

MARKOV CHAIN AND TIME-DELAY REDUCED
MODELING OF NONLINEAR SYSTEMS

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Kawin Nimsaila

December 2009

MARKOV CHAIN AND TIME-DELAY REDUCED
MODELING OF NONLINEAR SYSTEMS

Kawin Nimsaila

APPROVED:

Dr. Ilya Timofeyev, Chairman

Dr. Matthew Nicol

Dr. Giovanna Guidoboni

Dr. Christian Franzke
British Antarctic Survey, UK

Dean, College of Natural Sciences
and Mathematics

ACKNOWLEDGMENTS

First of all, I would like to thank Dr. Ji who helped and supported me during the first year of my study. My special thank goes to my advisor, Dr. Timofeyev, for guiding and supporting me throughout my study. Without his help, my dissertation would not have been completed. I would like to thank my committee members, Dr. Nicol, Dr. Guidoboni, and Dr. Franzke, for reading this dissertation and giving valuable comments. I am thankful to my friend, Arjun, for some useful discussions related to my research. Finally, I would like to thank my parents for giving me the opportunity to study abroad and encouraging me during my study.

MARKOV CHAIN AND TIME-DELAY REDUCED
MODELING OF NONLINEAR SYSTEMS

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Kawin Nimsaila

December, 2009

ABSTRACT

Multiscale modeling problems have become an active research area in recent years. There are many systems involving a large set of variables and these variables mostly behave in largely different time scales. It is necessary to derive proper effective models when one needs to obtain dynamical models that reproduce statistical properties of essential variables without wasting the computational time to compute non-essential variables in high dimensional systems.

In this dissertation, we develop two new approaches for stochastic effective models. The Markov chain stochastic parameterization technique is proposed for the effective models in the first part of this dissertation. This is a numerically oriented approach where some parts of the right hand side of essential variables are modeled by conditional Markov chains. It is shown that, under the proper conditioning scheme, statistical properties of essential variables from effective models have a good agreement with full models. Furthermore, we illustrate that the implementation of effective models including the conditioning scheme and the estimation of the transition probability matrices is simple and straightforward.

For the second part of this dissertation, we propose effective models using a stochastic delay differential equation. The memory part in stochastic delay models is a simple linear combination of essential variables with finite number of delays. We apply this technique to the Truncated Burgers-Hopf equation and show that the effective model reproduces statistical behaviours of the full model.

Contents

Acknowledgments	iii
Abstract	v
List of Tables	viii
List of Figures	xi
1 Introduction and Background	1
1.1 Background	1
1.2 Dissertation Outline	4
2 Markov Chain Modeling Approach for Effective Models	5
2.1 Introduction	5
2.2 General Setup for Markov Chain Modeling	6
2.2.1 The Estimation of Transition Probability Matrices	9
2.2.2 The State Space of the Markov Chain	10
2.3 Prototype Models	12
2.3.1 Triad Model Coupled with 2 Independent Non-essential Variables	12
2.3.2 Triad Model Coupled with Dependent Non-essential Variable	31

2.4	Truncated Burgers-Hopf Model	36
2.4.1	The TBH Equation	37
2.4.2	The Effective Model for TBH	39
2.4.3	One Essential Variable, $\lambda = 1$	40
2.4.4	Two Essential Variables, $\lambda = 2$	44
2.4.5	The Effective Model for TBH with Non-zero Hamiltonian	47
2.5	Conclusion	48
3	Stochastic Delay Differential Equation for the Effective Model	51
3.1	Introduction	51
3.2	General Setup for the Stochastic Delay Model	52
3.2.1	Model Setup and Parameter Estimation	53
3.3	The Effective Model for the First Fourier Mode of TBH System	56
3.3.1	Numerical Results for One Fourier Mode	57
3.4	The Effective Model for the First Three Fourier Modes	71
3.5	Conclusion	78
	Appendix	79
	A Asymptotic Mode Elimination	79
A.1	Forward and Backward Equations	79
A.2	Asymptotic Mode Elimination	82
A.2.1	Averaging Method	82
A.2.2	Homogenization Method	85
	B Least Square Method	90
	Bibliography	92

List of Tables

2.1	Mean and variance of x_1, x_2 , and x_3 in the simulation of the full model defined in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The parameters of the stochastic triad system are defined in (2.13)	17
2.2	Mean and variance of x_1, x_2 , and x_3 in the simulation of the full model in (2.9), and the Markov chain effective model in (2.10). The Markov chain effective model has 11 states. The state spaces for the effective model are both discrete and continuous.	20
2.3	Mean and variance of x_1, x_2 , and x_3 from the full model in (2.9) and the Markov chain effective models in (2.10). The numbers of states in the Markov chain effective model are 5, 7, 11, and 15.	24
2.4	Means and variances of x_1, x_2 , and x_3 from the full model in (2.9) and the Markov chain effective model in (2.10). The time steps used in the effective model are 0.0001, 0.001, 0.005, 0.01, and 0.1	26
2.5	Mean and variance of x_1, x_2 , and x_3 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10) The time steps for the Markov chain effective model are 0.01 and 0.1.	30

2.6	Mean and variance of essential variables from the full model defined in (2.18), the asymptotic reduced model defined in (2.20), and the Markov chain effective model with different conditioning schemes.	34
2.7	Mean, variance, and kurtosis of u_1 from the full model and the effective model.	42
2.8	Mean, variance, and kurtosis of u_1 and u_2 from the full model and the effective model with $\lambda = 2$	45
2.9	Mean, variance, and kurtosis of u_1 from the full model and the effective model. The TBH equation has non-zero Hamiltonian.	48
3.1	The coefficients of the stochastic delay differential equation defined in (3.16). The coefficient values are obtained by the least square method defined in (3.11).	60
3.2	The correlation time of e_1^{re} , e_1^{im} , u_1^{re} , and u_1^{im}	62
3.3	Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the stochastic delay effective model defined in (3.16).	64
3.4	The coefficients of the stochastic delay effective model defined in (3.16) with two time delays. The coefficient values are obtained by the least square method defined in (3.11).	66
3.5	The coefficients of the stochastic delay effective model defined in (3.16) with ten delays. The coefficient values are obtained by the least square method defined in (3.11).	67
3.6	Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the effective model defined in (3.16). The effective model consists of two time delays.	67
3.7	Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the effective model defined in (3.16). The effective model consists of ten time delays.	68

3.8	Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the effective model with five time delays. The effective model is numerically integrated with different time steps.	70
3.9	Mean, variance, and kurtosis of u_i^{re} and u_i^{im} for $i = 1, 2, 3$ from the full model and the effective model.	74

List of Figures

2.1	Auto-correlation functions of essential variables, x_1, x_2, x_3 , and non-essential variables, y_1, y_2 , from a full model in equation (2.9).	16
2.2	Auto-correlation function of x_1 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The Markov chain effective model has 11 states for the Markov chain. The partition for the state is defined in (2.14).	18
2.3	Auto-correlation functions of x_2 and x_3 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The Markov chain effective model has 11 states for the Markov chain. The partition for the state is defined in (2.14). . .	19
2.4	Comparison of the auto-correlation function of x_1 between discrete state space and continuous state space. The number of states is 11. .	21
2.5	Auto-correlation functions of x_1 from the full model in (2.9) and the Markov chain effective model in (2.10) with different numbers of states of the Markov chain.	23
2.6	Auto-correlation function of x_1 from the full model in (2.9) and the Markov chain effective model in (2.10). The time steps used in the Markov chain effective model are 0.0001, 0.001, 0.005, 0.01, and 0.1. Note that the Markov chain effective models with $\Delta t = 0.0001$ and 0.1 do not properly reproduce the auto-correlation function of x_1	25

2.7	Auto-correlation functions of x_1, x_2, x_3, y_1 , and y_2 from the full model in (2.9) with the parameters in (2.17).	28
2.8	Auto-correlation function of x_1 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The time steps for the Markov chain effective model are 0.01 and 0.1.	29
2.9	Auto-correlation function of x_1 from the full model, the asymptotic reduced model, and the Markov chain effective model with different conditioning schemes.	35
2.10	Auto-correlation functions of u_k , $k = 1, \dots, 5$. The TBH has $\Lambda = 20$, $E = 0.4$, and $H = 0$	39
2.11	Statistical properties of u_1 . The top left figure is the ACF of u_1^{re} . The top right is the ACF of u_1^{im} . The bottom left is the PDF of u_1^{re} and the bottom right is the PDF of u_1^{im}	43
2.12	Auto-correlation functions of u_1 and u_2 . The top left figure is the ACF of u_1^{re} . The top right is the ACF of u_1^{im} . The bottom left is the ACF of u_2^{re} and the bottom right is the ACF of u_2^{im}	46
2.13	Probability density functions of u_1 and u_2 . The top left figure is the PDF of u_1^{re} . The top right is the PDF of u_1^{im} . The bottom left is the PDF of u_2^{re} and the bottom right is the PDF of u_2^{im}	46
2.14	Statistical properties of u_1 . The top left figure is the ACF of u_1^{re} . The top right is the ACF of u_1^{im} . The bottom left is the cross-correlation between u_1^{re} and u_1^{im} . The bottom right is the PDF of u_1^{re}	49
3.1	Cross-correlation function between g_1^{re} and u_1 . The solid line is the cross-correlation between g_1^{re} and u_1^{re} . The dashed line is the cross-correlation between g_1^{re} and u_1^{im}	59

3.2 Cross-correlation function between g_1^{im} and u_1 . The solid line is the cross-correlation between g_1^{im} and u_1^{im} . The dashed line is the cross-correlation between g_1^{im} and u_1^{re} 59

3.3 Cross-correlation functions between noises (e_1^{re}, e_1^{im}) and essential variables (u_1^{re}, u_1^{im}) 61

3.4 Auto-correlation functions of e_1^{re} and u_1^{re} . Note that the auto-correlation of e_1^{re} decorrelates faster than the auto-correlation of u_1^{re} 63

3.5 Comparison of statistical properties of u_1 between the full model and the stochastic delay effective model defined in (3.16). The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right. 65

3.6 Comparison of statistical properties of u_1 between the full model and the stochastic delay effective model defined in (3.16). The effective model has two time delays. The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right. 68

3.7 Comparison of statistical properties of u_1 between the full model and the stochastic delay effective model defined in (3.16). The effective model has ten time delays. The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right. 69

3.8 Auto-correlation functions of u_1^{re} from the effective model. The effective model is numerically integrated with different time steps. 71

3.9 Correlation functions between g_i^{re} and $u_i^{re}, i = 1, 2, 3$ 72

3.10 The auto-correlation functions of e_i^{re} and $u_i^{re}, i = 1, 2, 3$. Note that the decorrelation time of e_i is faster than u_i , for $i = 1, 2, 3$ 73

3.11 Statistical properties of u_1 from the full model and the stochastic delay effective model. The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right. 75

3.12 Statistical properties of u_2 from the full model and the stochastic delay effective model. The auto-correlation of u_2^{re} is on the top left. The auto-correlation of u_2^{im} is on the top right. The cross-correlation between u_2^{re} and u_2^{im} is on the bottom left. The probability density function of u_2^{re} is on the bottom right. 76

3.13 Statistical properties of u_3 from the full model and the stochastic delay effective model. The auto-correlation of u_3^{re} is on the top left. The auto-correlation of u_3^{im} is on the top right. The cross-correlation between u_3^{re} and u_3^{im} is on the bottom left. The probability density function of u_3^{re} is on the bottom right. 77

Chapter 1

Introduction and Background

1.1 Background

In many high-dimensional systems, the system variables evolve on largely different time scales. Often, only a small subset of variables is of interest. These variables are called the essential variables (also sometimes called slow variables or resolved variables) in the multiscale modeling literature. The essential variables mainly represent the large-scale behaviour of the system whereas the non-essential variables (fast variables or unresolved variables) represent the small-scale behaviour of the system. To obtain statistical properties of essential variables, these systems have to be fully solved for all variables. Most of the computational time in numerical simulations is wasted on resolving the non-essential variables. Furthermore, since most of non-essential variables evolve on a fast time scale, a small time step must be utilized in the numerical integration in order to have a reasonable accuracy of a solution. Thus, it is not efficient to fully solve the full system in order to obtain statistical properties of essential variables. Therefore, there is a need for effective models that describe the dynamics of only essential variables. However, in many applications, the main objective of the effective models is not to reproduce the trajectory of essential variables

exactly, but to develop the effective models that reproduce statistical behaviours of essential variables of full systems.

The multiscale models are often found in many applications. For example, it is common in the atmospheric-oceanic models [3, 5, 6] that the variables vary on very different time scales. The atmosphere variables dominantly evolve on the time scale of days or weeks. On the other hand, the ocean variables evolve on a longer time scale such as months or years [5, 3]. To study the climate change problem, it is necessary to consider coupled atmospheric-oceanic system. However, the numerical integration is extremely inefficient since the time step is limited by atmosphere variables.

Another application of multiscale modeling is chemical reaction models [22, 23, 24]. Many chemical reactions often occur on largely different time scales. The fast reactions are fired more frequently than the slow reactions. Both of fast and slow reactions involve some of the same chemical species. In general, the slow reaction plays an important role of the system. However, most of computational time is spent on the fast reaction. Therefore, from the computational point of view, there is a need for the effective models that can describe behaviours of the slow reactions.

The derivation of effective models for essential variable has been actively studied in recent years. Several approaches have been developed for the effective models. These approaches can be categorized as semi-analytical approaches and numerically oriented approaches. Some semi-analytical approaches are based on Mori-Zwanzig formalism [31, 25]. It is based on the projection of all variables to the space of essential variables. The dynamics of essential variables obtained from Mori-Zwanzig formalism is called the generalized Langevin equation. It generally consists of three parts. The first part depends on the current values of essential variables. Therefore, it is referred to Markovian term. The second part involves the past values of essential variables and, therefore, represents the memory effect. The last part can be viewed as a noise term of the effective models. The techniques based on Mori-Zwanzig formalism include the

optimal prediction framework [10, 11, 12, 13, 8, 9]. Other semi-analytical approaches include the stochastic mode reduction technique [1, 2, 3, 4, 5] and the coarse-graining of spin-flip stochastic models [33, 34, 35]. These techniques are based on results of convergence of stochastic processes [42, 43, 44, 45].

Some of the stochastic mode reduction techniques utilize the asymptotic mode elimination approach. The asymptotic approach is the method that eliminates non-essential variables based on an infinite time scale separation. It is assumed that essential variables and non-essential variables are in very different regimes. Generally, to derive effective models, non-essential variables are replaced by their average quantities. This technique is called the averaging method. The average quantities can be constant values or functions that depend on essential variables. However, if the average quantities are zero, this approach cannot correctly reproduce the statistical properties of the full model. In this case, one needs to utilize the homogenization method which assumes a faster time scale of non-essential variables. In principle, both averaging and homogenization methods require an infinite scale separation. However, in many systems, the time scales of essential and non-essential variables might overlap and, therefore, there is no clear separation of time scales in this case. The effective models from the asymptotic approach may not be able to reproduce the statistical properties of the full system.

The numerically oriented approaches include methods for estimation of stochastic differential equations from numerical or observational data [36, 16, 17, 15], Heterogeneous Multiscale Methods [28, 29, 30, 14], and application of Hidden Markov Models [26, 32]. These techniques use numerical or observational data to estimate the effective model.

1.2 Dissertation Outline

In this dissertation, we propose two approaches to obtain the effective models. These two approaches can be categorized as numerically oriented approaches. For the first approach, we introduce the Markov chain variables in the effective models. We parameterize the interaction of non-essential variables appearing in the dynamics of essential variables by conditional Markov chains. Markov chains are conditioned on the state of essential variables. The details of Markov chain parameterization are given in chapter 2. We give an algorithm to construct Markov chains from numerical data in section 2.2. The construction of a Markov chain is simple and straightforward. It can be easy to implement for effective models. We apply this algorithm to the stochastic triad model and the Truncated Burgers-Hopf equation.

Another approach is the linear stochastic delay differential equation model. Generally, the effective models that obtain from the projection of non-essential variables consist of the Markovian term, the memory effect term, and the noise term. In our approach, we choose the memory term as a finite linear combination of the past history of the essential variables and approximate the noise term by some stochastic processes. The coefficients in a linear combination term are obtained by fitting equations with a least square method. This approach does not assume any knowledge of the model and is entirely data driven. If there is a strong time scale separation between essential variables and residuals, then residuals from least square method can be approximated by white noise. The white noise approximation is based on the asymptotic mode elimination technique [37, 38]. This approach is given in details in chapter 3.

Chapter 2

Markov Chain Modeling Approach for Effective Models

2.1 Introduction

As we mentioned in Chapter 1, the Markov chain effective models are based on a numerically oriented approach. The main idea of this approach is to construct a Markov chain that represents the behaviours of the interactions of non-essential variables. The main difference from other reduction techniques is that the time scale separation is not essential for this approach. This chapter describes the construction of the Markov chain effective models including the estimation of transition probability matrix and the estimation of the value of the Markov chain. The rest of this chapter is organized as follows. In section 2.2, we describe the construction of the Markov chain in general models. We describe how to setup Markov chain effective models from the numerical data obtained from the full model. Parameter estimation is also described in this section. Then, we apply the concept of Markov chain effective models to prototype models in section 2.3. We study the factors that affect the behaviours of the Markov chain effective models. These factors include the time step used in the

parameter estimation and numerical integration of the effective models, numbers of states in Markov chain models, and the number of conditional Markov chains. We also apply the Markov chain effective models to a high-dimensional deterministic system such as the Truncated Burgers-Hopf (TBH) equation. The details are provided in section 2.4. Finally, conclusions are included in section 2.5

2.2 General Setup for Markov Chain Modeling

We consider the dynamical systems where the dependent variables can be categorized into two groups: the essential variables, x , and the non-essential variables, y . The essential variables are typically the variables of interest. They represent the average behaviour of the system. Let us consider the system of the following form

$$\frac{dx}{dt} = f(x) + g(x, y), \quad (2.1a)$$

$$\frac{dy}{dt} = h(x, y). \quad (2.1b)$$

Both x and y are vector-valued variables where the dimension of x is normally less than the dimension of y . The main objective of this chapter is to obtain the effective model that describes the dynamics of the essential variables by replacing the interaction of the non-essential variables and the essential variables with the appropriate set of conditional Markov chains. We would like to emphasize that the purpose of the effective model is to reproduce the long term statistical behaviours of essential variables, not the original trajectory of essential variables. Thus, in (2.1), $g(x, y)$ is replaced by the Markov chain. For the effective model, we seek the equation of the following form

$$\frac{dX}{dt} = f(X) + m(Z; X) \quad (2.2)$$

where $m(Z; X)$ represents the collection of conditional Markov chains. These Markov chains are conditioned on the states of essential variables. The future values of the

Markov chain not only depend on the current values of the Markov chain, but also depend on the current values of the essential variables. Since the Markov chain modeling approach is an empirical approach, the data are obtained by sampling from the continuous model with a discrete time step. Thus, we consider the discrete version of (2.2) which can be written equivalently as follows

$$X_{n+1} = X_n + f(X_n)\Delta t + Z_{n+1}\Delta t. \quad (2.3)$$

Here $Z_{n+1} = m(Z_n; X_n)$ is the Markov chain value at time step $n + 1$ generated by conditioning on the value of Z_n and X_n . $X_n := X(n\Delta t)$ represents the time series of the essential variables.

To construct the conditional Markov chain, the discrete data of $g(x, y)$ has to be obtained from the full system. This can be done by integrating the full system with time step δt and sampling the data with time step Δt . In general, we assume that $\delta t \leq \Delta t$. The time step, Δt , is the same time step used in the numerical integration of the effective model (2.3).

For our simulations, we are interested in stationary statistical quantities of essential variables. If the process is ergodic, stationary statistical quantities can be obtained by time averaging instead of ensemble average [40]. In our simulations, we use time averaging to obtain statistical properties and we denote T as a time length for computing statistics by time averaging. Typically, T is a large number in order to get accurate statistical quantities.

To obtain the time series of the full model, we sample data with a time length of T_0 . Generally, T_0 is a short length of time such that the data are obtained enough to construct the effective model. T_0 is typically much less than T ($T_0 \ll T$).

Let $g_n = g(x_n, y_n) = g(x(n\Delta t), y(n\Delta t))$ be a time series of $g(x, y)$ sampled at time step Δt and $x_n = x(n\Delta t)$ a time series of x sampled with time step Δt . Let $N_0 = \lfloor \frac{T_0}{\Delta t} \rfloor$ where $\lfloor x \rfloor$ denotes the greatest integer that is less than or equal to x . Thus, N_0 is the length of time series that we use in the construction of the Markov chain. Since x has

a continuous state space, it requires an infinite number of conditional Markov chains in order to condition on every state space value of x . This is not possible in a practical situation. To avoid this problem, we define a finite partition on x and construct the conditional Markov chain based on this partition. The number of subintervals in this partition plays an important role in the statistical properties of the effective model as we will show in section 2.3 and section 2.4 of this chapter.

In general, if x is a vector-valued variable, a Markov chain can be conditioned on one or more components of x . However, for simplicity, we illustrate the construction of the conditional Markov chain based on one component of x . The case where the Markov chain is conditioned on two or more components of x is similar to the case where the Markov chain is conditioned on one component of x , but the conditioning scheme is more complicated. In this section we assume that x is a one-dimensional variable. Let $\mathbb{J} = \{-\infty = x^0 < x^1 < \dots < x^M = \infty\}$ be a partition on x and let \mathbb{J}_k denote the interval $(x^{k-1}, x^k]$. Then, the number of subintervals is M . This number, M , equals to the number of conditional Markov chains. Each conditional Markov chain consists of its own transition probability matrix. The detail for estimating the transition probability matrix is given in the next subsection. Similarly, since we assume that the number of state spaces of the Markov chain is finite, it also requires a finite partition of $g(x, y)$. Let N be the number of states of the Markov chain and let $\mathbb{I} = \{-\infty = g^0 < g^1 < \dots < g^N = \infty\}$ be the partition of $g(x, y)$ and let \mathbb{I}_i denote the subinterval $(g^{i-1}, g^i]$. The size of the transition probability matrix is, therefore, $N \times N$. Based on the number of conditional Markov chains and the number of states, it requires MN^2 coefficients to be estimated for the effective model. As we will see later, the typical values of M and N are in the order of 5 to 10 and the empirical result shows that the time series with $10^4 - 10^5$ sample points provides a good estimate for the effective model. The partitions \mathbb{I} and \mathbb{J} are empirically determined. After we define these partitions, the transition probability matrices have to be estimated. The

next subsection gives the detail on the estimation of transition probability matrices.

Note that, in the case of vector-valued variable, x , the conditioning scheme of the Markov chain is more complicated. We give an example for a two-dimensional case, i.e. $x = (x_1, x_2)$. Assume that the Markov chains in the effective model depend on both x_1 and x_2 . Also, assume that the partitions of x_1 and x_2 have M_1 and M_2 subintervals, respectively. Then, the number of conditional Markov chains in this case is $M_1 M_2$. The number of required coefficients is, therefore, $M_1 M_2 N^2$. It requires more sample points in order to have an accurate estimation of the Markov chain. This general idea can be applied to the conditioning scheme on two or more essential variables. However, it may not be possible in practical situations since it requires more data as the dimension of essential variables increases.

2.2.1 The Estimation of Transition Probability Matrices

The method that we use to estimate the transition probability matrices is a simple counting method. The method counts the number of state changes and normalizes this number by the total sum of state changes to obtain the transition probability. Let Q_k be the transition probability matrix for the k^{th} conditional Markov chain. This matrix is used to generate the value of the Markov chain in the effective model when the value of X is in the subinterval $(x^{k-1}, x^k]$. Let q_{kij} denote the element of Q_k at row i and column j . By the definition of the transition probability matrix, we write

$$q_{kij} = \Pr(g_{n+1} \in \mathbb{I}_j \mid g_n \in \mathbb{I}_i, x_n \in \mathbb{J}_k). \quad (2.4)$$

Equation (2.4) is the probability that the Markov chain changes from state i to state j conditioned on the essential variable, x , being in state k . Thus, to estimate this probability, we count the number of data points such that $g_n \in (g^{i-1}, g^i]$, $g_{n+1} \in$

$(g^{j-1}, g^j]$, and $x_n \in (x^{k-1}, x^k]$. Define the indicator function as follows

$$\mathbf{1}_{kij}(g_{n+1}, g_n, x_n) = \begin{cases} 1, & \text{if } g_{n+1} \in (g^{j-1}, g^j] \text{ and } g_n \in (g^{i-1}, g^i] \text{ and} \\ & x_n \in (x^{k-1}, x^k] \\ 0, & \text{otherwise.} \end{cases} \quad (2.5)$$

The counting method that we described above can be expressed in mathematical notation as follows

$$\hat{q}_{kij} = \frac{\sum_{n=0}^{N_0-1} \mathbf{1}_{kij}(g_{n+1}, g_n, x_n)}{\sum_{j=1}^N [\sum_{n=0}^{N_0-1} \mathbf{1}_{kij}(g_{n+1}, g_n, x_n)]}. \quad (2.6)$$

Equation (2.6) can be viewed as the number of changes that g changes from \mathbb{I}_i to \mathbb{I}_j where x is in \mathbb{J}_k . Then, this number is normalized by the total number of changes from state \mathbb{I}_i to all of the states. Thus, the summation over each row of Q_k equals to 1 and Q_k satisfies the stochastic matrix property. Recall that there are two components for the Markov chain. One is the transition probability matrix and the other is the value of the state space of the Markov chain. The procedure to obtain transition probability matrix was already described in this subsection. The next subsection gives the detail on how to generate a state space value. The state space can be either discrete state space or continuous state space. However, based on the numerical experiment in section 2.3, both discrete state space method and continuous state space method provide similar statistical results for the effective model. We will comment on the state space later in section 2.3.

2.2.2 The State Space of the Markov Chain

We provide two approaches for the state space of the Markov chain: the discrete state space and the continuous state space. The comparison result between these two approaches is presented in section 2.3. The details of these two approaches are described as follows.

Discrete State Space

In the effective model, the Markov chain is constructed from the time series of $g(x, y)$ obtained from the full model. The state space of $g(x, y)$ from the full model is continuous. For the discrete state space, the value of the Markov chain in each state is simply a single value. This value must be chosen such that it approximates $g(x, y)$ over the particular subinterval. Recall that the partition $\mathbb{I} = \{\mathbb{I}_i = (g^{i-1}, g^i]\}$, the value of the state space for state i can be chosen between g^{i-1} and g^i . In the discrete state space scheme, we adopt the average value of $g(x, y)$ with respect to the density of $g(x, y)$ over the subinterval $(g^{i-1}, g^i]$ as the value of the Markov chain at state i .

Let $m(Z; X)|_{ki}$ denote the value of the Markov chain at state i with X being in state k , then $m(Z; X)|_{ki}$ can be computed from the time series of $g(x, y)$ as follows

$$m(Z; X)|_{ki} = \frac{\sum_{n=0}^{N_0-1} g_n \mathbf{1}_{ki}(g_n, x_n)}{\sum_{n=0}^{N_0-1} \mathbf{1}_{ki}(g_n, x_n)}, \quad (2.7)$$

where

$$\mathbf{1}_{ki}(g_n, x_n) = \begin{cases} 1, & \text{if } g_n \in (g^{i-1}, g^i] \text{ and } x_n \in (x^{k-1}, x^k] \\ 0, & \text{otherwise.} \end{cases} \quad (2.8)$$

Equation (2.7) is simply the average value of $g(x, y)$ over the subinterval $(g^{i-1}, g^i]$ when X is in state k . Note that it is possible to choose the value of the Markov chain by other methods. For example, one can choose the mid-point of each subinterval, \mathbb{I}_i , as a state space value of the Markov chain.

Continuous State Space

In this scheme, we assume that the value of the Markov chain at state i can be any number in the subinterval $(g^{i-1}, g^i]$. We generate the value of the Markov chain by using a uniform distribution over $(g^{i-1}, g^i]$. If the state space of the Markov chain involves $\pm\infty$, we adopt the exponential distribution over that subinterval for generating the value of the Markov chain. The parameter of the exponential distribution

is obtained from the standard maximum likelihood estimator.

2.3 Prototype Models

We consider the stochastic triad system coupled with non-essential variables. The triad model can be viewed as the simple model of the wave interaction in the fluid dynamics. In this section, we apply the Markov chain modeling approach to the two different triad systems. For the first triad model, non-essential variables are independent of essential variables whereas, in the second model, non-essential variable depends on one of the essential variables. We illustrate some aspects that impact the statistical results of the effective models. These aspects include the time step used in the effective models, the number of states of the Markov chain, the number of conditional Markov chains, and the type of the state space of the Markov chain.

2.3.1 Triad Model Coupled with 2 Independent Non-essential Variables

We start with the simplest model of the triad system. Since non-essential variables are independent of essential variables, the Markov chain does not have to be conditioned on essential variables. Thus, it simplifies the Markov chain model.

Before we step into the detail of the Markov chain modeling approach. Let us summarize the major benefits of the Markov chain modeling approach as follows

- The number of the state space is typically in the order of 10. The small number of the state space requires small sample data for the estimation of the Markov chain. Thus, the effective models can be constructed from a short simulation of the full model.
- The Markov chain modeling approach is robust to the type of the state space.

We show that both discrete state space method and continuous state space method provide a good statistical behaviour for the effective model.

- The time step in the numerical integration of the effective model is generally greater than the time step used in the full model. Thus, the computational time in the effective model is generally faster than the computational time in the full model.
- The Markov chain modeling approach provides a good result on both strong scale separation and weak scale separation. Unlike the asymptotic approach where the method can not provide a good result under the weak scale separation.

For a stochastic triad system, let us consider the following system

$$\begin{aligned}
dx_1 &= A_1 x_2 x_3 dt + \alpha y_1 y_2 dt - b_1 x_1^3 dt \\
dx_2 &= A_2 x_1 x_3 dt - b_2 x_2 dt + \sigma_2 dB_2 \\
dx_3 &= A_3 x_1 x_2 dt - b_3 x_3 dt + \sigma_3 dB_3 \\
dy_1 &= -\gamma y_1 dt + \sigma dW_1 \\
dy_2 &= -\gamma y_2 dt + \sigma dW_2 \\
0 &= A_1 + A_2 + A_3.
\end{aligned} \tag{2.9}$$

$W_1, W_2, B_2,$ and B_3 are independent Brownian motions. $A_1, A_2,$ and A_3 are known coefficients. The parameters $b_1, b_2, b_3, \sigma_1,$ and σ_2 control the stationarity of the system. The essential variables consist of $x_1, x_2,$ and x_3 . These essential variables are coupled with the non-essential variables, y_1 and y_2 . y_1 and y_2 are the Ornstein-Uhlenbeck processes which are independent of essential variables. The time scale separation between essential variables and non-essential variables is determined by γ . A large value of γ gives a strong scale separation. We simulate the effective model both under the strong scale separation and weak scale separation. From (2.9), x_1 interacts directly with y_1 and y_2 . Thus, we have $g(x, y) = y_1 y_2$. In the effective model, $y_1 y_2$

is replaced by the Markov chain. Since non-essential variables do not depend on the essential variables, the Markov chain modeling approach has only one Markov chain. The Markov chain effective model can be written as follows

$$\begin{aligned}
dX_1 &= A_1 X_2 X_3 dt + \alpha m(Z) dt - b_1 X_1^3 dt \\
dX_2 &= A_2 X_1 X_3 dt - b_2 X_2 dt + \sigma_2 dB_2 \\
dX_3 &= A_3 X_1 X_2 dt - b_3 X_3 dt + \sigma_3 dB_3 \\
0 &= A_1 + A_2 + A_3.
\end{aligned} \tag{2.10}$$

Since we are interested in the average quantities of the essential variables, we compare the statistical properties of the Markov chain effective model to the statistical properties of the full model. We measure mean and variance of the essential variables. We also measure the auto-correlation function (ACF) and the probability density function (PDF) of the essential variables. The auto-correlation function of a zero mean stationary process, $X(t)$, is defined as

$$\text{ACF}_X(\tau) = \mathbb{E}[X(t)X(t + \tau)], \quad \forall t \geq 0.$$

We also compare the Markov chain effective model with the reduced model derived from the asymptotic approach. The asymptotic reduced model is briefly described as follows. To derive the reduced model, ϵ is introduced in the full model as follows

$$\begin{aligned}
dx_1 &= A_1 x_2 x_3 dt + \frac{1}{\epsilon} \alpha y_1 y_2 dt - b_1 x_1^3 dt \\
dx_2 &= A_2 x_1 x_3 dt - b_2 x_2 dt + \sigma_2 dB_2 \\
dx_3 &= A_3 x_1 x_2 dt - b_3 x_3 dt + \sigma_3 dB_3 \\
dy_1 &= -\frac{1}{\epsilon^2} \gamma y_1 dt + \frac{1}{\epsilon} \sigma dW_1 \\
dy_2 &= -\frac{1}{\epsilon^2} \gamma y_2 dt + \frac{1}{\epsilon} \sigma dW_2 \\
0 &= A_1 + A_2 + A_3.
\end{aligned} \tag{2.11}$$

We consider the limiting case where $\epsilon \rightarrow 0$. This corresponds to the infinite scale separation between x_1, x_2, x_3 and y_1, y_2 . The reduced model can be derived by

using the homogenization method provided in the appendix A. Thus, the asymptotic reduced model can be written as follows

$$\begin{aligned}
dx_1 &= A_1 x_2 x_3 dt - b_1 x_1^3 dt + \frac{\alpha \sigma^2}{2\gamma^{\frac{3}{2}}} dB_1 \\
dx_2 &= A_2 x_1 x_3 dt - b_2 x_2 dt + \sigma_2 dB_2 \\
dx_3 &= A_3 x_1 x_2 dt - b_3 x_3 dt + \sigma_3 dB_3 \\
0 &= A_1 + A_2 + A_3.
\end{aligned} \tag{2.12}$$

Note that, in the asymptotic reduced model, y_1 and y_2 are approximated by Brownian motion with a specific coefficient.

Numerical Settings

In this simulation, the parameters of (2.9) are as set follows

$$\begin{aligned}
A_1 &= -1.0, & A_2 &= 0.4, & A_3 &= 0.6, \\
\alpha &= 1.0, & b_1 &= 1.0, & b_2 &= 0.4, & b_3 &= 0.5, \\
\sigma_2 &= 0.6, & \sigma_3 &= 0.7, \\
\gamma &= 50.0, & \sigma &= 10\sqrt{2}.
\end{aligned} \tag{2.13}$$

The value of γ is relatively large. Thus, this case corresponds to the strong scale separation between essential variables and non-essential variables. Figure (2.1) shows the auto-correlation functions of essential and non-essential variables. The auto-correlation functions of non-essential variables decrease much faster than the auto-correlation functions of essential variables. In the simulation, we use the Euler-Maruyama method for the numerical integration [21]. The time step used in the numerical integration for the full model and the asymptotic reduced model is 10^{-4} . We sample data at the time step of 10^{-3} for the period of 200 ($T_0 = 200$). Thus, the number of sample points is $200/10^{-3} = 200,000$.

For the Markov chain effective model, the number of states is 11. We measure the

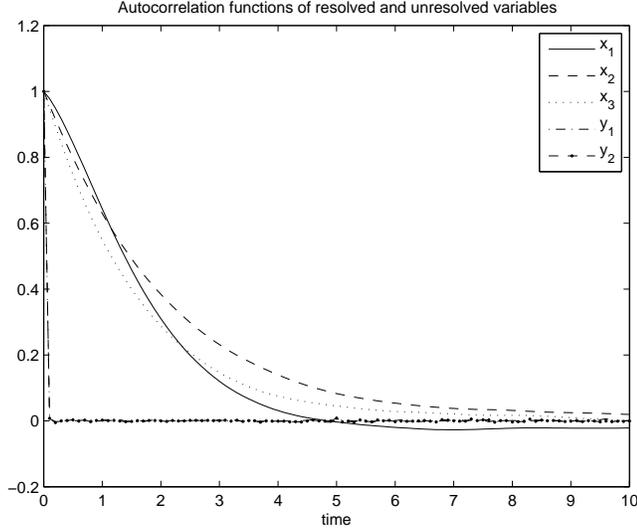


Figure 2.1: Auto-correlation functions of essential variables, x_1, x_2, x_3 , and non-essential variables, y_1, y_2 , from a full model in equation (2.9).

mean and standard deviation of the sample data, $g_n = y_1(n\Delta t)y_2(n\Delta t)$, and construct the partition for the state of the Markov chain as follows

$$\begin{aligned} \mathbb{I} = \{ & -\infty < \mu_g - 3.0\sigma_g < \mu_g - 2.0\sigma_g < \\ & \mu_g - 1.2\sigma_g < \mu_g - 0.6\sigma_g < \mu_g - 0.1\sigma_g < \\ & \mu_g + 0.1\sigma_g < \mu_g + 0.6\sigma_g < \mu_g + 1.2\sigma_g < \\ & \mu_g + 2.0\sigma_g < \mu_g + 3.0\sigma_g < \infty \} \end{aligned} \quad (2.14)$$

where $\mu_g = -0.03$ and $\sigma_g = 2.03$ are obtained from the time series of g_n . Then, the effective model is integrated with the time step of 0.001 and we choose the discrete state space scheme for the Markov chain.

We compute statistical behaviour of three systems: full model, asymptotic reduced model, and Markov chain effective model. We integrate numerically all of these three systems for $T = 20,000$ and compute the statistical properties of all three systems. Table (2.1), Figure (2.2) and (2.3) show the comparison of the statistical results from

these three models. The auto-correlation functions of x_1, x_2 , and x_3 from the Markov chain effective model have a good agreement with the auto-correlation functions from the full model and the asymptotic reduced model. The relative errors of the variances of the essential variables are relatively small (less than 5%). This result shows that the Markov chain effective model is a good representative for the full model. However, as we mentioned earlier, there are many factors that impact on the results of the Markov chain effective model. We will illustrate all of these impacts in later subsections.

	Full Model	Asymptotic	Error(%)	Markov Chain	Error(%)
μ_{x_1}	4.778×10^{-3}	-8.435×10^{-3}		-3.855×10^{-2}	
μ_{x_2}	1.011×10^{-2}	1.537×10^{-2}		1.093×10^{-2}	
μ_{x_3}	-1.598×10^{-3}	1.139×10^{-2}		1.150×10^{-2}	
$\sigma_{x_1}^2$	1.905×10^{-1}	1.908×10^{-1}	0.2	1.814×10^{-1}	4.8
$\sigma_{x_2}^2$	4.022×10^{-1}	4.183×10^{-1}	4.0	4.080×10^{-1}	1.4
$\sigma_{x_3}^2$	4.489×10^{-1}	4.346×10^{-1}	3.2	4.468×10^{-1}	0.5

Table 2.1: Mean and variance of x_1, x_2 , and x_3 in the simulation of the full model defined in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The parameters of the stochastic triad system are defined in (2.13)

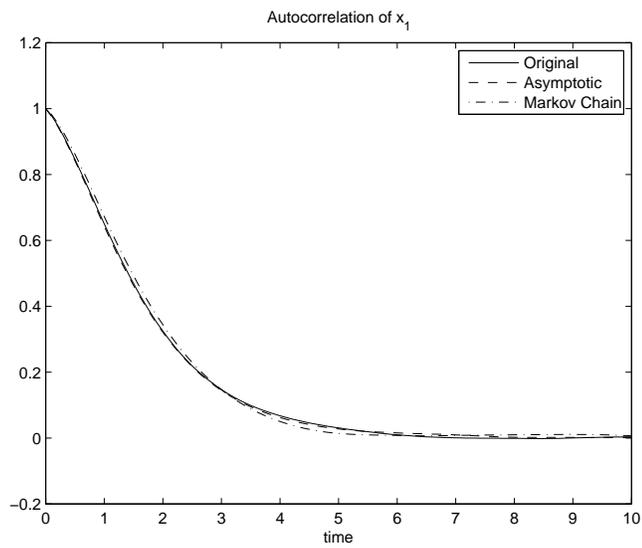


Figure 2.2: Auto-correlation function of x_1 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The Markov chain effective model has 11 states for the Markov chain. The partition for the state is defined in (2.14).

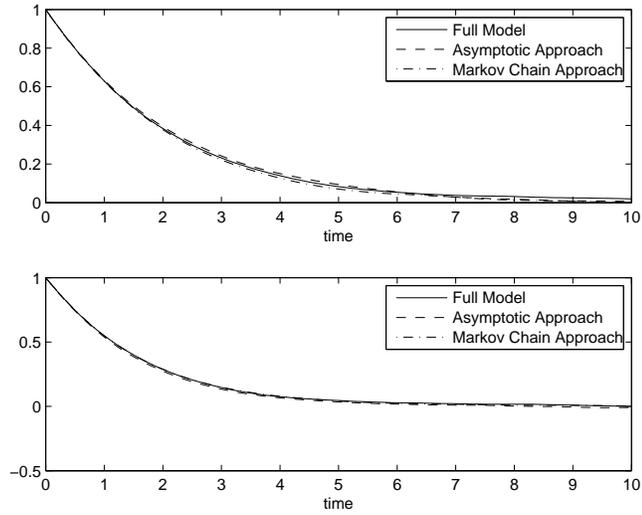


Figure 2.3: Auto-correlation functions of x_2 and x_3 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The Markov chain effective model has 11 states for the Markov chain. The partition for the state is defined in (2.14).

Note that non-essential variables, y_1 and y_2 , interact directly with x_1 only. Thus, the statistical results of x_2 and x_3 are impacted less by these non-essential variables. Thus, for the rest of this section, we concentrate on the statistical behaviours of x_1 .

Discrete State Space V.S. Continuous State Space

In section (2.2.2), we provide two approaches for the state space of the Markov chain; discrete state space and continuous state space. In this section we perform two simulations for the effective model; one with the discrete state space and the other with the continuous state space. Table (2.2) and Figure (2.4) show statistical values of the effective model with two different approaches for the state space. There is no significant difference between discrete state space and continuous state space. The effective model from the Markov chain modeling approach has a robustness over the state space scheme. However, the implementation of the discrete state space Markov chain is easier and less complex than the continuous state space Markov chain, we will use the discrete state space Markov chain for the rest of this section.

	Full Model	Continuous State Space	Error(%)	Discrete State Space	Error(%)
μ_{x_1}	4.778×10^{-3}	-3.385×10^{-2}		-3.855×10^{-2}	
μ_{x_2}	1.011×10^{-2}	1.040×10^{-2}		1.093×10^{-2}	
μ_{x_3}	-1.598×10^{-3}	1.259×10^{-2}		1.150×10^{-2}	
$\sigma_{x_1}^2$	1.905×10^{-1}	1.831×10^{-1}	3.9	1.814×10^{-1}	4.8
$\sigma_{x_2}^2$	4.022×10^{-1}	4.101×10^{-1}	2.0	4.080×10^{-1}	1.4
$\sigma_{x_3}^2$	4.489×10^{-1}	4.475×10^{-1}	0.3	4.468×10^{-1}	0.5

Table 2.2: Mean and variance of x_1, x_2 , and x_3 in the simulation of the full model in (2.9), and the Markov chain effective model in (2.10). The Markov chain effective model has 11 states. The state spaces for the effective model are both discrete and continuous.

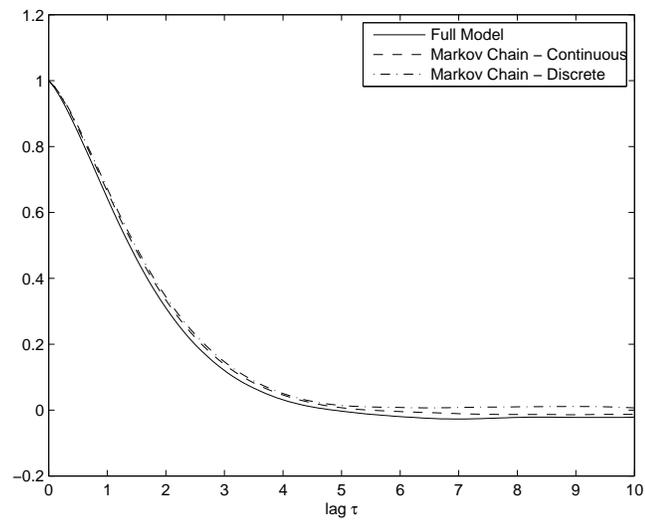


Figure 2.4: Comparison of the auto-correlation function of x_1 between discrete state space and continuous state space. The number of states is 11.

Number of States of the Markov Chain

We study the property of the effective model with different numbers of states of the Markov chain. It is expected that the Markov chain effective model with a large number of states will provide a good agreement with the full model. In this section, the numbers of states of the Markov chain are chosen to be $N = 5, 7, 11,$ and 15 . The partition, \mathbb{I} , for each N will be as follows.

For $N = 5$, the partition is

$$\mathbb{I} = \{-\infty < \mu_g - 2.1\sigma_g < \mu_g - 0.5\sigma_g < \mu_g + 0.5\sigma_g < \mu_g + 2.1\sigma_g < \infty\}.$$

For $N = 7$, the partition is

$$\mathbb{I} = \{-\infty < \mu_g - 2.1\sigma_g < \mu_g - 1.0\sigma_g < \mu_g - 0.3\sigma_g < \mu_g + 1.0\sigma_g < \mu_g + 2.1\sigma_g < \mu_g + 0.3\sigma_g < \infty\}.$$

For $N = 11$, the partition is

$$\begin{aligned} \mathbb{I} = \{ &-\infty < \mu_g - 3.0\sigma_g < \mu_g - 2.0\sigma_g < \\ &\mu_g - 1.2\sigma_g < \mu_g - 0.6\sigma_g < \mu_g - 0.1\sigma_g < \\ &\mu_g + 0.1\sigma_g < \mu_g + 0.6\sigma_g < \mu_g + 1.2\sigma_g < \\ &\mu_g + 2.0\sigma_g < \mu_g + 3.0\sigma_g < \infty\} \end{aligned} \quad (2.15)$$

Lastly, for $N=15$, the partition is

$$\begin{aligned} \mathbb{I} = \{ &-\infty < \mu_g - 3.0\sigma_g < \mu_g - 2.3\sigma_g < \mu_g - 1.7\sigma_g < \\ &\mu_g - 1.2\sigma_g < \mu_g - 0.8\sigma_g < \mu_g - 0.4\sigma_g < \\ &\mu_g - 0.1\sigma_g < \mu_g + 0.1\sigma_g < \mu_g + 0.4\sigma_g < \\ &\mu_g + 0.8\sigma_g < \mu_g + 1.2\sigma_g < \mu_g + 1.7\sigma_g < \\ &\mu_g + 2.3\sigma_g < \mu_g + 3.0\sigma_g < \infty\}. \end{aligned}$$

Table (2.3) and Figure (2.5) show the statistical properties of the effective model with different numbers of states of the Markov chain. The auto-correlation functions of x_1 from the effective models with different N are similar to the auto-correlation function of x_1 from the full model. However, the variances of x_1 from the effective models are different from the variance of x_1 from the full model when N is small. From the table, as N increases the relative error of the variance of x_1 decreases. For $N \geq 11$, relative errors are less than 5%. The statistical results from the effective model with $N = 11$ and $N = 15$ provide a good agreement with the full model. However, a large value of N requires more sample data in order to accurately estimate the transition probability matrix. The required number of data is proportional to N^2 . In general, the value of N between 7 to 11 provides a good agreement with the full model.

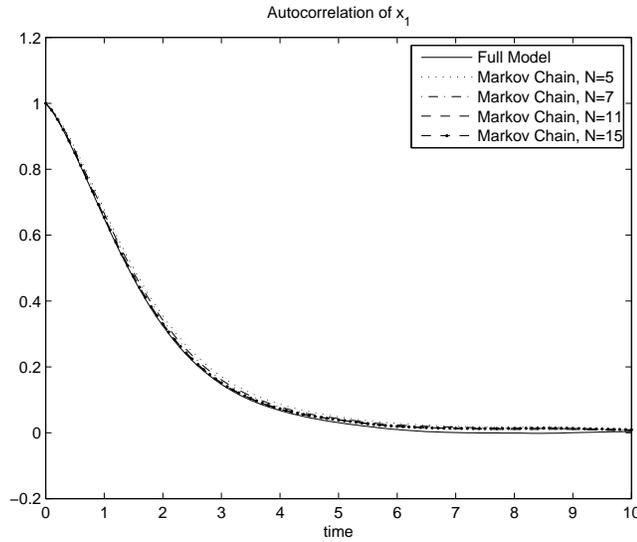


Figure 2.5: Auto-correlation functions of x_1 from the full model in (2.9) and the Markov chain effective model in (2.10) with different numbers of states of the Markov chain.

	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	$\sigma_{x_3}^2$
Full Model	1.905×10^{-1}	4.022×10^{-1}	4.489×10^{-1}
MC (N=5)	1.666×10^{-1}	4.055×10^{-1}	4.445×10^{-1}
Rel Error(%)	12.5	0.8	1.0
MC (N=7)	1.750×10^{-1}	4.070×10^{-1}	4.452×10^{-1}
Rel Error(%)	8.1	1.2	0.8
MC (N=11)	1.814×10^{-1}	4.080×10^{-1}	4.468×10^{-1}
Rel Error(%)	4.8	1.4	0.5
MC (N=15)	1.859×10^{-1}	4.088×10^{-1}	4.475×10^{-1}
Rel Error(%)	2.4	1.6	0.3

Table 2.3: Mean and variance of x_1 , x_2 , and x_3 from the full model in (2.9) and the Markov chain effective models in (2.10). The numbers of states in the Markov chain effective model are 5, 7, 11, and 15.

The Effect of the Time Step

In this section, we investigate the result of the Markov chain effective model when the time step is varied. The referred time step is the time step used in the numerical integration of the Markov chain effective model defined in (2.3). The time step used in the numerical integration is the same as the time step used in the estimation of the Markov chain. We simulate the Markov chain effective model with five different time steps: 0.0001, 0.001, 0.005, 0.01, and 0.1. Statistical results are shown in Table (2.4) and Figure (2.6).

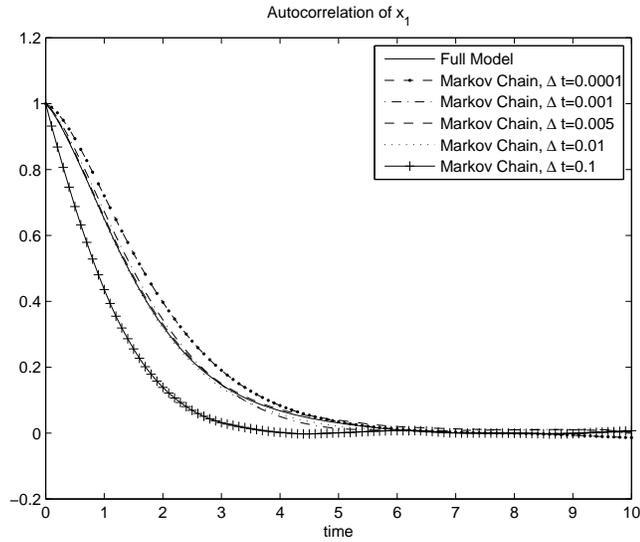


Figure 2.6: Auto-correlation function of x_1 from the full model in (2.9) and the Markov chain effective model in (2.10). The time steps used in the Markov chain effective model are 0.0001, 0.001, 0.005, 0.01, and 0.1. Note that the Markov chain effective models with $\Delta t = 0.0001$ and 0.1 do not properly reproduce the auto-correlation function of x_1 .

	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	$\sigma_{x_3}^2$
Full Model	1.905×10^{-1}	4.022×10^{-1}	4.489×10^{-1}
MC ($\Delta t = 10^{-4}$)	1.598×10^{-1}	3.959×10^{-1}	4.355×10^{-1}
Rel Error(%)	16.1	1.6	3.0
MC ($\Delta t = 10^{-3}$)	1.814×10^{-1}	4.080×10^{-1}	4.468×10^{-1}
Rel Error(%)	4.8	1.4	0.5
MC ($\Delta t = 5 \times 10^{-3}$)	1.890×10^{-1}	4.081×10^{-1}	4.466×10^{-1}
Rel Error(%)	0.8	1.5	0.5
MC ($\Delta t = 10^{-2}$)	1.927×10^{-1}	4.068×10^{-1}	4.504×10^{-1}
Rel Error(%)	1.2	1.1	0.3
MC ($\Delta t = 10^{-1}$)	3.224×10^{-1}	4.406×10^{-1}	4.741×10^{-1}
Rel Error(%)	69.2	9.5	5.6

Table 2.4: Means and variances of x_1 , x_2 , and x_3 from the full model in (2.9) and the Markov chain effective model in (2.10). The time steps used in the effective model are 0.0001, 0.001, 0.005, 0.01, and 0.1

From Table (2.4), the variances of the essential variables are significantly different from the full model when the time step is too small ($\Delta t = 10^{-4}$). When the time step is small, the transition probability matrix constructed from the time series is close to the identity matrix. The off-diagonal entries are close to zero and, thus, the accuracy of the estimation is lost with finite number of data. On the other hand, if the time step in the numerical integration is too large, there will be a problem on the stability and accuracy of the numerical integration. This problem can cause the large error in the effective model. From the empirical results, we expect that the optimal time step for the Markov chain effective model depends on the correlation time of $g(x, y)$ of the full model. In this case, the correlation time is $\frac{1}{2\gamma} = 0.01$. The proper value of the time step in this case is either 10^{-3} , 5×10^{-3} , or 10^{-2} . The numerical study shows that the optimal value of the time step is

$$\Delta t_{opt} \in [0.1\tau_c, 0.5\tau_c], \quad (2.16)$$

where τ_c is the correlation time of $g(x, y)$. We will mention about the time step again in the next subsection.

Weak Scale Separation

One major benefit of the effective model using the Markov chain modeling approach is that the Markov chain effective model can reproduce the statistical behaviour of the full model in the case of the weak scale separation where as the asymptotic reduced model can not provide a good result under the weak scale separation. In this section, we consider values of γ that correspond to the weak scale separation. The parameters

are set as follows

$$\begin{aligned}
 A_1 &= -1.0, & A_2 &= 0.4, & A_3 &= 0.6, \\
 \alpha &= 1.0, & b_1 &= 1.0, & b_2 &= 0.4, & b_3 &= 0.5, \\
 \sigma_2 &= 0.6, & \sigma_3 &= 0.7, \\
 \gamma &= 2.5, & \sigma &= \sqrt{10}.
 \end{aligned}
 \tag{2.17}$$

We set $\sigma = \sqrt{10}$ to maintain the same energy level of non-essential variables as in the case of the strong scale separation, i.e., $\text{Var}\{y_1\} = \text{Var}\{y_2\} = 2$ in both strong and weak scale separations. The auto-correlation functions of essential variables and non-essential variables are shown in Figure (2.7). The correlation times of y_1 and y_2 are approximately 0.38 and the correlation time of x_1 is 0.91. Thus, in this case, the correlation times of essential and non-essential variables are in the same order. For

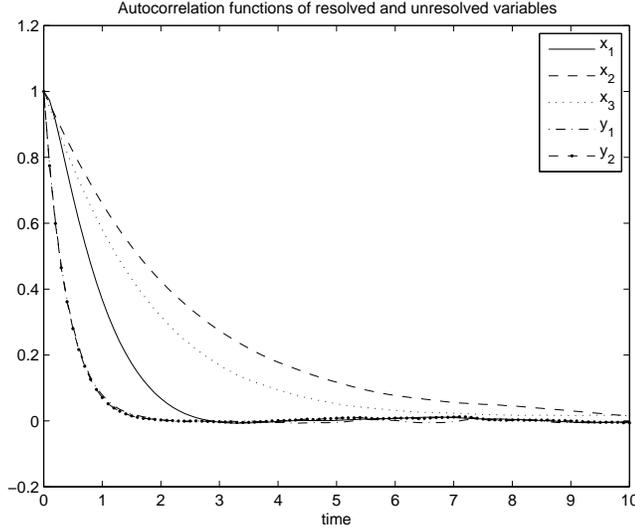


Figure 2.7: Auto-correlation functions of x_1, x_2, x_3, y_1 , and y_2 from the full model in (2.9) with the parameters in (2.17).

the Markov chain effective model, the Markov chain has 11 states. The partition is defined in (2.14). We simulate the effective model with two time step, $\Delta t = 0.01$ and

$\Delta t = 0.1$. The correlation time of $g(x, y)$ in this case is $\frac{1}{2 \times 2.5} = 0.2$. Therefore, the proper Δt should be in the range of 0.02 to 0.1. Table (2.5) and Figure (2.8) show the statistical results of the Markov chain effective model with two time steps, $\Delta t = 0.01$ and 0.1.

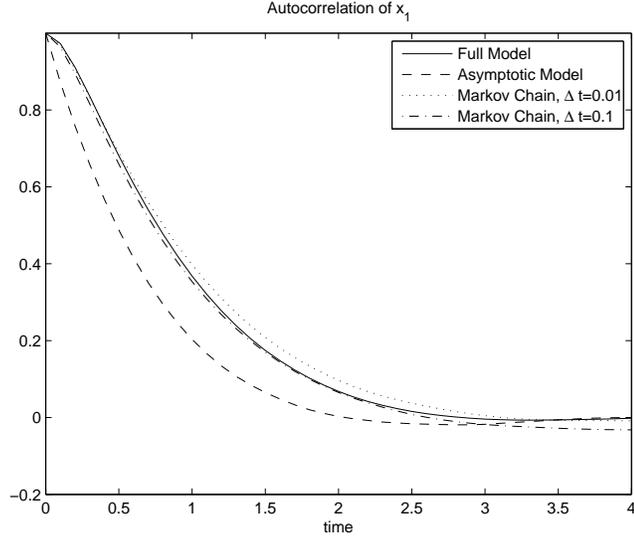


Figure 2.8: Auto-correlation function of x_1 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The time steps for the Markov chain effective model are 0.01 and 0.1.

Note that the asymptotic reduced model does not produce a good agreement with the full model. The variance of x_1 is significantly different from the full model and the auto-correlation function of x_1 is far from the auto-correlation of x_1 from the full model. This is caused by the weak scale separation between x_1, x_2, x_3 and y_1, y_2 in the original model. However, the Markov chain effective model has a good agreement with the full model when the time step is 0.1. Thus, with the appropriate time step, the Markov chain modeling approach can be a good effective model for the case of weak scale separation.

We would like to mention that in many systems, for example, the TBH system

	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	$\sigma_{x_3}^2$
Full Model	4.859×10^{-1}	4.713×10^{-1}	5.067×10^{-1}
Asymptotic Approach	5.953×10^{-1}	4.795×10^{-1}	5.112×10^{-1}
Rel Error(%)	22.5	1.7	0.9
MC ($\Delta t = 10^{-2}$)	4.198×10^{-1}	4.409×10^{-1}	4.988×10^{-1}
Rel Error(%)	13.6	6.5	1.6
MC ($\Delta t = 10^{-1}$)	4.852×10^{-1}	4.603×10^{-1}	5.169×10^{-1}
Rel Error(%)	0.1	2.3	2.0

Table 2.5: Mean and variance of x_1, x_2 , and x_3 from the full model in (2.9), the asymptotic reduced model in (2.12), and the Markov chain effective model in (2.10). The time steps for the Markov chain effective model are 0.01 and 0.1.

described in the next section, the scale separation between the essential variables and the non-essential variables can not be clearly defined. Thus, the asymptotic reduced model may encounter the problem of the weak scale separation. However, the Markov chain modeling approach resolves the problem and provides a good statistical result in this stochastic triad case. Therefore, we expect that the Markov chain modeling approach can perform well in models without strong scale separation.

2.3.2 Triad Model Coupled with Dependent Non-essential Variable

In section 2.3.1, the non-essential variables are independent of the essential variables. Thus, the Markov chain in the effective model does not need to be conditioned on the essential variables. In this section, we consider the effective model for the triad where the non-essential variables depend on the essential variable. The prototype equations are described as follows

$$\begin{aligned}
 dx_1 &= A_1 x_2 x_3 dt + \alpha y dt - b_1 x_1^3 dt \\
 dx_2 &= A_2 x_1 x_3 dt - b_2 x_2 dt + \sigma_2 dB_2 \\
 dx_3 &= A_3 x_1 x_2 dt - b_3 x_3 dt + \sigma_3 dB_3 \\
 dy &= -\gamma_1 (y - x_1) dt + \sigma_1 dW_1 \\
 0 &= A_1 + A_2 + A_3,
 \end{aligned} \tag{2.18}$$

where B_2, B_3 , and W_1 are independent Brownian motion processes. We add the damping and forcing terms to the right-hand side of x_2 and x_3 in order to have a stationary process. The essential variables are x_1, x_2 , and x_3 and the non-essential variable is y . Note that y is coupled with x_1 . From the previous section, the factors that impact the behaviour of the effective model are the number of states of the Markov chain and the time step used in both estimation of the transition probability matrix and numerical integration of the effective model. In this section, we also consider how conditioning of the Markov chain affects the performance of the effective model. In particular, we investigate several effective models with different numbers of conditional Markov chains. The statistical results from effective models are compared with the results from the full model and the asymptotic reduced model. The asymptotic

reduced model is derived from the full model by introducing ϵ as follows

$$\begin{aligned}
dx_1 &= A_1 x_2 x_3 dt + \alpha y dt - b_1 x_1^3 dt \\
dx_2 &= A_2 x_1 x_3 dt - b_2 x_2 dt + \sigma_2 dB_2 \\
dx_3 &= A_3 x_1 x_2 dt - b_3 x_3 dt + \sigma_3 dB_3 \\
dy &= -\frac{1}{\epsilon} \gamma_1 (y - x_1) dt + \frac{1}{\sqrt{\epsilon}} \sigma_1 dW_1 \\
0 &= A_1 + A_2 + A_3.
\end{aligned} \tag{2.19}$$

Using the averaging method in Appendix A, the asymptotic reduced model can be written as follows

$$\begin{aligned}
dx_1 &= A_1 x_2 x_3 dt + \alpha x_1 dt - b_1 x_1^3 dt \\
dx_2 &= A_2 x_1 x_3 dt - b_2 x_2 dt + \sigma_2 dB_2 \\
dx_3 &= A_3 x_1 x_2 dt - b_3 x_3 dt + \sigma_3 dB_3 \\
0 &= A_1 + A_2 + A_3.
\end{aligned} \tag{2.20}$$

Note that the asymptotic reduced model works well under the condition of strong scale separation. We will demonstrate the results of the effective model under both strong scale and weak scale separation.

Before we go into details of the Markov chain modeling approach, let us summarize its benefits.

- The Markov chain approach for this case has the same properties as in the previous case. The number of states of the Markov chain is in the order of 10 states. The time step of the effective model also depends on the correlation time of $g(x, y)$ and the state space of the Markov chain can be either discrete or continuous. The statistical results from these two state space schemes are very similar.
- The effective model with moderate numbers of conditional Markov chains provides a good statistical result. In the simulation, it is shown that the effective

model with four conditional Markov chains reproduces the statistical behaviour of essential variables well.

Numerical Settings

In this section, we begin with the strong scale separation case. The parameters are set as follows

$$\begin{aligned}
 A_1 &= -1.0, & A_2 &= 0.4, & A_3 &= 0.6, \\
 \alpha &= 1.0, & b_1 &= 1.0, & b_2 &= 0.4, & b_3 &= 0.5, \\
 \sigma_2 &= 0.6, & \sigma_3 &= 0.7, \\
 \gamma_1 &= 100.0, & \sigma_1 &= 50.0.
 \end{aligned}
 \tag{2.21}$$

We illustrate three effective models having different partitions of x_1 . The first effective model is not conditioned on x_1 . The second model has a partition on x_1 with two subintervals and the last model has four subintervals on x_1 . The partition of x_1 with two subintervals is written as follows

$$\mathbb{J}_{x_1} = \{-\infty < \mu_{x_1} < \infty\}
 \tag{2.22}$$

and the partition of x_1 with four subintervals is

$$\mathbb{J}_{x_1} = \{-\infty < \mu_{x_1} - \sigma_{x_1} < \mu_{x_1} < \mu_{x_1} + \sigma_{x_1} < \infty\}.
 \tag{2.23}$$

The values of μ_{x_1} and σ_{x_1} are calculated from the time series of x_1 from the full model. The time length for the sample data is equal to $T_0 = 200.0$. We sample the data with the time step, $\Delta t = 0.001$. Thus, the time series of x_1 has 200,000 sample points. Note that we use this time step in both estimation of transition probability matrices and numerical integration of the effective model. In this simulation, we have $\mu_{x_1} = 0.124$ and $\sigma_{x_1} = 0.832$. For each conditional Markov chain, the number of states is set to 11 states. The partition for each conditional Markov chain consists of

11 subintervals and can be written as follows

$$\begin{aligned} \mathbb{I} = & \{-\infty < \mu_g - 2.7\sigma_g < \mu_g - 1.9\sigma_g < \\ & \mu_g - 1.2\sigma_g < \mu_g - 0.6\sigma_g < \mu_g - 0.1\sigma_g < \\ & \mu_g + 0.1\sigma_g < \mu_g + 0.6\sigma_g < \mu_g + 1.2\sigma_g < \\ & \mu_g + 1.9\sigma_g < \mu_g + 2.7\sigma_g < \infty\}. \end{aligned}$$

Table (2.6) and Figure (2.9) show the comparison of the statistical properties of the full model, the Markov chain effective model, and the asymptotic reduced model. Note that even in the case of strong scale separation, the asymptotic approach using the averaging method does not have a good agreement with the full model on the auto-correlation function of x_1 .

	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	$\sigma_{x_3}^2$
Full Model	6.868×10^{-1}	5.012×10^{-1}	5.587×10^{-1}
Asymptotic	7.065×10^{-1}	5.215×10^{-1}	5.799×10^{-1}
Relative Error(%)	2.8	4.2	3.8
MC (1 subinterval)	2.407×10^{-1}	4.350×10^{-1}	4.573×10^{-1}
Relative Error(%)	64.9	13.2	18.2
MC (2 subintervals)	6.254×10^{-1}	5.206×10^{-1}	5.559×10^{-1}
Relative Error(%)	9.0	4.0	0.7
MC (4 subintervals)	6.424×10^{-1}	5.055×10^{-1}	5.415×10^{-1}
Relative Error(%)	6.5	0.2	3.0

Table 2.6: Mean and variance of essential variables from the full model defined in (2.18), the asymptotic reduced model defined in (2.20), and the Markov chain effective model with different conditioning schemes.

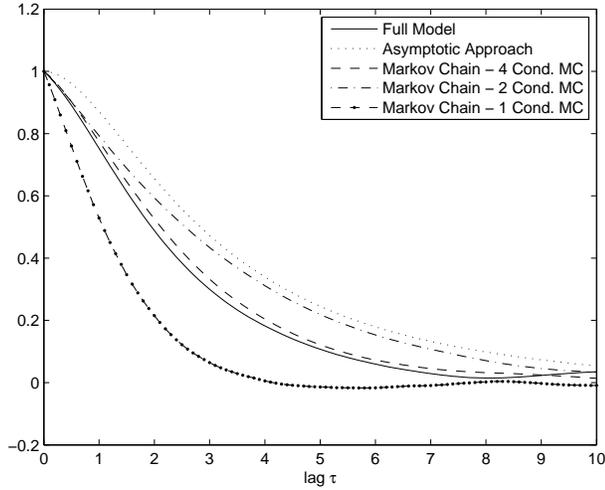


Figure 2.9: Auto-correlation function of x_1 from the full model, the asymptotic reduced model, and the Markov chain effective model with different conditioning schemes.

The effective model with one Markov chain cannot reproduce statistical properties of essential variables. The relative errors of variances of essential variables are very large. The auto-correlation function of x_1 does not resemble the auto-correlation of x_1 from the full model. With two conditional Markov chains, the variance of x_1 from the effective model improves significantly. However, there is a discrepancy between the auto-correlation function of x_1 from the Markov chain effective model and the full model. The effective model with four conditional Markov chains reproduces the statistical properties well. Both variance and auto-correlation function from the effective model agree with the ones from the full model. It is expected that the effective model with many conditional Markov chains is able to reproduce statistical properties of essential variables. However, this requires more data in order to estimate the transition probability matrices. From the empirical observation, four to eight conditional Markov chains should be sufficient to reproduce statistical behaviours of essential variables.

2.4 Truncated Burgers-Hopf Model

As we have seen in the previous section, the Markov chain effective model reproduces the statistical behaviour of the stochastic triad system under both strong and weak scale separations. In this section, we apply the Markov chain modeling approach to the Truncated Burgers-Hopf (TBH) system. It is known that the correlation times of the Fourier coefficients in TBH equation are approximately inversely proportional to the wave numbers [1, 2, 4]. Therefore, it is difficult to define a time scale separation between essential variables and non-essential variables. In this section, we apply the Markov chain modeling technique to derive effective models for one and two Fourier coefficients. The benefits of the Markov chain modeling approach can be summarized as follows

- The Markov chain approach is simple to implement. The number of states of the Markov chain is in the order of 10 states or less.
- In general, the Markov chains have to be conditioned on, possibly, all essential variables. However, based on the weak correlation among Fourier modes, each Markov chain can be conditioned on its corresponding Fourier coefficient. Thus, it simplifies the conditioning scheme of the Markov chain. Moreover, in the case of zero Hamiltonian, it is shown in [7] that the real and image parts of the Fourier coefficients are uncorrelated. Thus, the Markov chain can be conditioned only on the real or image part of that Fourier mode.
- For the non-zero Hamiltonian case, the Markov chain is conditioned on both real and image of the corresponding Fourier mode. However, the partition for the secondary variable requires a partition with a few subintervals. In particular, the partition with two subintervals for secondary variable significantly improves the statistical behaviour of the effective model. This result is shown in details later in this section.

- The time step used in the numerical integration of the effective model is generally larger than the time step used in the full model. Thus, the computational cost is reduced significantly.

In the next section, we describe some basic descriptions and properties of the TBH equation. Then, we apply the Markov chain approach for the effective model. We consider two effective models: the effective model with one Fourier mode and the effective model with two Fourier modes.

2.4.1 The TBH Equation

We consider the spectral projection of Burgers-Hopf equation with a periodic solution. In particular, the model can be written as follows

$$(U_\Lambda)_t + \frac{1}{2}P_\Lambda(U_\Lambda^2)_x = 0 \quad (2.24)$$

where U_Λ and $P_\Lambda U$ represent the finite Fourier projection of U defined as follows

$$P_\Lambda U(t, x) \equiv U_\Lambda(t, x) = \sum_{k=-\Lambda}^{\Lambda} u_k(t) e^{ikx}. \quad (2.25)$$

Λ is the degree of freedom for the TBH system. Since the TBH system has a periodic solution, we assume $U_\Lambda(t, 0) = U_\Lambda(t, 2\pi)$ without loss of generality. We consider the case of real solution of $U(t, x)$ such that $u_{-k}(t) = u_k(t)^*$ where u^* denotes the complex conjugate of u . Substituting (2.25) into (2.24), equation (2.24) can be recasted as a system of the ordinary differential equations with $|k| \leq \Lambda$ as follows

$$\frac{d}{dt}u_k(t) = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \Lambda}} u_p^* u_q^*. \quad (2.26)$$

The TBH system has some constant quantities. These quantities are momentum, energy and Hamiltonian. The momentum is defined as follows

$$M = \frac{1}{2\pi} \int_0^{2\pi} U_\Lambda dx = u_0, \quad (2.27)$$

the energy is defined as

$$E = \frac{1}{4\pi} \int_0^{2\pi} U_\Lambda^2 dx = \frac{1}{2} |u_0|^2 + \sum_{k=1}^{\Lambda} |u_k|^2, \quad (2.28)$$

and the Hamiltonian is

$$H = \frac{1}{12\pi} \int_0^{2\pi} P_\Lambda U_\Lambda^3 dx = \frac{1}{6} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \Lambda}} u_k u_p u_q. \quad (2.29)$$

From (2.26), the equation for u_0 is trivial and, thus, u_0 is constant. Without loss of generality, we assume $u_0 = 0$.

For a large Λ , the TBH system tends to have a stochastic behaviour. Statistical properties were studied extensively in [1, 2, 4]. It was shown that, with $H \approx 0$, the Fourier coefficients obtain an equipartition of energy and the joint probability density function of Fourier coefficients follow the normal distribution

$$\pi(u_1, u_2, \dots, u_\Lambda) = C e^{-\beta E}, \quad (2.30)$$

where C is the normalized coefficient. The energy per Fourier coefficient is β^{-1} .

One interesting statistical property of the TBH is the auto-correlation function. The correlation times of Fourier coefficients are inversely proportional to the Fourier numbers. Figure (2.10) shows the auto-correlation functions of the first five Fourier coefficients. Each Fourier coefficient represents the behaviour of the system at a different time scale. However, it is difficult to define a scale separation between slow and fast variables.

We are interested in the average behaviour of the system. The average behaviour is generally characterized by the low Fourier modes. In this section, we are interested in deriving the effective models for the first few Fourier modes. We describe the effective models for the TBH system with one and two essential variables in the next subsection.

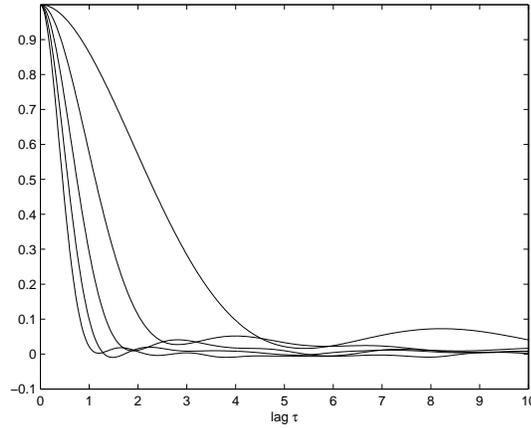


Figure 2.10: Auto-correlation functions of u_k , $k = 1, \dots, 5$. The TBH has $\Lambda = 20$, $E = 0.4$, and $H = 0$.

2.4.2 The Effective Model for TBH

Let λ be the number of the essential variables, i.e., u_1, \dots, u_λ are the essential variables and $u_{\lambda+1}, \dots, u_\Lambda$ are the non-essential variables. The right-hand side of the equation (2.26) can be rewritten as a summation of 2 terms as follows

$$\frac{d}{dt}u_k(t) = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \lambda}} u_p^* u_q^* - ik \left(\sum_{p=\lambda-k+1}^{\Lambda-k} u_p^* u_{p+k} \right). \quad (2.31)$$

The first term of the right-hand side consists of the interaction of the essential variables. The second term represents the interaction of non-essential variables. It involves two types of interactions - (i) essential variables and non-essential variables and (ii) non-essential variables with themselves. For the Markov chain effective model, we replace the second term by a collection of conditional Markov chains. The Markov chain is conditioned on the essential variables. The equation for the Markov chain effective model is written as follows

$$\frac{d}{dt}u_k = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \lambda}} u_p^* u_q^* + m_k. \quad (2.32)$$

Here, m_k represents the Markov chain corresponding to k^{th} essential variable. We consider the effective models with one essential variable ($\lambda = 1$) and two essential variables ($\lambda = 2$). The details are described as follows.

2.4.3 One Essential Variable, $\lambda = 1$

From the equation (2.32), the first term vanishes for $k = 1$. Thus, the right-hand side of u_1 consists of the Markov chain term. The dynamics of u_1 for the Markov chain effective model in the real form is written as follows

$$\frac{d}{dt}u_1^{re} = m_1^{re}(u_1^{re}, u_1^{im}) \quad (2.33a)$$

$$\frac{d}{dt}u_1^{im} = m_1^{im}(u_1^{re}, u_1^{im}), \quad (2.33b)$$

where $m_1^{re}(u_1^{re}, u_1^{im})$ represents the Markov chain conditioned on u_1^{re} and u_1^{im} . To construct m_1^{re} , the partitions of u_1^{re} and u_1^{im} need to be defined. Let M_1 be the number of subintervals for u_1^{re} and M_2 the number of subintervals for u_1^{im} . The number of conditional Markov chains for m_1^{re} is equal to M_1M_2 . We assume that each conditional Markov chain has N states. Thus, it requires $M_1M_2N^2$ coefficients to be estimated for the transition probability matrices. The estimation of the transition probability matrices was already described in section (2.2). The construction of m_1^{im} is similar to m_1^{re} . In this section, we apply the Markov chain effective model for the TBH system with zero Hamiltonian. Since there is a weak correlation between u_k^{re} and u_k^{im} [7], the Markov chain, m_1^{re} can be conditioned on u_1^{re} and, similarly, m_1^{im} can be conditioned on u_1^{im} , i.e.,

$$m_1^{re} = m_1^{re}(u_1^{re}), \quad m_1^{im} = m_1^{im}(u_1^{im}).$$

Thus, the required coefficients is reduced to M_1N^2 coefficients and the number of conditional Markov chains is M_1 .

Numerical Settings

For the full TBH system, it consists of 20 Fourier modes ($\Lambda = 20$). The total energy is 0.4 ($E = 0.4$). Thus, the energy per Fourier mode is $\frac{0.4}{20} = 0.02$ and the energy on the real part and image part for each u_k is 0.01. We use the Pseudo-Spectral method with fourth-order Runge-Kutta method to numerically integrate the full model. The time step is chosen to be 0.001. This time step is small enough such that the relative error of the total energy is less than 10^{-5} . We integrate the full model with length $T_0 = 20,000$ and sample the data with time step $\Delta t = 0.1$. Thus, the time series obtained from the full model has a length of 200,000 sample points. This time series is used to construct the Markov chain effective model. The Markov chain effective model is integrated with the same time step, $\Delta t = 0.1$. The Markov chain is conditioned on u_1^{re} (or u_1^{im}). The partition of the essential variable has 6 subintervals and is written as follows

$$\mathbb{J}_u = \{-\infty < \mu_u - 1.4\sigma_u < \mu_u - 0.6\sigma_u < \mu_u < \mu_u + 0.6\sigma_u < \mu_u + 1.4\sigma_u < \infty\} \quad (2.34)$$

where $u = u_1^{re}, u_1^{im}$. Each conditional Markov chain has 7 states and the partition for each conditional Markov chain is written as follows

$$\begin{aligned} \mathbb{I}_m = \{ &-\infty < \mu_m - 2.0\sigma_m < \mu_m - 1.0\sigma_m < \\ &\mu_m - 0.3\sigma_m < \mu_m + 0.3\sigma_m < \\ &\mu_m + 1.0\sigma_m < \mu_m + 2.0\sigma_m < \infty\}. \end{aligned} \quad (2.35)$$

We simulate the effective model with time step, $\Delta t = 0.1$, and integrate the model with a total time length of $T = 200,000$. We measure statistical properties such as mean, variance, kurtosis, ACF and PDF of u_1^{re} and u_1^{im} . Table (2.7) and Figure (2.11) show the comparison of statistical properties between the full model and the Markov chain effective model. The mean and variance of u_1 from the effective model are very close to the value from the theory. The kurtosis from the effective model is

	Theory	Full Model	Effective Model
Mean of u_1^{re}	0.0	-6.169×10^{-4}	1.616×10^{-4}
Mean of u_1^{im}	0.0	-7.013×10^{-4}	8.983×10^{-5}
Variance of u_1^{re}	0.01	9.854×10^{-3}	1.019×10^{-2}
Variance of u_1^{im}	0.01	9.845×10^{-3}	1.022×10^{-2}
Kurtosis of u_1^{re}	3.0	2.885	3.425
Kurtosis of u_1^{im}	3.0	2.867	3.386

Table 2.7: Mean, variance, and kurtosis of u_1 from the full model and the effective model.

higher than the theoretical value. However, the PDF of u_1 from the effective model is similar to the one from the full TBH model. The ACF of u_1 from the effective model decorrelates at the same rate as the ACF of u_1 from the full model. However, there is a slight discrepancy for time lag $5 \leq \tau \leq 10$. To compare the performance of the Markov chain effective model with another method, we consider the stochastic mode reduction method used in [4]. The Markov chain effective model provides a better result for the auto-correlation function compared with the auto-correlation function obtained from [4]. (See [4] on page 787.) The ACF in [4] decays exponentially whereas the ACF from the full model does not. This is because the derivation of the reduced model in [4] is based on the asymptotic approach where the infinite scale separation is assumed. However, as we mentioned earlier, the TBH system does not have an infinite scale separation. For this numerical results, the Markov chain approach performs well under the weak scale separation.

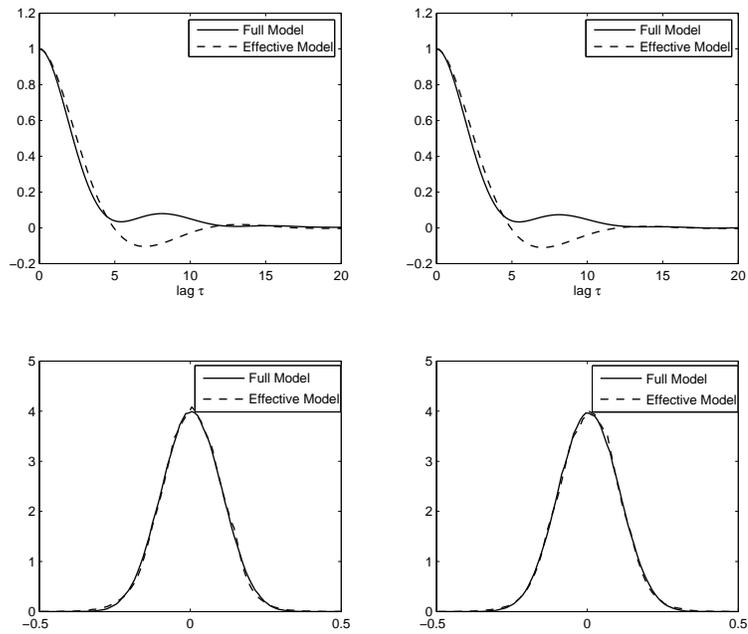


Figure 2.11: Statistical properties of u_1 . The top left figure is the ACF of u_1^{re} . The top right is the ACF of u_1^{im} . The bottom left is the PDF of u_1^{re} and the bottom right is the PDF of u_1^{im} .

2.4.4 Two Essential Variables, $\lambda = 2$

The effective model consists of essential variables, u_1 and u_2 ($u_1^{re}, u_1^{im}, u_2^{re}, u_2^{im}$). The dynamics of the essential variables is obtained from (2.32) by setting $\lambda = 2$. The Markov chain effective model for TBH with $\lambda = 2$ is written as follows

$$\begin{aligned}\frac{d}{dt}u_1^{re} &= (u_1^{re}u_2^{im} - u_1^{im}u_2^{re}) + m_1^{re} \\ \frac{d}{dt}u_1^{im} &= (-u_1^{re}u_2^{re} - u_1^{im}u_2^{im}) + m_1^{im} \\ \frac{d}{dt}u_2^{re} &= 2u_1^{re}u_1^{im} + m_2^{re} \\ \frac{d}{dt}u_2^{im} &= (u_1^{im})^2 - (u_1^{re})^2 + m_2^{im}.\end{aligned}$$

All conditional Markov chains are assumed to be conditioned on the essential variables. However, from the empirical data, the right-hand side of u_1 has a weak correlation with u_2 and, similarly, the right-hand side of u_2 has a weak correlation with u_1 . Thus, we can model m_1 as a function of u_1 and m_2 as a function of u_2 . Since the Hamiltonian is zero, we can model m_1^{re} and m_1^{im} , separately, as

$$m_1^{re} = m_1^{re}(u_1^{re}), \quad m_1^{im} = m_1^{im}(u_1^{im}). \quad (2.36)$$

Similarly, m_2 can be modeled as

$$m_2^{re} = m_2^{re}(u_2^{re}), \quad m_2^{im} = m_2^{im}(u_2^{im}). \quad (2.37)$$

Note that the weak correlation of Fourier modes simplifies the structure of the Markov chain. If there is a strong correlation between Fourier modes or strong correlation between real and image part, the Markov chain has to be conditioned on one or more essential variables. We will mention on this aspect again in the case of non-zero Hamiltonian. The partition for the essential variables is defined as the same as previous section. The time step is set to $\Delta t = 0.1$. All parameters are the same as the case of $\lambda = 1$.

	Theory	Full Model	Effective Model
Mean of u_1^{re}	0.0	-6.169×10^{-4}	3.459×10^{-4}
Mean of u_1^{im}	0.0	-7.013×10^{-4}	4.543×10^{-4}
Mean of u_2^{re}	0.0	-3.351×10^{-4}	-1.531×10^{-3}
Mean of u_2^{im}	0.0	-2.225×10^{-4}	2.314×10^{-4}
Variance of u_1^{re}	0.01	9.854×10^{-3}	9.882×10^{-3}
Variance of u_1^{im}	0.01	9.845×10^{-3}	9.968×10^{-3}
Variance of u_2^{re}	0.01	9.789×10^{-3}	1.001×10^{-2}
Variance of u_2^{im}	0.01	9.720×10^{-3}	1.005×10^{-2}
Kurtosis of u_1^{re}	3.0	2.885	3.247
Kurtosis of u_1^{im}	3.0	2.867	3.158
Kurtosis of u_2^{re}	3.0	2.882	3.360
Kurtosis of u_2^{im}	3.0	2.876	3.381

Table 2.8: Mean, variance, and kurtosis of u_1 and u_2 from the full model and the effective model with $\lambda = 2$.

From Table (2.8), Figure (2.12), and Figure (2.13), the effective model reproduces the statistical properties well. The variances of u_1 and u_2 are close to the values from the theory. The auto-correlation functions of both u_1 and u_2 have the same decay rate as the ones obtained from the full model. However, there is an oscillation in the auto-correlation function of essential variables in the effective model when $5 \leq \tau \leq 10$. The probability density functions from the effective model are similar to those obtained from the full model.

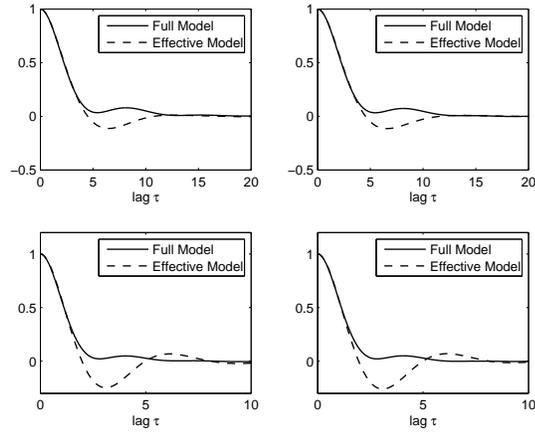


Figure 2.12: Auto-correlation functions of u_1 and u_2 . The top left figure is the ACF of u_1^{re} . The top right is the ACF of u_1^{im} . The bottom left is the ACF of u_2^{re} and the bottom right is the ACF of u_2^{im} .

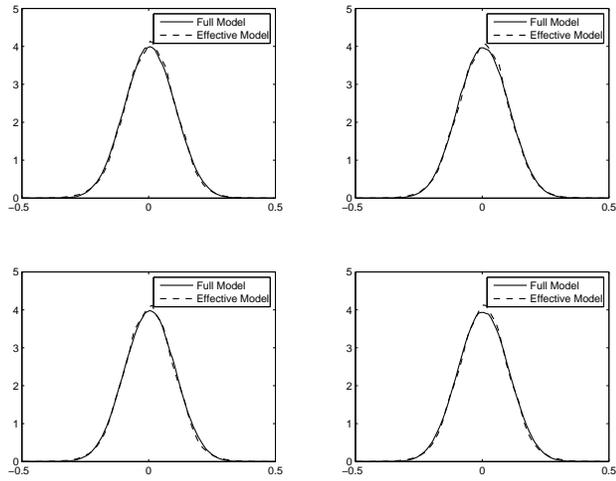


Figure 2.13: Probability density functions of u_1 and u_2 . The top left figure is the PDF of u_1^{re} . The top right is the PDF of u_1^{im} . The bottom left is the PDF of u_2^{re} and the bottom right is the PDF of u_2^{im} .

2.4.5 The Effective Model for TBH with Non-zero Hamiltonian

For zero Hamiltonian, the conditioning scheme for the Markov chain effective model is simple: the Markov chain is conditioned on one essential variable. For non-zero Hamiltonian, it is shown in [7] that the correlation between the real part and the image part of the Fourier modes is not negligible. If the Markov chain is conditioned on only one variable, then the effective model may not be able to reproduce the cross-correlation between the real and image parts of the essential variables. Thus, it is essential to model the Markov chain to be conditioned on both real part and image part of the essential variables. We demonstrate this issue for the case of one Fourier mode. For m_1^{re} , we consider u_1^{re} as a primary variable and introduce u_1^{im} as a secondary variable. Similarly, for m_1^{im} , we consider u_1^{im} as a primary variable and introduce u_1^{re} as a secondary variable. Thus, both m_1^{re} and m_1^{im} can be modeled as

$$m_1^{re} = m_1^{re}(u_1^{re}, u_1^{im}), \quad m_1^{im} = m_1^{im}(u_1^{im}, u_1^{re}). \quad (2.38)$$

The partition on the primary variable is the same as we defined in (2.34) which is

$$\mathbb{J}_u = \{-\infty < \mu_u - 1.4\sigma_u < \mu_u - 0.6\sigma_u < \mu_u < \mu_u + 0.6\sigma_u < \mu_u + 1.4\sigma_u < \infty\}$$

where u stands for the primary variable. The partition for the secondary variable is defined as

$$\mathbb{J}_v = \{-\infty < \mu_v < \infty\} \quad (2.39)$$

where v is the secondary variable. The number of subintervals for the primary variable is 6 subintervals and the number of subintervals for the secondary variable is 2 subintervals. Thus, in this case, the effective model consists of 12 conditional Markov chains. Also, we note that the number of required coefficients in the estimation of the transition probability matrices increases twice.

Table (2.9) and Figure (2.14) show the comparison of the statistical results between the full model and the effective model where the Markov chains are conditioned

on both primary and secondary variables. By introducing the secondary variable to the Markov chain, the effective model reproduces the cross-correlation between u_1^{re} and u_1^{im} . All other statistical properties are in a good agreement with the full model.

	Theory	Full Model	Effective Model
Mean of u_1^{re}	0.0	1.884×10^{-5}	-1.054×10^{-4}
Mean of u_1^{im}	0.0	3.225×10^{-4}	-1.173×10^{-4}
Variance of u_1^{re}	0.01	1.030×10^{-2}	9.852×10^{-3}
Variance of u_1^{im}	0.01	1.035×10^{-2}	9.811×10^{-3}
Kurtosis of u_1^{re}	3.0	2.800	3.188
Kurtosis of u_1^{im}	3.0	2.809	3.290

Table 2.9: Mean, variance, and kurtosis of u_1 from the full model and the effective model. The TBH equation has non-zero Hamiltonian.

2.5 Conclusion

We studied the Markov chain modeling of the effective models for some deterministic and stochastic systems. From numerical simulations, statistical properties of essential variables can be reproduced using the Markov chain modeling approach. The factors that impact statistical properties of effective models depend on the number of states of the Markov chain, the number of conditional Markov chains, and the time step used in the simulation of the Markov chain. In our simulations, the number of states is typically in the order of 10. Increasing the number of states improves statistical properties of the effective model but it also requires more data in order to estimate the transition probability matrix accurately. The number of conditional Markov chains depends on the structure of the system. If non-essential variables do not depend on essential variables, then Markov chains do not have to be conditioned

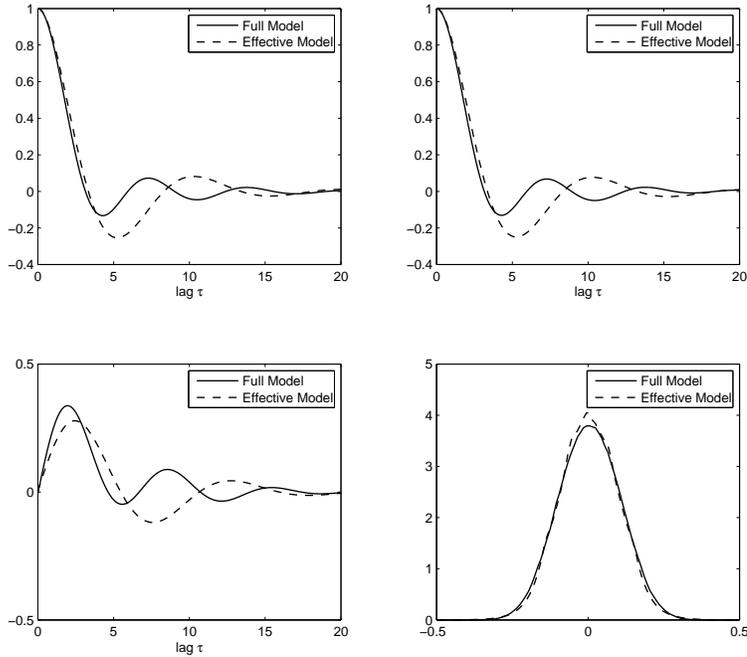


Figure 2.14: Statistical properties of u_1 . The top left figure is the ACF of u_1^{re} . The top right is the ACF of u_1^{im} . The bottom left is the cross-correlation between u_1^{re} and u_1^{im} . The bottom right is the PDF of u_1^{re} .

on essential variables. Markov chains have to be conditioned on essential variables if non-essential variables depend on essential variables. This increases the complexity of the structure of the conditional Markov chain if there are many essential variables. However, in most systems, the conditional Markov chain can be conditioned on one or two essential variables. In this case, the complexity is reduced significantly. The simulations in section (2.3.2) and (2.4) have shown that the effective models with 4 to 7 conditional Markov chains reproduce statistical properties of essential variables. The time step is also one important factor for the effective model. With an appropriate time step, the effective model reproduce statistical properties of the full model. The proper values of the time step are generally in the range of 0.1 to 0.5 of the

correlation time of $g(x, y)$ defined in (2.1a).

One important benefit of the Markov chain modeling approach is that it works well with the system both strong and weak scale separations. As we have seen the numerical results in this chapter, the Markov chain effective models outperform the asymptotic reduced models under the weak scale separation. This is an important benefit from the Markov chain approach since, in many systems, the scale separation between essential variables and non-essential variables cannot be clearly separated.

Chapter 3

Stochastic Delay Differential

Equation for the Effective Model

3.1 Introduction

In this chapter, we propose a stochastic delay model for the dynamics of essential variables. It is a numerically oriented approach. The effective model is constructed from the time series of essential variables. The effective model mainly consists of three parts. The first part is the Markovian term. The second part is the memory term and the last part is the noise term. Generally, the dynamics of essential variables in the effective model depends on the current and past values of themselves and is driven by noise. The coefficients of the stochastic delay effective model are obtained by simple least square method. The details are given later in this chapter.

The rest of this chapter is organized as follows. Section 3.2 describes the general idea of delay models. It includes the estimation of parameters in the effective model from the time series of essential variables. In section 3.3, we apply the stochastic delay model to the TBH equation. In this case, we assume that the effective model consists of the first Fourier mode. We also describe the approximation of the noise

term in the effective model by white noise approximation. We also investigate the case where the effective model consists of the first three Fourier modes in section 3.4. Finally, we conclude this chapter in section 3.5.

3.2 General Setup for the Stochastic Delay Model

We consider the equation of the following form

$$\frac{dx}{dt} = g_1(x, y) \tag{3.1a}$$

$$\frac{dy}{dt} = g_2(x, y) \tag{3.1b}$$

where x and y are the essential and the non-essential variables, respectively. We assume that the dimension of x is generally much less than the dimension of y . We are particularly interested in the dynamics of essential variables. The previous chapter has shown that the effective model can be obtained by replacing some parts of the right hand side of essential variables with the appropriate set of Markov chains. However, this method requires the knowledge of the right hand side of essential variables. In some cases, the dynamics of models is unknown, i.e., the right hand side of essential variables is not known explicitly. We can only observe the time series of essential variables. The objective of this chapter is to construct the effective system of essential variables from these numerical data. The numerical values of the right hand side of essential variables can be estimated from discrete consecutive observational data utilizing, for instance, the Euler discretization scheme. The main idea of this approach is based on the fact that there are correlations between essential variables and time derivatives of essential variables. Thus, the right hand side of essential variables depends on the current value and the past value of essential variables. In particular, we seek the following form of the effective model

$$\frac{dX(t)}{dt} = G(X(t), X(t - \tau)) + e(t). \tag{3.2}$$

Here, $X(t - \tau)$ represents the past values of essential variables and $e(t)$ is a random force in the effective system. For simplicity of the effective model, we choose the linear delay model with finite number of time delays for the effective equation. The model is written as follows

$$\frac{dX(t)}{dt} = \sum_{i=1}^n a_i X(t - t_i) + e(t), \quad (3.3)$$

where $e(t)$ represents the stochastic behaviour of the system. Without loss of generality, we assume $0 = t_1 < t_2 < \dots < t_n$. We simply set $t_1 = 0$ to represent the current value of the essential variable X . In some cases where there is a strong scale separation between $X(t)$ and $e(t)$, $e(t)$ can be approximated by a white noise (time derivative of Brownian motion). Thus, the effective models are represented by stochastic delay differential equations. In the next section, we describe the model in details and provide the method for parameter estimation.

3.2.1 Model Setup and Parameter Estimation

We assume that there are m essential variables and the time series of x are observed at known time step, Δt . For simplicity of the notation, we use the matrix representation for discrete data. Let \mathbf{X}_i be a row vector consisting of the time series of x_i with length N . Let \mathbf{X} be a matrix that contains $\mathbf{X}_i, \forall i = 1, \dots, m$. The dimension of \mathbf{X} is, therefore, m by N . The element, $\mathbf{X}_{i,j}$, represents the value of the i^{th} essential variable at time $j\Delta t$, i.e.,

$$\mathbf{X}_{i,j} = x_i(j\Delta t). \quad (3.4)$$

To estimate the right-hand side of the effective model which is the time derivative of essential variables, we let \mathbf{Y} be the matrix consisting of the approximation of the time derivative of essential variables. Then, the time derivative can be approximated from the second order finite difference scheme as follows

$$\mathbf{Y}_{i,j} = \frac{\mathbf{X}_{i,j+1} - \mathbf{X}_{i,j-1}}{2\Delta t}, \quad j = 1, \dots, N, \quad (3.5)$$

where we assume that $\mathbf{X}_{i,0} = \mathbf{X}_{i,N+1} = 0, \forall i \in \{1, \dots, m\}$. We assume that the number of time delays for each essential variable is n , and each essential variable can have a different time delay. Let t_{kj} be the j^{th} time delay of the k^{th} essential variable. We seek the linear delay model of the following form

$$\frac{dx_i(t)}{dt} = \sum_{k=1}^m \sum_{j=1}^n a_{kj}^i x_k(t - t_{kj}) + e_i(t), \quad i \in \{1, \dots, m\}. \quad (3.6)$$

From equation (3.6), the changes of essential variables are linear functions of themselves at different time delays. $e_i(t)$ represents a random force for an essential variable. The coefficients a_{kj}^i is estimated by the method of least square. The details are as follows.

Method of Least Square

Let $\tau_{kj} = \lfloor t_{kj}/\Delta t \rfloor$ where $\lfloor x \rfloor$ denotes the largest integer that is less than x . The value of τ_{kj} represents the corresponding discrete time of t_{kj} . Let

$$\tau_{max} = \max_{\substack{1 \leq k \leq m, \\ 1 \leq j \leq n}} \tau_{kj} \quad (3.7)$$

be the maximum discrete time delay of the effective model. To estimate a_{kj}^i , numerical data at different time delays are required. One can obtain the required data with specified time delay by shifting the time series. For example, to obtain the time series of $x_1(t - t_1)$, we can use the sample data from $\mathbf{X}_{1,1}$ to $\mathbf{X}_{1,N}$ as the time series of $x_1(t - t_1)$ and the sample data from $\mathbf{X}_{1,(1+\tau_1)}$ to $\mathbf{X}_{1,(N+\tau_1)}$ as the time series of $x_1(t)$ where $\tau_1 = \frac{t_1}{\Delta t}$. Thus, to estimate the parameters for the effective model from the data of length N , we need the time series with length $N + \tau_{max}$, and numerical data of length N for $x_k(t - t_{kj})$ are obtained from the time series of x_k from $\mathbf{X}_{k,(1+\tau_{max}-\tau_{kj})}$ to $\mathbf{X}_{k,(N+\tau_{max}-\tau_{kj})}$. Note that, in this case, $\mathbf{X}_{k,(1+\tau_{max})}$ corresponds to the sample point for $x_k(t)$. To put the time series of x_k with different time delays in a matrix form, we define \mathbf{U}^k as n by N matrix where the row j^{th} of \mathbf{U}^k represents the time series of

$x_k(t - t_{kj})$, i.e.,

$$\mathbf{U}^k = \begin{bmatrix} \mathbf{X}_{k,(1+\tau_{max}-\tau_{k1})} & \cdots & \mathbf{X}_{k,(N+\tau_{max}-\tau_{k1})} \\ \mathbf{X}_{k,(1+\tau_{max}-\tau_{k2})} & \cdots & \mathbf{X}_{k,(N+\tau_{max}-\tau_{k2})} \\ \vdots & & \vdots \\ \mathbf{X}_{k,(1+\tau_{max}-\tau_{kn})} & \cdots & \mathbf{X}_{k,(N+\tau_{max}-\tau_{kn})} \end{bmatrix}. \quad (3.8)$$

Let $\mathbf{U} \in \mathbb{R}^{mn \times N}$ be the matrix where the j^{th} row consists of \mathbf{U}^k . Thus, \mathbf{U} contains the time series of all essential variables with all time delays defined in (3.6). Next, we define a time series of the right-hand side of essential variables. We define a matrix $\mathbf{V} \in \mathbb{R}^{m \times N}$ where the component of \mathbf{V} is defined as follows

$$\mathbf{V}_{i,j} = \mathbf{Y}_{i,j+\tau_{max}}, \forall i \in \{1, \dots, m\}, \forall j \in \{1, \dots, N\}. \quad (3.9)$$

Each row of \mathbf{V} corresponds to the time series of the right hand side of $x_k(t)$. Equation (3.6) can be rewritten in a matrix notation as follows

$$\mathbf{V} = \mathbf{A}^T \mathbf{U} + \mathbf{E}, \quad (3.10)$$

where $\mathbf{A} \in \mathbb{R}^{mn \times m}$. In the context of method of least square, we seek a matrix \mathbf{A} such that the value of $\|\mathbf{E}\|^2 = \|\mathbf{V} - \mathbf{A}^T \mathbf{U}\|^2$ is minimized where $\|\mathbf{E}\|^2 = \sum_{i=1}^m \sum_{k=1}^N \mathbf{E}_{ij}^2$. Using the method of least square provided in Appendix B, \mathbf{A} is obtained as follows

$$\mathbf{A} = (\mathbf{U}\mathbf{U}^T)^{-1}(\mathbf{U}\mathbf{V}^T). \quad (3.11)$$

The coefficient a_{kj}^i of the right-hand side of x_i in equation (3.6) is $\mathbf{A}_{(k-1)n+j,i}$. The random force, $e_i(t)$, in (3.6) is obtained directly from the residual from least square method as follows

$$\mathbf{E} = \mathbf{V} - \mathbf{A}^T \mathbf{U}. \quad (3.12)$$

Here, the i^{th} row of \mathbf{E} , \mathbf{E}_i , is the time series of $e_i(t)$. Note that it is possible to approximate the random force, $e_i(t)$, by white noise approximation if there is a strong scale separation between $e_i(t)$ and $x_i(t)$. We illustrate the white noise approximation for the TBH case in the next section.

3.3 The Effective Model for the First Fourier Mode of TBH System

We consider the case where the effective model consists of the first Fourier mode. We write u_1 as a real part, u_1^{re} , and an image part, u_1^{im} , and assume that the delay model consists of n time delays. Thus, the delay effective model can be written as follows

$$du_1^{re} = \left[\sum_{i=1}^n a_i u_1^{re}(t - t_i) + \sum_{i=1}^n b_i u_1^{im}(t - t_i) \right] dt + e_1^{re} dt \quad (3.13a)$$

$$du_1^{im} = \left[\sum_{i=1}^n c_i u_1^{re}(t - t_i) + \sum_{i=1}^n d_i u_1^{im}(t - t_i) \right] dt + e_1^{im} dt. \quad (3.13b)$$

The coefficients a_i, b_i, c_i , and d_i are estimated from equation (3.11). The processes e_1^{re} and e_1^{im} are constructed from the time series of residuals obtained from (3.12). In practice, the random forces behave stochastically in a fast time scale. Thus, it is possible to model these random forces by white noises. We assume that the random forces are stationary and ergodic processes with zero mean. We assume that the random forces are written as follows

$$de_1^{re} = f_1(e_1^{re})dt + h_1(e_1^{re})dW_1 \quad (3.14a)$$

$$de_1^{im} = f_2(e_1^{im})dt + h_2(e_1^{im})dW_2, \quad (3.14b)$$

where W_1 and W_2 are independent Brownian motions. Combining (3.13) and (3.14), and introducing ϵ , the system of equations can be written as follows

$$du_1^{re} = \left[\sum_{i=1}^n a_i u_1^{re}(t - t_i) + \sum_{i=1}^n b_i u_1^{im}(t - t_i) \right] dt + \frac{1}{\epsilon} e_1^{re} dt \quad (3.15a)$$

$$du_1^{im} = \left[\sum_{i=1}^n c_i u_1^{re}(t - t_i) + \sum_{i=1}^n d_i u_1^{im}(t - t_i) \right] dt + \frac{1}{\epsilon} e_1^{im} dt \quad (3.15b)$$

$$de_1^{re} = \frac{1}{\epsilon^2} f_1(e_1^{re})dt + \frac{1}{\epsilon} h_1(e_1^{re})dW_1 \quad (3.15c)$$

$$de_1^{im} = \frac{1}{\epsilon^2} f_2(e_1^{im})dt + \frac{1}{\epsilon} h_2(e_1^{im})dW_2. \quad (3.15d)$$

We apply the homogenization method provided in Appendix A to (3.15) to obtain the stochastic delay effective model. The stochastic delay effective model is written as follows

$$du_1^{re} = \left[\sum_{i=1}^n a_i u_1^{re}(t - t_i) + \sum_{i=1}^n b_i u_1^{im}(t - t_i) \right] dt + \sqrt{2}\sigma_1 dB_1 \quad (3.16a)$$

$$du_1^{im} = \left[\sum_{i=1}^n c_i u_1^{re}(t - t_i) + \sum_{i=1}^n d_i u_1^{im}(t - t_i) \right] dt + \sqrt{2}\sigma_2 dB_2, \quad (3.16b)$$

where B_1 and B_2 are independent Brownian motions and the values of σ_1 and σ_2 are defined as follows

$$\sigma_1^2 = \int_0^\infty \mathbb{E}[e_1^{re}(t)e_1^{re}(0)]dt \quad (3.17a)$$

$$\sigma_2^2 = \int_0^\infty \mathbb{E}[e_1^{im}(t)e_1^{im}(0)]dt. \quad (3.17b)$$

Here, $\mathbb{E}[\cdot]$ denotes the expectation with respect to the invariant measure generated by e_1 . In practice, σ_1 and σ_2 can be computed numerically from the time series defined in (3.12). The expected value in (3.17) is computed by using the time average on a single realization. Thus, we write $\mathbb{E}[e_1^{re}(t)e_1^{re}(0)]$ as follows

$$\mathbb{E}[e_1^{re}(t)e_1^{re}(0)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T e_1^{re}(t+s)e_1^{re}(s)ds, \quad (3.18)$$

and (3.17) can be rewritten as

$$\sigma_1^2 = \int_0^\infty \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T e_1^{re}(t+s)e_1^{re}(s)ds \right] dt \quad (3.19a)$$

$$\sigma_2^2 = \int_0^\infty \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T e_1^{im}(t+s)e_1^{im}(s)ds \right] dt. \quad (3.19b)$$

3.3.1 Numerical Results for One Fourier Mode

We simulate the full TBH equation with parameters as follows. The number of Fourier modes is 20. ($\Lambda = 20$). The initial values of Fourier modes are chosen randomly such that the total energy of the system is 0.4. Thus, the energy for the real and

image of each Fourier mode is 0.01. We also choose the initial values such that the Hamiltonian of the system is not zero. The full model is numerically integrated by Pseudo Spectral method combining with the fourth-order Runge-Kutta method. The time step in the numerical integration of the full model is 10^{-3} . To obtain the time series of the essential variable, we sample data at time step $\Delta t = 0.01$ with the total time of 2,000. Thus, the length of the time series is 200,000 points. For convenience of notation, we assume that the dynamics of u_1 is written as follows

$$\begin{aligned}\frac{du_1^{re}}{dt} &= g_1^{re} \\ \frac{du_1^{im}}{dt} &= g_1^{im},\end{aligned}$$

where g_1^{re} and g_1^{im} depends on the current and past values of u_1^{re} and u_1^{im} . The numerical data of g_1^{re} and g_1^{im} are obtained by the method described in section (3.2.1). Figure (3.1) and (3.2) show the cross-correlation between essential variables and the right hand side of essential variables. In this effective model, there are two essential variables which are u_1^{re} and u_1^{im} . Both g_1^{re} and g_1^{im} are correlated with u_1^{re} and u_1^{im} . The cross-correlations are significant when the time lag is between 0 and 4. The cross-correlations decay to zero as the time lag increases. This suggests that g_1^{re} and g_1^{im} depend on the current and past values of u_1^{re} and u_1^{im} with a short period of memory. In this particular model, we choose a maximum time delay $t_{max} = 4$.

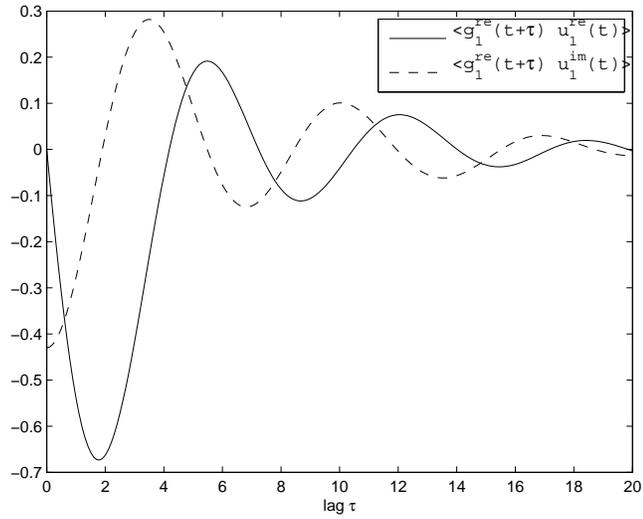


Figure 3.1: Cross-correlation function between g_1^{re} and u_1 . The solid line is the cross-correlation between g_1^{re} and u_1^{re} . The dashed line is the cross-correlation between g_1^{re} and u_1^{im} .

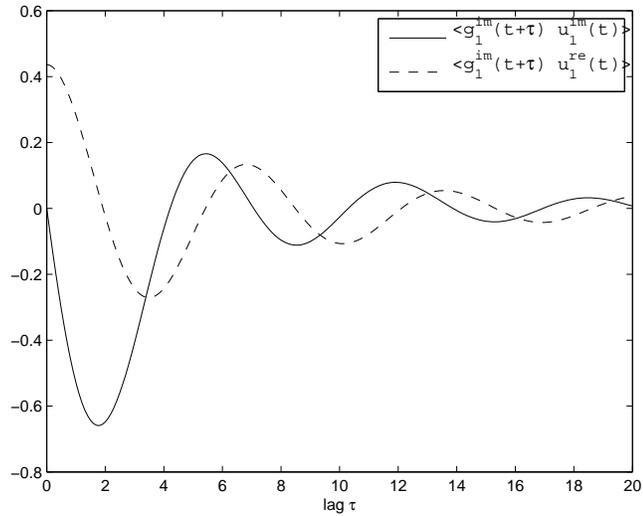


Figure 3.2: Cross-correlation function between g_1^{im} and u_1 . The solid line is the cross-correlation between g_1^{im} and u_1^{im} . The dashed line is the cross-correlation between g_1^{im} and u_1^{re} .

For the effective model, we set the number of delays to 5 ($n = 5$). Figures (3.1) and (3.2) show that the cross-correlations are strong at the time lag between 0 and 4. We choose the time delays for the effective model at time lags of 0.0, 1.0, 2.0, and 4.0. Note that the time lag $t_1 = 0.0$ corresponds to the current value of essential variables. Table (3.1) shows the coefficient values in (3.16) obtained from the least square method.

i	τ_i	a_i	b_i	c_i	d_i
1	0.0	1.1088	-0.1408	0.1358	1.1136
2	1.0	-1.5861	0.1902	-0.1888	-1.5959
3	2.0	0.4874	-0.1299	0.1374	0.5006
4	3.0	-0.0956	0.0663	-0.0744	-0.0996
5	4.0	0.0193	-0.0290	0.0319	0.0189

Table 3.1: The coefficients of the stochastic delay differential equation defined in (3.16). The coefficient values are obtained by the least square method defined in (3.11).

Once the coefficients are obtained, the time series of e_1^{re} and e_1^{im} can be obtained from the residual defined in (3.12). Note that the time series of e_1^{re} and e_1^{im} has to be uncorrelated with u_1^{re} and u_1^{im} since, by the concept of least square method, the residuals are orthogonal to the independent variables. Figure (3.3) shows the cross-correlation functions between noises (e_1^{re}, e_1^{im}) and essential variables (u_1^{re}, u_1^{im}) . From the figure, the cross-correlation functions are almost identically zero.

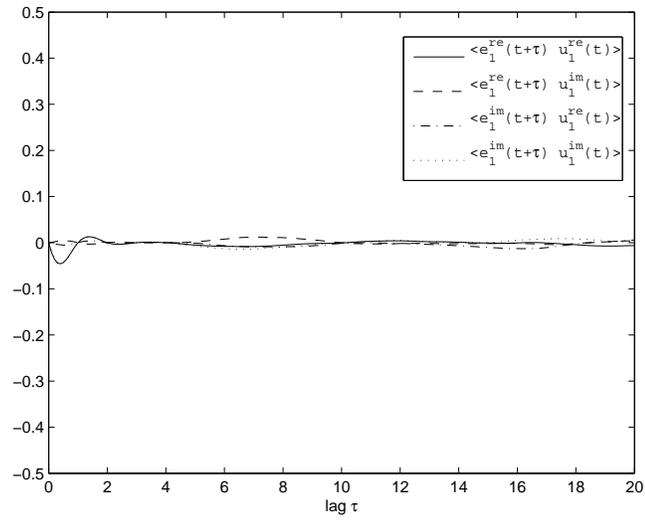


Figure 3.3: Cross-correlation functions between noises (e_1^{re}, e_1^{im}) and essential variables (u_1^{re}, u_1^{im}) .

To approximate e_1^{re} and e_1^{im} by white noise, a time scale separation between e_1^{re} , e_1^{im} and u_1^{re} , u_1^{im} has to be assumed. Their auto-correlation functions are illustrated in Figure (3.4). The auto-correlation function of e_1^{re} decays faster than the auto-correlation function of u_1^{re} . The correlation time, which is defined as

$$\text{correlation time of } X = \frac{\int_0^\infty \mathbb{E}[X(t+\tau)X(t)]d\tau}{\mathbb{E}[X(t)]^2}, \quad (3.20)$$

represents how fast the stochastic processes decorrelate. In this particular case, the correlation times of e_1^{re} , e_1^{im} , u_1^{re} and u_1^{im} are shown in Table (3.2). The correlation times of e_1^{re} and e_1^{im} are approximately one tenth of the correlation times of u_1^{re} and u_1^{im} .

	Correlation time
e_1^{re}	0.1440
e_1^{im}	0.1573
u_1^{re}	1.4738
u_1^{im}	1.5378

Table 3.2: The correlation time of e_1^{re} , e_1^{im} , u_1^{re} , and u_1^{im} .

In practice, the diffusion coefficients defined in (3.19) need to be approximated numerically from the time series of E defined in (3.12). The diffusion coefficients are calculated by the following equation

$$\sigma_i^2 = \sum_{m=1}^M \left[\frac{1}{N-M} \sum_{n=1}^{N-M} E_{i,n+m} E_{i,n} \right] \Delta t, \quad i \in \{1, 2\}, \quad (3.21)$$

where $M = \lfloor \frac{T_C}{\Delta t} \rfloor$ and $T_C = 10$. In principle, T_C has to be a large number but, from the numerical data, the auto-correlation function of e_1 decays to zero quickly. Thus, it is sufficient to set $T_C = 10$. The values of σ_1 and σ_2 computed from (3.21) are

$$\sigma_1 = 9.37 \times 10^{-3}$$

$$\sigma_2 = 9.75 \times 10^{-3}.$$

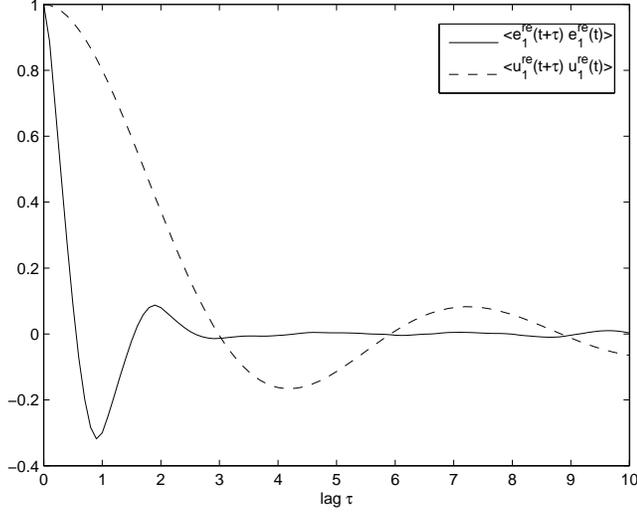


Figure 3.4: Auto-correlation functions of e_1^{re} and u_1^{re} . Note that the auto-correlation of e_1^{re} decorrelates faster than the auto-correlation of u_1^{re} .

To perform the simulation of the stochastic delay effective model, the equations in (3.16) are integrated by Euler-Maruyama method [21]. The discretized equations of (3.16) are written as follows

$$u_1^{re}(t + \Delta t) = u_1^{re}(t) + \left[\sum_{i=1}^n a_i u_1^{re}(t - t_i) + \sum_{i=1}^n b_i u_1^{im}(t - t_i) \right] \Delta t + \sqrt{2}\sigma_1 \sqrt{\Delta t} z_1 \quad (3.22a)$$

$$u_1^{im}(t + \Delta t) = u_1^{im}(t) + \left[\sum_{i=1}^n c_i u_1^{re}(t - t_i) + \sum_{i=1}^n d_i u_1^{im}(t - t_i) \right] \Delta t + \sqrt{2}\sigma_2 \sqrt{\Delta t} z_2, \quad (3.22b)$$

where z_1 and z_2 are independent standard normal distribution random variables. For simplicity, the time step, Δt , is chosen such that $\frac{t_i}{\Delta t}$, $i = 1, \dots, n$, are integer. Here, we choose $\Delta t = 10^{-2}$. Note that in the context of numerical programming, we have to store the past values of u_1 in a memory array. The length of the array equals to τ_{max} defined in (3.7). In this case, $\tau_{max} = 4/0.01 = 400$. This array is updated every time the value of u_1 is computed. We integrate the effective model with $T = 200,000$

and measure mean, variance, and kurtosis of u_1^{re} and u_1^{im} . We also measure the auto-correlation functions of u_1^{re} and u_1^{im} , the cross-correlation function between u_1^{re} and u_1^{im} , and the probability density functions of u_1^{re} and u_1^{im} .

		mean	variance	kurtosis
Theory		0.0	10^{-2}	3.0
Full model	u_1^{re}	-2.23×10^{-3}	1.01×10^{-2}	2.792
	u_1^{im}	-1.64×10^{-3}	1.01×10^{-2}	2.922
Effective model	u_1^{re}	-1.69×10^{-3}	1.01×10^{-2}	2.996
	u_1^{im}	-4.53×10^{-4}	1.03×10^{-2}	2.941

Table 3.3: Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the stochastic delay effective model defined in (3.16).

The stochastic delay effective model provides a good agreement with the full model. The relative error of the variance of u_1 is less than 5%. The kurtosis approximately equals to 3. The auto-correlation functions of u_1^{re} and u_1^{im} from the stochastic delay effective model are very close to the auto-correlation functions from the full model. There is a little discrepancy of the cross-correlation function. Overall, the effective model reproduces statistical behaviours of the full model both one-point and two-points statistics.

The Effect of the Number of Time Delays

In this section, we investigate the effect of the number of time delays in the effective model. We study two cases for the numerical simulations. The first effective model has two time delays ($n = 2$) and the time delays are at 0.0 and 1.0. The second effective model has ten time delays ($n = 10$) and the time delays are at 0.0, 1.0, ... , 9.0.

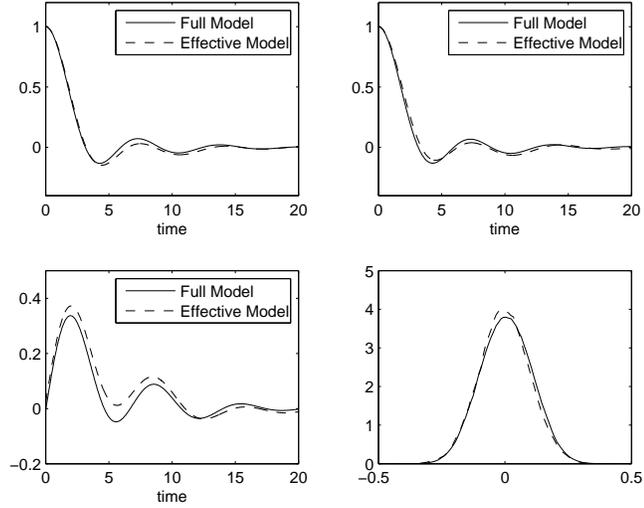


Figure 3.5: Comparison of statistical properties of u_1 between the full model and the stochastic delay effective model defined in (3.16). The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right.

Table (3.4) and (3.5) show the coefficients of stochastic delay models with two and ten delays, respectively. Statistical results from the effective model with two time delays are shown in Figure (3.6) and Table (3.6).

The variances of u_1^{re} and u_1^{im} from the stochastic delay effective model with two time delays are significantly larger than the variance from the full model. Both auto-correlation function and cross-correlation function from the effective model behave differently from the full model. There is a large oscillation in both auto-correlation and cross-correlation functions of the effective model. Overall, the effective model with two time delays is unable to reproduce the statistical behaviours of the full model. The result shows that the effective model requires more time delays in order to reproduce the statistical properties of the system. On the other hand, the effective

i	τ_i	a_i	b_i	c_i	d_i
1	0.0	0.8301	-0.1537	0.1582	0.8297
2	1.0	-0.9765	0.1431	-0.1436	-0.9667

Table 3.4: The coefficients of the stochastic delay effective model defined in (3.16) with two time delays. The coefficient values are obtained by the least square method defined in (3.11).

model with ten time delays reproduces the statistical properties of the full model. These statistical properties are shown in Figure (3.7) and and Table (3.7). Note that increasing the number of time delays from five to ten does not significantly improve statistical properties of the essential variables. The statistical behaviours from the effective models with five and ten time delays are similar. Moreover, the stochastic delay effective model with ten delays is more complex than the effective model with five delays. The computation needs more memory arrays to store the past values of essential variables. It is sufficient to model the stochastic delay effective model with five time delays.

i	τ_i	a_i	b_i	c_i	d_i
1	0.0	1.1155	-0.1300	0.1354	1.1116
2	1.0	-1.5872	0.1751	-0.1759	-1.5732
3	2.0	0.4881	-0.1215	0.1140	0.4712
4	3.0	-0.1044	0.0766	-0.0526	-0.0867
5	4.0	0.0030	-0.0650	0.0247	0.0173
6	5.0	-0.0064	0.0418	0.0057	0.0012
7	6.0	0.0071	-0.0191	-0.0129	-0.0034
8	7.0	-0.0161	0.0059	0.0141	-0.0067
9	8.0	0.0052	0.0030	-0.0115	0.0088
10	9.0	0.0017	-0.0022	0.0023	-0.0038

Table 3.5: The coefficients of the stochastic delay effective model defined in (3.16) with ten delays. The coefficient values are obtained by the least square method defined in (3.11).

		mean	variance	kurtosis
Theory		0.0	10^{-2}	3.0
Full model	u_1^{re}	-2.23×10^{-3}	1.01×10^{-2}	2.792
	u_1^{im}	-1.64×10^{-3}	1.01×10^{-2}	2.922
Effective model	u_1^{re}	-1.39×10^{-3}	3.03×10^{-2}	3.001
	u_1^{im}	-7.28×10^{-4}	3.01×10^{-2}	2.955

Table 3.6: Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the effective model defined in (3.16). The effective model consists of two time delays.

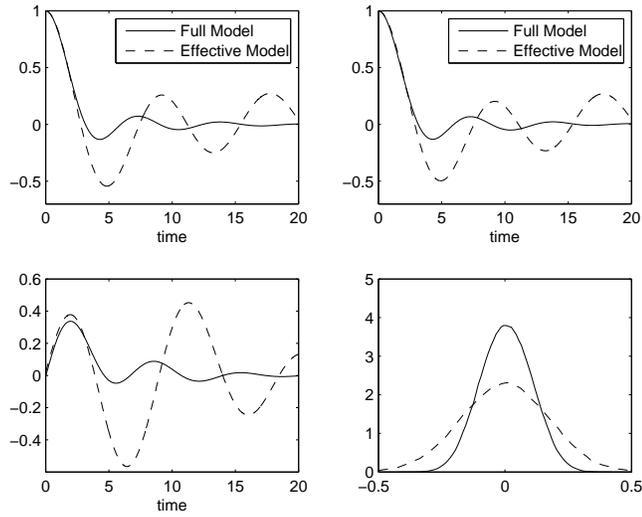


Figure 3.6: Comparison of statistical properties of u_1 between the full model and the stochastic delay effective model defined in (3.16). The effective model has two time delays. The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right.

		mean	variance	kurtosis
Theory		0.0	10^{-2}	3.0
Full model	u_1^{re}	-2.23×10^{-3}	1.01×10^{-2}	2.792
	u_1^{im}	-1.64×10^{-3}	1.01×10^{-2}	2.922
Effective model	u_1^{re}	-8.15×10^{-4}	1.05×10^{-3}	2.938
	u_1^{im}	1.14×10^{-4}	1.01×10^{-2}	3.007

Table 3.7: Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the effective model defined in (3.16). The effective model consists of ten time delays.

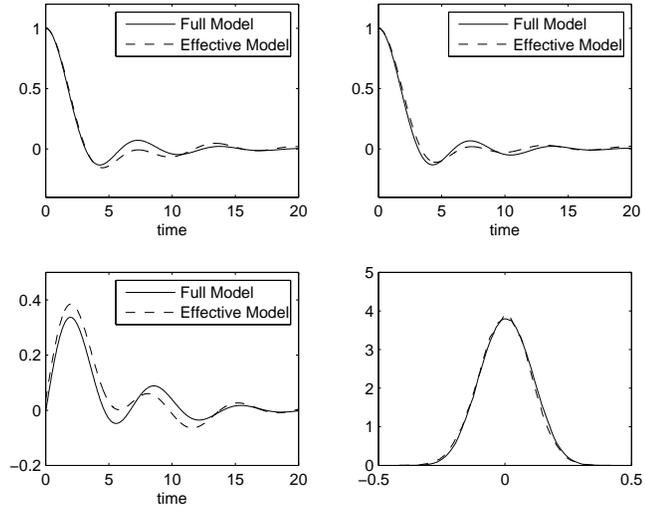


Figure 3.7: Comparison of statistical properties of u_1 between the full model and the stochastic delay effective model defined in (3.16). The effective model has ten time delays. The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right.

The Effect of the Time Step in Numerical Integration

In this section, we study the effect of the time step used in the numerical integration of the stochastic delay effective model. We choose the stochastic delay model with five time delays as an effective model. The coefficients of the effective model are shown in Table (3.1). We simulate the stochastic delay effective models with four different time steps, $\Delta t = 0.01, 0.02, 0.05, 0.1$. Note that these time steps are not related to the time step used in the parameter estimation. The time step used in the parameter estimation is 0.01 for all cases.

		mean	variance	kurtosis
Theory		0.0	10^{-2}	3.0
Full model with $\Delta t = 10^{-3}$	u_1^{re}	-2.23×10^{-3}	1.01×10^{-2}	2.792
	u_1^{im}	-1.64×10^{-3}	1.01×10^{-2}	2.922
Effective model with $\Delta t = 10^{-2}$	u_1^{re}	-1.69×10^{-3}	1.01×10^{-2}	2.996
	u_1^{im}	-4.53×10^{-4}	1.03×10^{-2}	2.941
Effective model with $\Delta t = 2 \times 10^{-2}$	u_1^{re}	-1.20×10^{-3}	1.02×10^{-2}	3.042
	u_1^{im}	3.95×10^{-5}	1.02×10^{-2}	3.005
Effective model with $\Delta t = 5 \times 10^{-2}$	u_1^{re}	-1.99×10^{-3}	1.02×10^{-2}	2.991
	u_1^{im}	1.47×10^{-5}	1.03×10^{-2}	2.999
Effective model with $\Delta t = 10^{-1}$	u_1^{re}	-1.24×10^{-3}	1.09×10^{-2}	3.005
	u_1^{im}	2.72×10^{-4}	1.10×10^{-2}	3.005

Table 3.8: Mean, variance, and kurtosis of u_1^{re} and u_1^{im} from the full model and the effective model with five time delays. The effective model is numerically integrated with different time steps.

Table (3.8) shows the one-point statistics of the effective model with different time steps. The variances of u_1^{re} and u_1^{im} from the effective model with $\Delta t = 0.01, 0.02,$

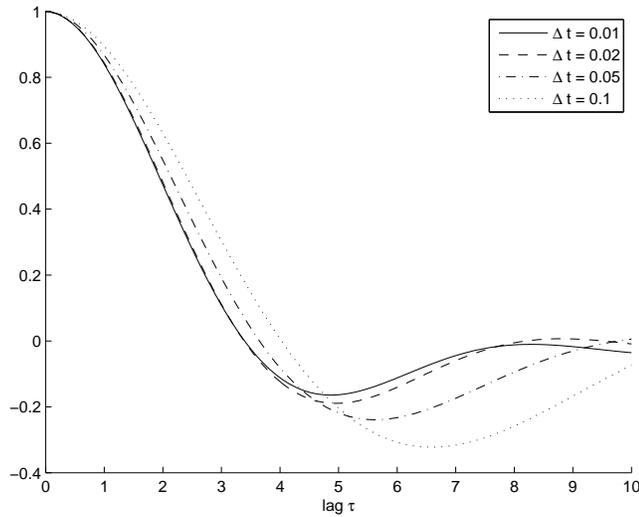


Figure 3.8: Auto-correlation functions of u_1^{re} from the effective model. The effective model is numerically integrated with different time steps.

and 0.05 are very close to the theoretical value. For $\Delta t = 0.1$, the variance is a little higher than the theoretical value. However, the auto-correlation function of u_1^{re} from the effective model with $\Delta t = 0.1$ behave differently from the others. In this simulation, the effective model with $0.01 \leq \Delta t \leq 0.05$ reasonably reproduces statistical properties of the full model. It is possible to increase the accuracy of the result by using the higher order numerical integration scheme for the stochastic delay model. The details of numerical solution of delay differential equations can be found in [20, 41].

3.4 The Effective Model for the First Three Fourier Modes

In this section, we consider the case where the essential variables consist of the first three Fourier modes. As we have seen in section (3.3.1), the effective model with five

time delays reproduces statistical properties of the full model. Thus, we use $n = 5$ for the effective model in this section. However, these three Fourier modes have different correlation times. The auto-correlation functions decorrelate at different rates. The effective model should have different time delays for each Fourier mode. Figure (3.9) shows the cross-correlation between g_i^{re} and u_i^{re} , $i = 1, 2, 3$, where g_i^{re} is the right-hand side of u_i^{re} .

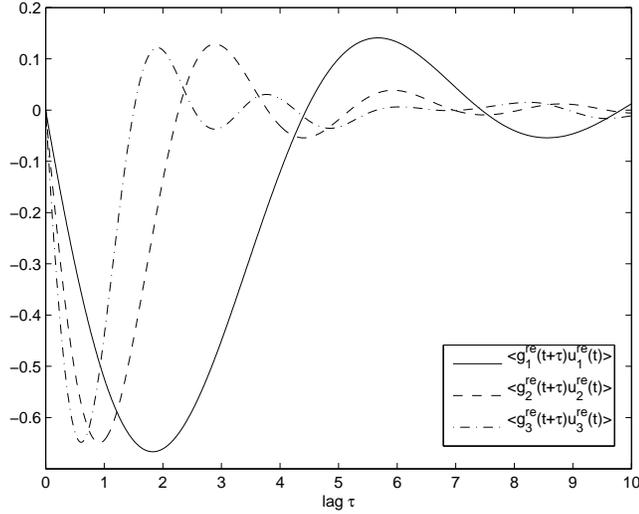


Figure 3.9: Correlation functions between g_i^{re} and u_i^{re} , $i = 1, 2, 3$.

From Figure (3.9), there is a strong correlation between g_1^{re} and u_1^{re} when $0 \leq \tau \leq 4$. Thus, for the first Fourier mode, we choose time delays at 0.0, 1.0, 2.0, 3.0, and 4.0. For the second Fourier mode, the cross-correlation function between g_2^{re} and u_2^{re} is significant when $0 \leq \tau \leq 4$. We choose the time delays within this interval. The time delays are at 0.0, 0.5, 1.0, 1.5, and 2.0. Similarly, the time delays for the third Fourier mode are chosen at time 0.0, 0.4, 0.8, 1.2, and 1.6.

To estimate the parameter of the effective model, \mathbf{U} is constructed such that its dimension is 30 by 200,000. Recall that the effective model has three Fourier modes with real and image. Thus, it has six essential variables. Each essential variable has

five time delays. Therefore, the number of rows in \mathbf{U} is $3 \times 2 \times 5 = 30$. The number of columns in \mathbf{U} is simply equal to the length of a time series which is 200,000. \mathbf{V} is a matrix with a dimension of 6 by 200,000. The coefficients of the effective model and the residual are obtained from (3.11) and (3.12), respectively. Figure (3.10) shows

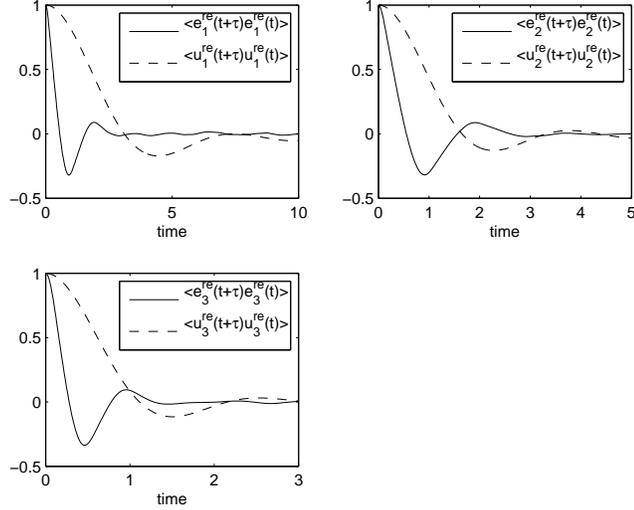


Figure 3.10: The auto-correlation functions of e_i^{re} and u_i^{re} , $i = 1, 2, 3$. Note that the decorrelation time of e_i is faster than u_i , for $i = 1, 2, 3$.

the comparison between the auto-correlation functions of the residual and the auto-correlation functions of the essential variables. Note that the residuals decorrelate faster than the essential variables. Thus, we apply the homogenization method to obtain the diffusion coefficients of the effective model. Table (3.9) shows the mean, variance, and kurtosis of u_i^{re} and u_i^{im} for $i = 1, 2, 3$. The variances of u_i , $i = 1, 2, 3$, are close to the theoretical value. The relative errors of the variances are less than 10 %. The kurtosis from the effective model approximately equals to 3. Figure (3.11) - (3.13) show the comparison of the correlation functions between the full model and the effective model. There is a slight discrepancy of the auto-correlation and cross-correlation functions of u_i , $i = 1, 2, 3$. However, the decorrelation times

		mean	variance	kurtosis
Theory		0.0	10^{-2}	3.0
Full model	u_1^{re}	-2.23×10^{-3}	1.01×10^{-2}	2.792
	u_1^{im}	-1.64×10^{-3}	1.01×10^{-2}	2.922
	u_2^{re}	-4.05×10^{-5}	1.05×10^{-2}	2.829
	u_2^{im}	-1.51×10^{-3}	1.03×10^{-2}	2.823
	u_3^{re}	-1.59×10^{-4}	1.03×10^{-2}	2.831
	u_3^{im}	2.47×10^{-5}	1.03×10^{-2}	2.822
Effective model	u_1^{re}	-1.23×10^{-5}	1.00×10^{-2}	3.011
	u_1^{im}	2.91×10^{-5}	1.01×10^{-2}	2.942
	u_2^{re}	2.27×10^{-4}	1.04×10^{-2}	3.002
	u_2^{im}	6.76×10^{-5}	1.07×10^{-2}	3.040
	u_3^{re}	9.13×10^{-5}	9.38×10^{-3}	3.012
	u_3^{im}	-6.18×10^{-5}	1.03×10^{-2}	2.992

Table 3.9: Mean, variance, and kurtosis of u_i^{re} and u_i^{im} for $i = 1, 2, 3$ from the full model and the effective model.

of $u_i, i = 1, 2, 3$, from these two models are approximately the same. Overall, the effective model reproduces the statistical behaviours of the full model.

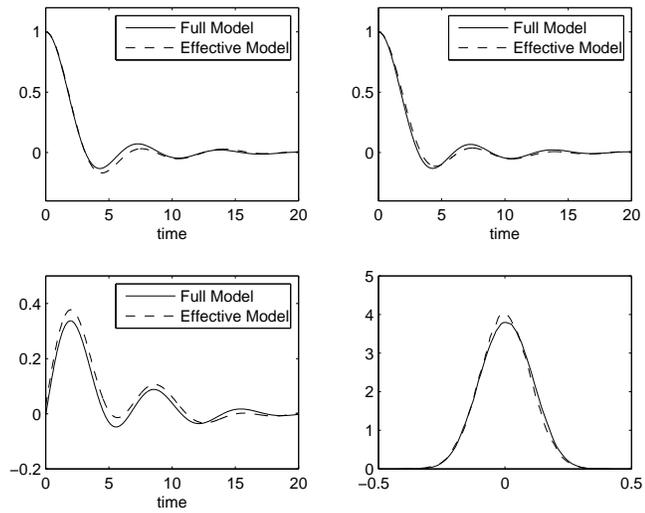


Figure 3.11: Statistical properties of u_1 from the full model and the stochastic delay effective model. The auto-correlation of u_1^{re} is on the top left. The auto-correlation of u_1^{im} is on the top right. The cross-correlation between u_1^{re} and u_1^{im} is on the bottom left. The probability density function of u_1^{re} is on the bottom right.

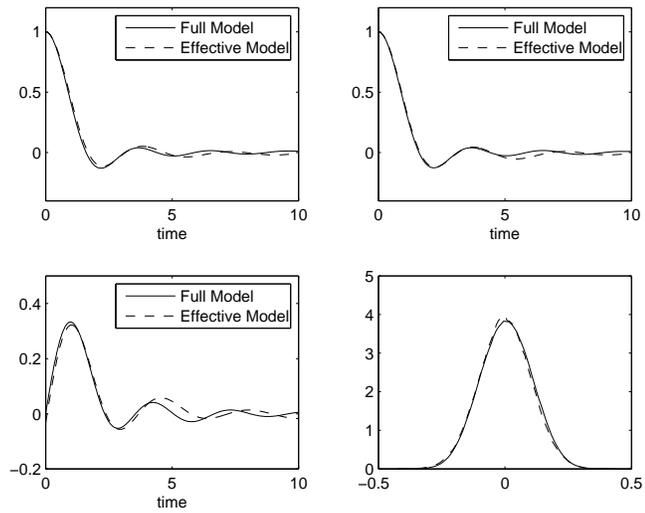


Figure 3.12: Statistical properties of u_2 from the full model and the stochastic delay effective model. The auto-correlation of u_2^{re} is on the top left. The auto-correlation of u_2^{im} is on the top right. The cross-correlation between u_2^{re} and u_2^{im} is on the bottom left. The probability density function of u_2^{re} is on the bottom right.

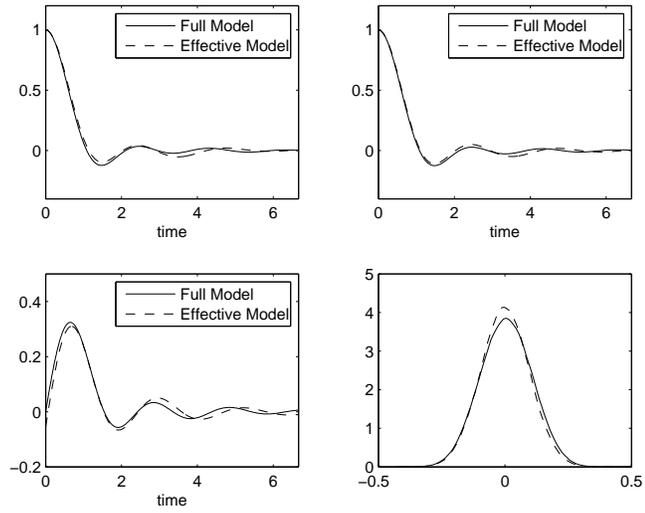


Figure 3.13: Statistical properties of u_3 from the full model and the stochastic delay effective model. The auto-correlation of u_3^{re} is on the top left. The auto-correlation of u_3^{im} is on the top right. The cross-correlation between u_3^{re} and u_3^{im} is on the bottom left. The probability density function of u_3^{re} is on the bottom right.

3.5 Conclusion

We have shown in this chapter that the stochastic delay effective model reproduced statistical behaviours of essential variables of TBH system. The stochastic delay effective model is simple and straightforward. It is a linear combination of essential variables at different time delays. One benefit of this effective model is that it does not require the knowledge of the dynamics of essential variables. The stochastic delay effective model can be obtained from a single realization of essential variables. Parameters of the stochastic delay effective model are obtained from least square method. We also showed that the factor that impact the effective model is mainly the number of time delays. It is shown that effective models with five or more time delays provided a good agreement with the full model. However, as we mentioned earlier in this chapter, the numerical simulation requires more memory arrays as the number of time delays increases but the statistical results are not significantly improved. Thus, it is best to employ minimal number of time delays in the stochastic delay effective model. We also showed that, under the strong scale separation assumption, the random force in the effective model can be approximated by Brownian motion. This simplifies the structure of the stochastic delay effective model.

Appendix A

Asymptotic Mode Elimination

A.1 Forward and Backward Equations

Consider a one-dimensional time homogeneous SDE

$$\begin{aligned} dx(t) &= \mu(x)dt + \sigma(x)dW(t), \\ x(0) &= x. \end{aligned} \tag{A.1}$$

Here $\mu(x)$ and $\sigma(x)$ are the drift and diffusion coefficients, respectively. Let $f(x)$ be a twice differentiable function. Define the operator L as follows

$$Lf(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}[f(x(t))] - f(x)}{t}. \tag{A.2}$$

It can be shown that

$$Lf(x) = \mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}. \tag{A.3}$$

To see this, consider the Ito's formula for $f(x)$

$$f(x(t)) - f(x) = \int_0^t \left(\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} \right) du + \int_0^t \sigma \frac{\partial f}{\partial x} dW(u). \tag{A.4}$$

By taking the expectation of (A.4) and taking the limit $t \rightarrow 0$, we have

$$\begin{aligned}
Lf(x) &= \lim_{t \rightarrow 0} \frac{\mathbb{E}[f(x(t))] - f(x)}{t} = \lim_{t \rightarrow 0} \frac{\mathbb{E} \int_0^t (\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}) du}{t} \\
&= \mathbb{E} \lim_{t \rightarrow 0} \frac{\int_0^t (\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}) du}{t} \\
&= \mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}. \tag{A.5}
\end{aligned}$$

Now consider the function of the form $v(t, x) = \mathbb{E}[f(x(t)) | x(0) = x] = \mathbb{E}^x[f(x(t))]$.

Applying the operator L to $v(t, x)$, then

$$\begin{aligned}
Lv(t, x) &= \lim_{h \rightarrow 0} \frac{\mathbb{E}^x[v(t, x(h))] - v(t, x)}{h} \\
&= \lim_{h \rightarrow 0} \frac{\mathbb{E}^x[\mathbb{E}^{x(h)}[f(x(t))] - v(t, x)]}{h} \\
&= \lim_{h \rightarrow 0} \frac{\mathbb{E}^x[\mathbb{E}^x[f(x(t+h)) | \mathbb{F}_h] - v(t, x)]}{h} \\
&= \lim_{h \rightarrow 0} \frac{\mathbb{E}^x[f(x(t+h))] - v(t, x)}{h} \\
&= \lim_{h \rightarrow 0} \frac{v(t+h, x) - v(t, x)}{h} = \frac{\partial v}{\partial t}. \tag{A.6}
\end{aligned}$$

Thus, the evolution of $v(t, x)$ can be represented by the following PDE.

$$\frac{\partial v}{\partial t} = \mu \frac{\partial v}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 v}{\partial x^2} \tag{A.7a}$$

$$v(0, x) = f(x). \tag{A.7b}$$

Equation (A.7) is called the Chapman-Kolmogorov equation. More details can be found in [39]. The solution of the Chapman-Kolmogorov equation can be written in the semi-group notation as follows

$$v(t, x) = e^{Lt} f(x). \tag{A.8}$$

Another equation that is related to the Chapman-Kolmogorov equation is called the Fokker-Planck equation. It represents the evolution of the probability density function of $x(t)$. To obtain the Fokker-Planck equation, we again consider the Ito's formula

(A.4). We consider the following equation

$$\begin{aligned}
\int_{\mathbb{R}} f(x) \frac{\partial \rho(t, x)}{\partial t} dx &= \lim_{h \rightarrow 0} \int_{\mathbb{R}} f(x) \frac{\rho(t+h, x) - \rho(t, x)}{h} dx \\
&= \lim_{h \rightarrow 0} \frac{\mathbb{E}[f(x(t+h))] - \mathbb{E}[f(x(t))]}{h} \\
&= \mathbb{E} \left[\lim_{h \rightarrow 0} \frac{\int_t^{t+h} (\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}) du}{h} \right] \\
&= \mathbb{E} \left[\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} \right] \\
&= \int_{\mathbb{R}} (\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}) \rho(t, x) dx. \tag{A.9}
\end{aligned}$$

Using the integration by part and the fact that $\lim_{x \rightarrow \pm\infty} \rho(t, x) = 0$ and $\lim_{x \rightarrow \pm\infty} \frac{\partial \rho(t, x)}{\partial x} = 0$. It can be shown that

$$\int_{\mathbb{R}} (\mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}) \rho(t, x) dx = \int_{\mathbb{R}} f(x) (-\frac{\partial(\mu\rho)}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2\rho)}{\partial x^2}) dx. \tag{A.10}$$

Equation (A.9) can be written as

$$\int_{\mathbb{R}} f(x) \frac{\partial \rho(t, x)}{\partial t} dx = \int_{\mathbb{R}} f(x) (-\frac{\partial(\mu\rho)}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2\rho)}{\partial x^2}) dx. \tag{A.11}$$

The above equation is true for any $f(x)$. Thus,

$$\frac{\partial \rho(t, x)}{\partial t} = -\frac{\partial(\mu\rho(t, x))}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2\rho(t, x))}{\partial x^2}. \tag{A.12}$$

Therefore, the probability density function of $x(t)$ can be solved from

$$\frac{\partial \rho(t, x)}{\partial t} = -\frac{\partial(\mu\rho(t, x))}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2\rho(t, x))}{\partial x^2} \tag{A.13a}$$

$$\rho(0, x) = \rho_0(x). \tag{A.13b}$$

Equation (A.13) is called Fokker-Planck equation. Similarly, the solution $\rho(t, x)$ can be written in semi-group notation as $\rho(t, x) = e^{L^*t} \rho_0(x)$ where

$$L^* \rho = \frac{\partial(\mu\rho)}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2\rho)}{\partial x^2}. \tag{A.14}$$

L^* is called the adjoint operator of L and satisfies $\langle v, L^* \rho \rangle = \langle Lv, \rho \rangle$ where $\langle f, g \rangle = \int_{\mathbb{R}} f(x)g(x)dx$. Another important relationship between v and ρ is

$$\int_{\mathbb{R}} e^{Lt} f(x) \rho_0(x) dx = \int_{\mathbb{R}} e^{L^*t} \rho_0(x) f(x) dx. \tag{A.15}$$

If $x(t)$ is the ergodic process, then the invariant density function which is defined as $\rho_\infty(x) = \lim_{t \rightarrow \infty} e^{L^*t} \rho_0(x)$ exists and satisfies $L^* \rho_\infty = 0$.

A.2 Asymptotic Mode Elimination

Let us consider the system of $m + n$ variables

$$dx = f(x, y)dt + a(x, y)dB \tag{A.16a}$$

$$dy = g(x, y)dt + b(x, y)dW, \tag{A.16b}$$

where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. B and W are independent multivariate Wiener processes with dimensions d_1 and d_2 , respectively. The functions $f(x, y)$ and $g(x, y)$ are real-valued vector functions. We also assume that $a(x, y) \in \mathbb{R}^{m \times d_1}$ and $b(x, y) \in \mathbb{R}^{n \times d_2}$. We also assume that x and y have a strong scale separation where x is in a slow regime and y is in a fast regime. The main objective of this section is to obtain the reduced equation for x . Particularly, we seek the reduced equation for x of the following form

$$dX = F(X)dt + A(X)dB. \tag{A.17}$$

By assuming scale separation between x and y , the asymptotic approach can be used to obtain the reduced equation. Two methods of the asymptotic approach are described in this section; the averaging method and the homogenization method. The appropriate method is determined from the structure of the full system (A.16).

A.2.1 Averaging Method

For the averaging method, the scale separation between x and y is defined by the parameter ϵ . We assume $\epsilon \ll 1$. We introduce ϵ to the full system as follows

$$dx = f(x, y)dt + a(x, y)dB \tag{A.18a}$$

$$dy = \frac{1}{\epsilon}g(x, y)dt + \frac{1}{\sqrt{\epsilon}}b(x, y)dW. \tag{A.18b}$$

The corresponding Chapman-Kolmogorov equation for the system (A.18) can be written as follows

$$\begin{aligned} \frac{\partial u}{\partial t} = & f(x, y) \cdot \nabla_x u + \frac{1}{2} \text{Trace}(a(x, y)a^T(x, y)\nabla_x(\nabla_x u)) + \\ & \frac{1}{\epsilon}(g(x, y) \cdot \nabla_y u + \frac{1}{2} \text{Trace}(b(x, y)b^T(x, y)\nabla_y(\nabla_y u)), \end{aligned} \quad (\text{A.19})$$

where $\nabla_x u$ denotes the gradient of u with respect to x , i.e., $\nabla_x u = [\frac{\partial u}{\partial x_1} \dots \frac{\partial u}{\partial x_m}]^T$, and $\nabla_x(\nabla_x u)$ denotes the second order partial derivative of u , i.e., $(\nabla_x(\nabla_x u))_{i,j} = \frac{\partial^2 u}{\partial x_i \partial x_j}$.

Define the operator L_0 and L_1 as

$$\begin{aligned} L_0 &= f(x, y) \cdot \nabla_x + \frac{1}{2} \text{Trace}(a(x, y)a^T(x, y)\nabla_x(\nabla_x)) \\ L_1 &= g(x, y) \cdot \nabla_y + \frac{1}{2} \text{Trace}(b(x, y)b^T(x, y)\nabla_y(\nabla_y)). \end{aligned}$$

Then, (A.19) can be written in an operator notation as follows

$$\frac{\partial u}{\partial t} = L_0 u + \frac{1}{\epsilon} L_1 u. \quad (\text{A.21})$$

We assume that u has the following expansion

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \dots \quad (\text{A.22})$$

Substituting (A.22) into (A.21), we have

$$\frac{\partial u_0}{\partial t} + \epsilon \frac{\partial u_1}{\partial t} + \dots = \frac{1}{\epsilon} L_1 u_0 + (L_0 u_0 + L_1 u_1) + \epsilon(L_0 u_1 + L_1 u_2) + \dots \quad (\text{A.23})$$

Comparing the coefficients on the scale ϵ^{-1} and ϵ^0 , we have

$$L_1 u_0 = 0 \quad (\text{A.24a})$$

$$\frac{\partial u_0}{\partial t} = L_0 u_0 + L_1 u_1. \quad (\text{A.24b})$$

Since L_1 consists of the derivatives with respect to y . Equation (A.24a) implies that u_0 is a constant function with respect to y . Thus, $u_0 \equiv u_0(t, x)$. Let $\rho(y; x)$ be the invariant density function related to the operator L_1 , i.e, $\rho(y; x)$ satisfies $L_1^* \rho = 0$,

where L_1^* is the adjoint operator of L_1 . Here y is treated as a variable and x is treated as a constant. Define the averaging operator \mathbb{P}_y as follows

$$\mathbb{P}_y u = \int_{\mathbb{R}^n} u \rho(y; x) dy. \quad (\text{A.25})$$

Applying the averaging operator \mathbb{P}_y to (A.24b), we have

$$\mathbb{P}_y \frac{\partial u_0}{\partial t} = \mathbb{P}_y L_0 u_0 + \mathbb{P}_y L_1 u_1. \quad (\text{A.26})$$

Since

$$\mathbb{P}_y \frac{\partial u_0}{\partial t} = \int_{\mathbb{R}^n} \frac{\partial u_0}{\partial t} \rho(y; x) dy = \frac{\partial u_0}{\partial t} \int_{\mathbb{R}^n} \rho(y; x) dy = \frac{\partial u_0}{\partial t}$$

and

$$\mathbb{P}_y L_1 u_1 = \int_{\mathbb{R}^n} L_1 u_1 \rho(y; x) dy = \int_{\mathbb{R}^n} u_1 (L_1^* \rho) dy = 0,$$

(A.26) can be rewritten as

$$\frac{\partial u_0}{\partial t} = \mathbb{P}_y L_0 u_0 = \mathbb{P}_y f(x, y) \cdot \nabla_x u_0 + \frac{1}{2} \mathbb{P}_y \text{Trace}(a(x, y) a^T(x, y) \nabla_x (\nabla_x u_0)). \quad (\text{A.27})$$

Equation (A.27) is the Chapman-Kolmogorov equation that represents the dynamics of x . In an infinite scale separation, $u \rightarrow u_0$ as $\epsilon \rightarrow 0$. Thus, the reduced equation can be obtained from (A.27)

Example

Consider the following system

$$dx = -y dt + dB_1 \quad (\text{A.28})$$

$$dy = -\alpha(y - x) dt + \sigma dB_2, \quad (\text{A.29})$$

where x and y are one dimensional stochastic processes. B_1 and B_2 are independent Wiener processes. We assume x and y have a strong scale separation. We introduce ϵ to the equation as follows

$$dx = -y dt + dB_1 \quad (\text{A.30})$$

$$dy = -\frac{1}{\epsilon} \alpha(y - x) dt + \frac{1}{\sqrt{\epsilon}} \sigma dB_2. \quad (\text{A.31})$$

The corresponding Chapman-Komolgorov equation can be written as follows

$$\frac{\partial u}{\partial t} = L_0 u + \frac{1}{\epsilon} L_1 u, \quad (\text{A.32})$$

where

$$L_0 u = -y \frac{\partial u}{\partial x} + \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \quad (\text{A.33})$$

$$L_1 u = -\alpha(y-x) \frac{\partial u}{\partial y} + \frac{1}{2} \sigma^2 \frac{\partial^2 u}{\partial y^2}. \quad (\text{A.34})$$

Using the expansion, $u = u_0 + \epsilon u_1 + \dots$ and substituting u into (A.32), equation (A.27) for this example can be written as follows

$$\frac{\partial u_0}{\partial t} = \mathbb{P}_y L_0 u_0 = \mathbb{P}_y \left(-y \frac{\partial u_0}{\partial x} + \frac{1}{2} \frac{\partial^2 u_0}{\partial x^2} \right), \quad (\text{A.35})$$

where $\mathbb{P}_y g(y, x) = \int_{\mathbb{R}} g(y, x) \rho(y; x) dy$ and $\rho(y; x)$ is the invariant density function

$$\rho(y; x) = \sqrt{\frac{\alpha}{\pi \sigma^2}} e^{-\frac{\alpha(y-x)^2}{\sigma^2}}. \quad (\text{A.36})$$

Equation (A.35) can be simplified as follows

$$\frac{\partial u_0}{\partial t} = - \int_{\mathbb{R}} y \rho(y; x) dy \frac{\partial u_0}{\partial x} + \frac{1}{2} \frac{\partial^2 u_0}{\partial x^2} = -x \frac{\partial u_0}{\partial x} + \frac{1}{2} \frac{\partial^2 u_0}{\partial x^2}. \quad (\text{A.37})$$

Equation (A.37) has a corresponding stochastic differential equation

$$dX = -X dt + dW. \quad (\text{A.38})$$

Equation (A.38) is, thus, the reduced equation for (A.28).

A.2.2 Homogenization Method

The averaging method dose not apply if $\mathbb{P}_y L_0 u_0 = 0$ in (A.27). In this case, the time of fast variables has to be rescaled to the order of ϵ^{-2} , the epsilon form of (A.16) is given as follows

$$dx = f_0(x, y) dt + \frac{1}{\epsilon} f_1(x, y) dt + a(x, y) dB \quad (\text{A.39a})$$

$$dy = \frac{1}{\epsilon^2} g(x, y) dt + \frac{1}{\epsilon} b(x, y) dW, \quad (\text{A.39b})$$

where we decompose $f(x, y)$ as $f(x, y) = f_0(x, y) + f_1(x, y)$.

The Chapman-Kolmogorov equation of (A.39) can be written as follows

$$\begin{aligned} \frac{\partial u}{\partial t} &= f_0(x) \cdot \nabla_x u + \frac{1}{2} \text{Trace}(a(x, y) a^T(x, y) \nabla_x (\nabla_x u)) + \frac{1}{\epsilon} f_1(x, y) \cdot \nabla_x u + \\ &\quad \frac{1}{\epsilon^2} (g(x, y) \cdot \nabla_y u + \frac{1}{2} \text{Trace}(b(x, y) b^T(x, y) \nabla_y (\nabla_y u))). \end{aligned} \quad (\text{A.40})$$

Define the operators L_0 , L_1 , and L_2 as

$$\begin{aligned} L_0 &= f_0(x, y) \cdot \nabla_x + \frac{1}{2} \text{Trace}(a(x, y) a(x, y)^T \nabla_x (\nabla_x)) \\ L_1 &= f_1(x, y) \cdot \nabla_x \\ L_2 &= g(x, y) \cdot \nabla_y + \frac{1}{2} \text{Trace}(b(x, y) b^T(x, y) \nabla_y (\nabla_y)). \end{aligned}$$

Then (A.40) can be rewritten as

$$\frac{\partial u}{\partial t} = L_0 u + \frac{1}{\epsilon} L_1 u + \frac{1}{\epsilon^2} L_2 u. \quad (\text{A.42})$$

Expanding u as (A.22) and substituting it into the above equation, we have

$$\frac{\partial u_0}{\partial t} + \epsilon \frac{\partial u_1}{\partial t} + \dots = \frac{1}{\epsilon^2} L_2 u_0 + \frac{1}{\epsilon} (L_1 u_0 + L_2 u_1) + (L_0 u_0 + L_1 u_1 + L_2 u_2) + \dots \quad (\text{A.43})$$

Comparing the coefficients on the scale ϵ^{-2} , ϵ^{-1} and ϵ^0 , we have

$$L_2 u_0 = 0 \quad (\text{A.44a})$$

$$L_1 u_0 + L_2 u_1 = 0 \quad (\text{A.44b})$$

$$\frac{\partial u_0}{\partial t} = L_0 u_0 + L_1 u_1 + L_2 u_2. \quad (\text{A.44c})$$

Equation (A.44a) implies that u_0 is in the null space of L_2 . Thus, u_0 is a constant function with respect to y , i.e., $u_0 \equiv u_0(t, x)$. Let $\rho(y; x)$ be the invariant density function related to the operator L_2 , i.e., $\rho(y; x)$ satisfies $L_2^* \rho = 0$, where L_2^* is the adjoint operator of L_2 . Define the averaging operator \mathbb{P}_y as

$$\mathbb{P}_y u = \int_{\mathbb{R}^n} u \rho(y; x) dy. \quad (\text{A.45})$$

Applying the averaging operator, \mathbb{P}_y to (A.44b), we have

$$\mathbb{P}_y L_1 u_0 + \mathbb{P}_y L_2 u_1 = \mathbb{P}_y L_1 u_0 = 0. \quad (\text{A.46})$$

Equation (A.46) is the condition for applying the homogenization method. In other word, one can apply the homogenization method if

$$\int_{\mathbb{R}^n} f_1(x, y) \rho(y; x) dy = 0. \quad (\text{A.47})$$

From (A.44b), u_1 can be rewritten as $u_1 = -L_2^{-1} L_1 u_0$. Substituting $u_1 = -L_2^{-1} L_1 u_0$ into (A.44c) and applying the averaging operator, we have

$$\mathbb{P}_y \frac{\partial u_0}{\partial t} = \mathbb{P}_y L_0 u_0 + \mathbb{P}_y [-L_1 L_2^{-1} L_1 u_0] + \mathbb{P}_y L_2 u_2. \quad (\text{A.48})$$

Since $\mathbb{P}_y \frac{\partial u_0}{\partial t} = \frac{\partial u_0}{\partial t}$ and $\mathbb{P}_y L_2 u_2 = 0$, the above equation can be simplified as

$$\frac{\partial u_0}{\partial t} = \mathbb{P}_y L_0 u_0 + \mathbb{P}_y [-L_1 L_2^{-1} L_1 u_0]. \quad (\text{A.49})$$

To interpret the inverse operator L_2^{-1} , we consider the following problem

$$-L_2^{-1} f = g, \quad \text{and} \quad \int f \rho(y; x) dy = 0. \quad (\text{A.50})$$

If we let $g = \int_0^\infty e^{L_2 t} f dt$. Then

$$\begin{aligned} -L_2 g &= - \int_0^\infty L_2 e^{L_2 t} f dt \\ &= - \int_0^\infty \frac{\partial}{\partial t} e^{L_2 t} f dt \\ &= f - \lim_{t \rightarrow \infty} e^{L_2 t} f = f - \int f \rho(y; x) dy = f. \end{aligned} \quad (\text{A.51})$$

Thus,

$$-L_2^{-1} f = \int_0^\infty e^{L_2 t} f dt. \quad (\text{A.52})$$

The term $\mathbb{P}_y [-L_1 L_2^{-1} L_1 u_0]$ in (A.49) can be computed as

$$\begin{aligned} \mathbb{P}_y [-L_1 L_2^{-1} L_1 u_0] &= \int_{\mathbb{R}^n} \rho(y; x) \left(\int_0^\infty (\mathbb{E}[\text{Trace}(f_1(x, y) f_1(x, y(t))^T \nabla_x (\nabla_x u_0)]) + \right. \\ &\quad \left. \mathbb{E}[f_1(x, y) \nabla_x f_1(x, y(t))^T] \cdot \nabla_x u_0) dt \right) dy \\ &= \int_0^\infty \left(\int_{\mathbb{R}^n} \rho(y; x) \mathbb{E}[\text{Trace}(f_1(x, y) f_1(x, y(t))^T \nabla_x (\nabla_x u_0))] dy + \right. \\ &\quad \left. \int_{\mathbb{R}^n} \rho(y; x) \mathbb{E}[f_1(x, y) \nabla_x f_1(x, y(t))^T] \cdot \nabla_x u_0 dy \right) dt. \end{aligned} \quad (\text{A.53})$$

The expectation and the averaging under the invariant density function in (A.53) is, in fact, redundant. Thus, equation (A.53) can be rewritten as follows

$$\begin{aligned} \mathbb{P}_y [-L_1 L_2^{-1} L_1 u_0] &= \int_0^\infty (\mathbb{E}[\text{Trace}(f_1(x, y) f_1(x, y(t))^T \nabla_x (\nabla_x u_0))] + \\ &\quad \mathbb{E}[f_1(x, y) \nabla_x f_1(x, y(t))^T] \cdot \nabla_x u_0) dt. \end{aligned} \quad (\text{A.54})$$

Equations (A.49) and (A.54) together represent the Chapman-Kolmogorov equation related to x variable. Thus, the reduced equation for x can be obtained from (A.49) and (A.54). We give an example of the homogenization method here.

Example

Consider the following system

$$dx = -xdt + ydt \quad (\text{A.55})$$

$$dy = -\alpha ydt + \sigma dB, \quad (\text{A.56})$$

where x and y are one dimensional stochastic processes, and B is a one dimensional Brownian motion. We assume x and y have a strong scale separation. We introduce ϵ to the equation as follows

$$dx = -xdt + \frac{1}{\epsilon} ydt \quad (\text{A.57})$$

$$dy = -\frac{1}{\epsilon^2} \alpha ydt + \frac{1}{\epsilon} \sigma dB. \quad (\text{A.58})$$

The corresponding Chapman-Kolmogorov equation can be written as follows

$$\frac{\partial u}{\partial t} = L_0 u + \frac{1}{\epsilon} L_1 u + \frac{1}{\epsilon^2} L_2 u \quad (\text{A.59})$$

where

$$L_0 u = -x \frac{\partial u}{\partial x} \quad (\text{A.60})$$

$$L_1 u = y \frac{\partial u}{\partial x} \quad (\text{A.61})$$

$$L_2 u = -\alpha y \frac{\partial u}{\partial y} + \frac{1}{2} \sigma^2 \frac{\partial^2 u}{\partial y^2}. \quad (\text{A.62})$$

Using the expansion, $u = u_0 + \epsilon u_1 + \dots$ and substituting u into (A.59), we have

$$\begin{aligned} \frac{\partial u_0}{\partial t} &= \mathbb{P}_y L_0 u_0 + \mathbb{P}_y [-L_1 L_2^{-1} L_1 u_0] \\ &= \mathbb{P}_y \left(-x \frac{\partial u_0}{\partial x} \right) + \left(\int_0^\infty \mathbb{E}[y(0)y(t)] dt \right) \frac{\partial^2 u_0}{\partial x^2}, \end{aligned} \quad (\text{A.63})$$

where the invariant density function related to L_2 in this case is

$$\rho(y) = \sqrt{\frac{\alpha}{\pi \sigma^2}} e^{-\frac{\alpha y^2}{\sigma^2}}. \quad (\text{A.64})$$

We note that

$$\int_0^\infty \mathbb{E}[y(0)y(t)] dt = \int_0^\infty \frac{\sigma^2}{2\alpha} e^{-\alpha t} dt = \frac{\sigma^2}{2\alpha^2}. \quad (\text{A.65})$$

Thus, the Chapman-Kolmogorov equation for u_0 is written as follows

$$\frac{\partial u_0}{\partial t} = -x \frac{\partial u_0}{\partial x} + \frac{\sigma^2}{2\alpha^2} \frac{\partial^2 u_0}{\partial x^2} \quad (\text{A.66})$$

and its corresponding stochastic differential equation is written as follows

$$dX = -X dt + \frac{\sigma}{\alpha} dW. \quad (\text{A.67})$$

Equation (A.67) is the reduced equation for (A.55). Readers interested in asymptotic mode elimination can find more details in [36, 37, 38].

Appendix B

Least Square Method

In this appendix, we assume that x and y are vector-valued variables where $x \in \mathbb{R}^{n \times 1}$ and $y \in \mathbb{R}^{m \times 1}$. We also assume that the relationship between x and y is

$$y = A^T x \quad (\text{B.1})$$

where $A \in \mathbb{R}^{n \times m}$ and A^T denotes the transpose of A . Let X be a time series of x with length N , i.e., $X \in \mathbb{R}^{n \times N}$. Similarly, we let Y denote a time series of y with length N , i.e., $Y \in \mathbb{R}^{m \times N}$. For the ease of notation, X_j denotes a row vector of X at row j and Y_i denotes a row vector of Y at row i . To estimate A from time series X and Y , we minimize the following error function

$$E = \frac{1}{2} \sum_{k=1}^N \sum_{i=1}^m \left(\sum_{j=1}^n A_{ji} X_{jk} - Y_{ik} \right)^2 = \frac{1}{2} \sum_{i=1}^m \sum_{k=1}^N \left(\sum_{j=1}^n A_{ji} X_{jk} - Y_{ik} \right)^2. \quad (\text{B.2})$$

There are mn coefficients to estimate for this case. By taking the derivative with respect to A_{ts} , $1 \leq t \leq n$, $1 \leq s \leq m$, and setting the derivative to zero, we have

$$\frac{\partial E}{\partial A_{ts}} = \sum_{k=1}^N \left(\sum_{j=1}^n A_{js} X_{jk} - Y_{sk} \right) X_{tk} = 0. \quad (\text{B.3})$$

Equation (B.3) can be rewritten as follows

$$\sum_{j=1}^n A_{js} (X_j X_t^T) = Y_s X_t^T. \quad (\text{B.4})$$

Equation (B.4) is true for $1 \leq t \leq n$, $1 \leq s \leq m$. In fact, (B.4) can be written in a matrix notation as follows

$$A^T (X X^T) = (Y X^T) \tag{B.5}$$

and the solution of (B.5) is

$$A = (X X^T)^{-1} (Y X^T). \tag{B.6}$$

Bibliography

- [1] A. J. MAJDA AND I. TIMOFEYEV, *Remarkable statistical behavior for truncated Burgers-Hopf dynamics*, Proc. Nat. Acad. Sci. USA, 97 (2000), pp. 12413–12417.
- [2] A. MAJDA, AND I. TIMOFEYEV, *Statistical mechanics for truncations of the Burgers-Hopf equation: a model for intrinsic stochastic behaviour with scaling*, Milan Journal of Mathematics, 70 (2002), pp. 39–96.
- [3] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *A mathematical framework for stochastic climate models*, Comm. Pure Appl. Math., 54 (2001), pp. 891–974.
- [4] A. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *Stochastic models for selected slow variables in large deterministic systems*, Nonlinearity, 19 (2006), pp. 769–794.
- [5] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *Systematic strategies for stochastic mode reduction in climate*, Journal of the Atmospheric Sciences, 60 (2003), p. 1705–1722.
- [6] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *Models for stochastic climate prediction*, Proc. Natl. Acad. Sci. USA, 96 (1999), p. 14687–14691.

- [7] R. ABRAMOV, G. KOVACIC, AND A. J. MAJDA, *Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation*, Comm. Pure Appl. Math., 56 (2003), pp. 1–46.
- [8] D. BERNSTEIN, *Optimal prediction of Burgers’s equation*, Multiscale Model. Simul., 6 (2007), pp. 27–52.
- [9] A. J. CHORIN, *Conditional expectations and renormalization*, Multiscale Model. Simul., 1 (2003), pp. 105–118.
- [10] A. J. CHORIN, O. H. HALD, AND R. KUPFERMAN, *Optimal prediction and the Mori-Zwanzig representation of irreversible processes*, Proc. Nat. Acad. Sci. USA, 97 (2000), pp. 2968–2973.
- [11] A. J. CHORIN, O. H. HALD, AND R. KUPFERMAN, *Optimal prediction with memory*, Physica D, 166 (2002), pp. 239–257.
- [12] A. J. CHORIN, A. P. KAST, AND R. KUPFERMAN, *Optimal prediction of underresolved dynamics*, Proc. Nat. Acad. Sci. USA, 95 (1998), pp. 4094–4098.
- [13] A. J. CHORIN, A. P. KAST, AND R. KUPFERMAN, *Unresolved computation and optimal prediction*, Comm. Pure Appl. Math., 52 (1999), pp. 1231–1254.
- [14] G. ARIEL, B. ENGQUIST, AND R. TSAI, *Numerical multiscale methods for coupled oscillators*, Multiscale Model. Simul., 7 (2008), pp. 1387–1404.
- [15] S. KRAVTSOV, D. KONDRASHOV, AND M. GHIL, *Multilevel regression modeling of nonlinear processes: derivation and applications to climatic variability*, Journal of Climate, 18 (2005), pp. 4404–4424.
- [16] D. T. CROMMELIN, AND E. VANDEN-EIJNDEN, *Fitting timeseries by continuous-time Markov chains: a quadratic programming approach*, Journal of Computational Physics, 217 (2006), pp. 782–805.

- [17] D. T. CROMMELIN, AND E. VANDEN-EIJNDEN, *Reconstruction of diffusions using spectral data from timeseries*, Comm. Math. Sci., 4 (2006), pp. 651–668.
- [18] D. CROMMELIN AND E. VANDEN-EIJNDEN, *Subgrid-scale parameterization with conditional Markov chains*, J. Atmos. Sci., 65 (2008), pp. 2661–2675.
- [19] T. D. FRANK, AND P. J. BEEK, *Stationary solutions of linear stochastic delay differential equations: applications to biological systems*, Physical Review E, 64 (2001), pp. 021917-1–021917-12.
- [20] U. KÜCHLER, AND E. PLATEN, *Strong discrete time approximation of stochastic differential equations with time delay*, Mathematics and Computers in Simulation, 54 (2000), pp. 189–205.
- [21] P. E. KLOEDEN, AND E. PLATEN, *Numerical solution of stochastic differential equations*, Springer-Verlag, 2000.
- [22] Y. CAO, D. T. GILLESPIE, AND L. R. PETZOLD, *The slow-scale stochastic simulation algorithm*, Journal of Chemical Physics, 122 (2005), pp. 014116-1–014116-18.
- [23] W. E, D. LIU, AND E. VANDEN-EIJNDEN, *Nested stochastic simulation algorithms for chemical kinetic systems with disparate rates*, J. Chem. Phys., 123 (2005), p. 194107.
- [24] K. TAKAHASHI, K. KAIZU, B. HU, AND M. TOMITA, *A multi-algorithm, multi-timescale method for cell simulation*, Bioinformatics, 20 (2004), p. 538–546.
- [25] E. DARVE, J. SOLOMON, AND A. KIA, *Computing generalized Langevin equations and generalized Fokker-Planck equations*, Proc. Natl. Acad. Sci., submitted.

- [26] C. FRANZKE, D. CROMMELIN, A. FISCHER, AND A. MAJDA, *A Hidden Markov Model perspective on regimes and metastability in atmospheric flows*. J. Climate, submitted, 2007.
- [27] X. SAN LIANG AND R. KLEEMAN, *A rigorous formalism of information transfer between dynamical system components. II. Continuous flow*, Physica D, 227(2) (2007), pp. 173–182.
- [28] E. VANDEN-EIJNDEN E W., D. LIU, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.
- [29] W. E AND B. ENGQUIST, *The heterogeneous multiscale methods*, Comm. Math. Sci., 1(1) (2003), pp. 87–132.
- [30] W. E, B. ENGQUIST, X. LI, W. REN, AND E. VANDEN-EIJNDEN, *Heterogeneous multiscale methods: A review*, Comm. Comput. Phys., 2(3) (2007), pp. 367–450.
- [31] C. HIJON, P. ESPANOL, E. VANDEN-EIJNDEN, AND R. DELGADO-BUSCALIONI, *Mori-Zwanzig formalism as a practical computational tool*, Preprint, 2009.
- [32] I. HORENKO AND C. SCHÜTTE, *Likelihood-based estimation of multidimensional Langevin models and its application to biomolecular dynamics*, Multiscale Model. Simul., 7 (2008), pp. 731–773.
- [33] M.A. KATSOULAKIS, A.J. MAJDA, AND A. SOPSAKIS, *Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 1, deterministic closures*, Comm. Math. Sci., 2 (2004), pp. 255–294.

- [34] M.A. KATSOULAKIS, A.J. MAJDA, AND A. SOPASAKIS, *Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 2, stochastic closures*, Comm. Math. Sci., 3 (2005), pp. 453–478.
- [35] M.A. KATSOULAKIS, A.J. MAJDA, AND A. SOPASAKIS, *Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems*, Nonlinearity, 19 (2006).
- [36] G. A. PAVLIOTIS AND A. STUART, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
- [37] D. GIVON, R. KUPFERMAN, AND A. STUART, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, 17 (2004), pp. R55–R127.
- [38] G. A. PAVLIOTIS AND A. STUART, *Multiscale methods: averaging and homogenization*, Springer 1st edition, 2008.
- [39] B. OKSENDAL, *Stochastic differential equations: an introduction with applications*, Springer 6th edition, 2007.
- [40] A. PAPOULIS, *Probability, random variables, and stochastic processes*, McGraw-Hill 3rd edition, 1991.
- [41] A. BELLEN AND M. ZENNARO, *Numerical methods for delay differential equations*, Oxford University Press 1st edition, 2003.
- [42] R. S. ELLIS AND M. A. PINSKY, *The first and second fluid approximations to the linearized Boltzmann equation*, J. Math. Pures Appl., 54 (1975), pp. 125–126.
- [43] R. Z. KHASHMINSKY, *Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion*, Theory Probab. Appl., 8 (1963), pp. 1–21.

- [44] T. G. KURTZ, *A limit theorem for perturbed operator semigroups with applications to random evolutions*, J. Functional Analysis, 12 (1973), pp. 55–67.
- [45] G. PAPANICOLAOU, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain J. Math., 6 (1976), pp. 653–673.