of Synchronization of the Lorenz System.

Brown method for the Lorenz system. We choose $\epsilon = 1$ and 2 to be the perturbation of the initial conditions.

Figure (5.4) shows the x-coordinates plotted from Eq. (5.1, 5.2). The difference of the initial conditions between two x-coordinates is 2. The dashed trajectory is from the unperturbed Lorenz system, and the solid trajectory is from the perturbed Lorenz system. The synchronization is observed. Figure (5.5) is a simultaneous plot of one trajectory from the time series from Fig. (5.4). The x-axis is the x-coordinate of the unperturbed Lorenz system, and the y-axis is the x-coordinate of the perturbed Lorenz system. We can see that the plot lines up along

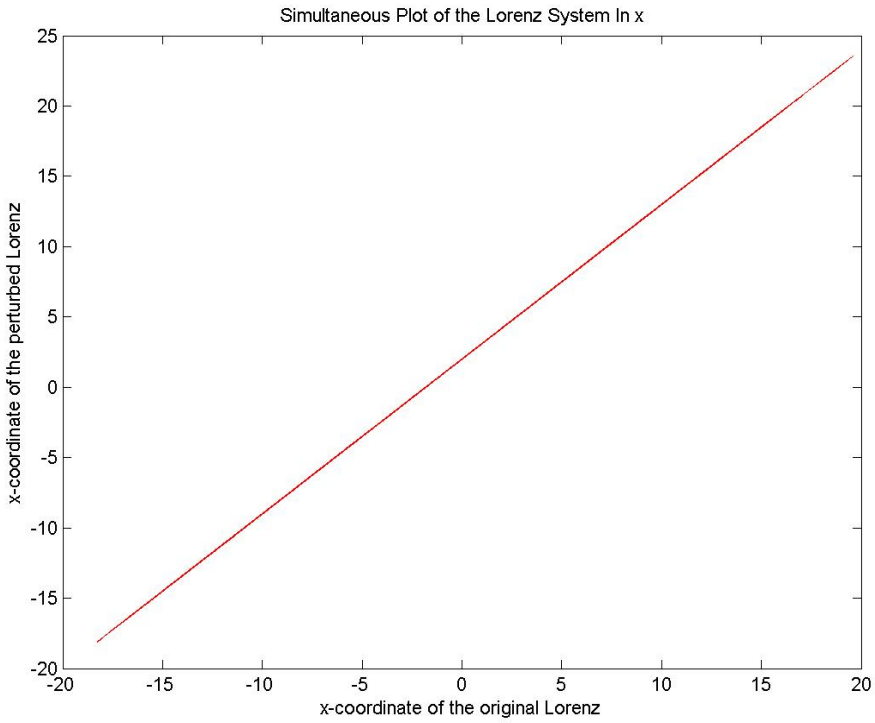

Figure 5.5: Simultaneous Plot of the x-Coordinates of Lorenz System and the Perturbed Lorenz System.

the diagonal since the two trajectories are synchronized.

5.3 Time Delay Embedding and Alternate Coordinate

In this section, we will give an example on time delay embedding with application of global modeling method. Consider fitting data from the Rössler system [2]. Proceeding naively to fit the best general quadratic model to the data set shown in figure (5.6), produced by the following equations,

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c). \end{aligned} \tag{5.18}$$

For the Rössler system, $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, P)$ in Eq. (5.18), we can make the delay plot by the vector of delay coordinates [57, 1, 23]. Here we assume that the data from the Rössler system, $(x(t), y(t), z(t)) \in \mathcal{X}$. The delay embedding data can be obtained by the time series of the x-oscillator of the system according to the following form of the vector,

$$w(t) = [x(t), x(t - T), x(t - 2T)], \tag{5.19}$$

for each time t in the time series. We assume $w(t) \in \mathcal{W}$. Instead of fitting the original data from the Rössler system, we do the global modeling method to the delay embedding data from this system to see whether a good fit can be found.

Figure (5.6) shows the original data of the Rössler system in exact x, y, z. Figure (5.7) shows the delay plot in delay variable

$$[w_1(t), w_2(t), w_3(t)] = [x(t), x(t - T), x(t - 2T)].$$

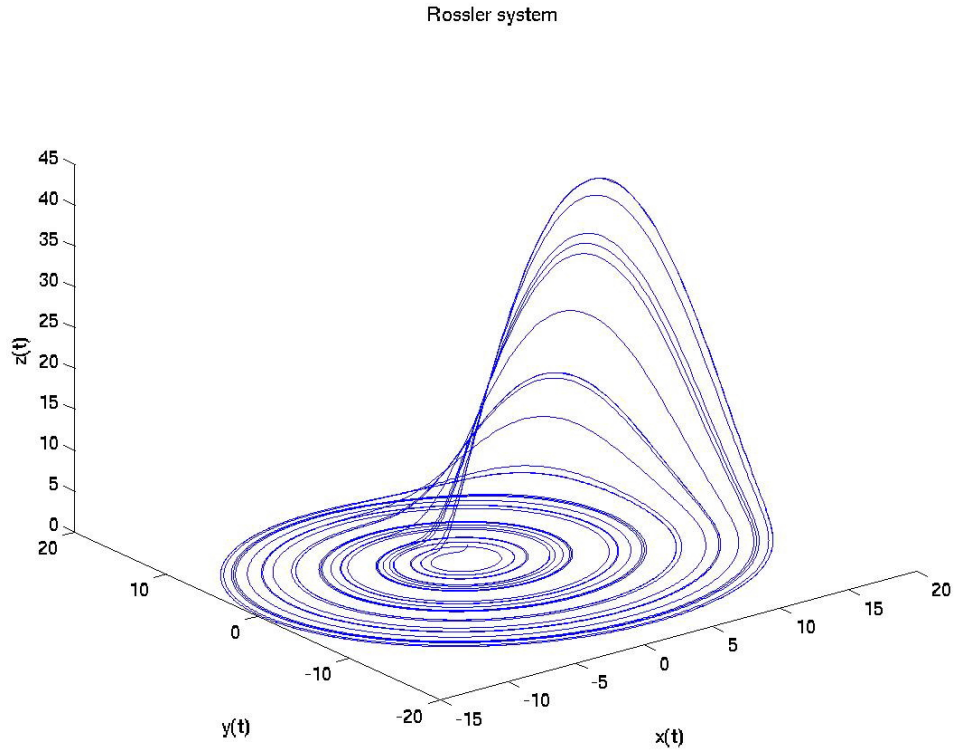

Figure 5.6: The Original Data of Rössler System.

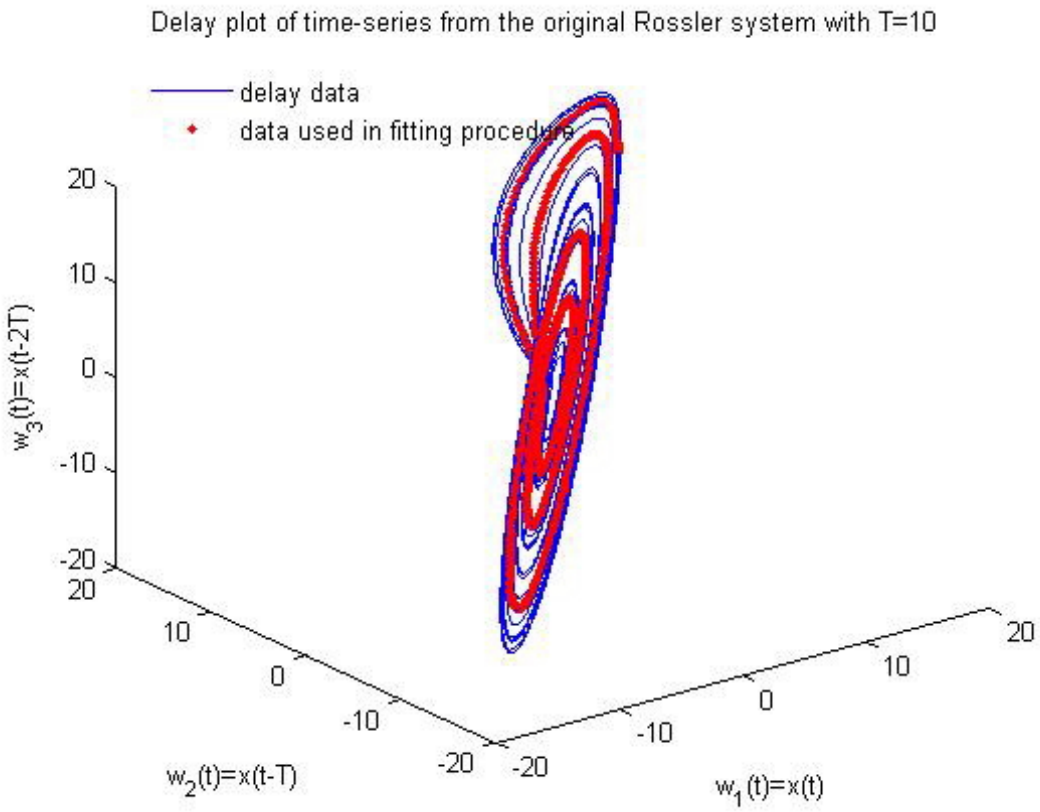

Figure 5.7: Time Delay Plot of Rössler System.

The dotted curve, which is for a short period of time, in the Figure (5.7) will be used in the modeling procedure since the error generated in the modeling procedure grows exponentially fast, thus only data of a short period of time is necessarily used to get a good fit with a small error.

The delay plot has the shape similar to the original data because of the embedding theory. We guess there exists a homeomorphism H between the original data and the delay embedding data. We denote the original space as $\mathcal{X}$ and the delay space as $\mathcal{W}$.

$$\begin{array}{ccccc} \Phi & : & \mathcal{X} & \rightarrow & \mathcal{X} \\ & H & \downarrow & & \downarrow \\ \phi & : & \mathcal{W} & \rightarrow & \mathcal{W} \end{array} \tag{5.20}$$

From Eq. (3.1), we know $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, P)$, which generates a flow on $\mathcal{X}$. Now we ask if in general we should expect if the original equations have a particular (quadratic) form, and the same simple form should carry over to governing equations of the conjugate variables. Assume there is an equation for the $\mathbf{w}$ variable such that,

$$\dot{\mathbf{w}} = \mathbf{g}(\mathbf{w}, P'), \tag{5.21}$$

which generates a flow on $\mathcal{W}$. Suppose there is a homeomorphism H between $\mathbf{x}$ and $\mathbf{w}$ and also $\mathbf{f}$ generates a flow on $\mathcal{X}$. Then, the ODE given in Eq. (5.21) can be expressed in the terms of the homeomorphism H in the following form,

$$\dot{\mathbf{w}} = \dot{H(\mathbf{x})} = \frac{dH}{d\mathbf{x}} \cdot \dot{\mathbf{x}} = \frac{dH}{d\mathbf{x}}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}, P) = \frac{dH}{d\mathbf{x}}(H^{-1}(\mathbf{w})) \cdot \mathbf{f}(H^{-1}(\mathbf{w}), P). \tag{5.22}$$

Therefore the equation obtained by the delay embedding data is a composition function of $\mathbf{f}$ and H^{-1}. If H is linear, the equation obtained by delay embedding

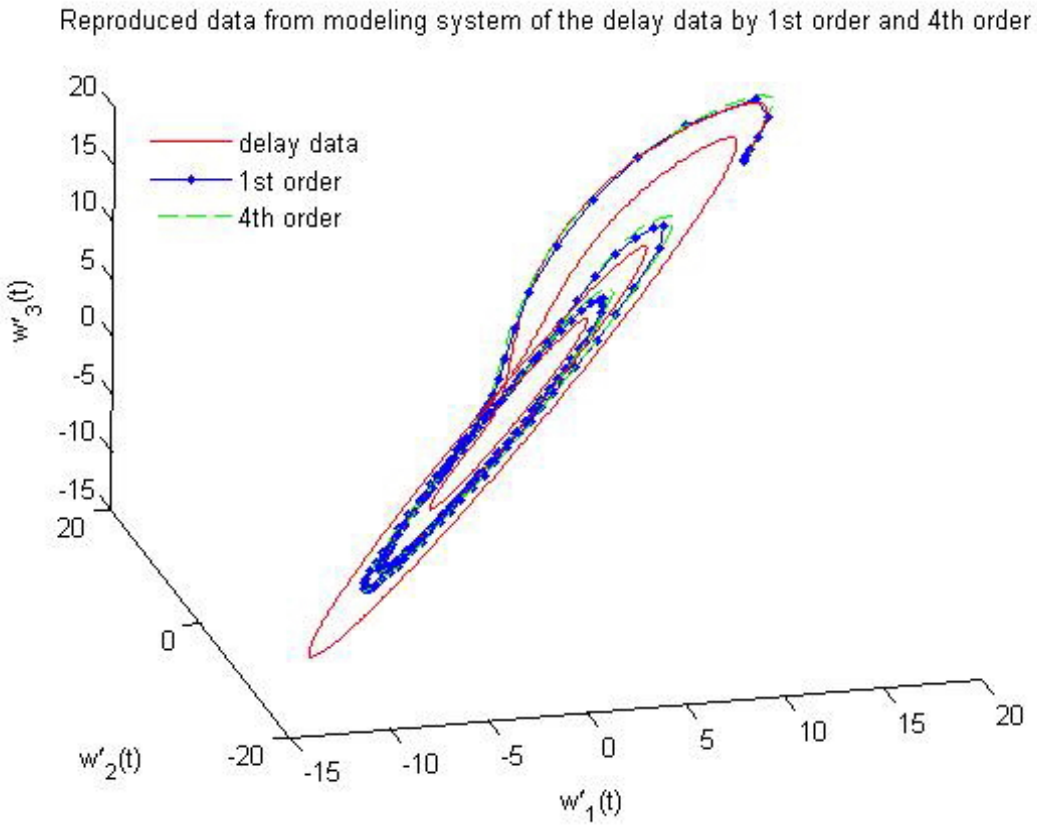

Figure 5.8: Delay Plot of Original Data and the Reproduced Data by 1st and 4th Order Schemes.

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 data fitting procedure.

Nonetheless, in the case of the simple Rössler example, we do find excellent fitted models. Figure (5.8) shows the data of delay variables and the reproduced data of the modeling system obtained by using quadratic fitting procedure with the

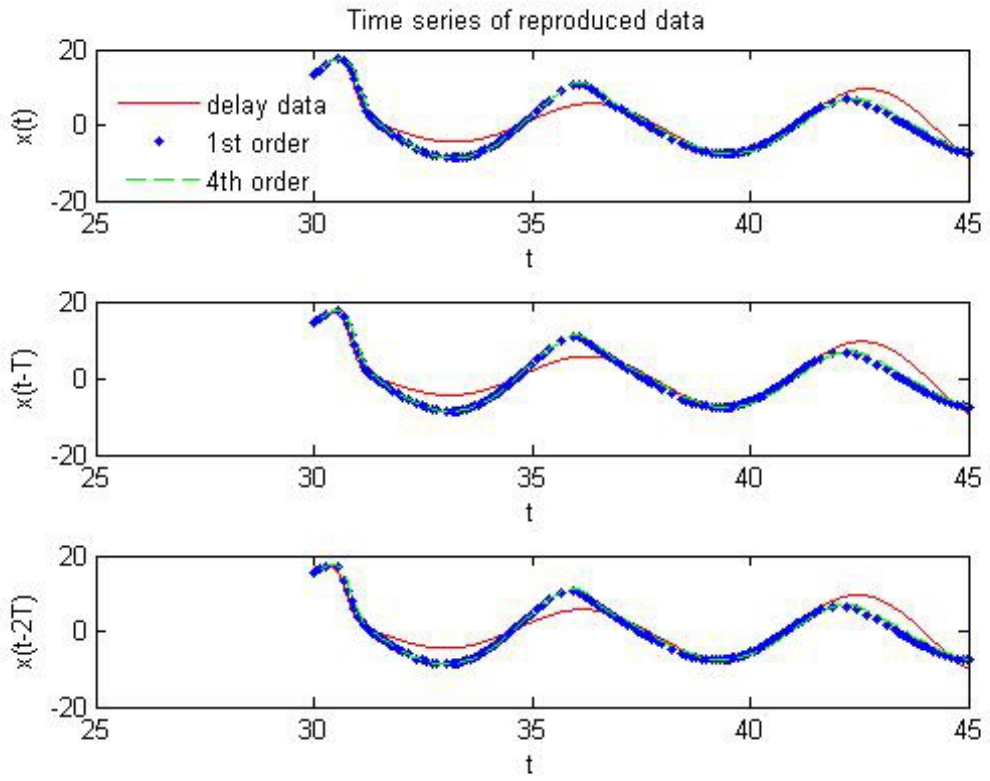

Figure 5.9: Time Series of Delay Data and Reproduced Data.

1st and 4th order schemes. In Figure (5.8), the solid curve is the delay data used in the modeling procedure, and dotted curve(.–) is the reproduced data by the forward Euler integration as the dashed green one by the 4th order scheme. Fig. (5.9) shows the time-series of each coordinate. The solid curve is the delay data, the dots and the dashed curve are reproduced data by the modeling system using 1st and 4th order schemes, respectively.

5.4 Summary

In this chapter, we have introduced several methods to estimate the quality of modeling in the view of the reproduced data of the modeling system. If the reproduced data approximates the low-dimensional data well, the analytical form given by the modeling technique can be considered as a closed form of the equations of the system that is embedded on the low-dimensional submanifold. To evaluate the error between the reproduced data and the given data numerically, we have shown that the Lyapunov analysis is capable of producing the growth rate of error, by which the local behavior of the error is able to be predicted.

Chapter 6
Modeling the Low-Dimensional Submanifolds

A major problem we will focus on in this chapter is how to globally model a low-dimensional manifold, which is expected to be embedded in some high-dimensional dynamical system, by using the ISOMAP method and the modeling techniques introduced in Chapter 2 and 3. As discussed in the Chapter 2, the ISOMAP method is a nonlinear analysis developed to deal with a discrete data set, and the output of the ISOMAP algorithm is also a data set. Therefore, a new problem arises: Given a set of high-dimensional randomly distributed empirical data points by some dynamical system, which are assumed to lie on a low-dimensional nonlinear manifold, can we find a low-dimensional analytical form of the equation that can reproduce the data set.

A common form of a system with a stable invariant manifold is well described in the context of a singularly perturbed system [12, 16], or fast-slow system. Consider a $m+n$ dimensional dynamical system of differential equations in form of,

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{F}(\mathbf{x},\mathbf{y}), \\ \epsilon\dot{\mathbf{y}} &= \mathbf{G}(\mathbf{x},\mathbf{y}). \end{aligned} \tag{6.1}$$

where $\mathbf{x} \in \Re^m$, $\mathbf{y} \in \Re^n$, $\mathbf{F} : \Re^m \times \Re^n \to \Re^m$, and $\mathbf{G} : \Re^m \times \Re^n \to \Re^n$. It is obvious that for $0 < \epsilon << 1$, the $\mathbf{y}$-equation runs fast, in relative to the slow dynamics of the $\mathbf{x}$-equation. Such systems are called singularly perturbed, since if $\epsilon = 0$, we obtain a differential-algebraic equation,

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{F}(\mathbf{x},\mathbf{y}), \\ \mathbf{G}(\mathbf{x},\mathbf{y}) &= 0. \end{aligned} \tag{6.2}$$

The second differential equation becomes an algebraic constraint. Under sufficient smoothness assumptions on the functions $\mathbf{F}$ and $\mathbf{G}$, the implicit function theorem can be applied in form of the Tokhonov theorem [53], there is a function, or $\epsilon = 0$ slow-manifold,

$$\mathbf{y} = H_\epsilon(\mathbf{x})|_{\epsilon=0},$$

such that

$$\mathbf{G}(\mathbf{x}, H_{\boldsymbol{\epsilon}}(\mathbf{x})|_{\epsilon=0}) = 0.$$

The singular perturbation theory concerns itself with continuation, and persistence of stability of this manifold $H_\epsilon(\mathbf{x})$ within $\mathcal{O}(\epsilon)$ of $H_\epsilon(\mathbf{x})|_{\epsilon=0}$, for $0 < \epsilon << 1$ and even for larger ϵ. In this chapter, we will refer to a stable invariant manifold H embedded in $\Re^{m+n}$ as an expression,

$$\begin{array}{rccc} H: & \Re^m & \longrightarrow & \Re^n \\ & \mathbf{x} & \longmapsto & \mathbf{y} = H(\mathbf{x}) \end{array}, \tag{6.3}$$

if it exists. In such a case, the degenerated system becomes,

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, H(\mathbf{x})),$$

which is equivalent to Eq. (6.1), with the constraint given in Eq. (6.3). Therefore, the manifold equation is obtained in form of,

$$\epsilon \, \frac{dH}{d\mathbf{x}} \cdot \mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{G}(\mathbf{x}, \mathbf{H}(\mathbf{x})). \tag{6.4}$$

For the purpose of this chapter, given a dynamical system, we presume that we cannot approach the dimensionality reduction in an analytical form of equations because we only know the system through a data set. The goal of this chapter is to find an analytical form of the stable manifold $\mathbf{y} = H(\mathbf{x})$ through the

data set, meaning to model the low-dimensional invariant manifold embedded in a high-dimensional dynamical system.

In order to perform the modeling technique, a low-dimensional invariant manifold should be obtained first. As we know from Chapter 2, if the stable manifold is approximately flat, the linear method, such as POD, will be able to discover a flat, nearly invariant and stable subspace. However, if the data lies on a highly curved low-dimensional manifold, the linear method will overly simplify the topology of the system. Moreover, for some typical fast-slow systems discussed in [8], the POD method may be overlooked when a change of units or parameter occurs. Therefore, the ISOMAP method will be applied in the following examples to avoid the shortcomings of the POD method.

6.1 Lorenz System

We consider a three dimensional dynamical system, the Lorenz system [2, 32], as an example,

$$\begin{aligned} \dot{x} &= -\sigma x + \sigma y \\ \dot{y} &= rx - y - xz \\ \dot{z} &= xy - bz, \end{aligned} \tag{6.5}$$

where we choose $\sigma = 10$, $r = 28$, $b = \frac{8}{3}$. There is no apparent invariant manifold for this system, but there is a famous Lorenz butterfly, which is an attractor of the

Lorenz systems with specific values of parameters σ, r, and b. This attractor has a fractal dimension slightly larger than 2. In fact, it is known that the Lorenz attractor is better described as a branched manifold [5, 59]. Therefore, this system makes a good data set to perform dimensionality reduction methods that insist on treating the data set as if there was a manifold. Firstly, we will apply the ISOMAP method to ascertain the structure of the low-dimensional attractor. Secondly, we will try to model the ISOMAP projections, which are expected on a lower dimensional submanifold, to see if a good analytical form can be obtained.

6.1.1 Application of the ISOMAP Method

In Chapter 2, the Tenenbaum filling condition [4] was introduced to ensure that the given data set can be used in ISOMAP method to discover an embedded low-dimensional submanifold, if there exists one. For the study of how to apply the ISOMAP method to dynamical systems such as those with some low-dimensional global attractors, we will hereafter display some successful applications of the ISOMAP in the Lorenz system.

Computationally, we will require that the modeling of any embedding submanifold in the ambient space by ISOMAP use only finite set of data. Ideally, sampling data on a submanifold would be uniform with respect to the intrinsic variables. However, dynamical systems tend to sample their attracting invariant manifolds according to highly irregular, and even fractal, invariant measures. In

[4], an asymptotic convergence theorem is presented to address the data sampling convergence to manifold model issue. However, it is hard to show that the invariant density, if it exists, of the Lorenz system satisfies the properties required in the asymptotic convergence theorem, aforementioned, implying that the data set, even generated by a long period of time, is not constantly uniform. Therefore, a preprocessing of resampling the data set to collect a uniformly distributed subset is needed before the application of the ISOMAP.

In the typical scenario when the invariant density is not uniform, we will present a resampling algorithm to preprocess the data set in a way that produces data sets, which are appropriate for the hypothesis of the ISOMAP asymptotic convergence theorem. Moreover, we will explore a method involving the linear combinations of the edges on a graph, which consists of the points in the resampling data set and the ordered pairs of edges connecting the points, to allow us to discover the underlying relationship between the ISOMAP projections of the resampling data set, and points which are released in the preprocessing step of resampling.

Meshing Method

In Fig. 6.1, the Lorenz attractor is shown. There are two ubiquitous holes on the attractor, one on each wing. From the graph, we can see that the points generated computationally close to the hole are much denser than those far away. Therefore, a better data set with a uniform distribution is needed to sufficiently guarantee success of application of the ISOMAP method in order to understand the

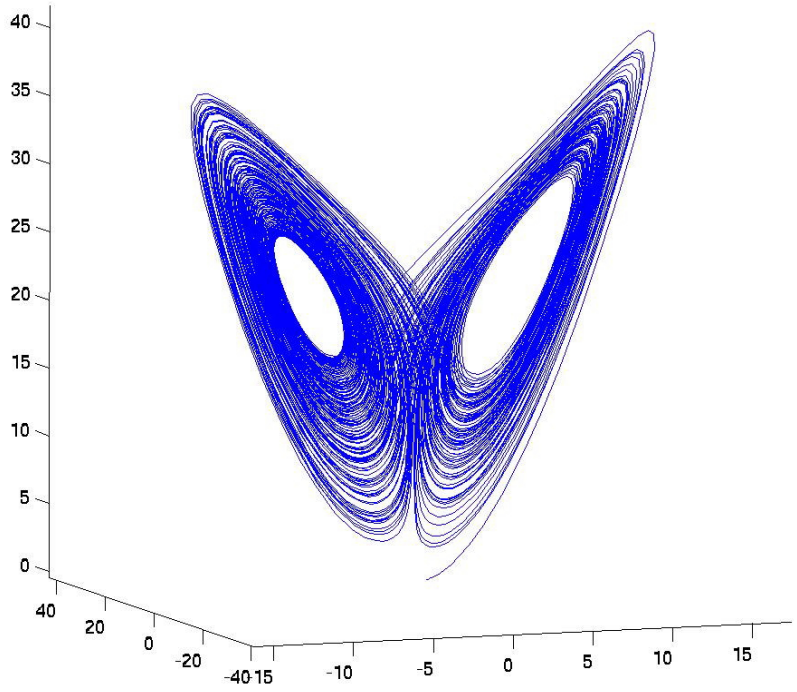

Figure 6.1: A Solution from Lorenz System Given in Eq. (5.5).

structure of the attractor.

In order to obtain a uniformly distributed data set, we need to generate a large data set experimentally, which consists of approximately 100,000 data points, so that the density of the points far away is not extremely less than that of the points close to the hole, even the data set is not uniformly distributed yet. In principle, it is possible to perform ISOMAP directly on this big data set, however, it is too computationally expensive to apply ISOMAP on such a large data set. In practice,

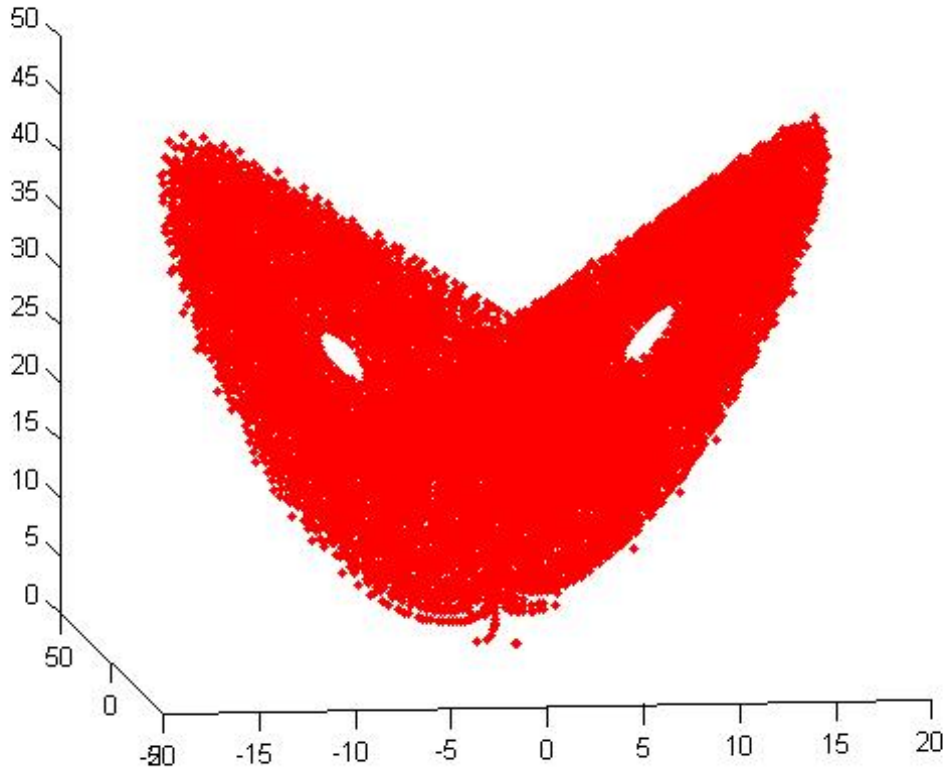

Figure 6.2: Uniformly Distributed Resampling Data Set Generated by the Meshing Method.

we can generate a resampling of the data set, which is almost uniformly distributed based on the aforementioned large data set, to diminish the size of the original data set without losing the good density. To do so, we mesh the three dimensional space where the attractor is located, into small cubes, inside each of which, the first point generated experimentally will be maintained, and afterwards we will then release the individual cube. Noticeably as the sizes of cubes become smaller, the resampling data set is apt to be more uniformly distributed. Consequentially, we are able to use a proper size of the cube to obtain a predicted uniformly distributed data set.

Fig. 6.2 reveals a uniformly distributed data set generated by the meshing method. The attractor contains approximately 12,500 data points, the size of which still remains too large to perform the ISOMAP. The reason that the size of the resampling data set is still large is that the points far away from the hole are not dense enough to build a good neighborhood graph, meaning that the properties required in the asymptotic convergence theorem can not be satisfied, when the ISOMAP is applied. Therefore, it is indispensable that a subsampling of such a uniformly distributed resampling data set will be employed on condition that the said subsampling must have similar statistics as the resampling data set.

We choose 5,000 points uniformly from 12,500 data points shown in Fig. 6.2 to make an illustration of Fig. 6.3 depicting a subsampling of data points, among which dark dots are the subsampling data. The set of blue points is sufficiently dense so that the approximations of the attractor by the discrete graph structure is

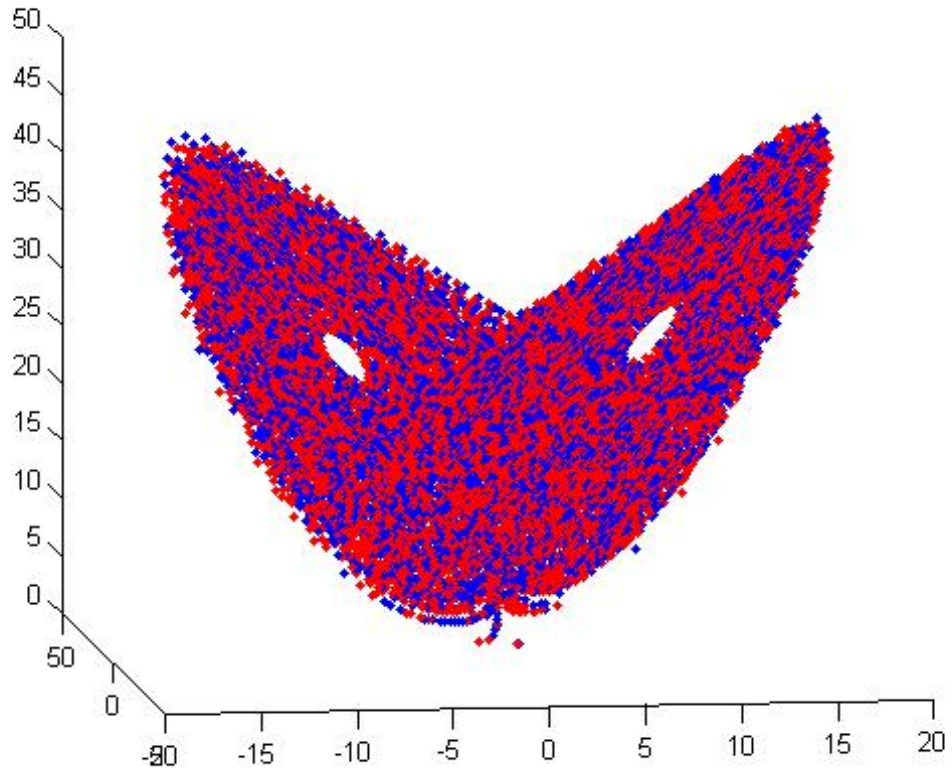

Figure 6.3: Uniformly Distributed Subsampling Data Set Generated by the Resampling Data Set.

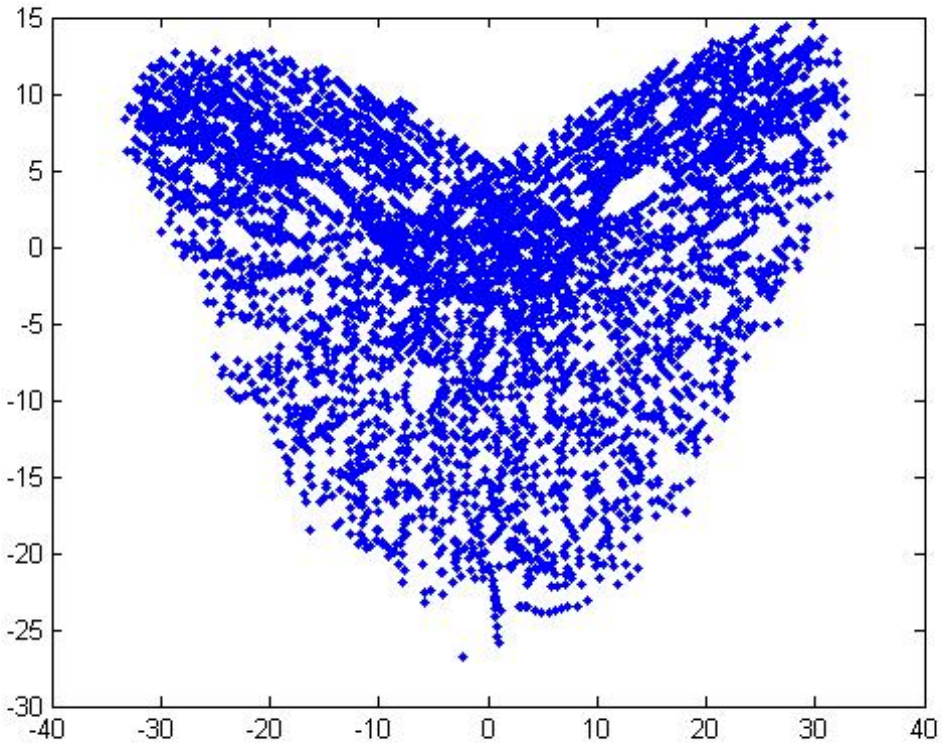

Figure 6.4: ISOMAP Projections of the Subsampling Data Set. The plot shows a two-dimensional nonlinear embedding of the subsampling data set given in Fig. (6.3).

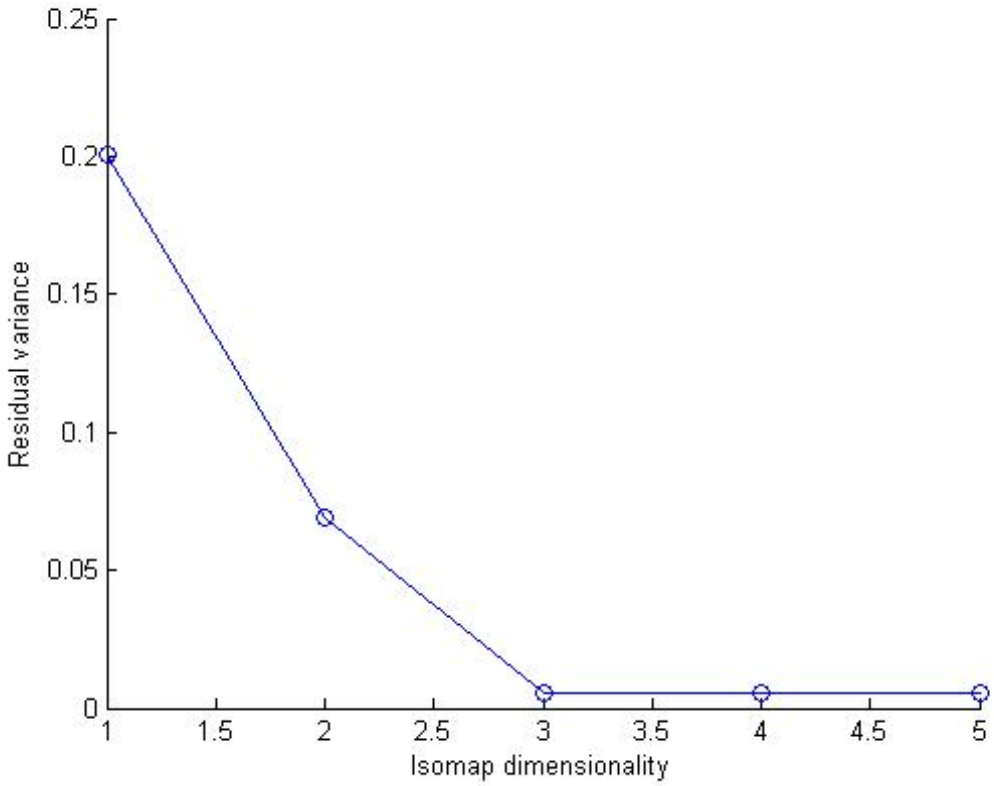

Figure 6.5: Dimensionality Found by the ISOMAP Method.

good. The outcome of applying the ISOMAP method on the set of blue points is presented in Fig. 6.4 . Fig. 6.5 shows the dimensionality found by the ISOMAP method applied to the subsampling data set given in Fig. 6.3. The x-axis is the test dimension and the y-axis is residual error in Eq. (2.4). It is rational that a 2-D manifold represents the 3-D attractor vividly. For the rest of this section, the outcome of ISOMAP is called as ISOMAP projection. In order to perform modeling globally on the lower dimensional data set, the ISOMAP projections of all data

points generated computationally have to be obtained. Therefore, another problem is raised thereafter whether we can find the ISOMAP projections of the red data set without performing the ISOMAP itself. We will introduce a technique involving linear combinations of edges on the neighborhood graph to explore this problem.

Method of Linear Combinations

The problem we will address in this part is how to find the ISOMAP projections of those points that are not part of the subsampling data set without applying the ISOMAP method directly. It is meant that the ISOMAP method is a one-to-one mapping, implying that the ISOMAP method gives the image of each point x, which is generated computationally by the Lorenz system in this case, by projecting points in high-dimensional coordinates into points in low-dimensional coordinates, and also preserves the geodesic distances, the fact of which means that the nearby points on the manifold keep nearby, and the far away points keep far away. Therefore, considering linear combinations of edges connecting near neighbors is a useful way to explore the ISOMAP projections of the points, which are not in the subsampling data set. For any point in the red data set, if it is also in the subsampling data set (blue data set), we will keep the ISOMAP projection of this point, which can be obtained in Fig. 6.4. If it is not in the blue data set, this point r can be represented by a linear combination of edges connecting blue points, which are closest to it. The ISOMAP method preserves the coefficients of the linear combination. Therefore, the ISOMAP projection of the point r is able to be found

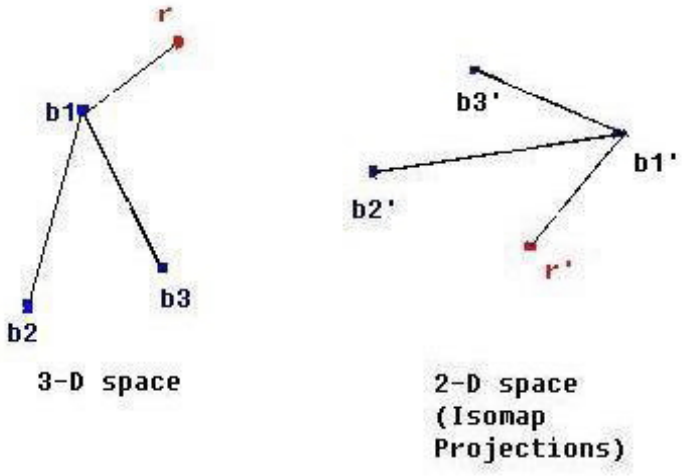

Figure 6.6: Example on Linear Combinations of Edges.

by the linear combination of the ISOMAP projections of blue points nearby with the same coefficients.

In Fig. 6.6, an example is shown. Here b_1, b_2, and b_3 are blue points in the subsampling data set. They are in 3-dimensional space and the ISOMAP method produces their ISOMAP projections b_1', b_2', and b_3', which are in 2-dimensional space. The red point r does not have a projection directly given by the ISOMAP method because it is not in the subsampling data set (blue). In Fig. 6.6(Left), we know there is a linear combination involving b_1, b_2, and b_3,

$$a_1 \cdot (r - b_1) + a_2 \cdot (b_2 - b_1) + a_3 \cdot (b_3 - b_1) = 0,$$

where, a_1, a_2, and a_3 are the coefficients of the linear combination, and can not all be 0. With the same value of the coefficients, we can derive an equation of the

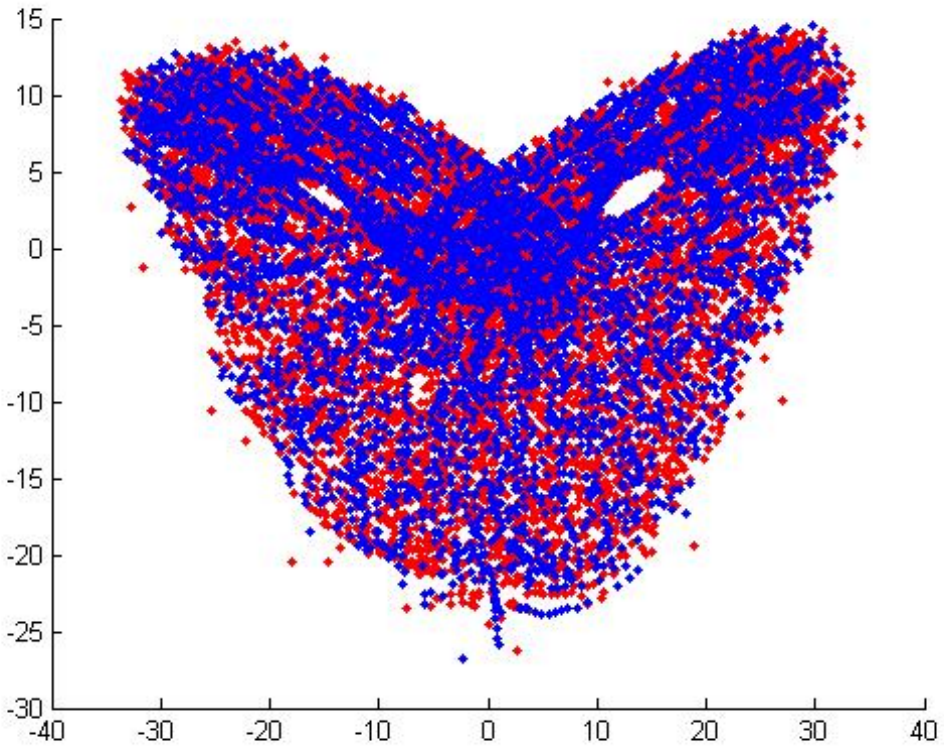

Figure 6.7: ISOMAP Projections of Uniformly Distributed Dataset.

ISOMAP projection of r,

$$a_1 \cdot (r' - b_1') + a_2 \cdot (b_2' - b_1') + a_3 \cdot (b_3' - b_1') = 0, \tag{6.6}$$

where b_1', b_2', and b_3' are the ISOMAP projections of b_1, b_2, and b_3. r' is the ISOMAP projection of r, which can be solved from Eq. (6.6). Therefore, for all points not in the subsampling data set, we can use the method of linear combinations to find their ISOMAP projections of those. In Fig. 6.7, the light points are the ISOMAP

projections of the light points in Fig. 6.3, which are not in the subsampling data set. Good results are given by the method of linear combinations.

The meshing method gives a way to find a data set with a specific density. In this case, we use a uniformly distributed data set. It is also possible to refine the mesh finer at particular locations of interest. By the method of linear combinations, the ISOMAP projections can be found without applying the ISOMAP directly to the whole data set, so that computations are saved.

6.1.2 Modeling the ISOMAP Projections

In this section, we will show an example on modeling the ISOMAP projections of a data set of the Lorenz system. As introduced in the previous part, for the purpose of saving computations, the meshing method and the method of linear combinations can be employed to obtain the ISOMAP projections when we deal with a very large data set, or a data set that does not satisfy the requirements of the asymptotic convergence theorem [4], for instance, the data set that is not uniformly distributed.

Given a set of data points, which are from laboratory experiments or generated from computations by the Lorenz system, it is known that a 2-dimensional manifold fits the attractor well. In the application of the ISOMAP method, assuming that the size of this data set is too large to apply ISOMAP directly, we can obtain a uniformly distributed subsampling data set, of which the ISOMAP projections are easy to be obtained. The ISOMAP projections of the original large data set are then able to be

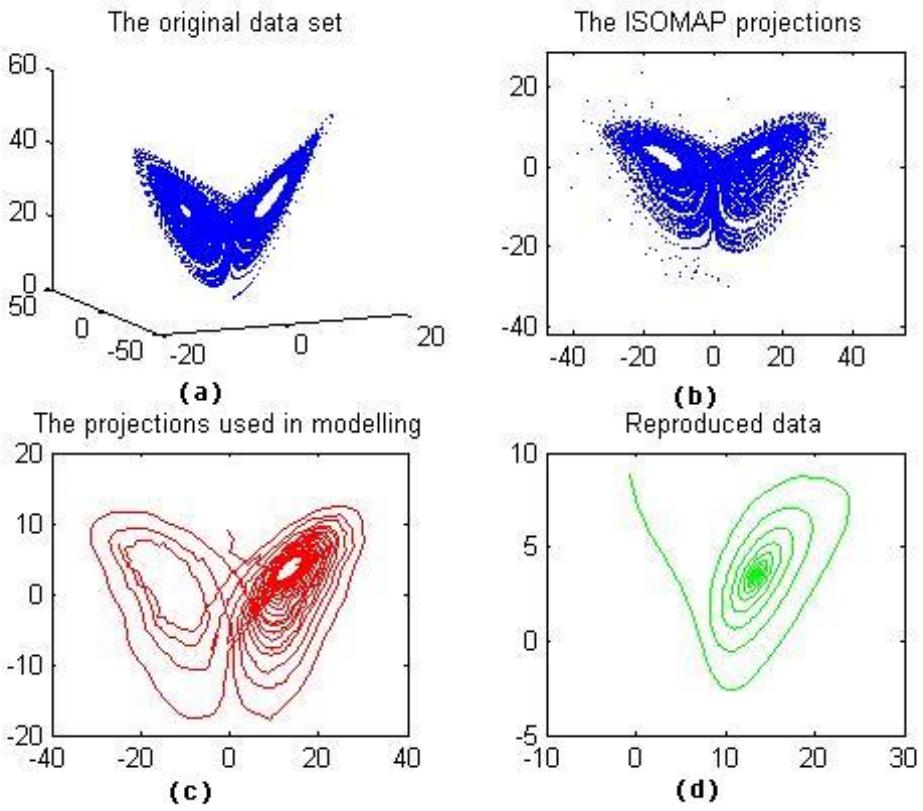

Figure 6.8: Example on Modelling a Low-Dimensional Submanifold of Lorenz System.

found by the method of the linear combinations. In fact, the ISOMAP projections of the data points of some particular period of time of interest will be needed in the following modeling procedure.

To investigate the problem of whether there exists an analytical form of some differential equations governing the ISOMAP projections, we perform the modeling technique on the 2-dimensional set of the ISOMAP projections.

Figure (6.8) reveals the numerical results of modeling the 2-dimensional submanifold of the Lorenz system by the ISOMAP method. Part (a) shows the original data set generated computationally by the Lorenz system. There are approximately 10,000 points on the attractor. In practice, it is computationally too expensive to perform the ISOMAP method directly. Moreover, the discrete graph constructed for the application of the ISOMAP is poor since the data set shown is not uniformly distributed. Hence, the subsampling data set is used in this case. Part (b) in Fig. (6.8) shows the ISOMAP projections of the data points given in part (a). In order to obtain this set of ISOMAP projections, the method of linear combinations of the ISOMAP projections of the subsampling data set, which is generated by the meshing method, is applied.

Part (c) in Fig. (6.8) shows the ISOMAP projections that are used in the modeling procedure. The reason why this particular period of time is focused on is that the initial point of the trajectory is not on either of two wings, and the trajectory of the 2-dimensional data points is moving from one wing to the other. Therefore, it might be interesting to investigate the underlying form of this part of the ISOMAP projections. To model the data set shown in part (c), we apply the second order scheme and the cubic fitting form, that is

$$\dot{X}(t) = \frac{X(t+1) - X(t-1)}{2h} = P_0 + P_1 X(t) + P_2 X(t)^2 + P_3 X(t)^3,$$

where the set of X is the 2-dimensional ISOMAP projections and $[P_0, P_1, P_2, P_3]$ is the parameter to be solved in the modeling procedure. Part (d) shows the reproduced

data by using the modeling system. As we see in the graph, the trajectory of the reproduced data is moving towards a spiral and attracted to the center. Therefore, the reproduced trajectory cannot describe the motion of the original trajectory for a long period of time. However, for a short period of time, the modeling system fits the original data set. The numerical results are shown in Table (6.1).

Table 6.1: Numerical results of the ISOMAP projections and the reproduced data by the modelling system

N	The ISOMAP projections used in modeling (Y_1, Y_2)	Reproduced data by the modeling system (X_1, X_2)
1	$(-0.6713, 8.9065)$	$(-0.6713, 8.9065)$
2	$(0.7453, 7.7378)$	$(0.2745, 7.1483)$
3	$(0.0860, 5.9368)$	$(1.1391, 6.1661)$
4	$(1.3110, 5.5719)$	$(1.8581, 5.5153)$
5	$(2.1638, 5.0355)$	$(2.4499, 5.0342)$
6	$(2.9918, 4.6221)$	$(2.9407, 4.6522)$
7	$(3.5774, 4.3693)$	$(3.3530, 4.3340)$
8	$(4.0875, 4.0503)$	$(3.7038, 4.0597)$

In the above example of the Lorenz system, we focus on the problem of how to obtain the ISOMAP projections of a very large data set, or a data set having a poor distribution. We also show some numerical results of the modeling procedure. In the following examples, we will focus on the set of the ISOMAP projections (intrinsic variables) and the modeling systems.

6.2 Fast-Slow Systems

In this section, we will consider a special case of fast-slow systems with one

coupling equation in form of,

$$\begin{array}{l} \dot{x} = f(x) \\ \epsilon\dot{y} = y - \alpha g(x) \end{array} . \tag{6.7}$$

For systems in this form, it is discussed in [8] that with a proper value of the coupling parameter α, there exists a slow-manifold,

$$y = \alpha g(x),$$

in a lower dimensional subspace when $\epsilon = 0$. The existence of the slow-manifold allows us to investigate the underlying structure of the system by modeling the low-dimensional invariant submanifold.

6.2.1 3-Dimensional Example on modeling ISOMAP Projections

First, we consider a problem, consisting of a simple 2nd order differential equation,

$$\ddot{x} + kx = 0, \tag{6.8}$$

contracting onto a stable nonlinear slow-manifold specified as a parabola in the w-variables. The form of this system can be written as a system of two 1st order differential equations by letting $\dot{x} = y$ with a slow-manifold in the w-variables,

$$\begin{array}{l} \dot{x} = y \\ \dot{y} = -kx \\ \epsilon\dot{w} = w - \alpha(x^2 + y^2) \end{array} . \tag{6.9}$$

This is a 3-dimensional system embedded on a 2-dimensional nonlinear slow-manifold,

$$w = \alpha(x^2 + y^2).$$

If we choose, $k = 0.5$, $\alpha = 0.5$, and $\epsilon = 0.001$, we get the data set shown wrapped onto a paraboloid, as in Fig. (6.9).

As an example of application of the ISOMAP method, we take the data $(x(t), y(t), w(t))$, which is a $3 \times n$ matrix, shown as the blue curve in $\Re^3$. Also shown on the xy-plane in black, is the data of the x, and y components, which are solutions of the 2nd order differential equation given in Eq. (6.8). The ISOMAP method produces the ISOMAP projections of the data set. The green curve shown in the Fig. 6.9 is the set of the 2-dimensional ISOMAP projections, which well approximates the discrete graph constructed by the 3-dimensional data. Geometrically, the ISOMAP projections are able to be obtained by stretching the curve, which is the data set on the paraboloid in $\Re^3$, onto a flat plane. The geometrical explanation of the ISOMAP projections gives us some insight into the reason that the circle of the ISOMAP projections is bigger than the one of the x and y components.

modeling the ISOMAP projections, an analytical form of linear equations is expected since the system of two 1st order differential equations, which are embedded on the 2-dimensional submanifold, is in linear form. The small circles in red shown in Fig. (6.9) are reproduced data of the modeling system by fitting the ISOMAP projections into a linear form with the 2nd order scheme. The modeling system is given analytically in the following form,

$$\begin{bmatrix} \dot{x}_m \\ y_m \end{bmatrix} = \begin{bmatrix} -0.024 \\ 0.0564 \end{bmatrix} + \begin{bmatrix} 0.0273 & 0.762 \\ 0.6568 & -0.033 \end{bmatrix} \cdot \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \tag{6.10}$$

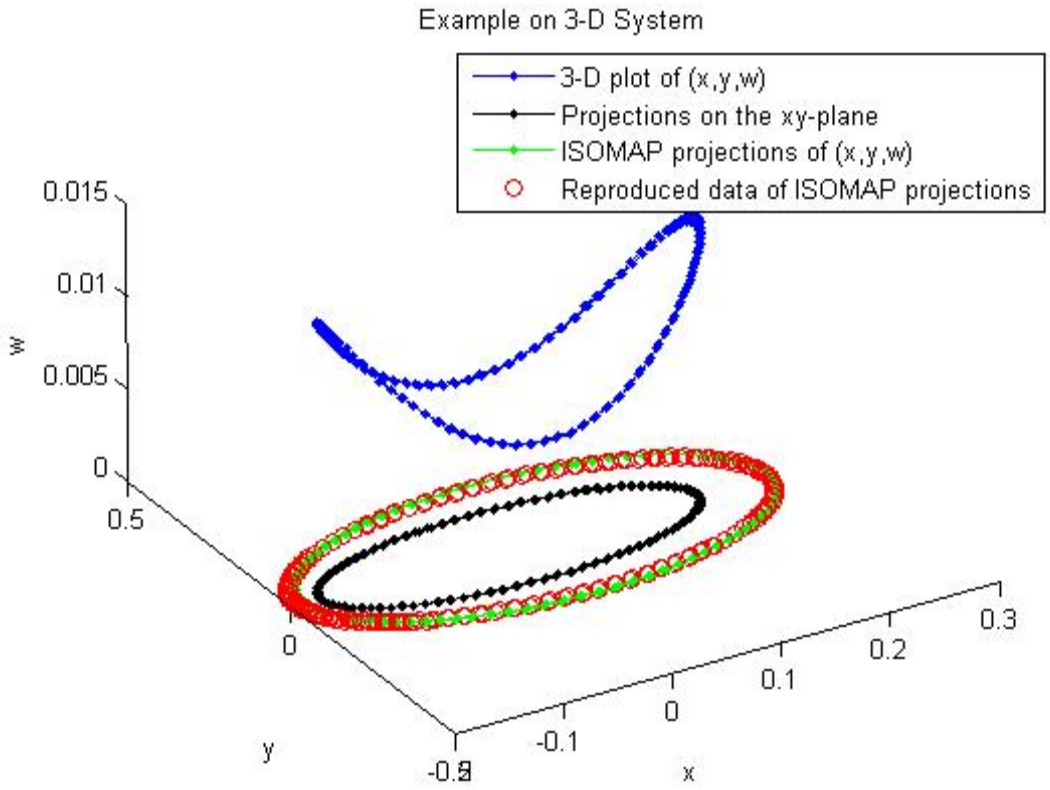

Figure 6.9: Example on a 3-D System.

where x_m and y_m are variables of the modeling system. In Fig. (6.9), it is shown that the modeling system fits the ISOMAP projections well, indicating that the modeling method analytically gives a good approximation of the system governing the ISOMAP projections. Comparing the modeling system given in Eq. (6.10) to the system of two 1st order differential equations, the form of which can be rewritten as

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k & 0 \end{bmatrix} \cdot \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \tag{6.11}$$

we observe a difference between parameters of those two systems. The coefficient of the constant term in Eq. (6.10) is nonzero and the coefficient of the linear term is not the same as, or even close to the one in Eq. (6.11). In fact, as discussed in the geometrical interpretation of the ISOMAP projections, the ISOMAP method deals with the geodesic distances between points along the manifold. Thus, the system governing the ISOMAP projections, which we will refer to as the projected system for the rest of this chapter, cannot be the same as, but related to, the original system. In the following example, we will focus on the projected system, where the ISOMAP projections come from, to reveal the relationship between the projected system and the original system.

6.2.2 2-Dimensional Example on Projected System

General Form

In order to investigate the relationship between the projected system and the

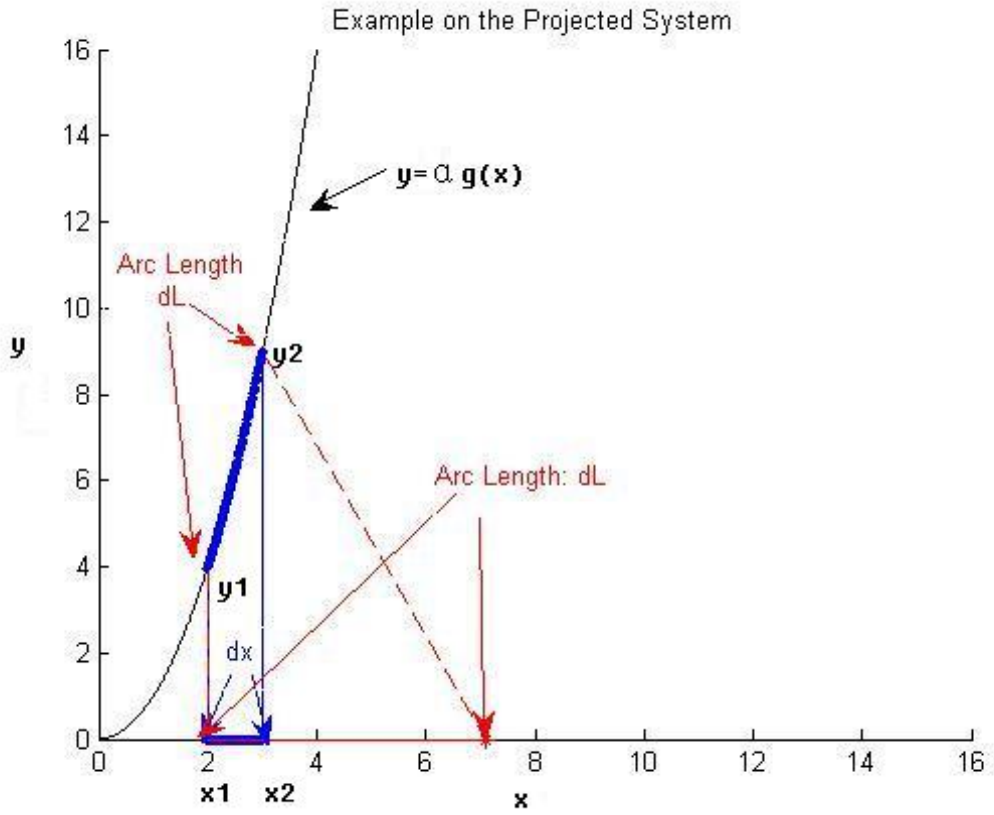

Figure 6.10: Example on the Projected System.

original system, we consider a benchmark example of 2-dimensional fast-slow system in form of Eq. (6.7),

$$\begin{aligned} \dot{x} &= f(x) \\ \epsilon\dot{y} &= y - \alpha g(x) \end{aligned} .$$

where a 1-dimensional slow manifold exists. Figure (6.10) shows the slow manifold $y = \alpha g(x)$ in black. The corresponding points on the manifold y_1 and y_2 are derived from the two data points x_1 and x_2. ISOMAP method generates a set of projections

by approximating the geodesic distances between points in the original data set. Therefore, it is necessary to concentrate on the geodesic distances when we are trying to investigate the projected system that is expected to govern the ISOMAP projections.

Geodesic Distance and Projected System

In this 2-dimensional case, the geodesic distance between points y_1 and y_2, is the arc length of the curve $y = \alpha g(x)$ when $x_1 \leq x \leq x_2$. The arc length function based on the applications of integration from the calculus [49] is given in form of,

$$L(x) = \int_a^x \sqrt{1 + [\alpha g'(u)]^2} du. \tag{6.12}$$

We can use Part 1 of the fundamental theorem of calculus to differentiate Eq. (6.12) to obtain the rate of change of arc length L with respect to x,

$$\frac{dL}{dx} = \sqrt{1 + [\alpha g'(x)]^2}.$$

Considering the rate of change of the arc length L w.r.t. time t, we are able to obtain a differential equation by chain rule,

$$\frac{dL}{dt} = \frac{dL}{dx} \cdot \frac{dx}{dt} = \sqrt{1 + [\alpha g'(x)]^2} \cdot f(x), \tag{6.13}$$

which describes the motion of the data points along the slow manifold. Hence, the system given in Eq. (6.13) is the expected projected system, which governs the ISOMAP projections.

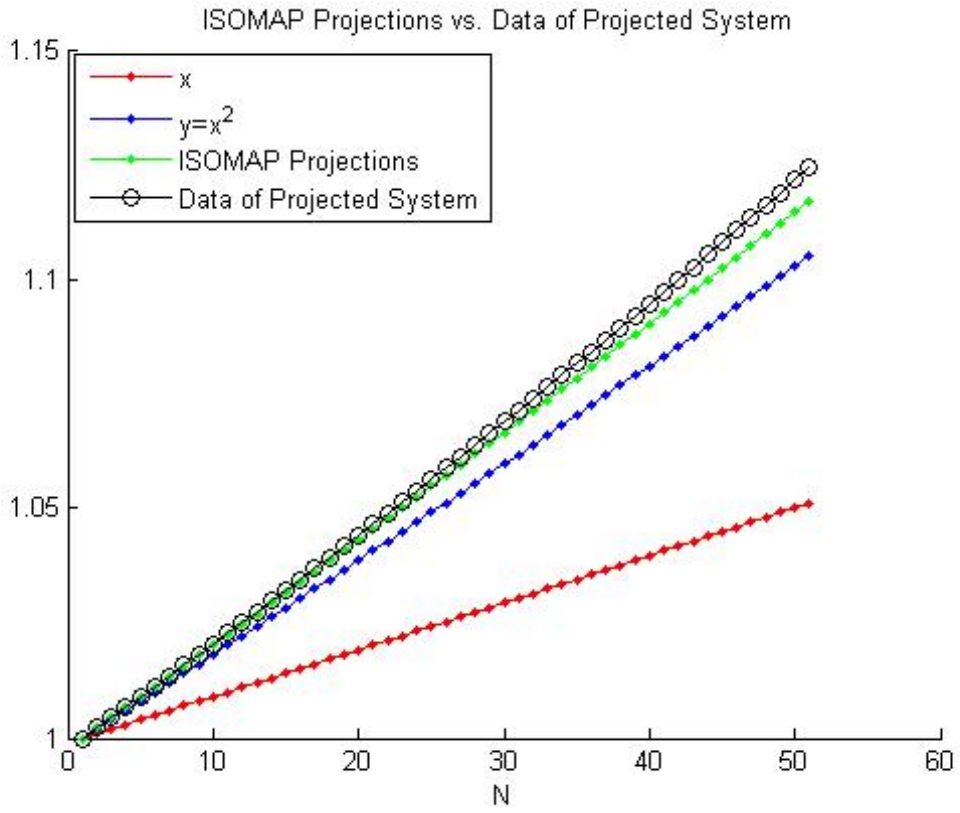

Figure 6.11: Example of 2-Dimensional Fast-Slow System on the ISOMAP Projections and the Projected System.

Linear Problem

Consider an explicit example of fast-slow system given in form of

$$\begin{aligned} \dot{x} &= ax \\ \epsilon\dot{y} &= y - \alpha x^2. \end{aligned} \tag{6.14}$$

The analytical solution of the x variables in Eq. (6.14) is $x = x_0 e^{at}$, where x_0 is the initial condition. The slow manifold exists and is in form of $y = \alpha x^2$. We choose $a = 0.1$, the coupling parameter $\alpha = 1$, $\epsilon = 0.001$, and the initial condition $x_0 = 1$.

The numerical solutions of x and y used to perform the ISOMAP method are shown in Fig. (6.11) as the red curve and the blue curve, respectively. The green curve shows the ISOMAP projections of the 2-dimensional system. The small black circles shown in Fig. (6.11) are the numerical solutions of the projected system constructed by Eq. (6.13). We see in Fig. (6.11) a clear match of the ISOMAP projections and the data of the projected system,

$$\frac{dL}{dt} = \sqrt{1 + 4\alpha^2 x^4} \cdot ax,$$

the fact of which justifies that the ISOMAP projections come from an expected projected system related to the original system in sense of arc length.

6.2.3 Rössler Oscillator on Paraboloid Fast-Slow System

In this section, we will consider a 4-dimensional fast-slow system, consisting of a Rössler oscillator [2] contracting onto a slow-manifold specified as a paraboloid in

the w-variables,

$$\begin{aligned}\dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c) \\ \epsilon\dot{w} &= w - \alpha(x^2 + y^2 + z^2).\end{aligned}$$

If we choose, $a = 0.1$, $b = 0.1$, $c = 18$, $\epsilon = 0.001$ and $\alpha = 0.002$, we get the chaotic data set shown in Fig. (6.12). The left graph shows a typical trajectory of x, y, and z components, which are the Rössler oscillator. The right is the trajectory shown in x, y,and w variables. Clearly, the data set $\{x_i, y_i, z_i\}$ of the trajectory lies on a 3-dimensional nonlinear submanifold, $w = \alpha(x^2 + y^2 + z^2)$. It is not a surprise that the ISOMAP method can approximately reveal structure of the true 3-dimensional submanifold. There are approximately 100,000 data points, which are not uniformly distributed in the space, shown in Fig. (6.12). As discussed in the example of the Lorenz system, therefore, a data set with a proper density has to be generated for the purpose of the application of the ISOMAP method.

The left graph in Fig. (6.13) shows a uniformly distributed resampling data set $\{x : u_i, y : u_i, z : u_i\}$ generated by the meshing method. The resampling data set has a sufficient data density to perform the ISOMAP method. The right graph in Fig. (6.13) is the set of the ISOMAP projections $\{x : iso_i, y : iso_i, z : iso_i\}$ of the resampling data set. The 3-dimensional projections serve sufficiently for revealing the underlying structure of the Rössler oscillator embedded on the slow-manifold.

However, in this particular case, the density of ISOMAP projections at some specified location, where a big hole is observed, is sparse. Therefore, a

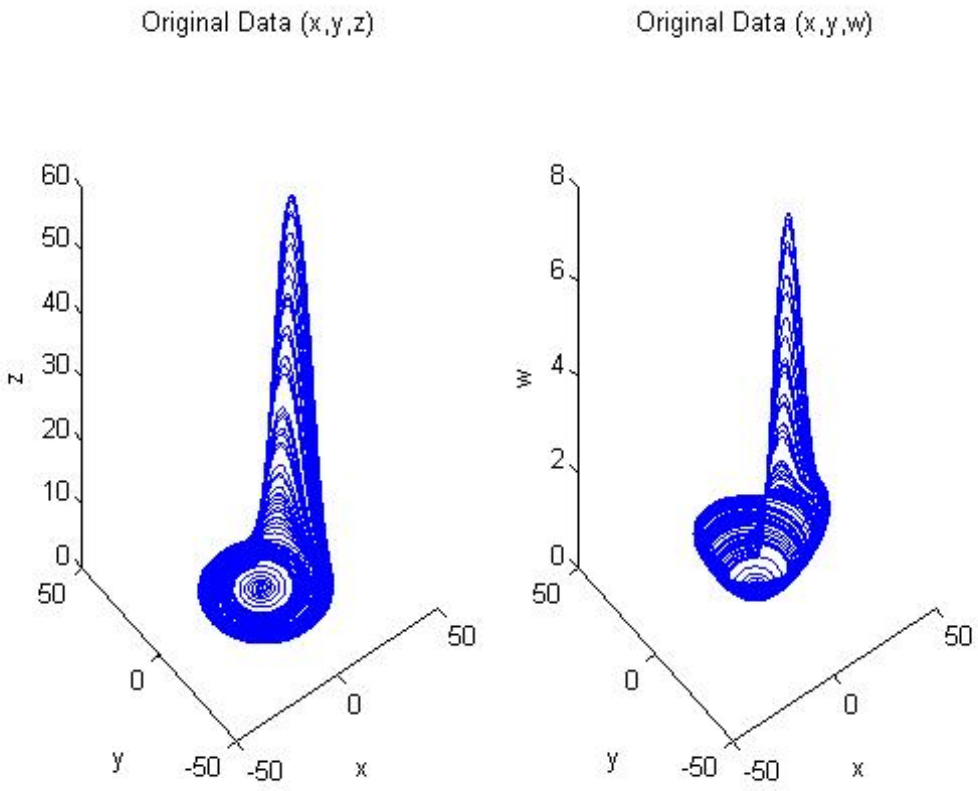

Figure 6.12: A Fast-Slow Rössler Oscillator on a Paraboloid Submanifold.

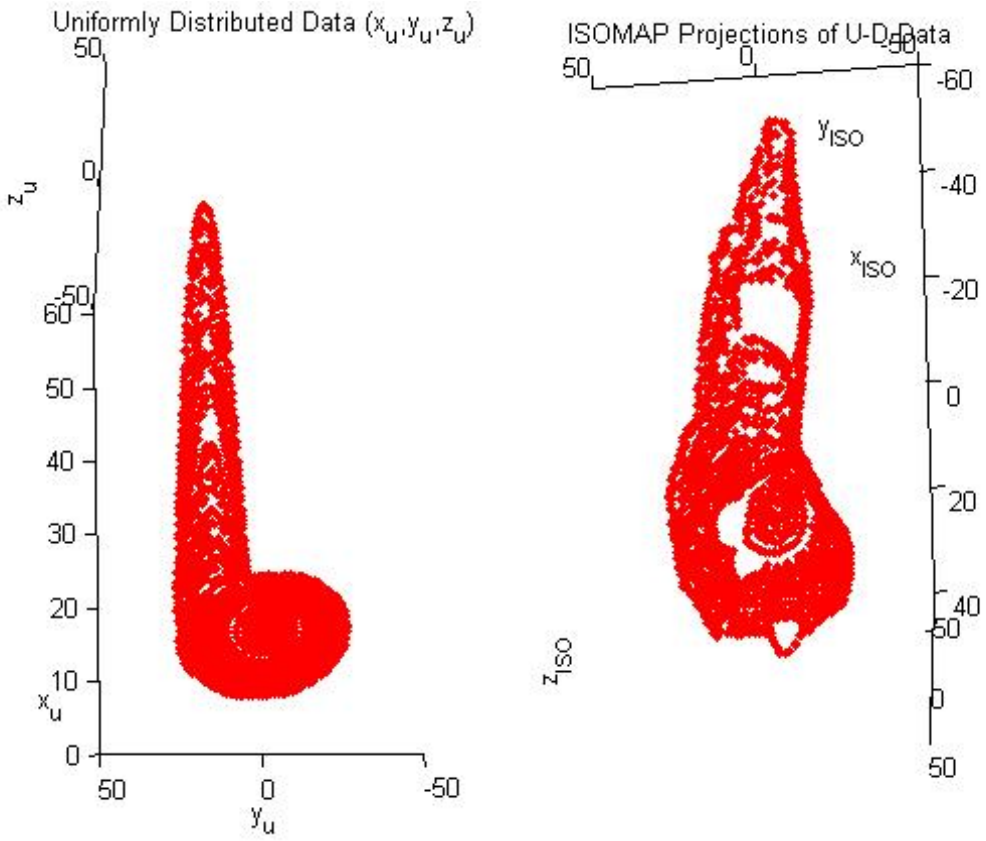

Figure 6.13: ISOMAP Projections of the Uniformly Distributed Resampling Data Set.

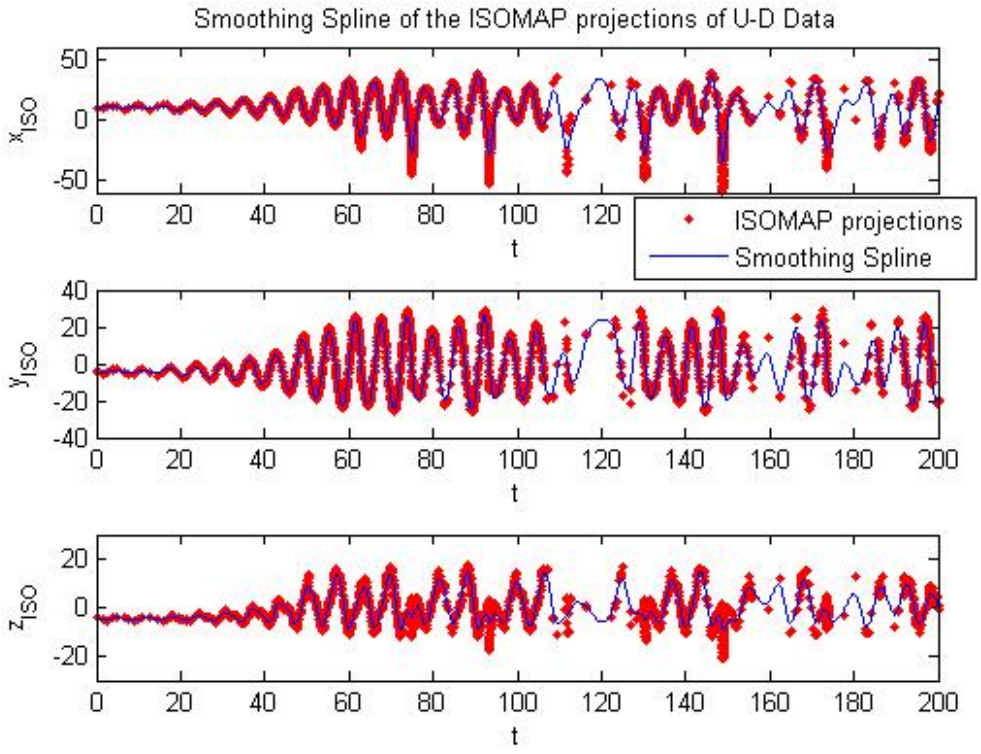

Figure 6.14: The ISOMAP Projections Regridded by the Smoothing Spline.

further operation of regridding to reduce the number of data points where they are particularly dense, and to increase data density where it is sparse, is needed, so that a uniformly distributed data set in time t can be generated for the purpose of modeling the ISOMAP projections. A traditional method of regridding uses multivariate splines. In this case, we use the smoothing spline to fit the ISOMAP projections. The smoothing spline minimizes the functional,

$$p\sum_i [x : iso_i - s(t_i)]^2 + (1-p)\int (\frac{d^2 s}{dt^2})^2 dt, \tag{6.15}$$

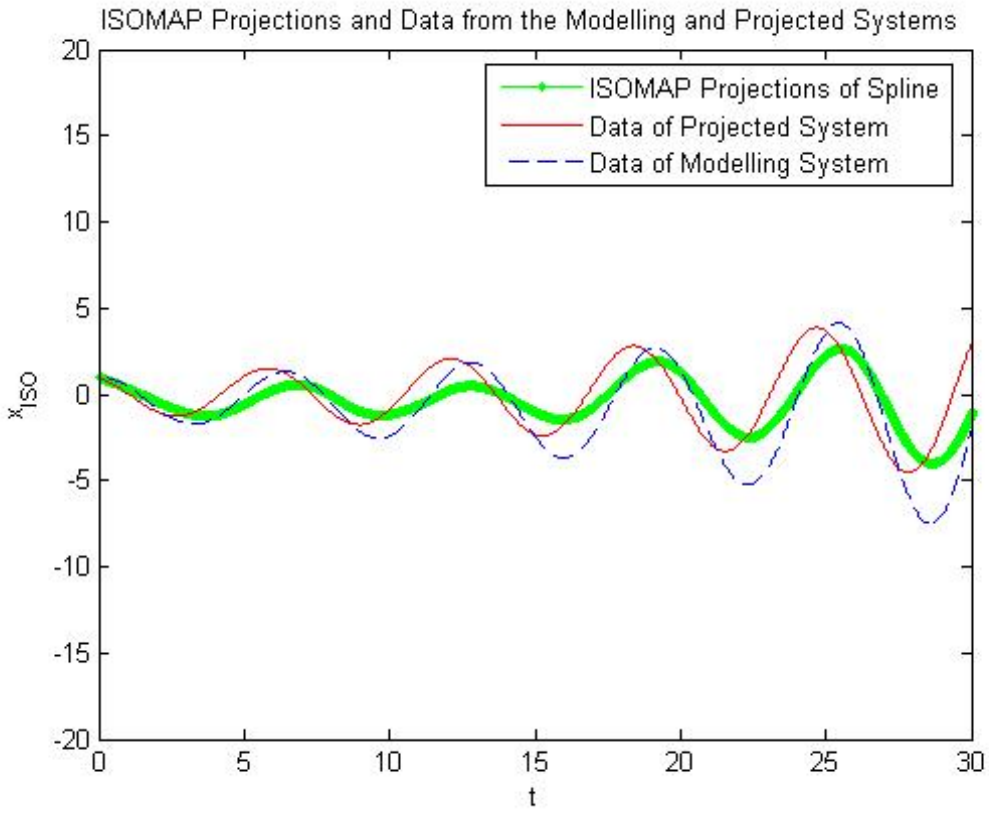

Figure 6.15: x-Component of the ISOMAP Projections Generated by the Smoothing Spline and the Data from the Modelling and the Projected Systems.

where $0 \leq p \leq 1$, $x : iso_i$ are the ISOMAP projections, and $s(t)$ is the smoothing function given by the smoothing spline. Notice that this functional is a balance between least squares straight line fit while $p = 0$, and a cubic spline interpolant while $p = 1$. We use the Matlab Spline Toolbox [34] of the smoothing spline to obtain the results shown in Fig. (6.14). The continuous curves, which are evaluated at a uniformly distributed time grids t_i, shown in blue are the functions given by the

smoothing spline. To use the 3-dimensional data set generated by the smoothing spline, $\{x : iso\_spline, y : iso\_spline, z : iso\_spline\}$, we can model the ISOMAP projections continuously in time t.

The 4th order scheme and the quadratic fitting form are applied when we perform the modeling procedure. The reproduced data of the $\{x : iso\_spline\}$ component from the modeling system is shown in Fig. (6.15). Also shown as (-) is the data from the projected system of this 4-dimensional fast-slow system for the purpose of comparison. The projected system of this 4-dimensional fast-slow system is generated by the arc length of the vector function r, which represents a point (x, y, z, w) in the fast-slow system.

$$r < x_1, x_2, x_3 >= x_1 e_1 + x_2 e_2 + x_3 e_3 + \alpha(x_1^2 + x_2^2 + x_3^2) e_4,$$

where e_i is a unit vector in $\Re^4$ that has 1 in the ith position. The vector r represents a point (x, y, z, w) in the fast-slow system. The form of the 3-dimensional projected system is provided by the arc length function in each direction and the analytical form of the Rössler system,

$$\begin{aligned} \dot{L_{x_1}} &= |\tfrac{dr}{dx_1}| \cdot f_1 = \sqrt{1 + 4\alpha^2 x^4} \cdot (-y - z) \\ \dot{L_{x_2}} &= |\tfrac{dr}{dx_2}| \cdot f_2 = \sqrt{1 + 4\alpha^2 y^4} \cdot (x + ay) \\ \dot{L_{x_3}} &= |\tfrac{dr}{dx_3}| \cdot f_3 = \sqrt{1 + 4\alpha^2 z^4} \cdot [b + z(x - c)] \end{aligned} \quad , \tag{6.16}$$

where f_i is the ith function of the Rössler system.

In this case, we extend the application of the ISOMAP method for some data set having spares density by applying the meshing method and the smoothing spline, and also show the relationship between the original system and the projected system

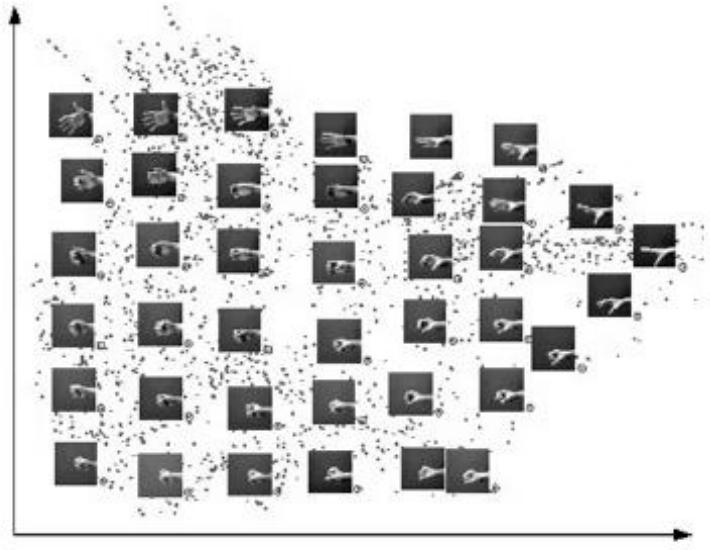

Figure 6.16: Example of ISOMAP Projections on an Image Space.

explicitly.

6.3 high-dimensional Example: Image Spaces and Video Sequences

In this section, we will explore the application of the ISOMAP method and the modeling technique in high-dimensional systems involving image spaces and video sequences. ISOMAP algorithm was developed initially for the purpose of discovering a low-dimensional representation of the space of possible images in image processing [4, 47, 52], the approach of which gives a new tool for the video analysis.

A collection of related images defines an image space, which is formed by video sequences in the examples shown in this section. Given an image having $m \times n$ pixels, we can consider it as an element in the mn-dimensional image space where each dimension is the intensity at a particular pixel. Therefore, a set of images can be used as a set of points in the mn-dimensional space to perform the ISOMAP method to seek lower dimensional projections of the mn-dimensional data points.

As discussed in [52], the ISOMAP projections of images, which describe the motions of hand shown in Fig. (6.16), reveal good results of the dimensionality reduction of this image space, particularly when there exists a two dimensional underlying subspace. The two major motions of the hand are wrist rotation considered as the x-direction in Fig. (6.16), and finger extension considered as the y-direction. The circled points are the ISOMAP projections of the shown hand images. As shown in the Fig. (6.16), the ISOMAP projections of those images, on which the hand is rotating while the fingers of the hand keep unstretched, are almost horizontal, and the ISOMAP projections of images, on which the fingers are stretching while the hand is unrotated, are almost vertical. Those are the 2 degrees of freedom apparently. ISOMAP is capable of discovering this fact from high-dimensional image sequences. Those properties of the 2-dimensional ISOMAP projections reveal the relationship among the images according to the two major axes.

Given a video,which can be considered as a collections of an ordering images establishing a trajectory, which is a low-dimension description of the video sequences, we are able to seek the relationship among the ISOMAP projections of images from the video to reveal the motion of the trajectory by the modeling procedure. In many cases, the form of the trajectory highlights important transitions in the scene or deviations from the normal scene appearance.

6.3.1 Example: Video Ball

First, we show a video of a ball generated by releasing it along a declining bench. In Fig. (6.17), the blue dots show the 2-dimensional ISOMAP projections of images recorded by the video. In practice, the ball is moving horizontally, the fact of which is illustrated by the relationship among the ISOMAP projections. As shown in Fig. (6.17), there are two major motions of the ball. One is along the declining bench, which is shown horizontally in x-direction; and the other is the rotation of the ball itself, which is shown as the small perturbations in y-direction. The reproduced data from the modeling system shown as small circles capture the major motion of the ball.

6.3.2 Example: Video Chicken

The second example we show in Fig. (6.18) is an application of ISOMAP method on a video generated by replacing the ball by a wooden chicken. As discussed in the first example, the chicken is moving along the declining bench while it

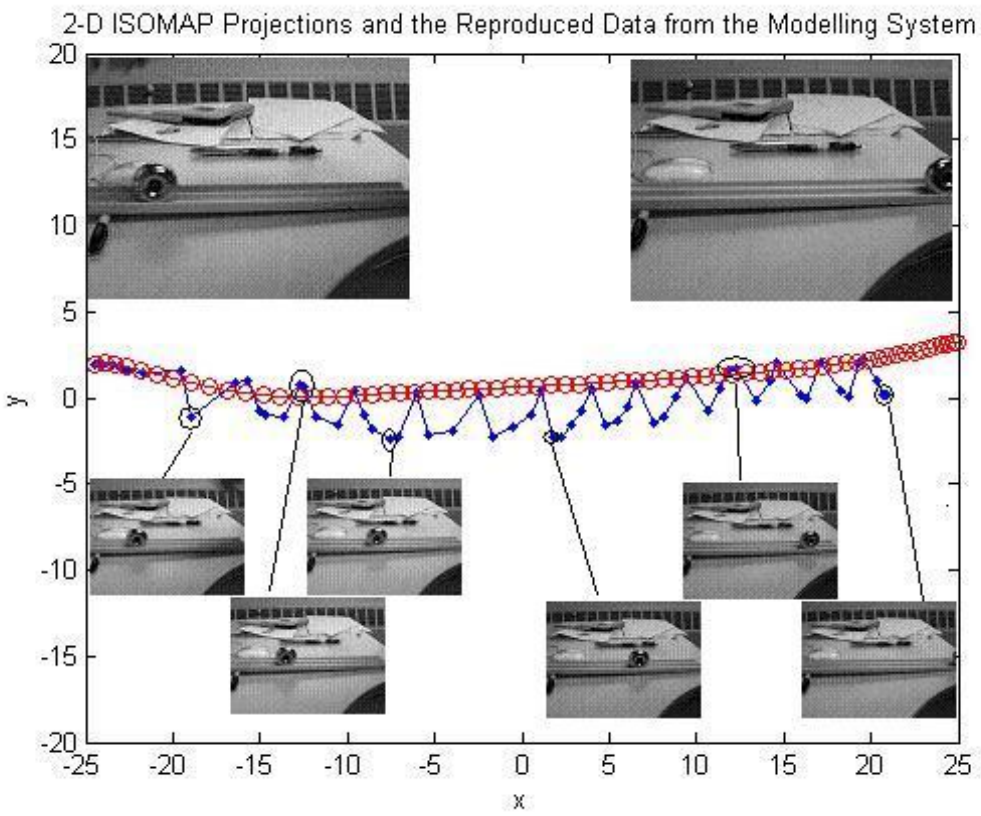

Figure 6.17: 2-D ISOMAP Projections of a Video of a Ball and the Reproduced Data.

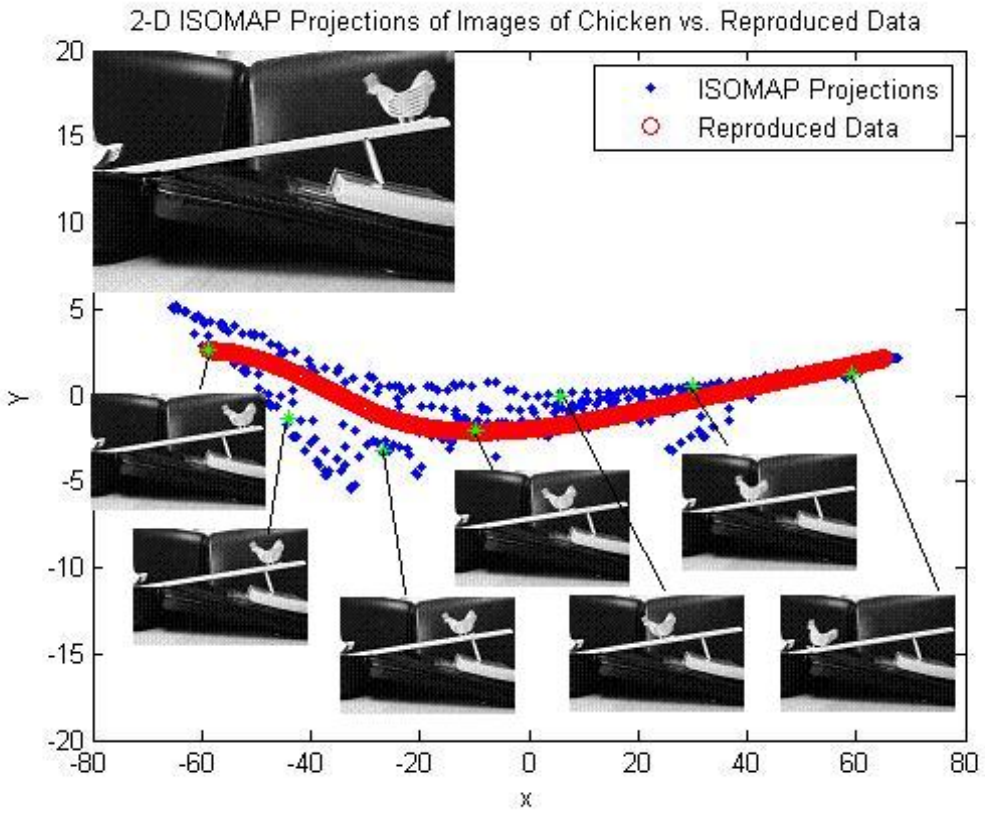

Figure 6.18: 2-D ISOMAP Projections of Video of Chicken and the Reproduced Data.

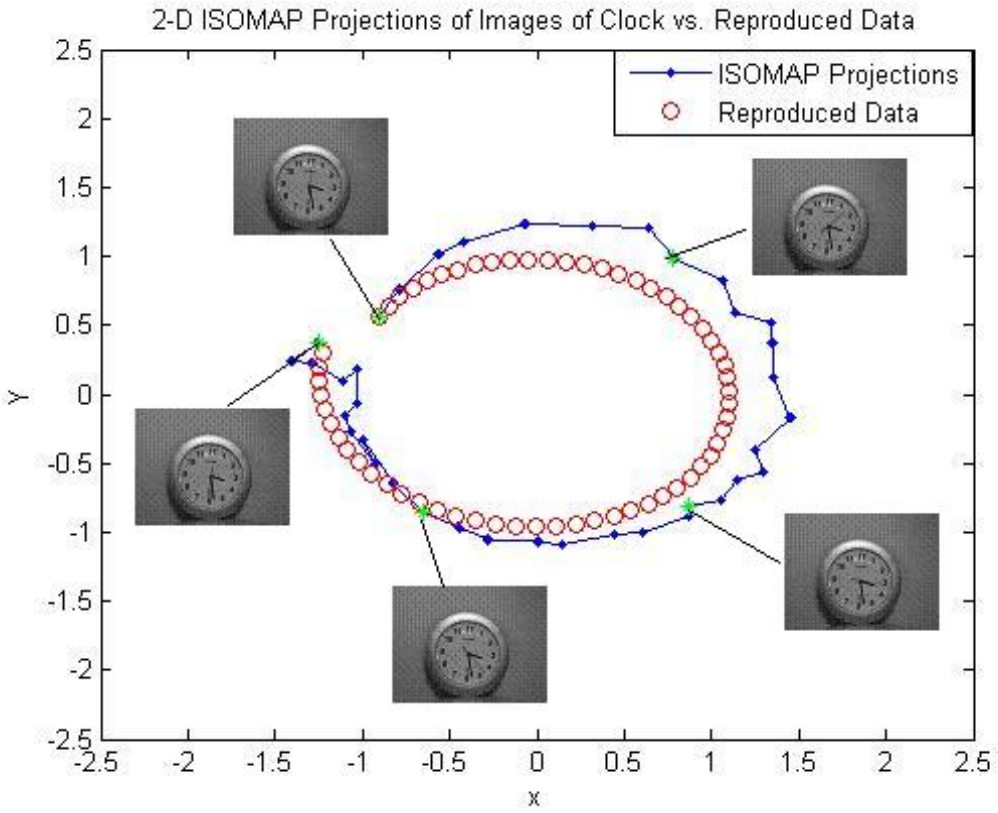

Figure 6.19: 2-D ISOMAP Projections of Video of a Clock vs. Reproduced Data.

oscillates vertically. The x-direction describes the motions along the bench, while the y-direction describes the oscillations of the chicken itself. The modeling system captures the major motion of the chicken.

6.3.3 Example: Video Clock

In the third example shown in Fig. (6.19), a video generated by recording a clock for thirty seconds is used to perform the ISOMAP method. The 2-dimensional

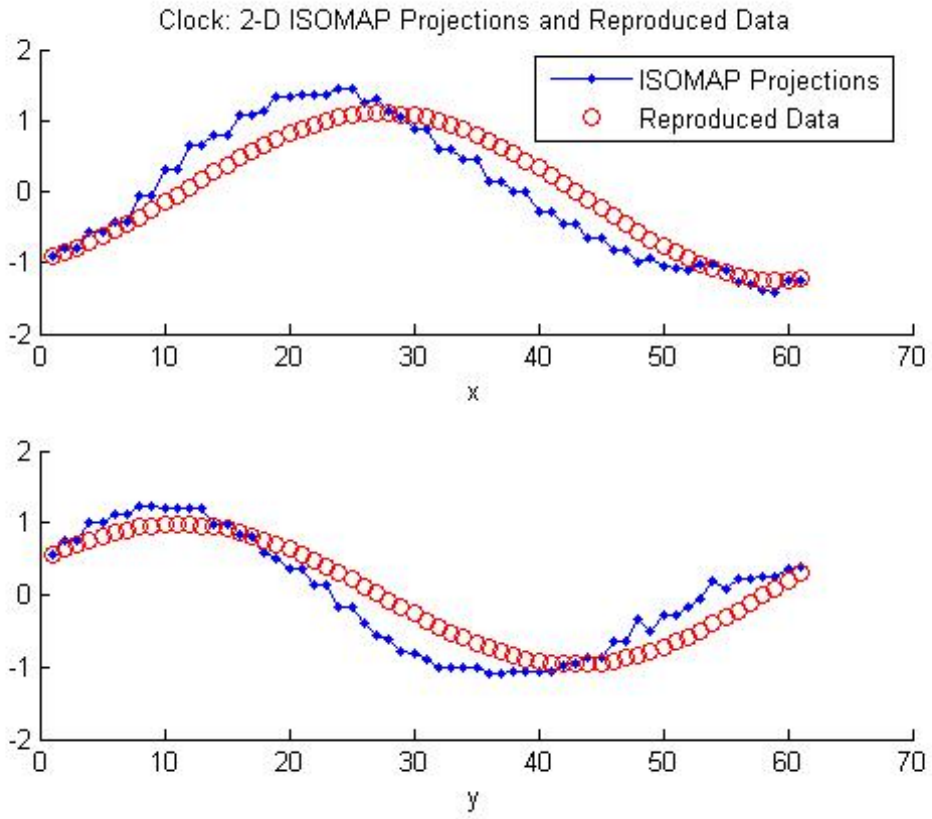

Figure 6.20: Time Series of ISOMAP Projections and Reproduced Data.

ISOMAP projections of images of this video reveal the trajectory of this set of images, which describe the movement of the second arm during the period of time. As shown in Fig. (6.20), the modeling system generating the reproduced data, the analytical form of which is given in the form of,

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -0.5686 \\ -0.1628 \end{bmatrix} + \begin{bmatrix} 0.7323 & 11.4365 \\ -8.5219 & -0.1833 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix},$$

fits the ISOMAP projections well.

6.4 Summary

In this chapter, we have performed the technique of modeling low-dimensional submanifolds in some examples, in which the projected systems, which can be constructed by arc length functions and analytical forms of original systems, and the modeling systems given by the modeling technique, have been shown for the purpose of investigating the underlying structures of the equations of the systems on the low-dimensional submanifolds, which are embedded in some high-dimensional dynamical systems, and even image spaces. In the examples of image spaces, we have shown some simple videos to give some insight into applications of the modeling technique in more complicated cases, such as video sequences generated by more complicated motions.

Chapter 7
Summary

Modeling high-dimensional process by a low-dimensional process is a fundamental problem in dynamical systems. In this dissertation, we presented a new modeling technique, in which an analytical form of equations on a low-dimensional invariant manifold, known only through a data set, is able to be obtained.

Given a data set generated from some unknown high-dimensional dynamical system, of which the asymptotic behavior may be restricted to a nonlinear submanifold, we performed the ISOMAP method, in which the meshing method and the method of linear combinations can be applied for the purpose of satisfying the sufficient conditions required by the asymptotic convergence theorem, to obtain a set of low-dimensional points, which is an approximation of embedding low-dimensional submanifold. In order to obtain an analytical form that dominates the empirical data on a low-dimensional submanifold, if there exists one, the global modeling technique was applied. As discussed in this dissertation, depending on the used fitting schemes, the analytical form developed by the modeling technique converges as the sampling rate, which was considered as the time step size h, approaches zero. Hence, the obtained analytical form can be considered as a closed-form representation of the system on the low-dimensional submanifold. With the analytical form, several useful processes in applied dynamical systems theory, such as prediction of the time series for a short period of time, investigations of the structure of the

low-dimensional submanifold, and even the relationship between the analytical form and the original high-dimensional system are able to be addressed.

Certain merits of this technique based on the ISOMAP method and the global modeling are highlighted below in comparison with what have been developed for the purpose of dimensionality reduction.

1. This technique requires only a data set to discover a low-dimensional submanifold, while some other methods, such as the theory of global attractors approaches, and singular perturbation theory, require an analytical form of the system.

2. This technique can deal with an invariant space that is intrinsically a highly nonlinear manifold, while the popular POD method is essentially linear in nature.

3. This technique produces an analytical form of the system on the low-dimensional submanifold, by which the properties of the system are able to be discovered. Moreover, for some cases in the fast-slow systems, the produced analytical form is related to the original system in the sense of the arc length.

4. This technique can be applied in various fields, such as the image processing.

7.1 Future Work

There are several directions in which we may move the work of this dissertation in

the future. Some of them are described as follows:

1. Extend the modeling technique to more dynamical systems. The examples discussed were systems with global attractors, or the fast-slow systems with a slow manifold. It is likely that there are other interesting dynamical systems with embedding low-dimensional manifold.

2. Explore the relationship between the initial distributions of the initial conditions of dynamical systems and the invariant measure on the invariant set.

3. Improve the global modeling technique by using more complicated basis functions, such as periodic functions, instead of polynomials. Moreover, the modeling technique can also be improved by using higher order fitting schemes.

4. Apply the modeling technique in some other fields. In this dissertation, simple examples of the image spaces were shown. We believe that the modeling technique can be used as a new tool for video analysis and manipulation in the field of image processing.

References

[1] H. Abarbanel, *Analysis of Observed Chaotic Data,* (New York: Springer, 1996).

[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] M. Bernstein, V. de Silva, J. C. Langford, J. B. Tenenbaum, "Graph Approximations to Geodesics on Embedded Manifolds", 2000.

[5] J. Birman, R. F. Williams, "Knotted Periodic Orbits in Dynamical System-I: Lorenz's Equations", *Topology,* 22(1) 47 82 (1983).

[6] E. Bollt, "Model selection, confidence, and scaling in predicting chaotic time-series", *International Journal of Bifurcation and Chaos (IJBC) in Applied Sciences and Engineering*, Vol. 10, No. 6, 1407-1422 (2000).

[7] E. Bollt, "Attractor Modeling and Empirical Nonlinear Model Reduction of Dissipative Dynamical Systems" to appear International Journal of Bifurcation and Chaos (IJBC) in Applied Sciences and Engineering, 17 4 (2007).

[8] E. Bollt, C. Yao, I. Schwartz, "Implications of Dynamical Data on Manifolds to Empirical KL Analysis of Fast-Slow Systems" Submitted to SIAM J. of Dynamical Systems (SIADS), (2007).

[9] J. L. Breeden, A. Hubler, "Reconstructing equations of motion from experimental data with unobserved variables", *Phys. Rev. A*, 42, 5817(1990).

[10] R. Brown, N. Rulkov, E. Tracy, "Modeling and synchronizing chaotic systems from time-series data", *Phys. Rev. E* 49, 5 (1994).

[11] R. L. Burden, J. Douglas Faires, A. C. Reynolds *Numerical analysis (second edition),* (Prindle, Weber & Schmidt).

[12] J. Carr, *Applications of Center Manifold Theory, (*Springer-Verlag, NY 1981).

[13] T. F. Cox, M. A. Cox, *Multidimensional Scaling*, (Chapman and Hall, London).

[14] J. D. Farmer, J. Sidorowich, "Predicting chaotic time-series", *Phys. Rev. Lett*. 59, 845 (1987).

[15] J. Farmer, J. Sidorowich, “Exploiting Chaos to Predict the Future and Reduce Noise”, appear in *Evolution, Learning, and Cognition,* (ed. Y.C. Lee, World Scientific Press, pp. 277, 1988).

[16] N. Fenichel “ Geometric singular perturbation theory for ordinary differential equations. J. Differential Equations”, 31(1):5389, 1979.

[17] P. Holmes, J. Lumley, G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, (Cambridge, UK, 1996).

[18] T. J.R. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, (Dover Publications).

[19] M. Giona, “Functional reconstruction and local prediction of chaotic time-series”, *Phys. Rev. A* 44, 3496(1991).

[20] G. H. Golub, C. F. Van Loan, *Matrix, Computations, third edition* (The Johns Hopkins University Press, 1996).

[21] G. Gouesbet, “Reconstruction of the vector fields of continuous dynamical systems from numerical scalar time-series”, *Phys. Rev. A* 43, 5321 (1994).

[22] J. Guckenheimer, P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, (Springer, NY, 1983).

[23] H. Kantz, T. Schrieber, *Nonlinear time-series analysis*, (Cambridge U. Press, New York, 1997).

[24] K. Karhunen, *Ann. Acad. Sci. Fennicae 34, Ser. A1, 37* (1946)

[25] J. K. Kevorkian, J. D. Cole, *Multiple Scale and Singular Perturbation Methods (Applied Mathematical Sciences)*, (Springer, 1996).

[26] B. Kitchens, *Symbolic Dynamics, One-sided, Two-sided and Countable State Markov Shifts,* (Springer, 1998).

[27] J. Krener, “The local convergence of the extended Kalman filter”, (2004). http://arxiv.org/abs/math.OC/0212255.

[28] W. Kosmala, *A Friendly Introduction to Analysis; Single and Multivariable. (Second Edition)*, (Pearson Education, NJ, 2004).

[29] E. J. Kostelich, J. A. Yorke, “Noise reduction in dynamical systems”, *Phys. Rev. A* 38, 1649 (1988).

[30] D. Lind, B. Marcus, *An Introduction to Symbolic Dynamics and Coding,* (Cambridge University Press, 1995).

[31] M. Loeve, *Probability Theory,* (Van Nostrand, Princeton, NJ, 1955).

[32] E. N. Lorenz, "Deterministic Nonperiodic Flows", *J. Atmospheric Science*, 20 130-141.

[33] J. L. Lumley, *Stochastic Tools in Turbulence, (*Academic, NY, 1970).

[34] Matlab 7, The MathWorks, (Natick, MA, 2005).

[35] J. Munkres, *Topology, 2nd Edition*, (Prentice-Hall 2000).

[36] G. Nakos, D. Joyner, *Linear Algebra with Applications*, (Brooks/ Cole Publishing Company, 1998)

[37] L. Perko, *Differential equations and dynamical systems (3rd Edition)*, (Springer-Verlag, New York, 2001).

[38] W. Press, B. Flannery, S. Teukolsky, W. Veterling *Numerical Recipes in C, the Art of Scientific Computing*, (Cambridge University Press, 1988).

[39] J. Rissanen, Ann. Stat. 11, 416(1983); Automatica 14,465 (1978); *Stochastic complexity in statistical inquiry*, (World Scientific, NJ, 1989).

[40] J. C. Robinson, "All possible chaotic dynamics can be approximated in three dimensions", *Nonlinearity*, 11 (1998) 529-545.

[41] J. Robinson, *Infinite-Dimensional Dynamical Systems (From Basic Concepts to Actual Calculations),* (Cambridge University Press, 2001).

[42] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos (second edition)*, (CRC Press LLC, 1999).

[43] S. Roweis, L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", *Science* 290: 2323-2326.

[44] H. L. Royden, *Real Analysis, third edition*, (Prentice Hall, NJ).

[45] W. J. Rugh, *Linear Systems Theory, Second Edition*, (Upper Saddle River, NJ: Prentic Hall, 1996).

[46] J. D. Skufca, E. M. Bollt, "A Concept of Homeomorphic Defect for Defining Mostly Conjugate Dynamical Systems," submitted to Physica D, (2007).

[47] V. de Silva, J. B. Tenenbaum, “Global versus Local Methods in Nonlinear Dimensionality Reduction”.

[48] L. Sirovich, “Chaotic Dynamical of Coherent Structure”, *Physica D 37 126-145 (1989)*.

[49] J. Stewart, *Calculus : Early Transcendentals (5th edition),* Brooks/Cole.

[50] S. H. Strogatz, *Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering,* (Westview, 1994).

[51] F. Takens, *Dynamical Systems and Turbulence*, edited by D. Rand and L.-S. Young, Lecture Notes in Mathematics No. 898 (Springer, Berlin), p.366, (1980).

[52] J. B. Tenenbaum, V. de Silva, J. C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, *Science* 290 5500 (2000).

[53] A. N. Tokhonov, A. B. Vasileva, A. G. Sveshnikov, *Differential Equations,* (Springer, 1985).

[54] I. Triandaf, I. B. Schwartz, “Karhunen-Loeve mode control of chaos in a reaction-diffusion process”, *Phys. Rev. E* 56 204 (1997).

[55] C. F. Van Loan, *Computational frameworks for the fast Fourier transform*, (SIAM Publications, Philadelphia, PA. 1992).

[56] C. F. Van Loan, N. Pitsianis, “Approximation with Kronecker products”. www.math.tkk.fi/ mhuh-tane/kronecker.ps

[57] A. S. Weigenbend, N. A. Gershenfeld, *Times Series Prediction: Forecasting the Future and Understanding the Past,* (eds. Reading, MA: Addison-Wesley, 1993).

[58] G. Welch, G. Bishop, “An Introduction to the Kalman Filter”, (2004).

[59] R. F. Williams, “The Structure of Lorenz Attractors”, *Publications Mathematiques de 1’IHS,* 50, 73-99.

[60] C. Yao, E. Bollt, “modeling and Nonlinear Parameter Estimation with Kronecker Product Representation for Coupled Oscillators and Spatiotemporal Systems”, *Physica D 277 1, 78-99* (2007).

[61] C. Yao, E. Bollt, “Application of ISOMAP in Dynamical Systems, Despite Non-Uniform Invariant Measure”, submitted to MCMDS, 2006.

[62] G. Young, A.S. Householder, “Discussion of a set of points in terms of their mutal distance”, *Psychometrika, 3, 19-22* (1938).