DATA-DRIVEN TECHNIQUES FOR ESTIMATION AND STOCHASTIC REDUCTION OF MULTI-SCALE 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

Ankita Jain

December 2012

DATA-DRIVEN TECHNIQUES FOR ESTIMATION AND STOCHASTIC REDUCTION OF MULTI-SCALE SYSTEMS

Ankita Jain

APPROVED:

Dr. Ilya Timofeyev (Chairman),
Department of Mathematics, University of Houston

Dr. Robert Azencott (Co-Chairman),
Department of Mathematics, University of Houston
Emeritus Professor, ENS, France

Dr. William Ott,
Department of Mathematics, University of Houston

Dr. Gregor Kovačič,
Department of Mathematics, RPI, Troy, New York

Dean, College of Natural Sciences and Mathematics

DATA-DRIVEN TECHNIQUES FOR ESTIMATION AND STOCHASTIC REDUCTION OF MULTI-SCALE 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

Ankita Jain

December 2012

Abstract

Parametric estimation of stochastic processes is one of the most widely used techniques for obtaining effective models, given a discrete dataset. Often, the experimental or observational datasets are not explicitly generated by the underlying stochastic model and are, thus, expected to agree with the model only in approximate statistical sense, also referred as Indirect Observability. Therefore, there may be an inherent difference between the data and the model, leading to the inconsistency of the parametric estimation.

The dissertation is presented in three parts. In first part (chapter 2), the goal is to develop an efficient and accurate parametric estimation procedure for the reduced model (SDE for slow variables alone) when the given data are the time series of the slow variables in the multi-scale high-dimensional Lorenz-96 model. The first estimator considered for the reduced model is approximate Maximum Likelihood estimator, which is highly dependent on the subsampling time-step of the given data. There is no feasible solution to compute optimal subsampling time-step for consistent approximate Maximum Likelihood estimator. Next, we look at the moment estimator which is weakly dependent on subsampling time-step of the given data, but is valid only if the mean of the slow variables in the full model is relatively large. Both estimators give acceptable values for the parameters in reduced model. Given the situation of non-linear multi-scale model to be reduced to stochastic model of slow variables alone, moment estimator is preferred if the mean of the slow variables is relatively large and the reduced model is not very sensitive to the change in parameters.

The second part is chapter 3, in which we consider the multi-scale model having energy conserving fast subsystem and the stochastic terms been added to the equation of slow variables alone. In such situations, the energy of the fast variables changes with time due to the coupling between the slow and fast dynamics and hence, considered as an additional hidden slow variable. We develop a stochastic mode reduction technique to derive an

efficient stochastic model for the original slow variables in full model and additional slow variable given by energy of the fast subsystem.

In the third part (chapter 4), similar parametric estimation procedure is studied under Indirect Observability as done for Lorenz-96 model in chapter 2 but here we consider multi-scale fast-oscillating potential model. We get linear reduced model which simplifies certain analytical computations and we specify explicit conditions for which the estimators of the reduced model are consistent under Indirect Observability. Another important aspect discussed in the chapter is estimation of an effective model from a dataset generated with a fixed but unknown value of the scale separation parameter ϵ.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction and Motivation

Stochastic modeling of discrete dynamic data has been an active area of research for many decades. There are many applications where it is desirable to fit a reduced stochastic description (e.g. an SDE) to data. Some of these applications include molecular dynamics [28, 45], atmosphere/ocean science [12, 26, 37], and econometrics [14, 74]. Parametric estimation of stochastic processes is one of the most widely used techniques for obtaining effective models. In such situations, the main objective is to estimate the model parameters to best fit the given observations. This is one of the main issues of the dissertation. Often, the experimental or observational datasets are not explicitly generated by the underlying stochastic model and are, thus, expected to agree with the model only in approximate statistical sense. Since, the data to be fitted are only an approximation to the model, we refer to such situations as Indirect Observability. The main concern is that the parametric estimation of stochastic models under Indirect Observability framework may not be robust.

In such scenarios, one must show caution in applying standard statistical techniques to obtain estimates for the model parameters such as Maximum Likelihood estimation.

The multi-scale nature of complex dynamics is common in applications of contemporary science, such as geophysical science and climate change prediction [15, 37, 40, 64]. Multi-scale systems are typically characterized by the time and space scale separation of patterns of motion, with (typically) fewer slowly evolving variables and much larger set of faster evolving variables. This time-space scale separation causes direct numerical simulation of the evolution of the dynamics to be computationally expensive, due both to the large number of variables and the necessity to choose a small discretization time-step in order to resolve the fast components of the dynamics.

In the climate change science the situation is further complicated by the fact that the climate is characterized by the long-term statistics of the slow variables, but the slow variables are influenced by small changes of physical parameters (such as the solar radiation forcing, greenhouse gas content, etc.). Therefore, climate change occurs over even longer time-scale than the dynamics of the slow variables themselves. In this situation, where long-term statistics of the slow motion patterns need to be captured, the direct forward time integration of the most comprehensive global circulation models (GCM) is subject to enormous computational expense.

It has long been recognized that a closed simplified model for the slow variables is a more computationally feasible alternative to a direct forward time integration of a complete multi-scale model. One could use this simplified low-dimensional model to efficiently simulate the long-term statistics of the slow variables. One example is the field of molecular dynamics, where it is desirable to find the effective models for low-dimensional phenomena (such as conformational dynamics, vacancy diffusion and so forth) which are embedded within higher dimensional time series. Another example is the ocean-atmosphere sciences where it

is desirable to find effective models for large-scale structures, while representing the small-scales stochastically. The multi-scale structure of the data in these problems renders the problem of parameter estimation very subtle and great care has to be taken in order to estimate the coefficients correctly.

In the last few decades the importance of multi-scale effects has been particularly emphasized in many applications. In this context, reduced modeling of complex PDEs has been one of the main motivations behind many techniques. For instance, the stochastic mode reduction technique (same as homogenization) based on the earlier works [35, 50, 51, 52, 53, 66] has succesfully modeled the dynamics of large-scale structures in system with time-scale separation [24, 36, 58, 59, 60, 61]; stochastic mode-reduction has been applied to the finite-difference discretizations of the Burgers equation [29, 30] to derive an effective equation for local spatial averages; an optimal prediction framework has enabled coarse-grained dynamic modeling of statistical descriptors [18, 19, 21]; microscopic spin-flip process have provided coarse-grained models of car and pedestrian traffic flow [7, 16, 17, 32, 41, 73]; reduced Markov chain models have been applied to prototype atmosphere-ocean interactions [24, 31, 63]; a framework for dimension reduction in metastable systems has been developed [42, 71]; and importance of multi-scale effects in data-driven stochastic modeling [39, 43, 44, 75] has been discussed.

1.1 Indirect Observability

The Indirect Observability context considered here is motivated primarily by stochastic modeling of large-scale structures in turbulent geophysical partial differential equations. The climate system is assumed to be in statistical equilibrium and the reduced system is derived using averaging or homogenization with respect to the equilibrium measure of fast variables. This approach can be formalized by splitting all dynamic variables into

two sets - slow and fast variables, properly introducing an artificial small parameter, ϵ, into the climate system, and deriving the effective reduced stochastic differential equation for the slow variables in the limit as $\epsilon \to 0$. This technique has been applied to some prototype models [8, 60, 61] and more realistic atmospheric models of various complexity [25, 37, 38]. The main disadvantage of this technique is that coefficients in the reduced model are computed using the stationary statistics of fast variables and the long simulation of the full model are typically necessary to estimate this data. Alternatively, it is possible to estimate the coefficients of the reduced model from the time series of the slow variables themselves. This introduces a mismatch between the estimated model and data, since slow variables are only a subset of full slow-fast dynamics. For instance, time series of the slow variables alone are non-Markovian and multi-scale effects from the fast unresolved variables can lead to inconsistencies in reduced stochastic modeling of the large-scale structures in the atmosphere-ocean applications [12, 26].

In practical situations the full slow-fast system is not particularly close to the limiting stochastic dynamics. For atmospheric dynamics, the scale separation ϵ is generally considered to be in the range of $[0.3 \cdots 0.5]$, but it is impossible to estimate the precise value of ϵ since it does not enter the model explicitly. Nevertheless, it is desirable to obtain a closed form reduced model for the time-evolution of the large-scale structures. Therefore, the main practical goal of stochastic modeling under Indirect Observability is to develop efficient and accurate parametric estimation procedure for SDEs when the data are generated by a multi-scale model for which the scale separation ϵ has a finite, but unknown value. In [9, 10, 11], authors show the asymptotic results for the development of bias-corrected estimators accurately reproducing the statistical properties of observed or simulated dynamic data sets. In next subsection, we discuss the precise mathematical framework of Indirect Observability for the centered stationary continuous time processes, shown in [9, 10, 11].

1.1.1 Mathematical framework of Indirect Observability

The mathematical description of the Indirect Observability framework is a formal way to address an important case when the nature of the observed process is not known exactly or too complex to use in numerical/analytical calculations; instead, it is desirable to approximate this process by a suitable stochastic process X_t with matching statistical features. The observable complex process is denoted as Y_t^ϵ where ϵ is the scale separation parameter in multi-scale systems. To test the consistency of the estimation procedure we consider examples when Y_t^ϵ is such that $Y_t^\epsilon \to X_t$ as $\epsilon \to 0$ in some suitable sense. The limit process of interest is denoted as X_t, but it is not observed directly; instead the parameters of the stochastic model for X_t are inferred from the data subsampled from Y_t^ϵ . Here Y_t^ϵ is also referred to as an approximating process.

Consider the parameter vector to be estimated in underlying process X_t as θ which needs to be estimated using the observations of Y_t^ϵ. The limiting behavior of Y_t^ϵ can be derived explicitly via a well known homogenization procedure. The homogenization procedure is used to derive the explicit equations for the limiting process and the "true" values of parameter vector θ . These "true" values depends on the moments of the fast variables in full model. The concrete target is to efficiently use the data of approximating process Y_t^ϵ to generate consistent estimators of the unknown "underlying" parameter vector θ. Hence, the "true" values, derived using the homogenization procedure, are only used to test the behavior of the estimators as $\epsilon \to 0$. Estimator for the unknown parameters vector θ computed using the data of Y_t^ϵ is the function of ϵ, denoted by $\hat{\theta}_\epsilon$. Since X_t is the limit process of Y_t^ϵ as $\epsilon \to 0$, it is desirable that the parameter estimator $\hat{\theta}_\epsilon$ computed from discrete data of Y_t^ϵ converge to the "true" parameter value of θ as $\epsilon \to 0$.

The homogenization procedure provides us with an explicit equation for X_t and the

"true" values of parameters. As part of the comparison procedure, equation of X_t is simulated and compared with the observable process Y_t^ϵ as $\epsilon \to 0$. We observe that at short time-scales, the trajectories Y_t^ϵ of the observable physical process are quite different from sample paths X_t but on longer time-scales, the behavior of Y_t^ϵ is well matched by X_t. This situation is typical for data generated by a numerical dynamical model, such as fluid dynamics PDEs. For very short times, there is an inconsistency in de-correlation times between X_t and Y_t^ϵ. Therefore, if $\hat{\theta}_\epsilon$ depends on the subsampling time lag Δ (e.g., the Maximum Likelihood estimator), $\hat{\theta}_\epsilon$ can loose it's consistency if one simply substitutes the data generated by Y_t^ϵ instead of the data X_t, especially if the observations of Y_t^ϵ are too dense in time. Hence, there is a need to subsample the data at an optimal time-step Δ such that the estimator converges to true parameter as $\epsilon \to 0$. This study highlights the necessity to subsample at adequate rates when the observations are not generated by the underlying stochastic model whose parameters are being estimated.

Here, we consider estimators computed from discrete observational samples. Consider discretely subsampled data $U_n = Y^\epsilon_{n\Delta}, n = 0, 1, 2, ...N$, from an observable process Y_t^ϵ which has complex dynamics. Our goal is to determine explicit conditions on the number of discrete observations $N = N(\epsilon)$, and the uniform time-step $\Delta = \Delta(\epsilon)$ such that, as $\epsilon \to 0$, the estimator $\hat{\theta}_\epsilon$ based on the observations U_n subsampled from the approximating process Y_t^ϵ converges to the true parameter value θ.

Subsampling strategies become essential when the parameters of an SDE driving X_t must be estimated using discrete data extracted from a process Y_t^ϵ quite close to X_t for small ϵ, but having higher trajectory smoothness than X_t. Subsampling approaches have, for instance, been studied for the homogenization problem [67, 68, 9, 10, 11].

Previous work. In [9, 10, 11], authors characterize efficient subsampling strategies with a complete determination of the optimal subsampling rates. The main results are

presented via a prototype example in which the underlying process X_t is a stationary Ornstein-Uhlenbeck (OU) process with unknown drift and diffusion coefficients γ and σ^2, given by

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{1.1}$$

where W_t is the standard Brownian motion and the unknown parameters γ, σ are strictly positive. It is assumed that the only available observations are generated by another stationary Gaussian process Y_t^ϵ, indexed by a small parameter $\epsilon > 0$. One example of the approximating process Y_t^ϵ is the Smoothed Ornstein-Uhlenback process given by

$$Y_t^\epsilon = \frac{1}{\epsilon}\int_{t-\epsilon}^{t} X_s ds. \tag{1.2}$$

It can be shown that $Y_t^\epsilon \to X_t$ in L_2 as $\epsilon \to 0$, and in particular, the correlation function of Y_t^ϵ converges to the correlation function of X_t as $\epsilon \to 0$. The process X_t is not directly observable and the only available information is N number of observations extracted from Y_t^ϵ by subsampling with a time-step Δ.

The goal is to consistently estimate the drift and diffusion parameters γ and σ of the non-observable OU process X_t, using $N(\epsilon)$ observations extracted by subsampling at time-step $\Delta(\epsilon)$, of the approximating process Y_t^ϵ which tends to X_t in L_2 as $\epsilon \to 0$. Estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ are considered for parameters γ and σ^2, based on the second-order covariance estimators of the underlying process. These estimators are also shown asymptotically equivalent to the Maximum Likelihood estimators for the OU SDE. The main goal is to determine necessary and sufficient conditions on parametric estimation under Indirect Observability, i.e. conditions to ensure that estimators $(\hat{\gamma}, \hat{\sigma}^2) \to (\gamma, \sigma^2)$ as $\epsilon \to 0$ when estimators are computed using the observations of Y_t^ϵ subsampled at time-step Δ.

In [10], the authors developed the necessary and sufficient conditions for the consistency of the Maximum Likelihood estimators based on adequately subsampled approximate data

and also investigated the optimal speed of convergence. In particular, the consistency conditions for vanishing lags estimators are formulated as

$$N(\epsilon) \to \infty, \quad \Delta(\epsilon) \to 0, \quad N(\epsilon)\Delta(\epsilon) \to \infty, \quad \epsilon/\Delta(\epsilon) \to 0. \tag{1.3}$$

If, for simplicity, the power-law relationships between N, Δ and ϵ are considered, then, the consistency requirements for the subsampling procedure becomes

$$\Delta = \epsilon^{\alpha}, \quad \alpha \in (0,1), \quad N = \epsilon^{-\beta}, \quad \beta > \alpha. \tag{1.4}$$

Previous work: Modeling a given dataset generated at a fixed but unknown scale. In [9, 10, 11], authors also considered the situation of estimating an effective stochastic model for the large-scale structures from a single dataset generated by an approximate multi-scale model with an unknown fixed parameter ϵ. The algorithm is developed to construct a new bias-corrected estimator. This is an important point, addressing a practical question of stochastic modeling of large-scale structures in multi-scale high-dimensional systems. Typically, in such situations the time series of the large-scales are available from numerical simulations of the full model, but the exact value of the multi-scale parameter is unknown.

Consider the Maximum Likelihood estimators $\hat{\gamma}, \hat{\sigma}^2$ which depend on the behavior of lagged covariance of large-scale observations. The goal is to develop an approach for constructing the bias-corrected estimators when the data are generated from a trajectory $Y^{\epsilon}_{n\Delta}$ with a fixed, but unknown value of the multi-scale parameter ϵ. The bias-corrected estimator is constructed by analysis of the curve $\hat{\gamma}(\Delta)$ vs Δ and concluded that the slope estimator is an unbiased estimator for the damping parameter γ, i.e. the bias-corrected estimators can be computed by linear regression of $\hat{\gamma}(\Delta)$ Δ versus Δ. If we have non-linear reduced model, the curve $\hat{\gamma}(\Delta)$ vs Δ can be more complicated. We show the results of [9]

on multi-scale fast-oscillating potential model in chapter 4 for which we have OU process as the reduced model.

1.2 Outline of Dissertation

Chapter 2: Parameter Estimation of Lorenz-96 Model

Our goal is to test estimation under Indirect Observability on a more realistic multi-scale model related to fluid dynamics. In chapter 2, the results of Indirect Observability in [9, 10, 11] are extended to the complex multi-scale Lorenz-96 model having 18 slow variables and total 360 fast variables. We consider the non-linear two-scale L96 model having chaotic behavior, forcing, and dissipation with dynamics ranging from weakly to strongly chaotic, and fully turbulent depending on choice of value of force. The L96 model is more complicated in comparison to OU and triad model considered in [9, 10, 11]. The main difference in the L96 model is that the reduced model is not linear.

The numerical simulation of the L96 model can be computationally very expensive due both to the large number of variables and the necessity to choose a small discretization step in order to resolve the fast variables of the model. Hence, the computationally feasible solution for studying the statistical behavior of the slow variables is approximating the slow variables of the L96 model by an explicit SDE model. The main goal of chapter 2 is to develop an efficient and accurate parametric estimation procedure for the reduced model (SDE for slow variables alone) when the given data is the time series of the slow-variables in the full L96 model.

We introduce a small parameter $\epsilon > 0$ to emphasize multi-scale effects in the full model. The homogenization procedure is used to derive an explicit reduced equation and the "true" values of parameters. The "true" values depend on lagged covariances of fast variables in the L96 model. Due to the complexity of the equation for the fast variables in the L96 model,

we can not evaluate the "true" values of parameters analytically. Hence, the "true" values are computed using one long simulation of the fast subsystem of the L96 model. Recall that the main goal is to estimate parameters of the reduced model using observations of the slow variables in the full model. Hence, the "true" values for the reduced model are only used to test the behavior of the estimators as the scale separation parameter $\epsilon \to 0$.

The first estimator considered here is Maximum Likelihood estimator. Since the reduced model is the limit process of the slow variables in the L96 model as $\epsilon \to 0$, it is desirable that the Maximum Likelihood parameter estimates computed from discrete data of slow variables in L96 converge to the "true" parameter values as $\epsilon \to 0$. Similar to [9, 10, 11], we observe that the Maximum Likelihood estimators depend on auto-correlation of slow variables and hence can lose their consistency if we use the data of slow variables of the full model at a very dense time-step. To this end, there is a need to subsample the data at an optimal time-step Δ such that the MLEs converge to true parameters as $\epsilon \to 0$. In [9, 10, 11], for all the examples, the reduced model was linear which considerably simplified analytical calculations and allowed authors to develop method for obtaining optimal subsampling time-step for consistency of Maximum Likelihood estimators. However, in the L96 example, the reduced model is highly non-linear, and hence we can not use the same strategy as the one developed before. It is analytically impossible to find the optimal subsampling time-step in the L96 model. Hence, the only solution is comparing the Maximum Likelihood estimators with "true" values of parameters and investigating numerically the existence of optimal subsampling time-step. Clearly, this is not the best solution. Therefore, the most valid next goal is to look for estimator which do not depend on subsampling time-step.

The second estimator considered is the moment estimator derived by applying the "Method of Moments" on reduced SDE model driving slow variables alone. Although,

the moment estimator is derived using the reduced equation, the main goal is to test its accuracy using the observations of slow variables in the full L96 model. One advantage of this estimator is its independence of subsampling time-step. We compare the two estimators in the Direct Observability context by computing the L^2 errors using the data of the reduced model. Overall, we conclude that if the mean of the slow variables in the multi-scale non-linear model is relatively large and the reduced model is not very sensitive to the change in parameters then the moment estimator is considered a more robust estimator compared with Maximum Likelihood estimator since it has no subsampling complexity.

Chapter 3: Stochastic Mode-reduction of Multi-scale Models with Energy as a Hidden Slow Variable

In this chapter, we continue the objective of deriving an effective stochastic model for the slow variables in the complex multi-scale model. In chapter 2, we consider the non-linear multi-scale L96 model with fast variables represented stochastically. Statistical theories and stochastic modeling with non-essential degrees of freedom represented stochastically provide computationally feasible alternatives for calculating the statistical evolution of the slow variables, and this topic received a lot of attention in recent years [6, 13, 20, 21, 27, 33, 34, 70, 72, 59, 58]. A systematic approach to stochastic mode elimination was developed by authors in [59, 58] which is generalized in [62] for large deterministic systems. In [62], the preliminary step in the stochastic mode elimination method in which the non-linear self-interactions of the fast degrees of freedom are represented stochastically, is avoided. Authors considered the truncated Burgers-Hopf (TBH) system which is deterministic and energy conserving model. Under the assumptions of ergodicity and mixing, they developed a procedure giving closed form stochastic model for slow variable in the limit of infinite separation of time-scales.

In our work, we generalize the procedure developed in [62] to the situation when stochastic terms are added to the slow variables in energy conserving non-linear systems. In such situation, since the noise only affect the slow variables, the fast subsystem is deterministic evolving on a sphere of constant energy. On the other hand, the radius of the sphere slowly changes due to the coupling between the slow and fast dynamics. Therefore, in order to apply the stochastic mode reduction techniques, we consider the energy of the fast subsystem as an additional hidden slow variable. Hence, we generalize the stochastic mode reduction technique in [62] to derive an efficient stochastic model for the original slow variables in full model and additional slow variable given by energy of the fast subsystem.

Also, we introduce a similar second procedure to derive the reduced model for slow variables but this time using Fokker-Planck equation of the reduced model and using the explicit knowledge of the stationary distribution of slow variables in full model. Both stochastic mode reduction techniques are shown on the generalization of prototype *Triad model* with the modification that we add the stochastic terms in slow variables instead of fast variables and the number of fast variables is increased to ensure the ergodicity and mixing in deterministic fast subsystem.

In the considered the non-linear full model, we show the explicit stationary distribution for the slow and fast variables which leads to the derivation of a new procedure of stochastic mode reduction by imposing the fact the the Fokker-Planck operator annihilates the joint stationary density. The second procedure which is generalization of the procedure derived in [62] uses the stationary distribution of fast variables in fast subsystem. The fast subsystem is deterministic and energy conserving, hence, we consider the fast variables to be uniformly distributed over the sphere of radius as initial energy of the system. Both methods provide the same reduced model for the considered non-linear full model.

One of the main concern is that the derived analytical formula of parameters in the

reduced model is dependent on the fourth-order two-point moments of fast variables in fast subsystem which on the other hand, is dependent on the initial energy of the fast subsystem. Since the radius of the sphere changes stochastically in the full model, we would need to simulate the fast subsystem with all possible values of energy and compute the required fourth-order two-point moments of fast variables which is clearly not feasible. Fortunately, the fast subsystem for our model is invariant under a particular rescaling which simplifies the fourth-order two-point moment to the corresponding moment of the fast subsystem with fixed energy n. The consequence of rescaling leads to simulation of the fast subsystem only once with fixed initial energy and hence, simplification of parameters in reduced model.

The derived reduced model is verified numerically by comparing with full model as $\epsilon \to 0$ by testing auto-correlation, density and kurtosis. Conclusion using numerical simulations is that reduced model is an accurate approximation of the slow process in full model.

Chapter 4: Parametric Estimation for Fast-oscillating Potential Model under Indirect Observability

In this chapter, we study the adequate data subsampling for consistent parametric estimation of unobservable stochastic differential equations (SDEs) under Indirect Observability, similar to the study done for the Lorenz-96 model in chapter 2. The reduced model for slow variables in the Lorenz-96 model is non-linear in chapter 2 but the reduced model is linear OU model in the case of fast-oscillating potential model in chapter 4. Thus, we can refer to the results in [9, 10, 11] shown for the multi-scale model having linear OU reduced model. As specified in section 1.1, authors in [9, 10, 11] have provided a rigorous foundation for the parameter estimation of linear stochastic model under Indirect Observability. The authors considered the asymptotic behavior of the Maximum Likelihood estimators for the unknown parameters of the stochastic model, using the observations of multi-scale approximating process as scale separation parameter $\epsilon \to 0$. In particular, they demonstrated

that for consistent estimation of the diffusion parameters the underlying dataset has to be subsampled with time-steps constrained by specific subsampling criteria, depending on the value of the multi-scale parameter ϵ. Otherwise, if these subsampling criteria are violated, the estimated underlying diffusion model will not reproduce the statistical features of the data and the corresponding parameter estimators will be biased even in the limit $\epsilon \to 0$.

We extend the results in [10, 11] on the model with the fast-oscillating potential to illustrate the subsampling problem. First, the numerical investigation of the subsampling criteria derived in [10, 11] in the context of homogenized models is performed.

Another important aspect discussed in the chapter is estimation of an effective model from a dataset generated with a fixed but unknown value of the scale separation parameter ϵ. This issue is important in practical situations, since there has been a considerable effort to efficiently parametrize a stochastic model for the large-scale structures from numerical simulations of various geophysical models. In [9], authors introduced a regression approach for constructing bias-corrected estimators from a single dataset generated by a multi-scale approximate dynamics with a fixed, but unknown value of the parameter ϵ. We show numerically that the regression approach introduced in [9] works for the multi-scale fast-oscillating potential model also.

CHAPTER 2

Parameter Estimation of the Lorenz-96 Model

2.1 Introduction

Original Lorenz-96 model

The Lorenz-96 model, introduced by Lorenz and Emanuel [55, 56] as a simple model with large-scale features of complex non-linear geophysical systems, is given as follows

$$\begin{aligned} \frac{d}{dt}x_k &= x_{k-1}(x_{k+1} - x_{k-2}) - \gamma x_k + F - \frac{k_x}{J}\sum_{j=1}^{J} y_{j,k}, \quad k = 1 \dots K, \\ \frac{d}{dt}y_{j,k} &= c\left(y_{j+1,k}\left(y_{j-1,k} - y_{j+2,k}\right) - y_{j,k}\right) + k_y x_k, \quad j = 1 \dots J. \end{aligned} \tag{2.1}$$

The L96 model describes the linked dynamics of a set of K slow, large-amplitude variables x_k, each of which is associated with J fast, small-amplitude variables $y_{j,k}$. The model is designed to mimic mid-latitude weather and climate behavior and to study the influence of multiple spatio-temporal scales on the predictability of atmospheric flows. Although the

L96 is a prototype model, the slow x_k and the fast $y_{j,k}$ variables in the model are analogous to some atmospheric quantities discretized along the latitude circle. The unit spatial scale between the discrete nodes in L96 model is regarded as a non-dimensional mid-altitude Rossby radius ≈ 800 km. In mid-altitude weather systems, the main "weather waves", the Rossby waves, have westward phase velocity, but eastward group velocity. For the values of constant forcing F ranging from 5 to 32, the L96 model has the band of linearly unstable waves, located roughly between the Fourier wave numbers 3 and 12. It is shown in [55, 56] and chapter 2 of [57] that this band of linearly unstable wave numbers has westward phase and eastward group velocities, just like actual Rossby waves.

L96 and its modifications have been used to test various statistical and stochastic techniques. In [4], authors developed and tested two novel computational algorithms for predicting the mean linear response of the chaotic L96 system to small changes in external forcing via the fluctuation-dissipation theorem (FDT): the short-time FDT and the hybrid FDT. Authors considered the large-scale dynamics of the L96 model alone. It was shown that the blended response algorithms have a high level of accuracy for the linear response of both mean state and variance throughout all the different chaotic regimes of the 40-mode model. However, in multi-scale dynamical systems with time-scale separation, the ST-FDT method developed in [4] can be vulnerable to the presence of fast variables, especially when the response is practically needed only for slow model variables, due to increased response errors at fast scales. Hence, in [1], the author developed an approximate algorithm based on averaged dynamics of multi-scale systems to predict the mean response of the chaotic L96 system to small changes in external forcing. The new method allows to compute the response operators directly for slow variables using existing FDT formulas, which improves numerical stability and reduces computational expense. The author tests this new algorithm on the deterministic L96 system with the addition of constant force in

the equation for the fast variables. The addition of constant forcing in small-scale variables is considered to induce chaotic behavior of fast variables. In [46], authors develop a filtering scheme for the chaotic signals with long memory depth and test it on the deterministic L96 model having slow variables alone. Filtering is the process of finding the best statistical estimate of the true signal if given noisy observations. The authors consider a generalization of the Mean Stochastic model with a diagonal autoregressive linear stochastic model in Fourier space as a filter model for chaotic signals with long memory depth. Using the deterministic L96 model, they show that the non-Markovian nature of this autoregressive model is an important feature in capturing highly oscillatory modes with long memory depth. In [77], the author consider multi-scale original L96 system and study stochastic parameterizations of unresolved fast variables. High-dimensional multi-scale system is reduced to low-dimensional system consisting of large-scale variables alone. The effects of the unresolved variables are parametrized and replaced by a non-linear stochastic function of resolved variables. This stochastic function is computed using regression of unresolved tendency on resolved variables. In [36], the authors consider the original multi-scale deterministic L96 model and develop numerical schemes to estimate the effective low-dimensional system for large-scale variables. The effects of the unresolved variables are replaced by an effective forcing term which accounts for the effect of fast variables on slow variables. The derived effective forcing is function of just one slow variable and is computed by averaging the coupling term in equation of slow variables, conditioned on a particular fixed value of slow variables. The authors also consider the L96 model with non-linear coupling and investigate the role of possible hidden slow variables as well as the additional effects arising on the diffusive time-scale. In [3], authors proposed a method of determining the closed model for slow variables alone, which requires only a single computation of appropriate statistics for the fast dynamics with a certain fixed state of the slow variables. The method

is based on the first-order Taylor expansion of the averaged coupling term with respect to the slow variables, which can be computed using the linear fluctuation-dissipation theorem. The method is tested on a two-scale deterministic L96 model, with additional forcing in fast variables. The effects of the unresolved variables are parametrized and replaced by linear deterministic function of resolved variable. The authors extended the method in [3] to the L96 model with non-linear and multiplicative coupling in [2].

Outline of our work on Lorenz-96 model

We consider a stochastic extension of the multi-scale L96 model with addition of white-noise forcing to fast variables to ensure ergodicity of the L96 model. In the absence of the stochastic terms, the behavior of fast variables may not be ergodic; hence we add the Brownian motion in the equation for the fast variables to ensure strong mixing and ergodicity of the system. In particular, mixing and ergodicity of the fast subsystem is essential for applying the homogenization technique. We also introduce a small parameter $\epsilon > 0$ into the model and refer to the resulting system as the "accelerated L96 model". The parameter $\epsilon > 0$ induces scale separation between the slow and fast variables since our goal is to study how the multi-scale nature of the data affects parametric estimation. The direct numerical simulation of the evolution of the dynamics in L96 is computationally expensive, due both to the large number of small-scale variables and the necessity to choose a small discretization time-step in order to resolve the fast components of L96. Hence, the valid goal is to derive a low-dimensional stochastic model for slow variables alone, referred as reduced model or limiting process. The reduced model should be such that the behavior of slow variables in the full model (L96) weakly converges to the corresponding statistics in the reduced model as the scale separation parameter $\epsilon \to 0$.

We derive a closed reduced model for slow variables using the homogenization procedure. The effect of the fast variables in the L96 model is replaced by linear damping and

diffusion terms in the reduced model. Homogenization procedure is used to derive the explicit equations for the limiting process and also, the "true" values of parameters in the reduced model. The "true" values of parameters can not be computed analytically since they depend on lagged covariance of the non-linear fast subsystem. Hence, true values of parameters are computed using numerical data of the fast subsystem which determines the behavior of the fast variables in the full model as $\epsilon \to 0$. Nevertheless, the main goal is to estimate parameters of the reduced model using observations of the slow variables in full model with a finite value of ϵ and to test the estimation procedure as $\epsilon \to 0$. Thus the "true" values derived using homogenization are only used to test the accuracy of the estimators.

We focus on parametric estimation to emphasize the Indirect Observability framework, and illustrate the impact of Indirect Observability on various estimators. The main goal is to prevent large errors in parametric estimation of reduced model when the available observations are subsampled from data of the slow variables in the full L96 model. We analyze the performance of two different estimators and use the true values of parameters only to test their accuracy. Since we know that the statistical behavior of the slow variables in the full model converges to the reduced equation, we expect parameter estimators to converge to the true values as $\epsilon \to 0$.

First, the method of approximate Maximum Likelihood estimation is used to obtain estimators for the damping and diffusion coefficients in the reduced model. The approximate Maximum Likelihood estimators depend on the auto-correlation of the slow variables in the full model. As mentioned in section 1.1, analytical properties of sample paths in the full and the reduced models are very different especially for small time-steps. This can have a potentially high impact on the accuracy of the approximate Maximum Likelihood estimators, since they depend on auto-correlations of the slow variables in the full model.

Next we analyze a moment estimator derived using method of moments. Although, the moment estimator is derived using discretization of the reduced model, the main goal is to test it's accuracy when this estimator is computed using observations of the slow variables in the full model. In contrast with approximate Maximum Likelihood estimators, the moment estimator depends only on one-point stationary moments of the slow variables and is independent of subsampling time-step. Thus, we compare and contrast the two estimators computed on data of the slow variables in the full model with different values of ϵ and subsampled at time-step Δ.

2.2 Accelerated Lorenz-96 Model

We consider a modified Lorenz-96 model with the scale separation parameter $\epsilon > 0$ explicitly controlling the scale separation between the slow and the fast variables. As discussed in the introduction, the L96 model considered here has generic features of climate-weather systems, such as the presence of linearly unstable waves, strong non-linearity, constant forcing, linear damping, dissipation, chaos, and mixing. The modified multi-scale Lorenz model is given by:

$$\begin{aligned} dx_k &= x_{k-1}(x_{k+1} - x_{k-2})dt - \gamma x_k dt + F dt - \frac{k_x}{\epsilon}\frac{B_k}{J}dt, \quad k = 1 \dots K \\ dy_{j,k} &= \frac{c}{\epsilon^2}\left(y_{j+1,k}\left(y_{j-1,k} - y_{j+2,k}\right) - y_{j,k}\right)dt + \frac{k_y}{\epsilon}x_k dt + \frac{s}{\epsilon}dW_{j,k}, \quad j = 1 \dots J \end{aligned} \tag{2.2}$$

where $B_k = \sum_{j=1}^{J} y_{j,k}$. $dW_{j,k}$ represent the increment in independent Brownian motions, $\epsilon > 0$ is the scale separation parameter, γ, k_x, k_y, c, s are known constants, and F is a constant forcing.

The L96 model, given by (2.2), is designed to mimic mid-altitude weather and climate behavior in which the slow x_k and fast $y_{j,k}$ variables in the model represent some atmospheric quantities discretized respectively into K and $K \times J$ sectors along the latitude circle.

The full model describes the linked dynamics of a set of K slow variables x_k, each of which is coupled to J fast variables $y_{j,k}$. For a fixed value of $k = 1$ to K, all the fast variables $y_{j,k}, j = 1 \dots J$, are coupled to each other and affected equally by the large-scale variable x_k, to which they belong to.

We assume that both x and y variables are periodic (or cyclic), i.e. $x_{k+K} = x_k$ and $y_{j,k+K} = y_{j,k}, y_{j+J,k} = y_{j,k+1}$ where K and J are the total number of x and y variables, respectively. Also, x_k variables are invariant under the index shift which implies that stationary statistics is also invariant under the index shift

$$\langle x_{i_1}(t_1)x_{i_2}(t_2)\dots x_{i_k}(t_k)\rangle = \langle x_{i_1+p}(t_1)x_{i_2+p}(t_2)\dots x_{i_k+p}(t_k)\rangle\,,$$

where $p \geq 0$ is the index shift. We will refer to this characteristic of slow variables as "stationary statistically invariant under the index shift" from now on in the dissertation.

We fix the total number of x-variables as $K = 18$ and the number of y-variables for each x_k as $J = 20$, so that there are 360 y-variables in total. Coupling parameters k_x and k_y affect the strength of the interaction between the slow and the fast sets of variables; we fix them to be $k_x = 4$ and $k_y = 1$. The coefficient c in the equation for y-variables, is fixed as $c = 1$.

Constant forcing F is one of the key bifurcation parameters in the L96 model. It is shown in [55, 56] and chapter 2 of [57] that the Lorenz model in (2.2), for the values of constant forcing F ranging from 5 to 32, has a band of linearly unstable waves which have westward phase and eastward group velocities, just like the main 'weather waves', the Rossby waves. Also, it's demonstrated in chapter 2 of [57] that the dynamical regime of the L96 model varies considerably with different values of the constant forcing term F. The model is in weakly chaotic regimes for $F = 5, 6$, strongly chaotic regime for $F = 8$, and turbulent regimes $F = 12, 16, 24$. We will present our results for some specific values of forcing F, for example, $F = 8, 10, 12, 24$.

In further sections, we show the results using various values of force F and linear damping coefficient γ but all other parameters in (2.2) are fixed as

$$k_x = 4, \quad k_y = 1, \quad c = 1, \quad s = 2.236, \quad K = 18, \quad J = 20. \tag{2.3}$$

Multi-scale systems like (2.2) are particularly challenging for direct numerical computations because a time-step of the order of ϵ^2 is necessary to resolve the fast variables y; therefore a total number of steps of the order of ϵ^{-2} is required to simulate the evolution of the slow variables x. To overcome these computational difficulties caused by the separation of time-scales, it is desirable to capture the statistical features of the slow variables in the multi-scale system (2.2) by fitting a reduced (single-scale) system of SDEs to the data set of slow variables x. Therefore, the main goal of this chapter is to develop efficient parametric estimation techniques for SDEs which rely on the data of large-scale variables alone. Since the data is generated by the multi-scale model L96 in (2.2), the small-scale effects in the data can play a significant role and yield incorrect estimation results.

To test our estimation procedure, we use analytical results of homogenization for SDEs where a limiting equation for the x-variables can be derived explicitly in the limit $\epsilon \to 0$, explained in detail in next section.

2.3 Homogenization for the L96 Model

Given observations of the large-scale variable x_k, our objective is to fit an effective SDE consistent with those observations and, hence, we need to estimate parameters of that SDE accurately and efficiently. Our main goal is to test how multi-scale effects in the data affect the estimation of parameters in the stochastic reduced model. To this end, we need to compute the "true" values of parameters and these true values will be used to verify the accuracy of parametric estimation.

Mode-reduction (a.k.a homogenization) provides a natural tool for semi-analytical computation of such "true" values of parameters. Homogenization technique is a natural analytical approach to obtain a reduced model of the full model with a multi-scale parameter ϵ in the limit of infinite scale separation (i.e., $\epsilon \to 0$). In this procedure, coefficients of the reduced model are estimated from the data of the fast variables. Thus, the homogenization will yield the "true" values for the coefficients of the effective model which can be used for the quantitative comparison with the results of parametric estimation.

The fast subsystem determines the behavior of y variables in full model, given by (2.2), as $\epsilon \to 0$. We will show in the this section that the fast subsystem plays a particularly important role in the homogenization procedure. In particular, the explicit knowledge of the stationary distribution for the fast subsystem significantly simplifies the analytical calculations in the mode reduction.

2.3.1 The fast subsystem and it's stationary distribution

The fast subsystem, consisting of y-variables of L96 model, corresponds to the $O(\epsilon^{-2})$ terms in the Fokker-Planck equation for the L96 SDE (2.2). It is given by:

$$dy_{j,k} = c(y_{j+1,k}(y_{j-1,k} - y_{j+2,k}) - y_{j,k})dt + sdW_{j,k}, \tag{2.4}$$

where $j = 1 \dots J,\ k = 1 \dots K$ and $W_{j,k}$ are independent Brownian motions where $dW_{j,k}$ represent increment in Brownian motion. The fast subsystem is a ring of $J \times K = 360$ variables.

We prove analytically that the joint invariant measure of all $y_{j,k}$ variables is a product measure and each $y_{j,k}$ variable in (2.4) follows a Gaussian distribution with the same mean and variance, i.e., $y_{j,k} \sim N\left(0, \frac{s^2}{2c}\right) \forall j = 1 \dots J,\ k = 1 \dots K$.

Let ρ be the invariant joint measure of $\{y_{j,k},\ j \dots J,\ k = 1 \dots K\}$. We prove analytically

that ρ is a product measure and is given by,

$$\rho = C \prod_{k=1}^{K} \prod_{j=1}^{J} exp\left(-\frac{2c}{s^2} y_{j,k}^2\right), \tag{2.5}$$

where C is a normalizing constant. The differential operator for the Fokker-Planck equation of (2.4) is

$$L = \sum_{k=1}^{K} \sum_{j=1}^{J} \left(-c(\partial_{y_{j,k}} y_{j+1,k}(y_{j-1,k} - y_{j+2,k}) - \partial_{y_{j,k}} y_{j,k}) + \frac{s^2}{2} \partial_{y_{j,k}^2}^2\right). \tag{2.6}$$

To prove that invariant joint measure of all $y_{j,k}$ variables in fast subsystem (2.4) is ρ (2.5), we show explicitly that $L\rho = 0$. First, we find the partial derivatives of ρ w.r.t. $y_{j,k}$ as:

$$\begin{aligned} \partial_{y_{j,k}} \rho &= \partial_{y_{j,k}} \left(C \prod_{k=1}^{K} \prod_{j=1}^{J} exp\left(-\frac{c}{s^2} y_{j,k}^2\right)\right) = -\frac{2c}{s^2} y_{j,k} \rho, \\ \partial_{y_{i,k}} (y_{j,k} \rho) &= \delta_{i,j} \rho - \frac{2c}{s^2} y_{i,k} y_{j,k} \rho, \\ \partial_{y_{i,k}^2}^2 \rho &= -\frac{2c}{s^2} \left(1 - \frac{2c}{s^2} y_{i,k}^2\right) \rho, \end{aligned} \tag{2.7}$$

where $\delta_{i,j}$ is the Kronecker delta function. Substituting partial derivatives of ρ in $L\rho$ (2.6) gives:

$$\begin{aligned} L\rho &= \sum_{k=1}^{K} \sum_{j=1}^{J} \left(-c\left(\partial_{y_{j,k}} y_{j+1,k}\left(y_{j-1,k} - y_{j+2,k}\right) - \partial_{y_{j,k}} y_{j,k}\right) + \frac{s^2}{2} \partial_{y_{j,k}^2}^2\right) \rho, \\ &= \sum_{k=1}^{K} \sum_{j=1}^{J} \left(-c y_{j+1,k}\left(y_{j-1,k} - y_{j+2,k}\right) \left(-\frac{2c}{s^2} y_{j,k} \rho\right) + c\left(1 - \frac{2c}{s^2} y_{j,k}^2\right) \rho \right. \\ &\quad \left. -c\left(1 - \frac{2c}{s^2} y_{j,k}^2\right) \rho\right), \end{aligned}$$

which further simplifies to

$$L\rho = \frac{2c^2}{s^2} \sum_{k=1}^{K} \sum_{j=1}^{J} y_{j,k}\, y_{j+1,k} \left(y_{j-1,k} - y_{j+2,k}\right) \rho. \tag{2.8}$$

Note that the assumption of periodicity of y-variables in the full model (2.2) follows in the fast subsystem also, i.e., $y_{j,k+K} = y_{j,k}, y_{j+J,k} = y_{j,k+1}$. Hence, we obtain

$$\sum_{k=1}^{K} \sum_{j=1}^{J} y_{j,k}\, y_{j+1,k} \left(y_{j-1,k} - y_{j+2,k}\right) = 0,$$

which simplifies equation (2.8) as

$$L\rho = 0. \tag{2.9}$$

$L\rho = 0$ in (2.9) shows that ρ in (2.5) is an invariant measure of all $y_{j,k}$ variables and since it's a product measure, their stationary one-point moments are independent of each other, hence, $y_{j,k} \sim N(0, \frac{s^2}{2c}),\ j = 1 \dots J,\ k = 1 \dots K$.

Empirical Moments: We observe low-order one-point moments and two-point covariance numerically. The fast subsystem model, (2.4), is simulated by keeping all parameters fixed as in (4.5) with the integration time-step $\delta t = 0.0001$ and the total time $T = 50000$. Note that since y variables are circulatory stationary hence all $y_{j,k}$ have equal stationary moments. We obtain low-order stationary moments of y-variable in the fast subsystem (2.4) as follows

$$\text{Mean} = \langle y_j \rangle = 0, \quad \text{Variance} = \left\langle y_j^2 \right\rangle - \langle y_j \rangle^2 = 2.49,$$

$$\text{Skewness} = \frac{\left\langle y_j^3 \right\rangle}{\left\langle y_j^2 \right\rangle^{3/2}} = -0.005, \quad \text{Kurtosis} = \frac{\left\langle y_j^4 \right\rangle}{3\left\langle y_j^2 \right\rangle^2} = 0.99,$$

and all other odd moments are approximately equal to zero.

We also observe numerically that the stationary lagged covariances are approximately zero for different indexes, i.e.

$$\langle y_i(0) y_j(\tau) \rangle \approx 0, \quad i \neq j, \tag{2.10}$$

where $\tau > 0$ is covariance time lag. This property of lagged covariance of mixed fast variables being zero, does not follow from the stationary distribution in (2.5). This is a stronger property of the fast subsystem.

2.3.2 Derivation of reduced model

Homogenization technique is a natural analytical approach to obtain a reduced model of the full model with a multi-scale parameter ϵ in the limit of infinite scale separation (i.e., $\epsilon \to 0$). In this subsection, we apply the homogenization procedure to the L96 model given by (2.2). Also, we derive the coefficients of the effective model as a function of the data of the fast subsystem (2.4).

Let $\vec{x} = \{x_1, x_2, \ldots, x_K\}$ and $\vec{y} = \{y_{j,k}, j = 1 \ldots J,\ k = 1 \ldots K\}$. The Kolmogorov backward equation associated to the full model, (2.2), is applied to an arbitrary function $u = u(t, \vec{x}, \vec{y})$, where

$$u(t, \vec{x}, \vec{y}) = \mathbb{E}[f(\vec{x}(t),\ \vec{y}(t)) \mid \vec{x}(0) = \vec{x},\ \vec{y}(0) = \vec{y}],$$

where f is an arbitrary test function. Here, $\vec{x}(t),\ \vec{y}(t)$ are vectors of the stochastic processes defined by (2.2) and $\vec{x},\ \vec{y}$ are vectors of initial conditions which also play the role of independent variables in the backward equation. Thus, the backward equation is given by

$$-\partial_t u = L_0 u + \frac{1}{\epsilon} L_1 u + \frac{1}{\epsilon^2} L_2 u, \tag{2.11}$$

where the differential operators L_i are given by:

$$\begin{aligned} L_0 &= \sum_{k=1}^{K} (x_{k-1}(x_{k+1} - x_{k-2}) - \gamma x_k + F)\partial_{x_k}, \\ L_1 &= \sum_{k=1}^{K} \left(-k_x \frac{B_k}{J} \partial_{x_k} + \sum_{j=1}^{J} k_y x_k \partial_{y_{j,k}} \right), \\ L_2 &= \sum_{k=1}^{K} \sum_{j=1}^{J} \left(c(y_{j+1,k}\,(y_{j-1,k} - y_{j+2,k}) - y_{j,k})\partial_{y_{j,k}} + \frac{s^2}{2} \partial^2_{y_{j,k}} \right), \end{aligned} \tag{2.12}$$

where $B_k = \sum_{j=1}^{J} y_{j,k}$.

Note that the differential operator L_2 is the infinitesimal generator for the fast subsystem

(2.4). As proved in section 2.3.1, the fast subsystem has an invariant measure of y_j variables given by:

$$\rho(\vec{y}) = C \prod_{k=1}^{K} \prod_{j=1}^{J} exp\left(-\frac{2c}{s^2} y_{j,k}^2\right),$$

where $\vec{y} = \{y_{i,j}, i = 1 \dots J, k = 1 \dots K\}$. Hence, the adjoint operator of L_2 applied to the invariant measure ρ gives:

$$L_2^* \rho = 0. \tag{2.13}$$

Let us consider the multi-scale expansion of u as

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + ..$$

Plugging in expansion of u into (2.11) and collecting different powers of ϵ in (2.11) gives the following relations:

$$\begin{aligned} \frac{1}{\epsilon^2} &: \quad L_2 u_0 = 0, \\ \frac{1}{\epsilon} &: \quad L_1 u_0 + L_2 u_1 = 0, \\ 1 &: \quad -\partial_t u_0 = L_0 u_0 + L_1 u_1 + L_2 u_2. \end{aligned} \tag{2.14}$$

Recall that L_2 is a differential operator w.r.t. all y_j variables, defined as the third equation in (2.12). Hence, since u_0 is arbitrary, the first equation in (2.14) implies that $u_0 = u_0(\vec{x}, t)$, i.e. u_0 is independent of the fast variables $\vec{y}$. Now, let $\mathbb{P}$ denote the expectation with respect to the invariant density $\rho(\vec{y})$ as

$$\mathbb{P}g = \int g(y)\ \rho(\vec{y})\ d\vec{y}. \tag{2.15}$$

where g is any bounded Borel function. Recall that ρ is invariant density for the fast subsystem, (2.4), i.e. satisfies the Fokker-Planck equation with $L_2^* \rho = 0$. Therefore, $\mathbb{P} L_2 f = 0$ for all smooth functions f with compact support.

Compatibility Condition: Applying the projection operator $\mathbb{P}$ to the second equation in (2.14) we obtain the compatibility condition

$$\mathbb{P}L_1 u_0 = 0, \tag{2.16}$$

since $\mathbb{P}L_2 \cdot = 0$. Since u_0 is an arbitrary function of $\vec{x}$, this implies that operator $\mathbb{P}L_1$ applied to any function of $\vec{x}$ must be zero. This compatibility condition, often written as $\mathbb{P}L_1\mathbb{P} = 0$, must hold in order for the homogenization approach to be applicable. Hence, we show that L96 model satisfies this compatibility condition.

When we consider $L_1 u_0$, then $\partial y_{j,k} u_0 = 0$ in the expression for L_1 in (2.14) since u_0 is the function of $\vec{x}$. Hence, the compatibility condition becomes

$$-\frac{k_x}{J}\sum_{k=1}^{K} \partial_{x_k} u_0 \int B_k \ \rho(\vec{y}) \ d\vec{y} = 0, \tag{2.17}$$

where $B_k = \sum_{j=1}^{J} y_{j,k}$. The above expression can be simplified as

$$\int y_{j,k} \ \rho(\vec{y}) \ d\vec{y} = 0, \tag{2.18}$$

which is true in L96 model. We showed in section 2.3.1 that means of $\vec{y}$-variables are zero. Therefore, the L96 model satisfies the compatibility condition and the homogenization procedure is valid in this case. This is equivalent to "Averaging = 0" condition.

Second equation in 2.14 implies that

$$u_1 = -L_2^{-1} L_1 u_0, \tag{2.19}$$

where the operator L_2^{-1} is given by

$$L_2^{-1} f(\vec{y}) = -\int_0^{\infty} e^{L_2 \tau} \ f(\vec{y}) \ d\tau. \tag{2.20}$$

Also, using the Feynmann-Kac technique, the solution can be obtained as the conditional expectation and hence we obtain

$$L_2^{-1} f(\vec{y}) = -\int_0^\infty \mathbb{E}[f(\vec{y}_\tau | y_0 = y)]\ d\tau, \tag{2.21}$$

where $\vec{y}_\tau$ is the solution of the fast subsystem (2.4) at time τ and $\mathbb{E}[f(\vec{y}_\tau|\vec{y}_0 = \vec{y})]$ is the conditional expectation with respect to $\vec{y}_\tau$ given initial value $\vec{y}_0 = \vec{y}$.

Applying the projection operator $\mathbb{P}$ to both sides of the third relation in (2.14), gives,

$$\mathbb{P}(-\partial_t u_0) \quad = \quad \mathbb{P}(L_0 u_0) + \mathbb{P}(L_1 u_1) + \mathbb{P}(L_2 u_2). \tag{2.22}$$

Since $u_0 = u_0(\vec{x}, t)$ and $\mathbb{P}$ is the expectation with respect to the invariant density $\rho(\vec{y})$, we obtain $\mathbb{P}(-\partial_t u_0) = -\partial_t u_0$. Similarly, definition of L_0 operator in (2.12) implies $\mathbb{P}(L_0 u_0) = L_0 u_0$. Lastly, $L_2^* \rho = 0$ simplifies the equation (2.22) as

$$-\partial_t u_0 \quad = \quad L_0 u_0 + \mathbb{P}(L_1 u_1). \tag{2.23}$$

Using the expression (2.19) for u_1 gives the backward equation for u_0 as

$$-\partial_t u_0 \quad = \quad L_0 u_0 - \mathbb{P}(L_1 L_2^{-1} L_1 u_0).$$

Substituting operator L_1 specified in (2.12) gives $-\mathbb{P}(L_1 L_2^{-1} L_1 u_0)$ as

$$\begin{aligned} -\mathbb{P}(L_1 L_2^{-1} L_1 u_0) \quad = \quad & \sum_{k=1}^{K} \int .. \int \left(-k_x \frac{B_k}{J} \partial_{x_k} + \sum_{j=1}^{J} k_y x_k \partial_{y_{j,k}} \right) L_2^{-1} \\ & \left(-k_x \frac{B_k}{J} \partial_{x_k} + \sum_{j=1}^{J} k_y x_k \partial_{y_{j,k}} \right)\ \rho(\vec{y})\ u_0\ d\vec{y}. \end{aligned}$$

where $B_k = \sum_{j=1}^{J} y_{j,k}$. The multiple integrals in above equation represent J number of integrals w.r.t. $\{y_{j,k},\ j = 1 \dots J\}$ variables for fixed k, which gets simplified to double integrals in next expression. The first equation in (2.14) implies that $u_0 = u_0(\vec{x}, t)$, hence, $\partial_{y_{j,k}} u_0 = 0$.

Also, partial derivatives of u_0 with respect to x_k can be pulled out of the integrals in the above expression since the integration is with respect to $\vec{y}$. Hence, $\mathbb{P}(L_1 L_2^{-1} L_1 u_0)$ simplifies to

$$\begin{aligned} -\mathbb{P}(L_1 L_2^{-1} L_1 u_0) &= \sum_{k=1}^{K} \frac{k_x^2}{J^2}\, \partial_{x_k}^2\, u_0 \sum_{i,j=1}^{J} \int\int y_{i,k}\; L_2^{-1}\; y_{j,k}\; \rho(y_i, y_j)\; dy_i\; dy_j \\ &\quad - \sum_{k=1}^{K} \frac{k_x k_y}{J}\; x_k\; \partial_{x_k} u_0 \sum_{i,j=1}^{J} \int\int \partial_{y_{j,k}}\; L_2^{-1}\; y_{i,k}\; \rho(y_i, y_j)\; dy_i\; dy_j, \end{aligned}$$

where $\rho(y_i, y_j) = exp\left[-\frac{2c}{s^2}\left(y_i^2 + y_j^2\right)\right]$. Since $\rho(\vec{y})$ in (2.5) is a product measure of $\vec{y}$ variables, integration with respect to $y_l,\ l \neq i, j$ can be separated and hence multiple integrals gets simplified to double integrals in above expression. Note that $\partial_{y_{j,k}} y_{i,k} = 0$ for any $j \neq i$, hence, we further obtain:

$$\begin{aligned} -\mathbb{P}(L_1 L_2^{-1} L_1 u_0) &= \sum_{k=1}^{K} \frac{k_x^2}{J^2}\, \partial_{x_k}^2 u_0 \sum_{i,j=1}^{J} \int\int y_{i,k}\; L_2^{-1}\; y_{j,k}\; \rho(y_i, y_j)\; dy_i\; dy_j \\ &\quad - \sum_{k=1}^{K} \frac{k_x k_y}{J}\; x_k\; \partial_{x_k} u_0 \sum_{j=1}^{J} \int \partial_{y_{j,k}}\; L_2^{-1}\; y_{j,k}\; \rho(y_j)\; dy_j. \end{aligned}$$

Note that integrals in the above expression do not depend on $\vec{x}$. Also, ρ is the invariant measure of the fast subsystem, therefore, the statistics should be computed from the fast subsystem, not from the full model.

Substituting the expression for $-\mathbb{P}(L_1 L_2^{-1} L_1 u_0)$ given by (2.24) in the backward equation (2.24) characterizes the reduced model as

$$dX_k = X_{k-1}(X_{k+1} - X_{k-2})dt - \gamma X_k dt + F dt - \alpha X_k dt + \sigma dW_k, \tag{2.24}$$

where $k = 1$ to K, W_k are independent Brownian motions, and parameters σ^2 and α are defined as

$$\begin{aligned} \sigma^2 &= 2\, \frac{k_x^2}{J^2} \sum_{i,j=1}^{J} \int\int y_{i,k}\; L_2^{-1}\; y_{j,k}\; \rho(y_i, y_j)\; dy_i\; dy_j, \\ \alpha &= \frac{k_x k_y}{J} \sum_{j=1}^{J} \int \partial_{y_j}\; L_2^{-1}\; y_{j,k}\; \rho(y_j)\; dy_j. \end{aligned} \tag{2.25}$$

Using the definition of the operator L_2^{-1} in the equation (2.21), σ^2 can be further written as

$$\sigma^2 \;=\; 2\,\frac{k_x^2}{J^2}\,\sum_{i,j}\int_{\tau=0}^{\infty}\langle y_{i,k}(t)\; y_{j,k}(t+\tau)\rangle\, d\tau.$$

In (2.10), we showed empirically that

$$\langle y_{i,k}(0)\; y_{j,k}(t)\rangle \approx 0 \quad \forall t \neq 0, \quad i \neq j,$$

which we obtained numerically by integrating the fast subsystem (2.4) with parameters $c = 1,\ s = 2.236$. We will use this numerical estimate for the two-point covariance in analytical simplification of the reduced model. Hence, σ^2 can be further simplified as

$$\sigma^2 \;=\; 2\frac{k_x^2}{J^2}\sum_{i}\int_{\tau=0}^{\infty}\langle y_{i,k}(t)\; y_{i,k}(t+\tau)\rangle\, d\tau. \tag{2.26}$$

In section 2.3.1, we proved that the joint invariant measure of $y_{j,k},\ j = 1\ldots J,\ k = 1\ldots K$ variables is a product measure and

$$y_{j,k} \sim N\left(0, \frac{s^2}{2c}\right),$$

for all values of $j = 1$ to J and $k = 1$ to K. This product measure is used to simplify the expression for α in (2.25) using integration by parts as

$$\alpha = \frac{k_x k_y}{J}\sum_{j=1}^{J}\int \partial_{y_j}\; L_2^{-1}\; y_{j,k}\; \rho(y_j)\; dy_j,$$

with the integration by parts defined as

$$u_\alpha = \rho(y_j) = \frac{\sqrt{2c}}{s\sqrt{2\pi}}e^{-c\frac{y_j^2}{s^2}}, \quad dv_\alpha = \partial_{y_j}L_2^{-1}y_j dy_j,$$
$$\Rightarrow du_\alpha = -\frac{2c}{s^2}y_j\rho(y_j), \quad v_\alpha = L_2^{-1}y_j.$$

Using the specified integration by parts and the fact that the Gaussian density $\rho(\vec{y})$ vanishes exponentially as $|y| \to \infty$, α simplifies to

$$\alpha \;=\; \frac{k_x k_y}{J}\frac{2c}{s^2}\sum_{j=1}^{J}\int_{\tau=0}^{\infty}\langle y_{j,k}(t)\; y_{j,k}(t+\tau)\rangle\, d\tau. \tag{2.27}$$

Finally substituting σ^2 expression from (2.26) and α from (2.27) in reduced model equation (2.24)and comparing α and σ^2 expressions, we obtain the following reduced model,

$$dX_k = X_{k-1}(X_{k+1} - X_{k-2})dt - \gamma X_k dt + Fdt - \alpha X_k dt + \sigma dW_k, \tag{2.28}$$

where $k = 1$ to K, W_k, $k = 1$ to K are independent Brownian motions and parameters α, σ^2 are given by,

$$\alpha = \frac{k_x k_y}{J} \frac{2c}{s^2} \sum_{j=1}^{J} \int_{\tau=0}^{\infty} \langle y_{j,k}(t)\ y_{j,k}(t+\tau) \rangle \, d\tau, \quad \sigma^2 = \frac{k_x s^2}{k_y c J} \alpha.$$

It is impossible to obtain the value of the lagged covariance of $\vec{y}$-variables analytically, hence, we use one long simulation of the fast subsystem to compute the lagged covariance numerically.

Reduced model

In section 2.3.2 , we showed using homogenization procedure that the slow variables x_k in L96 (2.2) converges weakly to X_k variables in the reduced model (2.29) as $\epsilon \to 0$ for each $k = 1$ to K. The effective stochastic reduced model for large-scale variables X_k, as shown in (2.28), is given by

$$dX_k = X_{k-1}(X_{k+1} - X_{k-2})dt - \gamma X_k dt + Fdt - \alpha X_k dt + \sigma dW_k. \tag{2.29}$$

where $k = 1$ to K, W_k are independent Brownian motions, and parameters α and σ can be computed explicitly as

$$\begin{aligned} \alpha &= \frac{k_x k_y}{J} \frac{2c}{s^2} \sum_{j=1}^{J} \int_{\tau=0}^{\infty} \mathrm{cov}_{y_{j,k}}(\tau) d\tau, \\ \sigma^2 &= \frac{k_x s^2}{J k_y c} \alpha, \end{aligned} \tag{2.30}$$

where $cov_{y_{j,k}}(\tau)$ is the stationary lagged covariance function of $y_{j,k}$ at time lag $\tau > 0$, defined as

$$\mathrm{cov}_{y_{j,k}}(\tau) = \mathbb{E}\left(y_{j,k}(t)\ y_{j,k}(t+\tau)\right). \tag{2.31}$$

The lagged covariance of $y_{j,k}$ variables, required for α and σ in (2.30), is computed numerically from the fast subsystem (2.4). Note that the equation for $y_{j,k}$ variables in the fast subsystem (2.4) does not depend on the force F and damping γ. Hence, we can conclude using formulas in (2.30) that the explicit true values of α and σ in the reduced stochastic model (2.29) do not depend on force F and damping γ in the equation for the slow variables. Therefore, the parameters α and σ can be estimated from a single simulation of the fast subsystem for all values of F and γ and it is not necessary to recompute them when F, γ change.

In the next subsection, we discuss the numerical computation of lagged covariance of $y_{j,k}$ variables and the resulting numerical "true" values of parameters α and σ.

2.3.3 Numerical estimates of true parameters of L96 limiting equation

In section 2.3.2, we derived the stochastic reduced model (2.29) for the slow variables x_k and its true parameters α and σ (2.30). Note that to compute the explicit value of α in (2.30), we need lagged covariance of $y_{j,k}$ variables which is not feasible to find analytically in the fast subsystem (2.4). In this subsection, we find numerical estimates of lagged covariance of $y_{j,k}$ variables using *Riemann sum*. Further, we compare the stationary statistics of the reduced model (2.29) with different numerically computed parameters α, σ. We show numerically that the reduced model is not sensitive to small changes in these parameters.

Since all $y_{j,k}$ variables have identical stationary moments in the fast subsystem (2.4), it's enough to find auto-covariance of just one $y_{j,k}$ variable i.e. for one fixed value of $j = 1$ and $k = 1$. We use Riemann sum to find the auto-correlation time of $y_{j,k}$ variables in (2.30). For that, we simulate one long trajectory of the fast subsystem (2.4) which requires only two fixed parameters c and s. Hence, keeping parameters same as in (4.5) i.e. $c = 1$ and $s = 2.236$, we simulate one long trajectory of the fast subsystem (2.4) with the integration

time-step $\delta t = 0.0001$ and total time $T = 50000$. We obtain auto-covariance of $y_{1,1}$ variable shown in Figure 2.1.

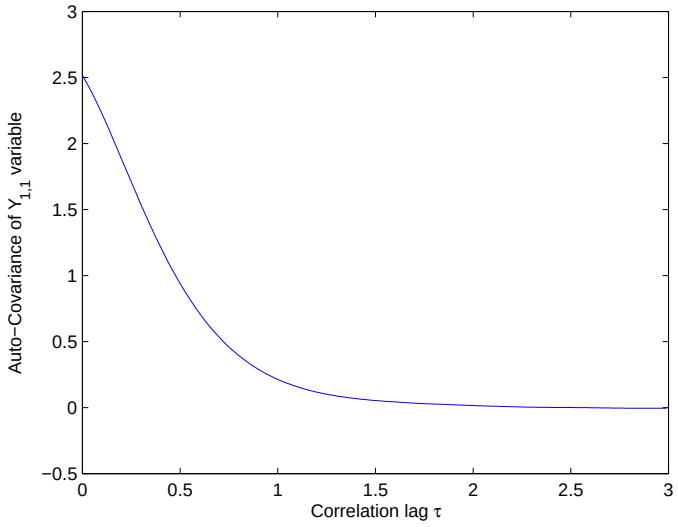

Figure 2.1: Auto-covariance of $y_{1,1}$ variable in the fast subsystem (2.4)

Figure 2.1 clearly shows that for fixed parameters, given by (4.5), auto-covariance of $y_{1,1}$ variable converges to zero for correlation time lag $\tau \geq 2$. Hence, to compute α in (2.30), it's sufficient to look at the integrated lagged covariance of $y_{j,k}$ variable until correlation lag time $\tau = 2$. Using Riemann sum technique, we approximate the integrated lagged covariance of $y_{j,k}$ variable in Figure 2.1 using partition of lag time-step as $\tau = 0 : 0.01 : 2$. The numerical estimate for the integrated lagged covariance of $y_{j,k}$ variables can be written as,

$$\int_{\tau=0}^{\infty} \mathrm{cov}_{y_{j,k}}(\tau) d\tau \quad = \quad \sum_{\tau=0,\Delta\tau=0.01}^{\tau=2} \mathrm{cov}_{y_{j,k}}(\tau)\Delta\tau. \tag{2.32}$$

where $\text{cov}_{j,k}(\tau)$ is lagged covariance of $y_{j,k}$ variables, given by (2.31). Therefore, the numerical estimate for parameters α and σ in (2.30) for this particular case is given by

$$\begin{aligned} \alpha &= \frac{k_x k_y}{J}\frac{2c}{s^2}\sum_{j=1}^{J}\sum_{\tau=0,\Delta\tau=0.01}^{\tau=2}\text{cov}_{y_{j,k}}(\tau)\Delta\tau, \\ \sigma^2 &= \frac{k_x s^2}{k_y c J}\alpha. \end{aligned} \tag{2.33}$$

We obtain integrated lagged covariance of all $y_{j,k}$ variables as

$$\sum_{\tau=0,\Delta\tau=0.01}^{\tau=2}\text{cov}_{y_{j,k}}(\tau)\Delta\tau = 1.1867.$$

and the corresponding numerical estimate of α and σ are $\alpha = 1.8987, \sigma = 1.3999$.

Numerical Estimates of α and σ: One concern of using the numerical estimation outlined above is possibility of various numerical errors affecting the estimates. The numerical estimate of α parameter given in (2.33) is a reliable estimate only if we show that the Riemann sum computation presented above is robust and statistics of the reduced model (2.29) is not sensitive to small numerical errors in empirical parameter estimates for α and σ.

We simulate 150 trajectories of the fast subsystem (2.4) by changing either seed for random number generation, integration time-step δt, correlation time-step τ, or total time T. The average and the standard deviation of the 150 numerical estimates of the parameter α are

$$\begin{aligned} &\text{mean}(\alpha) = 1.8553,\ \text{std}(\alpha) = 0.0344, \\ &\text{mean}(\sigma) = 1.362,\ \text{std}(\sigma^2) = 0.0344. \end{aligned} \tag{2.34}$$

Note that for the specific parameters in (4.5), we obtain $\sigma = \sqrt{\alpha}$, therefore, it's enough to mention the estimates of α. The minimum and maximum values of α are 1.8522 and 1.9208, respectively. Substituting either minimum or maximum estimate of α in the reduced model (2.29), gives numerically indistinguishable stationary moments from the reduced model

with parameter value $\alpha = 1.8553$. Note that the value of σ also changes with α using the relation in (2.30). Thus the reduced model is not sensitive to small numerical errors in parameters α and σ and hence numerical estimates of these parameters using the Riemann sum is sufficient to be used as true values of parameters in stochastic reduced model.

For the rest of this chapter, we use the true value of parameters α and σ as,

$$\alpha = 1.8553, \quad \sigma = 1.362, \tag{2.35}$$

for fixed parameters as given in (4.5). We emphasize that these parameters α and σ in the reduced model depend only on parameters c, k_x, k_y and s. Therefore, "true" values of parameters α and σ are valid for any choice of F and γ.

2.3.4 Comparison of the full and the reduced model using true parameters

In this subsection, we investigate numerically the convergence of x_k variables in the full model (2.2) to X_k in the reduced model (2.29) as $\epsilon \to 0$. Keeping all parameters fixed as in (4.5), we will consider the convergence of stationary correlation functions and stationary density of x_k in the full model to the corresponding statistics of X_k in the reduced model as $\epsilon \to 0$ for various values of force F. We would like to remind that the true values α and σ, in reduced model, will remain fixed all values of force F and damping γ.

Numerical Setting: Fix parameters as in (4.5). Convergence of stationary auto-correlation and stationary density is demonstrated for different values of force $F = 8,\ 10,\ 12, 16,$ and 24 and damping is fixed at $\gamma = 2$. We simulate the full model with $\epsilon = 1,\ 0.5,\ 0.3,\ 0.1$ and observe that the full model with $\epsilon = 0.1$ is sufficiently close to the reduced model. We use $\alpha = 1.8553,\ \sigma = 1.362$ in the reduced model for all values of force F. Table (2.1) shows the integration time-step and total time for various simulations of the full model and the reduced model.

Figure 2.2 depicts the convergence of the correlation function and the stationary density for $F = 10,\ \gamma = 2$. Figure 2.3 shows similar results for different values of force $F = 8,\ 12,\ 24$ and damping $\gamma = 2$.

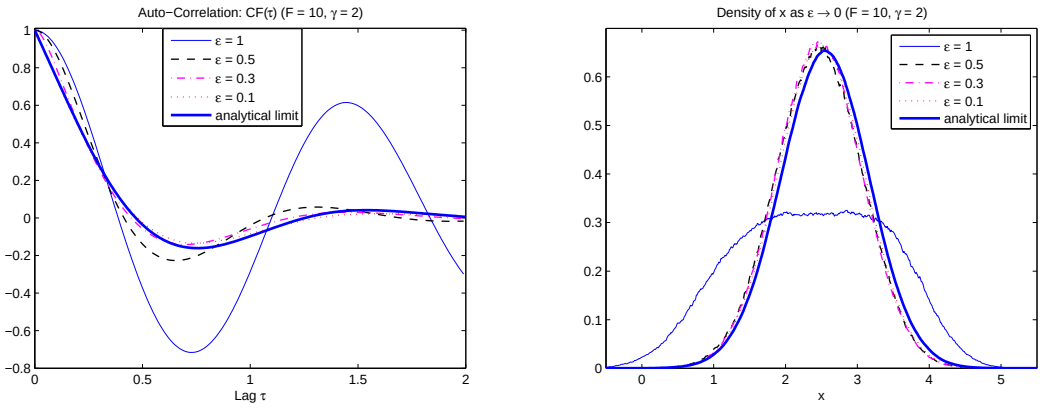

Figure 2.2: Left part - Convergence, as $\epsilon \to 0$, of auto-correlation function of x_k in full model (2.2) to the stationary auto-correlation function of X_k in reduced model (2.29). Right part - Convergence, as $\epsilon \to 0$, of the stationary density of x_k in full Model (2.2) to the stationary density of X_k in reduced model (2.29). Both parts are for parameters given by (4.5) and $F = 10, \gamma = 2$. Note in left part that $\epsilon = 0.1$ and analytical limit nearly overlap and thus, are not distinctively visible in the figure.

2.4 Numerical Method

We use a split-step method to integrate the full model (2.2) and stochastic reduced model (derived in section 2.3). We use the second-order Runge Kutta integrator for the deterministic part of the model and then use Euler discretization to add a Gaussian random variable which approximates dW, i.e., the increment of the Brownian motion. Numerical simulations of the full L96 model (2.2) get computationally expensive for small values of the scale separation parameter due to both the large number of variables and the necessity to choose a small discretization time-step in order to resolve the fast components of the

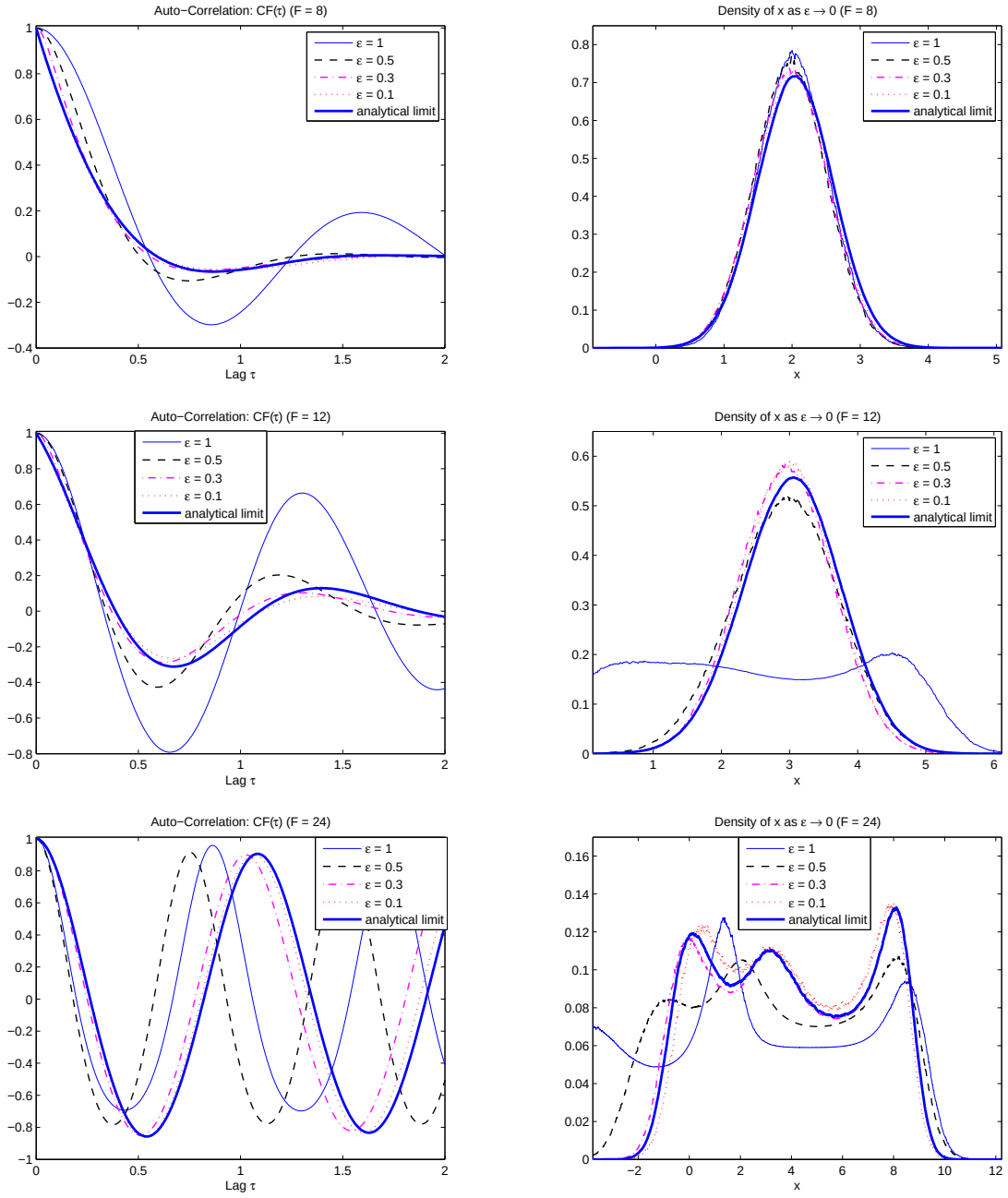

Figure 2.3: Convergence, as $\epsilon \to 0$, of auto-correlation function and density of x_k in full model (2.2) to the stationary auto-correlation function of X_k in reduced model (2.29). Left part - Stationary auto-correlation function, Right part - Stationary density. Top part - Force $F = 8$, damping $\gamma = 2$. Middle part - Force $F = 12$, damping $\gamma = 2$. Bottom part - Force $F = 24$, damping $\gamma = 2$.

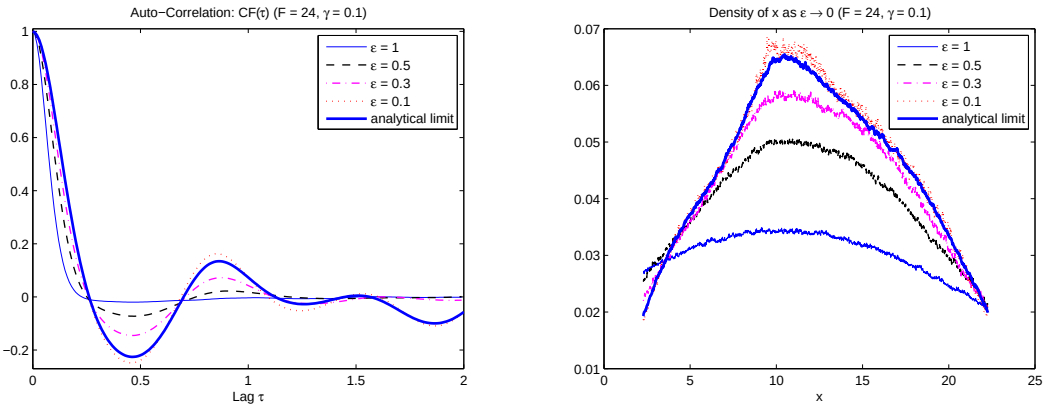

Figure 2.4: Left part - Convergence, as $\epsilon \to 0$, of auto-correlation function of x_k in full model (2.2) to the stationary auto-correlation function of X_k in reduced model (2.29). Right part - Convergence, as $\epsilon \to 0$, of the stationary density of x_k in full Model (2.2) to the stationary density of X_k in reduced model (2.29). Both parts are for parameters given by (4.5) and $F = 24, \gamma = 0.1$. Note in left part that $\epsilon = 0.1$ and analytical limit nearly overlap and thus, are not distinctively visible in the figure.

dynamics. Let δt represent the integration time-step for the trajectory, T represent the averaging time window for computing long-term time averages, and T_0 be the initial spin-up time, i.e. time skipped before computing averages to let the numerical trajectory reach stationary state. Table 2.1 provide values of δt, T and T_0 for each simulations of the full model for considered values of ϵ and the reduced model. Also, we specify the approximate running time of each simulation. All the trajectories of the full model and reduced model are simulated using Intel Core i5-2400 CPU @ 3.10 GHz and only one cpu core, i.e. no parallel computing.

Table 2.1: Description of numerical simulations of full model and reduced model and time taken for each simulation

Model	ϵ	Integration time step δt	Time T	Skipped time T_0	Approximate time for the simulation
Full Model	1	10^{-4}	30000	5000	4 hrs.
	0.5	2.5×10^{-5}	30000	5000	12 hrs.
	0.3	9×10^{-6}	30000	5000	1 day
	0.1	10^{-6}	10000	2000	5 days
	0.05	2.5×10^{-7}	10000	2000	10 days
Reduced Model	-	10^{-5}	50000	10000	6 hrs.

2.5 Parameter Estimation: Approximate Maximum Likelihood Approach

In many realistic situations, it is desirable to capture the statistical features of high-dimensional multi-scale dynamics by fitting a low-dimensional (single-scale) system of SDEs to the observed dataset. In such situations, generally the available data is only of large-scale variables since it is computationally expensive to obtain the small-scale data. Our main objective is to develop accurate and efficient estimation techniques, for parametric fitting SDEs to the observed data set, when the data of the large-scale variables is generated by multi-scale dynamics.

In this section, we derive estimators $\hat{\alpha}_{mle}$ and $\hat{\sigma}_{mle}$ of parameters α and σ, respectively, using approximate Maximum Likelihood approach (MLE) on the reduced model and use the true values of α and σ in (2.30) to test their accuracy. These estimators depend on empirical estimates of stationary moments of the data. Although the estimators are derived from the reduced model, the goal is to analyze the performance of $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$ when the data of the x_k variables from the full model at fixed $\epsilon > 0$ is used in computing the parameter values. We introduce ϵ in the notations of approximate Maximum Likelihood estimators as $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$ to explicitly emphasize their dependence on the observations of slow

variables x_k in full model. We refer to this framework as Indirect Observability, since the process X_k is not observable. The only data which are available is generated by x_k which is an approximation for X_k.

Reduced model: We introduce new notation $X_k(t)$ which represents variable X_k at time t. The reduced model in (2.29) can be rewritten as,

$$\begin{aligned} dX_k(t) &= X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))dt - (\gamma + \alpha)X_k(t)dt \\ &\quad + Fdt + \sigma dW_k, \end{aligned} \tag{2.36}$$

where $k = 1$ to K and W_k are independent Brownian motions. We can apply the approximate Maximum Likelihood technique to each X_k separately since the damping and diffusion coefficient matrices are assumed to be diagonal. The damping matrix, in (2.36), is αI and the diffusion matrix is σI where I is the $K \times K$ identity matrix. The transition density of $X_k(t)$ is not available analytically, so we compute an approximation to the Likelihood function based on the discretized version of the SDE.

In particular, given an initial value $X_k(t)$, we use the *Euler 1.0 discretization scheme* to approximate the solution $X_k(t + \Delta)$ of the reduced model (2.36) over a small time interval $[t, t + \Delta]$ by,

$$\begin{aligned} X_k(t + \Delta) &= X_k(t) + X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))\Delta - (\gamma + \alpha)X_k(t)\Delta + F\Delta \\ &\quad + \sigma\Delta W_k(t) + O(\Delta), \end{aligned} \tag{2.37}$$

where $\Delta W_k(t) = W_k(t + \Delta) - W_k(t)$ represents the increment of the Brownian motion over the time-interval $\Delta > 0$. Note that in the L96 model (2.2) and the reduced model (2.29), we have additive noise and hence, Euler 0.5 order discretization scheme is equivalent to Euler 1.0 order discretization scheme for our work.

Since increments of the Brownian motion are Gaussian, equation (2.37) gives a Gaussian approximation to the exact conditional transition density function associated with $X_k(t)$.

Rewriting the discretized equation in (2.37) as

$$\begin{aligned}\sigma(\Delta W_k(t)) &\approx X_k(t+\Delta) - X_k(t) + X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))\Delta \\ &\quad -(\gamma+\alpha)X_k(t)\Delta + F\Delta, \end{aligned} \tag{2.38}$$

we can construct the Likelihood function given the discrete observations of $X_k(t)$. The Likelihood estimators computed using the discretized equation (2.37) are in general not consistent (biased) for a fixed subsampling time-step. Hence, in the estimation of MLE we need to consider a small subsampling time-step $\Delta > 0$. To test estimation procedure we compare the approximate Maximum Likelihood estimators with the true values of parameters in (2.30).

We repeat that the main objective is to fit a stochastic model on an observed data so we assume that we have total of $N = T/\Delta + 1$ discretely sampled observations at equidistant time-steps, $0 = t_0 < t_1 < .. < t_N = T$, of length $\Delta > 0$, for each of the X_k, $k = 1 \dots K$, variable. Using the fact that the right hand side of the discretized equation, in (2.38), follows a Normal distribution with mean zero and variance $\sigma^2\Delta$, we can write the log-likelihood function as

$$\begin{aligned}\mathbb{L}_N(\alpha,\sigma) &= N\log\left(\frac{1}{\sqrt{2\pi}\sigma\sqrt{\Delta}}\right) + \frac{1}{2\sigma^2\Delta}\sum_{t=0:\Delta:T}\big(X_k(t+\Delta) - X_k(t) \\ &\quad - \left(X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right) - (\gamma+\alpha)X_k(t) + F\right)\Delta\big)^2. \end{aligned} \tag{2.39}$$

The principle of the approximate Maximum Likelihood is to maximize the probability of the observed data. This is achieved by maximizing the log-likelihood function with respect to the values of the unknown parameters. Hence, we search for the values of α and σ such that the gradient of log-likelihood function with respect to the parameters α and σ is zero. This determines a system of two equations for α and σ whose solutions provide an expression for the estimators denoted as $\hat{\alpha}_{mle}$ and $\hat{\sigma}_{mle}$, depending on the observations

and N, Δ. Taking partial derivative of the log-likelihood $\mathbb{L}_N(\alpha, \sigma)$ (2.39) w.r.t. parameter α gives the following relation

$$\begin{aligned}\hat{\alpha}_{mle} &= -\frac{\sum_{t=0:\Delta:T}\left(X_k(t+\Delta)X_k(t) - X_k^2(t)\right)}{\sum_{t=0:\Delta:T} X_k^2(t)\Delta} + \\ &\quad \frac{\sum_{t=0:\Delta:T}\left(X_{k-1}(t)X_k(t)\left(X_{k+1}(t) - X_{k-2}(t)\right) - \gamma X_k^2(t) + FX_k(t)\right)}{\sum_{t=0:\Delta:T} X_k^2(t)}. \quad (2.40)\end{aligned}$$

Assume that number of observations N in the given large-scale data is large enough to reach the stationary stage of the L96 model. Also, recall that the all X_k variables are symmetric in the reduced model since the reduced model is also invariant under the index shift. Hence, the third-moments will cancel in the above equation, since they are the identical third-moments shifted by one index. Therefore, in (2.40),

$$\sum_{t=0:\Delta:T} X_{k-1}(t)X_k(t)\left(X_{k+1}(t) - X_{k-2}(t)\right) \to 0, \quad \text{as } N \to \infty.$$

Thus, MLE of the parameter α in (2.40) can be rewritten as

$$\hat{\alpha}_{mle} = -\frac{\sum_{t=0:\Delta:T}\left(X_k(t+\Delta)X_k(t) - X_k^2(t)\right)}{\sum_{t=0:\Delta:T} X_k^2(t)\Delta} - \gamma + F\frac{\sum_{t=0:\Delta:T} X_k(t)}{\sum_{t=0:\Delta:T} X_k^2(t)}. \quad (2.41)$$

Similarly, taking the partial derivative of the log-likelihood function in (2.39) w.r.t. σ, assuming parameter α as fixed, gives

$$\begin{aligned}\hat{\sigma}_{mle}^2 &= -\frac{1}{N\Delta}\sum_{t=0:\Delta:T}\left(X_k(t+\Delta) - X_k(t) - \left(X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right) - \right.\right. \\ &\quad \left.\left.(\gamma+\alpha)X_k(t) + F\right)\Delta\right)^2. \quad (2.42)\end{aligned}$$

In the above equation for σ, we expand the terms on the right-hand side of the equation

and neglect higher-order terms w.r.t. Δ, thus, we obtain,

$$\begin{aligned}\hat{\sigma}^2_{mle} &= -\frac{2}{N\Delta}\sum_{t=0:\Delta:T}\left(X_k(t+\Delta)X_k(t) - X_k^2(t)\right) - \frac{2}{N}\sum_{t=0:\Delta:T}\left(X_k(t+\Delta) - X_k(t)\right) \\ &\quad \left(X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right) - (\gamma+\alpha)X_k(t) + F\right) + O(\Delta). \end{aligned} \tag{2.43}$$

Note that $(X_k(t+\Delta) - X_k(t))$ is $\mathrm{O}(\Delta)$, which can be observed by looking at the discretized equation given in (2.37). Hence, the expression (2.43) for $\hat{\sigma}^2$ can be further reduced to provide asymptotically equivalent estimators for small subsampling time-step Δ as

$$\hat{\sigma}^2_{mle} = -\frac{2}{N\Delta}\sum_{t=0:\Delta:T}\left(X_k(t+\Delta)X_k(t) - X_k^2(t)\right). \tag{2.44}$$

Lastly, corresponding to a discrete centered stationary process $X_k = \{X_k(t), t = 0:\Delta:T\}$ for a fixed arbitrary value of $k = 1$ to K, define the standard empirical covariance estimator $\hat{r}_1(N,\Delta)$, second-moment estimator $\hat{r}_0(N,\Delta)$ and mean estimator $\hat{\mu}(N,\Delta)$, as

$$\begin{aligned}\hat{r}_1(N,\Delta) &= \frac{1}{N}\sum_{t=0:\Delta:T} X_k(t+\Delta)X_k(t), \\ \hat{r}_0(N,\Delta) &= \frac{1}{N}\sum_{t=0:\Delta:T} X_k^2(t), \\ \hat{\mu}(N,\Delta) &= \frac{1}{N}\sum_{t=0:\Delta:T} X_k(t), \end{aligned} \tag{2.45}$$

where N is total number of observations in the data for each X_k variable, given by $N = T/\Delta + 1$. Then, the approximate Maximum Likelihood estimators of α and σ based on the observed data set and functions of total number of observations N and small subsampling time-step Δ is given by

$$\begin{aligned}\hat{\alpha}_{mle} &= -\frac{\hat{r}_1 - \hat{r}_0}{\hat{r}_0\Delta} - \gamma + F\frac{\hat{\mu}}{\hat{r}_0}, \\ \hat{\sigma}^2_{mle} &= -\frac{2}{\Delta}\left(\hat{r}_1 - \hat{r}_0\right), \end{aligned} \tag{2.46}$$

where $\hat{r}_1 = \hat{r}_1(N,\Delta), \hat{r}_0 = \hat{r}_0(N,\Delta)$ and $\hat{\mu} = \hat{\mu}(N,\Delta)$ are the standard empirical estimators defined in (2.46).

The auto-correlation of X_k variable in reduced model for time lag $\tau > 0$ is defined as

$$\text{Corr}_{X_k}(\Delta) = \frac{\mathbb{E}\big[(X_k(t)-\mu)(X_k(t+\Delta)-\mu)\big]}{\mathbb{E}\big[\,(X_k(t)-\mu)^2\big]}. \tag{2.47}$$

For the observed data $X_k(t), t = 0 : \Delta : T$, the empirical auto-correlation of X_k variable is equivalent to

$$\text{Corr}_{X_k}(\Delta) \approx \frac{\hat{r}_1 - \hat{\mu}^2}{\hat{r}_0 - \hat{\mu}^2}. \tag{2.48}$$

Getting motivation from the above correlation formula and using $log(1-x) \approx x$, the approximate Maximum Likelihood estimator of α can be rewritten for small Δ as

$$\hat{\alpha}_{mle} \;=\; -\frac{\hat{r}_0 - \hat{\mu}^2}{\hat{r}_0 \Delta} \log\left(\frac{\hat{r}_1 - \hat{\mu}^2}{\hat{r}_0 - \hat{\mu}^2}\right) - \gamma + F\frac{\hat{\mu}}{\hat{r}_0}. \tag{2.49}$$

We can rewrite approximate Maximum Likelihood estimator of σ^2 as a function of $\hat{\alpha}_{mle}$ by comparing the two equations in (2.46) as

$$\hat{\sigma}^2_{mle} \;=\; 2\left((\hat{\alpha}_{mle}+\gamma)\hat{r}_0 - F\hat{\mu}\right). \tag{2.50}$$

Recall that although the approximate Maximum Likelihood estimators are derived from equation of reduced model (2.29), our main objective is to estimate parameters based on data set of x_k in full model. Hence, next we summarize the formulas of approximate Maximum Likelihood estimators based on observed data set of x_k in full model, having multi-scale dynamics.

Summary of approximate Maximum Likelihood estimators under Indirect Observability

Assume we have been given a discrete set of observations of x_k from the full model (2.2), $\{x_k(t),\ t = 0 : \Delta : T,\ k = 1 \ldots K\}$, with a subsampling time-step $\Delta > 0$ at a fixed $\epsilon > 0$. Our objective is to develop accurate and efficient estimation techniques for fitting the effective SDE on this data set. The approximate Maximum Likelihood estimators for

parameters α and σ in the reduced model are given in (2.49) and (2.50), respectively. These estimators depend on two parameters: total number of observations N and subsampling time-step $\Delta > 0$. The goal is to understand the behavior of these estimators when we use the data of the slow variables in the full model in the estimation procedure. To this end, we define the empirical mean and covariance of the slow variables in the full model, for fixed $\epsilon > 0$, as

$$\hat{r}_1^\epsilon = \frac{1}{N}\sum_{t=0:\Delta:T} x_k(t+\Delta)x_k(t), \quad \hat{r}_0^\epsilon = \frac{1}{N}\sum_{t=0:\Delta:T} x_k^2(t), \quad \hat{\mu}_\epsilon = \frac{1}{N}\sum_{t=0:\Delta:T} x_k(t), \tag{2.51}$$

where $N = T/\Delta + 1$ is the total number of observations. Using the multi-scale data from the full model in the estimators (2.49) and (2.50) is equivalent to substituting the empirical moments (2.51) into the expressions of both estimators. Thus the estimators become

$$\begin{aligned} \hat{\alpha}_{mle}(\epsilon) &= -\frac{\hat{r}_0^\epsilon - \hat{\mu}_\epsilon^2}{\hat{r}_0^\epsilon \Delta}\log\left(\frac{\hat{r}_1^\epsilon - \hat{\mu}_\epsilon^2}{\hat{r}_0^\epsilon - \hat{\mu}_\epsilon^2}\right) - \gamma + F\frac{\hat{\mu}_\epsilon}{\hat{r}_0^\epsilon}, \\ \hat{\sigma}_{mle}^2(\epsilon) &= 2\big((\hat{\alpha}_{mle}(\epsilon)+\gamma)\hat{r}_0^\epsilon - F\hat{\mu}_\epsilon\big). \end{aligned} \tag{2.52}$$

where we have introduced ϵ into the formulas for the estimators to explicitly emphasize their dependence on the multi-scale data.

2.5.1 Necessary conditions for consistency of approximate Maximum Likelihood estimators using the data of reduced model

Consistency is one of the basic properties of any estimator. Although, the approximate Maximum Likelihood estimators are derived using Euler discretization of the reduced model, the main goal is to test the accuracy of $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$ when computed using observations of x_k in full model at fixed ϵ. This concept is referred to as Indirect Observability, i.e. estimating coefficients in the reduced stochastic model from the time series of an approximated process (slow variables in full model). Indirect Observability always introduce a mismatch between the estimated model and data, since slow variables are only a subset of full L96

model. The data recording the dynamics of the slow variables x_k are non-Markovian and multi-scale effects from the fast variables $\vec{y}$ can lead to inconsistencies in reduced stochastic modeling of the large-scale structures. Therefore, to ensure that the only contributions to the bias in approximate Maximum Likelihood estimators are from (i) Indirect Observability, (ii) finite number of observations N and subsampling time-step Δ, we begin by showing the consistency of the estimators under the Direct Observability.

In this subsection, we show the consistency of approximate Maximum Likelihood estimators under the Direct Observability concept, i.e., estimate coefficients in the reduced model using the time series of the reduced model. We use the "true" values of parameters, given by (2.30), to generate a time series of X_k in the reduced model and use that data to compute approximate Maximum Likelihood estimators.

Consistency of drift estimator $\hat{\alpha}_{mle}$ in the context of Direct Observability

Given discrete observations of X_k from the reduced model, we prove the consistency of the drift estimator $\hat{\alpha}_{mle}$, given by (2.52). We derive analytical expression of auto-correlation of X_k in the reduced model for small subsampling time-step $\Delta \to 0$ and then use that to show the convergence of $\hat{\alpha}_{mle}$ to true value α under specific conditions.

Let us be given observations as $\{X_k(t),\ t = 0 : \Delta : T\}$ for each $k = 1 \dots K$, where $\Delta > 0$ is a fixed subsampling time-step. The estimator $\hat{\alpha}_{mle}$ in (2.52) is the function of the number of observations $N = T/\Delta + 1$ and the subsampling time-step $\Delta > 0$. We derive conditions on the parameters N and Δ such that the estimator $\hat{\alpha}_{mle}$ is asymptotically consistent estimator of α, if computed based on observations of reduced model.

Theorem 2.5.1. ***Analytical correlation of the reduced model for small time lag $\Delta > 0$.*** *Analytical auto-correlation of the large-scale variable X_k in the reduced stochastic*

model (2.29) for small lag Δ *(or first-order expansion in* Δ*) is*

$$Corr_{X_k}(\Delta) = 1 + (F\mu - (\gamma + \alpha)r_0)\frac{\Delta}{r_0 - \mu^2}, \tag{2.53}$$

where $r_0 = \left\langle X_k^2 \right\rangle, \mu = \langle X_k \rangle$.

Proof. Reduced model, derived in (2.29), is given by,

$$\begin{aligned} dX_k(t) &= X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))dt - (\gamma + \alpha)X_k(t)dt \\ &\quad + Fdt + \sigma dW_k(t), \end{aligned} \tag{2.54}$$

where $k = 1$ to K, $X_k(t)$ represents the variable X_k value at time t and W_k are independent Brownian motions.

Given an initial value $X_k(t)$, we discretize the model in (2.54), using the *Euler 1.0 order scheme*, to approximate the solution $X_k(t)$ over a small time interval $[t, t + \Delta]$ as

$$\begin{aligned} X_k(t + \Delta) &= X_k(t) + X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right)\Delta - (\gamma + \alpha)X_k(t)\Delta \\ &\quad + F\Delta + \sigma(\Delta W_k(t)) + O(\Delta), \end{aligned} \tag{2.55}$$

where $\Delta W_k(t) = W_k(t + \Delta) - W_k(t)$ represents the increment of the Brownian motion over the time interval $[t, t + \Delta]$. Note that the discretization using Euler scheme of order 0.5 and 1.0 will represent same discretization for reduced model (2.29) since the noise is additive in the model.

Multiplying both sides of equation (2.55) by $X_k(t)$ and averaging both sides with respect to the invariant measure of $X_k(t)$, we obtain:

$$\begin{aligned} \langle X_k(t + \Delta)X_k(t)\rangle &= \left\langle X_k^2(t)\right\rangle + \langle X_k(t)X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right)\rangle \Delta \\ &\quad -(\gamma + \alpha)\left\langle X_k^2(t)\right\rangle \Delta + F\left\langle X_k(t)\right\rangle \Delta \\ &\quad +\sigma\left\langle X_k(t)\Delta W_k(t)\right\rangle \Delta + O(\Delta), \end{aligned} \tag{2.56}$$

where the third-moment, $\langle X_k(t)X_{k-1}(t)\left(X_{k+1}(t)-X_{k-2}(t)\right)\rangle = 0$ since $X_k(t)$ variables in the reduced model (2.29) are invariant under the index shift. Also, $\langle X_k \Delta W_k(t)\rangle = 0$ since $X_k(t)$ is independent of the increment of the Brownian motion $\Delta W_k(t)$. Therefore, expression in (2.56) is simplified as

$$\langle X_k(t+\Delta)X_k(t)\rangle - \left\langle X_k^2(t)\right\rangle \quad = \quad \left(F\left\langle X_k(t)\right\rangle - (\gamma+\alpha)\left\langle X_k^2(t)\right\rangle\right)\Delta + O(\Delta).$$

Defining the covariance function as $r_1 = \langle X_k(t+\Delta)X_k(t)\rangle$, second-moment as $r_0 = \left\langle X_k^2(t)\right\rangle$ and mean of large-scale variable as $\mu = \langle X_k(t)\rangle$ for any fixed arbitrary value of $k = 1$ to K, equation in (2.57) can be rewritten as

$$r_1 - r_0 \quad = \quad (F\mu - (\gamma+\alpha)r_0)\,\Delta + O(\Delta). \tag{2.57}$$

Auto-correlation of $X_k(t)$ for time lag Δ is defined as:

$$\mathrm{Corr}_{X_k}(\Delta) = \frac{r_1 - \mu^2}{r_0 - \mu^2}. \tag{2.58}$$

where r_1 is the lagged auto-covariance of $X_k(t)$ at time lag Δ and r_0, μ are stationary second and first-moments of $X_k(t)$ respectively. Comparing the expressions in equations (2.57), (2.58) and using the fact that the expression in (2.57) can be rewritten as

$$r_1 - r_0 = 1 + \frac{r_1 - r_0}{r_0 - \mu^2},$$

we conclude that the auto-correlation of $X_k(t)$ for small time lag Δ is

$$\mathrm{Corr}_{X_k}(\Delta) \approx 1 + (F\mu - (\gamma+\alpha)r_0)\,\frac{\Delta}{r_0 - \mu^2}. \tag{2.59}$$

□

Theorem 2.5.2. ***Necessary conditions for consistency of estimator*** $\hat{\alpha}_{mle}$ ***using the data of the reduced model***

Consider an observed variable data from the reduced model, (2.29)*, i.e., given* N *observations,* $\{X_k(t), t = 0 : \Delta : T\}$*, for each* $k = 1$ *to* K*, where* $\Delta > 0$ *is fixed subsampling time-step and* $N = T/\Delta + 1$*. Define the estimator* $\hat{\alpha}_{mle}$ *by formula derived in* (2.52)*. Then, under the conditions*

$$\Delta \to 0, N \to \infty \text{ and } N\Delta \to \infty, \tag{2.60}$$

the estimator $\hat{\alpha}_{mle}$ *is an asymptotically consistent estimator of* α *in the reduced model,* (2.29)*.*

Proof. In theorem (2.5.1) we derived first-order expansion of analytical auto-correlation of $X_k(t)$ for small lag $\Delta > 0$ in the reduced model (2.29). Taking logarithm of that expression, we obtain:

$$\log\left(\frac{r_1 - \mu^2}{r_0 - \mu^2}\right) = \log\left(1 + (F\mu - (\gamma + \alpha)r_0)\frac{\Delta}{r_0 - \mu^2}\right). \tag{2.61}$$

Expanding logarithm function using Taylor series, the expression in (2.61) becomes

$$\log\left(\frac{r_1 - \mu^2}{r_0 - \mu^2}\right) = \big(F\mu - (\gamma + \alpha)r_0\big)\frac{\Delta}{r_0 - \mu^2} + O(\Delta^2), \tag{2.62}$$

where $\mathrm{O}(\Delta^2)$ represents higher order terms in Δ. Re-arranging equation (2.62) to get an expression for α, we obtain

$$\alpha = \frac{r_0 - \mu^2}{\Delta r_0}\log\left(\frac{r_1 - \mu^2}{r_0 - \mu^2}\right) - \gamma + F\frac{\mu}{r_0} + O(\Delta^2). \tag{2.63}$$

Taking the limit in above expression,we obtain

$$\frac{r_0 - \mu^2}{\Delta r_0}\log\left(\frac{r_1 - \mu^2}{r_0 - \mu^2}\right) - \gamma + F\frac{\mu}{r_0} \to \alpha \qquad \text{as } \Delta \to 0. \tag{2.64}$$

This limit of the expression in (2.64) shows that if we know stationary covariance and moments of variable $X_k(t)$, i.e., if we know r_1, r_0 and μ, then expression in (2.64) connecting moments of $X_k(t)$ converges to the true value of α in the reduced model (2.29) as subsampling time-step $\Delta \to 0$.

Given observed data of X_k in the reduced model, (2.29), we can obtain empirical covariance $\hat{r}_1$ and moments $\hat{r}_0, \hat{\mu}$ of $X_k(t)$, which depend on the number of observations N and fixed subsampling time-step $\Delta > 0$ of the data. For each fixed $\Delta > 0$, the covariance estimator $\hat{r}_1(N, \Delta)$ is consistent and asymptotically efficient estimator of r_1 as $N \to \infty$. Similar result holds for moment estimators $\hat{r}_0$ and $\hat{\mu}$, which concludes that the estimator $\hat{\alpha}_{mle}$, (2.52), is an asymptotically consistent estimator of the parameter α as time-step $\Delta \to 0$. Hence, we proved that based on observed data set of the reduced model, i.e. Direct Observability, the estimator $\hat{\alpha}_{mle}(\epsilon)$ given by (2.52) is asymptotically consistent estimator of α under the conditions

$$\Delta \to 0,\ N \to \infty,\ N\Delta \to \infty.$$

□

Recall that the main objective was to study estimator $\hat{\alpha}_{mle}(\epsilon)$ computed using the data set of x_k from the full model, (2.2), i.e. under Indirect Observability. It is not feasible to prove the consistency of $\hat{\alpha}_{mle}(\epsilon)$ analytically, thus, we will investigate numerically the behavior of $\hat{\alpha}_{mle}$ in later sections.

Necessary conditions for consistency of diffusion estimator $\hat{\sigma}_{mle}$ using the data of the reduced model

Assume to be given an observed data from the reduced model, (2.29). Given $\{X_k(t),\ t = 0 : \Delta : T\}$ at fixed subsampling time-step $\Delta > 0$, with the total $N = T/\Delta + 1$ observations for each $k = 1$ to K. In previous subsection 2.5.1, we proved that the estimator $\hat{\alpha}_{mle}$ computed from the reduced model data is a consistent estimator of parameter α under conditions $\Delta \to 0,\ N\Delta \to \infty$, and $N \to \infty$. Similarly, we show the same conditions for consistency of estimator $\hat{\sigma}^2_{mle}$.

Theorem 2.5.3. ***Analytical relation between parameters α and σ using method***

of moments. *In the reduced model* (2.29) *parameters* σ *and* α *obey the following relation*

$$\sigma^2 = 2((\alpha+\gamma)r_0 - F\mu), \tag{2.65}$$

where $r_0 = \left\langle X_k^2(t)\right\rangle, \mu = \langle X_k(t)\rangle$ *and* $\langle . \rangle$ *is averaging under the invariant stationary measure of* $X_k(t)$.

Assume to be given $N = T/\Delta + 1$ *observations of* $X_k(t)$ *at a fixed subsampling time-step* $\Delta > 0$, *then the estimator* $\hat{\sigma}^2_{mle}$ *is a consistent estimator for parameter* σ^2 *under conditions* $\Delta \to 0, N\Delta \to \infty$ *and* $N \to \infty$.

Proof. The stochastic reduced model derived in section 2.3.2 is given by

$$\begin{aligned} dX_k(t) &= X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))dt - (\gamma+\alpha)X_k(t)dt \\ &\quad + Fdt + \sigma dW_k(t), \end{aligned} \tag{2.66}$$

where $k = 1$ to K and W_k are independent Brownian motions.

Let $f(X_k(t)) = X_k^2(t)$. Using Ito's lemma, derivative of function $f(X_k(t))$ is derived as

$$\begin{aligned} df(X_k(t)) &= \frac{\partial f}{\partial t}dt + \sum_{i=1}^{K} \frac{\partial f}{\partial X_i(t)} dX_i(t) \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{K} \frac{\partial^2 f}{\partial X_i(t)\partial X_j(t)} dX_i(t)dX_j(t). \end{aligned} \tag{2.67}$$

Since $f(X_k(t)) = X_k^2(t)$, partial derivatives of the function f required in (2.67) are given by

$$\frac{\partial f}{\partial t} = 0,$$

$$\frac{\partial f}{\partial X_i(t)} = \begin{cases} 2X_k(t), & \text{i = k,} \\ 0, & \text{i} \neq \text{k,} \end{cases}$$

and

$$\frac{\partial^2 f}{\partial X_i \partial X_j} = \begin{cases} 2, & \text{i, j = k,} \\ 0, & \text{i or j} \neq \text{k.} \end{cases}$$

Substituting above partial derivatives to (2.67) we obtain

$$df(X_k(t)) \;=\; 2X_k(t)dX_k(t) + \frac{1}{2}2(dX_k(t))^2, \tag{2.68}$$

where $dX_k(t)$ is given by (2.29). Using properties in the Ito's lemma as

$$(dt)^2 = 0,\ dW_k dt = 0$$

$$dW_k dW_{k'} = \begin{cases} dt, & \text{k = k'}, \\ 0, & \text{k} \neq \text{k'}, \end{cases}$$

and $dX_k(t)$ in reduced model (2.29), the expression in (2.68) becomes

$$\begin{aligned} df(X_k(t)) \;=\; & 2X_k(t)X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))dt - 2(\gamma+\alpha)X_k^2(t)dt \\ & +2FX_k(t)dt + 2\sigma X_k(t)dW_k + \frac{1}{2}2\sigma^2 dt. \end{aligned}$$

Averaging above expression for $d\left(X_k^2(t)\right)$ under the invariant stationary measure of $X_k(t)$, we obtain

$$\begin{aligned} \langle df(X_k(t))\rangle \;=\; & 2\left\langle X_k(t)X_{k-1}(t)(X_{k+1}(t) - X_{k-2}(t))\right\rangle dt - 2(\gamma+\alpha)\left\langle X_k^2(t)\right\rangle dt \\ & +2F\left\langle X_k(t)\right\rangle dt + 2\sigma\left\langle X_k(t)dW_k\right\rangle + \frac{1}{2}2\sigma^2 dt. \end{aligned} \tag{2.69}$$

We consider stationary moments, hence change in moment $\left\langle X_k^2(t)\right\rangle$ is considered zero, i.e. $\langle df(X_k(t))\rangle = 0$. Also, recall that equations for X_k in the reduced model are invariant under the index shift, hence,

$$\langle X_{k-1}(t)X_k(t)X_{k+1}(t)\rangle = \langle X_{k-2}(t)X_{k-1}(t)X_k(t)\rangle\,.$$

And lastly, since $X_k(t)$ and W_k are independent random variables and mean of change in Brownian motion is zero, we obtain

$$\langle X_k(t)dW_k\rangle = 0.$$

Hence, equation in (2.69) is simplified to

$$\begin{aligned} 0 = \left\langle d(X_k^2(t)) \right\rangle &= -2(\gamma+\alpha) r_0 dt + 2F\mu dt + \sigma^2 dt, \\ \Rightarrow \sigma^2 &= 2((\gamma+\alpha) r_0 - F\mu), \end{aligned} \tag{2.70}$$

where $r_0 = \left\langle X_k^2 \right\rangle, \mu = \langle X_k \rangle$ and $\langle \cdot \rangle$ is the averaging under the invariant measure of $X_k(t)$. The expression in (2.70) gives an explicit analytical relation between parameters σ and α in the reduced model (2.29).

Assume to be given an observed data of X_k from reduced model, (2.29). Consider $\{X_k(t),\ t = 0 : \Delta : T,\ k = 1 \dots K\}$, where $\Delta > 0$ is fixed subsampling time-step and $N = T/\Delta + 1$ is total number of observations for each k. We can obtain empirical moments $\hat{r}_0$ and $\hat{\mu}$ of the variable $X_k(t)$, which depend on the number of observations N and fixed subsampling time-step $\Delta > 0$ of the data. For each fixed $\Delta > 0$, as number of observations $N \to \infty$, the covariance estimator $\hat{r}_0(N, \Delta)$ is the consistent and asymptotically efficient estimator of r_0. Similar result holds for first-moment estimator $\hat{\mu}$. If true value of parameter α is known, then the only conditions for consistency of $\hat{\sigma}_{mle}$ is $N \to \infty$ under Direct Observability.

If true value of α is not known and need to be estimated using the given data of reduced model, then the consistency conditions of $\hat{\sigma}_{mle}$ depends on consistency conditions of $\hat{\alpha}_{mle}$ also. In theorem (2.5.2), we proved that as $N \to \infty, \Delta \to 0$ and $N\Delta \to \infty$, $\hat{\alpha}_{mle}$ is asymptotically consistent estimator of α. Hence, we conclude that if given the discrete observations of X_k in reduced model, $\hat{\sigma}^2_{mle}$ is an asymptotically consistent estimator of σ^2, under the conditions

$$\Delta \to 0,\ N\Delta \to \infty,\ N \to \infty$$

. □

Recall that the main objective is to compute the approximate Maximum Likelihood

estimator $\hat{\sigma}^2_{mle}(\epsilon)$ computed using the discrete observations of x_k from the full model. We investigate the consistency of approximate Maximum Likelihood estimators $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}^2_{mle}(\epsilon)$ numerically in later sections.

2.5.2 Relation between the approximate Maximum Likelihood estimators and the derivative of the auto-correlation functions of slow variables

In Theorem (2.5.1), we showed expansion of the analytical correlation of $X_k(t)$ in the reduced model for small time lag $\Delta > 0$. Let η be the coefficient of first-order of Δ term in the auto-correlation, (2.5.1). η can be also be considered equivalent to the derivative of the auto-correlation of $X_k(t)$ w.r.t. lag Δ as $\Delta \to 0$, i.e.,

$$\eta = \lim_{\Delta \to 0} \frac{\partial}{\partial \Delta} \text{Corr}_{X_k}(\Delta).$$

We will refer η to be the slope of auto-correlation of slow variables. Using the first-order expansion of auto-correlation derived in (2.5.1), we obtain

$$\begin{aligned} \eta &= \frac{F\mu - (\gamma + \alpha) r_0}{r_0 - \mu^2}, \\ \Rightarrow \alpha &= -\eta \frac{r_0 - \mu^2}{r_0} - \gamma + F\frac{\mu}{r_0}. \end{aligned} \tag{2.71}$$

which demonstrates a relation between parameter α and coefficient η of first-order Δ term in the auto-correlation of $X_k(t)$, as $\Delta \to 0$.

If we expand the auto-correlation of X_k used in estimator $\hat{\alpha}_{mle}$ to the first-order of Δ (i.e. neglect $\text{O}(\Delta^2)$ terms), then the expression for α in (2.71) is same as $\hat{\alpha}_{mle}$ (2.49).

Given the observations of slow variables x_k in full model for fixed $\epsilon > 0$ with small subsampling time-step Δ, approximate Maximum Likelihood estimator $\hat{\alpha}_{mle}(\epsilon)$ in (2.52)

can be rewritten as the function of the slope η^ϵ of auto-correlation of x_k as

$$\hat{\alpha}_{mle}(\epsilon) \quad = \quad -\eta^\epsilon \frac{\hat{r}_0^\epsilon - \hat{\mu}_\epsilon^2}{\hat{r}_0^\epsilon} - \gamma + F\frac{\hat{\mu}_\epsilon}{\hat{r}_0^\epsilon}. \tag{2.72}$$

We denote the slope of auto-correlation of x_k by η^ϵ to emphasize it's dependence on the full model.

In later sections, we will show the importance of slope η^ϵ while computing $\hat{\alpha}_{mle}(\epsilon)$ using observations of x_k in full model. Recall that in Indirect Observability, we are using observations of slow variables in full model to estimate the parameters for reduced model. Hence, if the slope of auto-correlation of slow variables in full model and reduced model are different for some time lag Δ then there might be an inconsistency in estimator $\hat{\alpha}_{mle}(\epsilon)$ computed using same time lag Δ. This inconsistency in approximate Maximum Likelihood estimator due to difference in auto-correlation of slow variables in full model and reduced model is referred as *Subsampling Issue*. We conclude that subsampling issue is one of the bias created due to Indirect Observability.

Similarly, we derive the relation between $\hat{\sigma}_{mle}(\epsilon)$ and the slope η^ϵ of auto-correlation of slow variables in full model. Comparing the analytical relation between parameters α and σ derived in equation (2.70) and (2.72), we obtain

$$\hat{\sigma}_{mle}^2(\epsilon) \quad = \quad -2\eta^\epsilon(\hat{r}_0^\epsilon - \hat{\mu}_\epsilon^2).$$

where η^ϵ is the slope of the auto-correlation of the slow variables x_k in the full model for fixed ϵ, equivalent to the derivative of the auto-correlation of $x_k(t)$ in (2.5.1) w.r.t. time-step Δ. Due to Indirect Observability and the difference in slopes of auto-correlation function of the slow variables in the full model and the reduced models, the subsampling issue also arises in estimating parameter σ.

2.5.3 Subsampling issue

The goal is to approximate the dynamics of key statistical descriptors of the turbulent high-dimensional, multi-scale L96 model by a closed form low-dimensional stochastic process, given by (2.29). In section 2.3.2 we derived a single-scale reduced model (2.29) with explicit values of parameters α, σ such that the stationary auto-correlation and density of slow variables x_k in the full model (2.2) converges to the corresponding stationary statistics of X_k in the reduced model (2.29) as $\epsilon \to 0$.

The main objective is to analyze estimators of parameters α and σ in the reduced model computed from the observations of the slow variables x_k in the full L96 model, also referred to as Indirect Observability. The parameter values derived in (2.30) depend on lagged covariance of $y_{j,k}$ variables. In section 2.5 we derive approximate Maximum Likelihood estimators $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$ which depend on stationary moments of x_k variables in the full model. We use the true parameters α and σ derived in (2.30) to test the approximate Maximum Likelihood estimators $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$, respectively. Note that $\hat{\alpha}_{mle}(\epsilon)$ in (2.52) depends on auto-correlation of x_k variable. We depict the difference of correlation of x_k in the full model and X_k in the reduced model for small correlation lag $\Delta \to 0$ which leads to the intuition for inconsistency of $\hat{\alpha}_{mle}(\epsilon)$ at small correlation lags.

We observe that at short time-scales, the trajectories of $x_k(t)$ in the full model (2.2) for even small ϵ are quite different from sample paths of $X_k(t)$ in the reduced SDE (2.29) but on longer time-scales the behavior of $x_k(t)$ in the full model as $\epsilon \to 0$ is well emulated by the behavior of $X_k(t)$ of the reduced model (2.29). The correlation function of x_k has very different curvature compared with the correlation function of X_k variable in the reduced model. This difference is due to the fact that trajectories of x_k in the full model are differentiable in time. On the other hand, trajectories of X_k in the reduced model (2.29) are not differentiable in time due to the presence of the Brownian motion in the equations

of X_k. In such scenario, the Maximum Likelihood estimator $\hat{\alpha}_{mle}(\epsilon)$, based on the discrete observations of the variable x_k, might not be consistent at certain correlation time lags Δ. Since estimator $\hat{\alpha}_{mle}(\epsilon)$, given by (2.52), depends on the auto-correlation of x_k, the difference in correlation of the full and the reduced model, for small lags will effect the consistency of the estimator $\hat{\alpha}_{mle}(\epsilon)$. Also, in (2.72), we derived a relation between $\hat{\alpha}_{mle}(\epsilon)$ and the slope η^ϵ of the auto-correlation of the slow variables. The difference in curvature of the auto-correlation of the slow variables x_k in the full and X_k in the reduced models will lead to difference in slope η of correlation for small time lag $\Delta \to 0$ concluding the inconsistency of $\hat{\alpha}_{mle}(\epsilon)$. In later section, we investigate numerically the conditions on the uniform subsampling time-step Δ such that as $\epsilon \to 0$, the estimator $\hat{\alpha}_{mle}(\epsilon)$ based on the observations from the full model converges to the true parameter value α.

Figure 2.2 showed the correlation of x_k in the full model converging to correlation of X_k in the reduced model as $\epsilon \to 0$. In this section, we show the same figure, but only for very small lags. In figure 2.5, we show the difference in correlation of X_k in the full model and corresponding correlation of the reduced model for lag close to zero and several small values of ϵ. We fix parameters as in (4.5), damping $\gamma = 2$ and force $F = 10$ in the full model (2.2). Then true parameters of reduced model are computed using homogenization are $\alpha = 1.8553$ and $\sigma = 1.362$. The full model is simulated for values of $\epsilon = 1, 0.5, 0.3, 0.1$.

Figure 2.5 shows that close to lag zero, correlation function of the large-scale x_k variable in the full model (2.2) is concave upward for all ϵ whereas the correlation of the reduced model (2.29) is concave downward. Therefore, the approximate Maximum Likelihood estimator $\hat{\alpha}_{mle}(\epsilon)$ will not be a consistent estimator of α if we use observations of the full model subsampled at a very small time lag. Also, the figure 2.5 clearly shows that the tangent of the correlation at $\Delta = 0$ is different for the slow variables in the full model (as $\epsilon \to 0$) and the reduced model. Recalling the relation between approximate Maximum Likelihood

estimator and slope η of the auto-correlation of the slow variables, given by (2.72), we conclude that $\hat{\alpha}_{mle}(\epsilon)$ is not consistent estimator for very small lag Δ. To obtain estimator $\hat{\alpha}_{mle}(\epsilon)$ value close enough to true α, we need to subsample the observations of the full model at an intermediate optimal time-step.

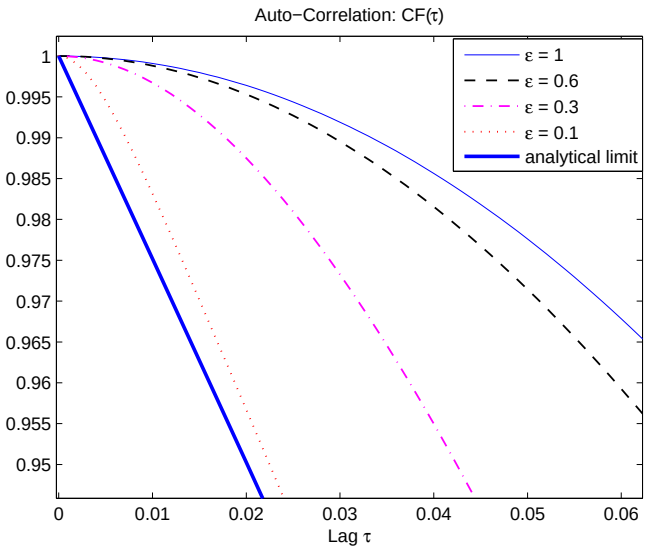

Figure 2.5: Correlation of x_k in the full model for $\epsilon = 1, 0.6, 0.3, 0.1$ compared to auto-correlation of X_k in reduced model, for small lag time $\tau > 0$. The parameters in full model are fixed as in (4.5), force $F = 10$ and damping $\gamma = 2$. Parameters of reduced model are derived using homogenization in (2.30) are given by $\alpha = 1.8553, \sigma = 1.362$. Bold solid line - auto-correlation of X_k in the reduced model and other lines represents the full model for various ϵ. This figure showed the difference in curvature of auto-correlation between full and reduced model.

In figure 2.6 we show the curvature difference in the correlation of x_k variable in the full and X_k in the reduced model more clearly by taking the log of correlation functions. For the same trajectories as taken in figure 2.5, figure 2.6 shows the expression $\Delta^{-1}\log(\mathrm{CF}_{x_k}(\Delta))$ which shows a clear difference in trajectories of x_k variable in the full and X_k in the reduced model for short time lags ($\Delta \to 0$). In figure 2.6, the full model is computed only for small $\epsilon = 0.05$. We would like to point out that the expression showed in figure 2.6 is the same

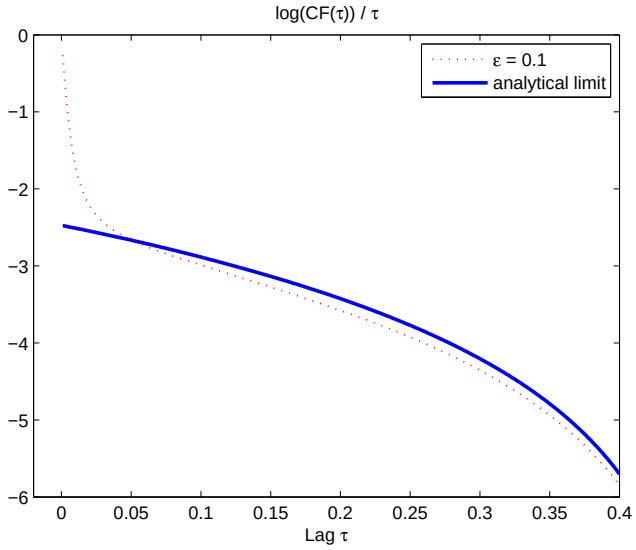

Figure 2.6: Expression $\Delta^{-1}\log(CF_{x_k})$ compared between x_k in the full model for small $\epsilon = 0.1$ and X_k in reduced model. For small $\tau = 0.001$, the comparison of given expression between both models gives 64% error. Dotted line - $\Delta^{-1}\log(CF_{x_k})$ computed using observations of x_k in the full model for $\epsilon = 0.01$, solid line - same expression computed on X_k in reduced model.

expression which is used in estimating parameter $\hat{\alpha}_{mle}(\epsilon)$ in (2.52). This gives a clear indication that computing estimator $\hat{\alpha}_{mle}(\epsilon)$ using discrete observations of the full model at small subsampling time lag will not give a correct estimator for parameter α in the reduced model.

2.5.4 Numerical comparison of approximate Maximum Likelihood estimator with true parameter under Indirect Observability

In this subsection, we compare $\hat{\alpha}_{mle}(\epsilon)$ with "true" value of α in (2.30) where $\hat{\alpha}_{mle}(\epsilon)$ is computed using discrete observations of x_k in the full model. The goal is to examine numerically whether $\hat{\alpha}_{mle}(\epsilon)$ is a good estimate of the true value of α using observations of x_k subsampled at small time-steps.

Numerical Setting: In the full model parameters are given by (4.5) and damping $\gamma = 2$, force $F = 10$. With these specific parameters, we have already shown in figure 2.2 that correlation of x_k in the full model converges to the correlation of X_k in the reduced model, with parameters $\alpha = 1.8553, \sigma = 1.362$. Moreover we observed that as $\epsilon \to 0$ and we observed that $\epsilon = 0.1$ is small enough scale separation parameter so that the full model is very close to the limiting reduced equation. For $\epsilon = 0.1$ in the full model, we simulate a trajectory of x_k using integration time-step $\delta t = 10^{-6}$ and total time $T = 10^3$. We use these discrete observations to compute estimator $\hat{\alpha}_{mle}(\epsilon)$ at different subsampled time lags Δ.

Results are presented in Figure 2.7. We observe two inconsistency regimes in the behavior of $\hat{\alpha}_{mle}(\epsilon)$. First regime is when subsampling time-step $\Delta < 0.004$. This regime is due to the subsampling issue created by the difference in auto-correlation of x_k in the full model and X_k in the reduced model for very small correlation lags. Second regime is when subsampling time-step $\Delta > 0.004$ due to the Euler discretization used to derive $\hat{\alpha}_{mle}(\epsilon)$.

Hence, we observe that $\hat{\alpha}_{mle}(\epsilon)$ diverges from the true value of α as Δ increases. Note that these both regimes are dependent on choice of ϵ in full model. For $\epsilon = 0.1$, we observe that $\hat{\alpha}_{mle}(\epsilon)$ is a correct estimator only if the discrete observations of x_k are subsampled at $\Delta = 0.004$.

Numerical results presented in this section implies that it is necessary to subsample the discrete observations of x_k from the full model at an optimal time-step to correctly estimate $\hat{\alpha}_{mle}(\epsilon)$. However, this optimal time-step will vary for different ϵ in the full model. Since it is impossible to find the optimal subsampling time-step analytically, the only solution is to observe $\hat{\alpha}_{mle}(\epsilon)$ numerically for a fixed ϵ and empirically determine the optimal subsampling time-step.

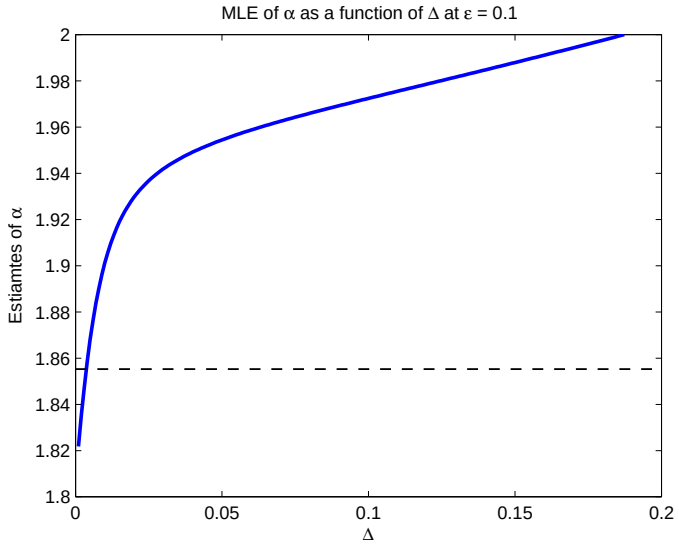

Figure 2.7: Approximate Maximum Likelihood estimator $\hat{\alpha}_{mle}(\epsilon)$ computed on discrete observations of x_k of the full model for $\epsilon = 0.1$. Estimator $\hat{\alpha}_{mle}(\epsilon)$ is compared with the true value of $\alpha = 1.8553$. Solid line - $\hat{\alpha}_{mle}(\epsilon)$ computed using the data of the full model with $\epsilon = 0.1$, dotted line - $\alpha = 1.8553$ derived using homogenization in (2.30). Figure shows that $\hat{\alpha}_{mle}(\epsilon)$ intersects true α for small correlation lag time which is considered "optimal time-step". However, there is potentially a large inconsistency in the estimator $\hat{\alpha}_{mle}(\epsilon)$ if computed using the observations subsampled at time-step Δ which is not optimal.

2.5.5 Optimal subsampling time-step for the approximate Maximum Likelihood estimator

In previous subsection, we showed that $\hat{\alpha}_{mle}(\epsilon)$ is an inconsistent estimator in two regimes if computed on discrete observations x_k, from the full model. First regime is when observations x_k are subsampled at a very dense time-step. The second regime for large Δ arises since $\hat{\alpha}_{mle}(\epsilon)$ is derived by applying the Maximum Likelihood technique on the Euler discretization of the reduced model.

Optimal subsampling time-step for fixed $\epsilon > 0$: Given observations $x_k,\ k = 1$ to K in the full model for the fixed $\epsilon > 0$, we subsample the data with various time-steps and can compute estimator $\hat{\alpha}_{mle}(\epsilon)$ with different subsampling time-steps Δ. Recall that the true value of α in (2.30) is numerically computed using observations of the fast subsystem (2.4) and can have small numerical errors. Hence, instead of just comparing $\hat{\alpha}_{mle}(\epsilon)$ with one explicit true value of α, we consider a 95% confidence interval to test the estimator $\hat{\alpha}_{mle}(\epsilon)$. We call a subsampling time-step as optimal for a fixed $\epsilon > 0$ if $\hat{\alpha}_{mle}(\epsilon)$ estimator lies in the 95% confidence interval around the true value α.

The 95% confidence interval for true parameter α: In subsection 2.3.3, we simulated 150 trajectories of the fast subsystem with various integration time-steps $\delta t > 0$, total number of observations N, covariance time lag $\Delta > 0$ and seed for random values. Therefore, we obtained 150 values of parameters α, σ. From the 150 trajectories, we obtained mean$(\alpha) =$ 1.8553 and standard deviation is std$(\alpha) = 0.0304$. Hence, the 95% interval for α using 150 trajectories of fast subsystem is $(1.7865, 1.9241)$. Therefore, the values of α in the range $(1.7865, 1.9241)$ are considered reliable to test the optimal subsampling time-step for estimator $\hat{\alpha}_{mle}$.

Given discrete observations of x_k in the full model for various $\epsilon > 0$, we compute estimator $\hat{\alpha}_{mle}(\epsilon)$ by subsampling observations at different time-step Δ. Figure 2.8 shows

comparison of the estimator $\hat{\alpha}_{mle}(\epsilon)$ for various ϵ and Δ, with the 95% confidence interval for the true value of α. In the full model, all parameters are fixed as in (4.5), force $F = 10$ and damping $\gamma = 2$. We consider discrete observations from the full model for $\epsilon = \{0.3, 0.2, 0.1\}$. For each ϵ, optimal subsampling time-step is considered as the range of subsampling time-step Δ such that $\hat{\alpha}_{mle}(\epsilon, \Delta)$ belongs to $(1.7865, 1.9241)$.

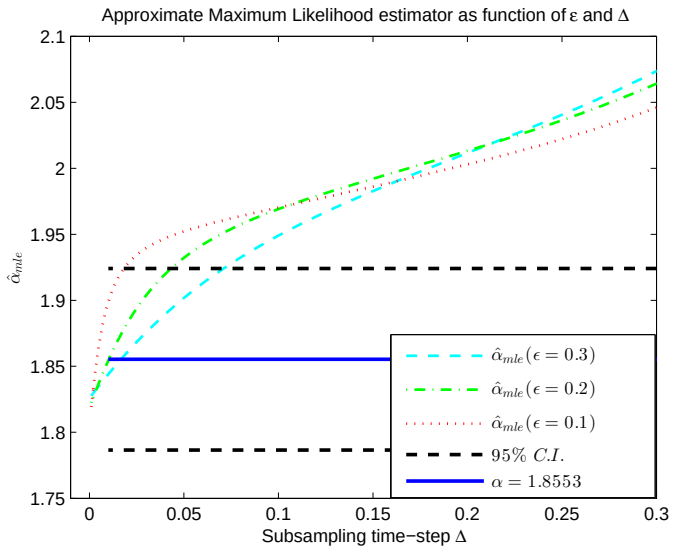

Figure 2.8: Estimator $\hat{\alpha}_{mle}(\epsilon)$ as function of the subsampling time-step Δ, computed using observations of x_k in the full model for various $\epsilon > 0$. Straight line represents the "true" value of α, given by (2.30), computed numerically using the fast subsystem. Top and bottom straight lines represent the 95% confidence interval, $(1.7865, 1.9241)$, for parameter α. Middle straight solid line - "true" value of α. For fixed $\epsilon \to 0$, $\hat{\alpha}_{mle}(\epsilon)$ is considered consistent estimator if it lies in 95% confidence interval of α.

Optimal subsampling time-step as $\epsilon \to 0$. We apply regression on optimal subsampling time-step for various ϵ to predict scaling of the optimal time-step as $\epsilon \to 0$. Figure 2.9 shows the optimal subsampling time interval for $\epsilon = \{0.03, 0.1, 0.2, 0.3, 0.6\}$, denoted as (Min_Δ, Max_Δ). Figure 2.9 shows that as $\epsilon \to 0$, the length of the optimal subsampling time-step scales as ϵ^2. Note that Min_Δ is same for all $\epsilon > 0$, but Max_Δ varies with value of ϵ. We use quadratic polynomial fitting on Max_Δ and numerically obtain an

expression of Max_Δ as function of ϵ

$$Min\_t = 0.0001,$$

$$Max\_t = 0.0042 + 0.1426\epsilon + 0.2233\epsilon.^2.$$

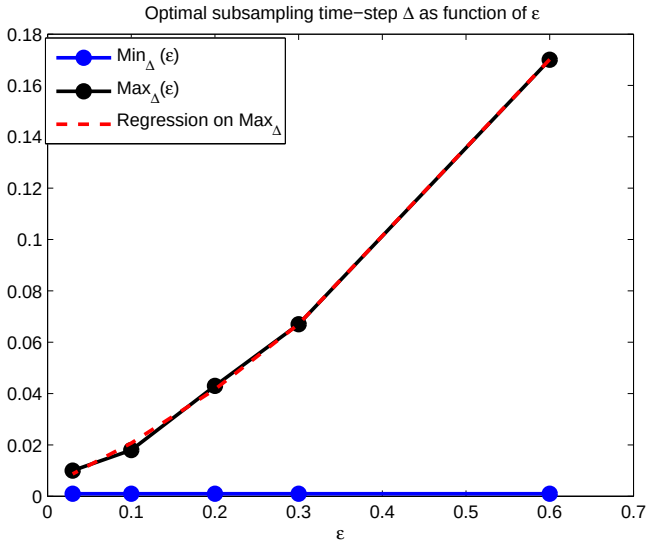

Figure 2.9: Optimal subsampling time-step range (Min_Δ, Max_Δ), of estimator $\hat{\alpha}_{mle}(\epsilon)$, varies with value of ϵ. The values of ϵ considered here are $\epsilon = \{0.03, 0.1, 0.2, 0.3, 0.6\}$. Regression is showed on curve of Max_Δ for varying value of ϵ. Solid line - Optimal subsampling time-step range (Min_Δ, Max_Δ) for $\hat{\alpha}_{mle}(\epsilon)$ as function of ϵ. Dotted line - Regression on Max_Δ.

Conclusion: An approximate Maximum Likelihood estimator is derived from the reduced model. We want to analyze the behavior of this estimator when $\hat{\alpha}_{mle}(\epsilon)$ is computed using the data of the slow variables in the full model. In section 4.2.1, we observed the subsampling issue in estimator $\hat{\alpha}_{mle}(\epsilon)$. We showed that $\hat{\alpha}_{mle}(\epsilon)$ is inconsistent estimator in two regimes. First regime is when subsampling time-step is very small. This regime is due to the subsampling issue created by the difference in auto-correlation of x_k in the full model and X_k in the reduced model for very small correlation lags. Second regime is when subsampling time-step is relatively large and this inconsistency is due to the Euler

discretization used to derive $\hat{\alpha}_{mle}(\epsilon)$. Main concern is that there is no feasible solution to find optimal subsampling time-step analytically for computation of consistent $\hat{\alpha}_{mle}(\epsilon)$. Therefore, our next goal is obtain an estimator which does not depend on subsampling time-step, i.e. it should not depend on two-point moments of given observations. In next section, we show a new parameter estimate of α which will not depend on subsampling time-step Δ and hence, it simplifies the complications of subsampling. It depends on only one-point moments of given observations.

2.6 Parameter Estimation using the Method of Moments

Given observations of the slow variables x_k from the full model, having multi-scale dynamics, our main goal is to fit a stochastic model which closely reproduces the statistical properties of the observed dataset. In this section, we derive a new estimator of α denoted as $\hat{\alpha}_{mom}$ using the method of moments. The moment estimator depends only on one-point stationary moments of the given observations. Hence, we expect this estimator to be more robust with respect to the subsampling time-step Δ.

Theorem 2.6.1. ***Derivation of moment estimator of parameter*** α**.** *In the reduced model given by* (2.29) *parameter* α *has the following relationship with one-point moments*

$$\alpha_{mom} = -\gamma + \frac{F+\rho}{\mu},$$

where ρ *is stationary mixed second-moment of* X_k*, defined as*

$$\rho = \langle X_{k-1}(X_{k+1} - X_{k-2})\rangle\,,$$

and μ *is stationary mean of* X_k *for any* $k = 1 \ldots K$*.*

Proof. Reduced model given in (2.29) is

$$\frac{d}{dt}X_k(t) \;=\; X_{k-1}(t)\,(X_{k+1}(t) - X_{k-2}(t)) - (\gamma+\alpha)X_k(t) + F + \sigma\dot{W}_k, \qquad (2.73)$$

where $k = 1 \ldots K$, W_k are independent Brownian motions and $X_k(t)$ represents the X_k value at time t.

Given an initial value $X_k(t)$, we discretize the reduced model (2.73), using the Euler-Maruyama scheme, to approximate the solution $X_k(t+\Delta)$ over a small time interval $[t, t+\Delta]$ as

$$\begin{aligned} X_k(t+\Delta) &= X_k(t) + X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right)\Delta - (\gamma+\alpha)X_k(t)\Delta \\ &\quad F\Delta + \sigma\Delta W_k(t) + O(\Delta), \end{aligned} \tag{2.74}$$

where $\Delta W_k(t)$ represents the increment of the Brownian motion.

Averaging both sides of the equation (2.74) with respect to the invariant measure of X_k, we obtain

$$\begin{aligned} \langle X_k(t+\Delta)\rangle &= \langle X_k(t)\rangle + \langle X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right)\rangle\Delta - (\gamma+\alpha)\langle X_k(t)\rangle\Delta \\ &\quad +F\Delta + \sigma\langle\Delta W_k(t)\rangle + O(\Delta). \end{aligned} \tag{2.75}$$

Note that the stationary mean of X_k variables is independent of time t, i.e. $\langle X_k(t+\Delta)\rangle = \langle X_k(t)\rangle$. Also, mean of increment of the Brownian motion is zero. Hence, above equation is further simplified to

$$0 = \langle X_{k-1}(t)\left(X_{k+1}(t) - X_{k-2}(t)\right)\rangle - (\gamma+\alpha)\langle X_k(t)\rangle + F. \tag{2.76}$$

We solve the above equation for α and obtain a relationship between α and one-point moments

$$\alpha_{mom} = -\gamma + \frac{F+\rho}{\mu},$$

where α_{mom} is notation for α derived using method of moments, ρ is the stationary second-moment given by, $\rho = \langle X_{k-1}(X_{k+1} - X_{k-2})\rangle$ and $\mu = \langle X_k\rangle$ is the stationary mean. Note above expression for α_{mom} is valid only if the stationary mean $\mu \neq 0$. □

Estimator for parameter σ^2: In theorem (2.5.3), we analytically derived the relation between parameters α and σ^2 in reduced model. The same analytical relation is followed here to compute σ^2 using α_{mom} and is given by

$$\sigma^2_{mom} \quad = \quad 2((\gamma + \alpha_{mom})\, r_0 - F\mu), \tag{2.77}$$

where r_0 is the stationary second-moment of X_k, defined as $r_0 = \langle X_k^2 \rangle$.

Consistency of moment estimator using data of X_k from the reduced model Assume to be given discrete observations of X_k from the reduced model as $\{X_k(t), t = 0 : \Delta : T\}$, for each $k = 1 \dots K$, with a fixed subsampling time-step $\Delta > 0$ and total time of observations as T. If empirical mean of the observations of X_k is not equal to zero, then define moment estimator as

$$\hat{\alpha}_{mom} \quad = \quad -\gamma + \frac{\hat{\rho} + F}{\hat{\mu}}, \tag{2.78}$$

where $\hat{\rho}$ and $\hat{\mu}$ are empirical estimates of the stationary second mixed moment and the mean of observations, respectively, i.e.

$$\hat{\rho} = \frac{1}{N} \sum_{t=0:\Delta:T} X_k(t) \left(X_{k+2}(t) - X_{k-1}(t) \right), \quad \hat{\mu} = \frac{1}{N} \sum_{t=0:\Delta:T} X_k(t),$$

and $N = T/\Delta + 1$ is the total number of observations of X_k for each $k = 1 \dots K$. As the number of observations $N \to \infty$, the mixed moment estimator $\hat{\rho}$ is an efficient and consistent estimator of ρ. Similar result holds for the first-moment estimator $\hat{\mu}$. Also, recall that α_{mom}, given by (2.77), is derived using Euler discretization of X_k in reduced model. Hence, we can assume that $\alpha_{mom} \to \alpha$ as subsampling time-step $\Delta \to 0$. Therefore, given observations of X_k from the reduced model, $\hat{\alpha}_{mom}$ is consistent estimator for α under conditions $\Delta \to 0, N \to \infty$ and $N\Delta \to \infty$.

Summary of moment estimators under Indirect Observability. Assume we have been given a discrete set of observations of x_k from the full model (2.2), $\{x_k(t),\ t =$

$0 : \Delta : T,\ k = 1 \dots K\}$, with a subsampling time-step $\Delta > 0$ at a fixed $\epsilon > 0$. Our objective is to develop accurate and efficient estimation techniques for fitting the effective SDE on this data set. The moment estimators for parameters α and σ in the reduced model are given by (2.78). These estimators depend on total number of observations N and weakly dependent on the subsampling time-step Δ. The goal is to understand the behavior of these estimators when we use the data of the slow variables in the full model in the estimation procedure. To this end, we define the empirical mean and covariance of the slow variables in the full model, for fixed $\epsilon > 0$, same as in (2.51) with addition to one-point mix moment given by

$$\hat{\rho}_\epsilon = \frac{1}{N} \sum_{t=0:\Delta:T} x_k(t) \left(x_{k+2}(t) - x_{k-1}(t) \right), \tag{2.79}$$

where $N = T/\Delta + 1$ is the total number of observations. Using the multi-scale data from the full model in the estimators (2.78) is equivalent to substituting the empirical moments (2.51) into the expressions of both estimators. Thus the estimators become

$$\begin{aligned} \hat{\alpha}_{mom}(\epsilon) &= -\gamma + \frac{\hat{\rho}_\epsilon + F}{\hat{\mu}_\epsilon}, \\ \hat{\sigma}^2_{mom}(\epsilon) &= 2\big((\hat{\alpha}_{mom}(\epsilon) + \gamma) \hat{r}^\epsilon_0 - F \hat{\mu}_\epsilon \big). \end{aligned} \tag{2.80}$$

where we have introduced ϵ into the formulas for the estimators to explicitly emphasize their dependence on the multi-scale data.

2.6.1 Moment estimator as a function of subsampling time-step Δ under Indirect Observability

The moment estimator $\hat{\alpha}_{mom}(\epsilon)$ in (2.80) depends only on one-point moments of the discrete observations of x_k, it does not depend directly on the subsampling time-step of the data. The dependence on the subsampling time-step Δ enters weakly though the use of Euler discretization in the derivation of $\hat{\alpha}_{mom}(\epsilon)$. Hence, the moment estimator $\hat{\alpha}_{mom}(\epsilon)$ can

be inconsistent if computed using the observations of the slow variables in the full model subsampled at large time-step Δ.

In this subsection, we analyze the weak dependence of $\hat{\alpha}_{mom}(\epsilon)$ on the subsampling time-step Δ numerically and compare it with the dependence of $\hat{\alpha}_{mle}(\epsilon)$ on Δ. We use the same numerical setting as in section 2.5.5. Given observations x_k, $k = 1$ to K in the full model for the fixed $\epsilon = 0.3$, we subsample the data with various time-steps Δ and compute estimators $\hat{\alpha}_{mle}(\epsilon)$ (2.52) and $\hat{\alpha}_{mom}(\epsilon)$ (2.80) with different subsampling time-steps Δ.

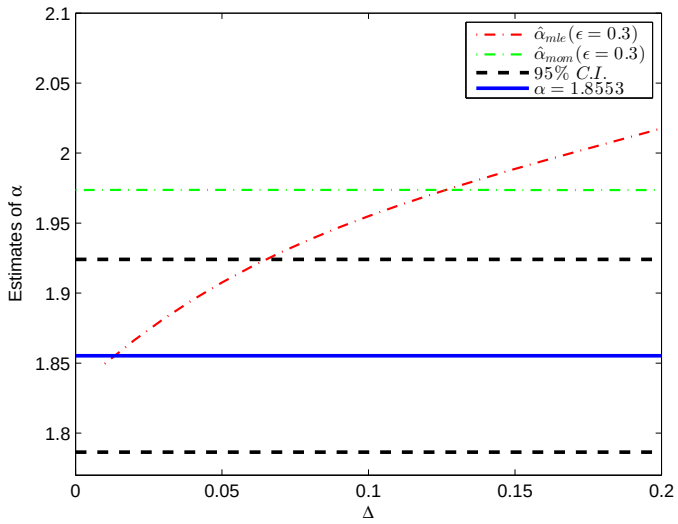

Figure 2.10: Moment estimator $\hat{\alpha}_{mom}(\epsilon)$ and approximate Maximum Likelihood estimator $\hat{\alpha}_{mle}(\epsilon)$ using the observations of x_k in the full model for $\epsilon = 0.3$. In the full model, parameters are fixed as in (4.5) and force and damping are fixed as $F = 10$ and $\gamma = 2$, respectively. Dotted dashed curve - $\hat{\alpha}_{mle}(\epsilon = 0.3)$ given by (2.52) as a function of Δ, dotted dashed straight line - $\hat{\alpha}_{mom}(\epsilon = 0.3)$ given by (2.80) as a function of Δ. Straight line represents the "true" value of α, given by (2.30), computed numerically using the fast subsystem. Top and bottom straight lines represent the 95% confidence interval, $(1.7865, 1.9241)$, for parameter α. Middle straight solid line - "true" value of α.

Figure 2.10 shows comparison of both estimators $\hat{\alpha}_{mle}(\epsilon = 0.3)$ and $\hat{\alpha}_{mom}(\epsilon = 0.3)$ for various subsampling time-steps Δ. Figure 2.10 also shows the 95% confidence interval $(1.7865, 1.9241)$ for the true value of $\alpha = 1.8553$. This figure shows that the moment

estimator $\hat{\alpha}_{mle}(\epsilon = 0.3)$ is not very sensitive to change in subsampling time-step Δ and is almost constant for Δ values as large as $\Delta = 0.2$. Hence, we consider $\hat{\alpha}_{mom}(\epsilon)$ to be approximately constant in later sections w.r.t. the subsampling time-step Δ and focus on it's dependence on ϵ alone.

2.6.2 Moment estimator as a function of scale parameter ϵ under Indirect Observability

In theorem (2.6.1), we derived an expression for α denoted as α_{mom} analytically, using Euler discretization of the reduced model. Next, we analyze the behavior of $\hat{\alpha}_{mom}$ computed using the data of the slow variables in the full model.

Numerical Setting: In the full model (2.2), we fix parameters as in (4.5), force $F = 10$ and damping $\gamma = 2$. We obtain discrete observation from the full model for values of $\epsilon = 1, 0.5, 0.3, 0.1$. The integration time-step δt and total time of observations T chosen for each ϵ in the full model are specified in Table 2.1. For each of these $\epsilon > 0$, we compute estimator $\hat{\alpha}_{mom}(\epsilon)$ in (2.80) using discrete observations of x_k from the full model,. Then we compare $\hat{\alpha}_{mom}(\epsilon)$ with true $\alpha = 1.8553$.

Figure 2.11 shows the estimator $\hat{\alpha}_{mom}(\epsilon)$ as a function of ϵ compared with the true value of $\alpha = 1.8553$. As $\epsilon \to 0$, we observe estimator $\hat{\alpha}_{mom}(\epsilon)$ converging towards α. And for smallest $\epsilon = 0.1$, we obtain $\hat{\alpha}_{mom} = 1.937$ which is 4.4% error compared to true value $\alpha = 1.8553$. Since both $\hat{\alpha}_{mom}$ (2.80) and α (2.30) are computed numerically, we assume 4.4% error to be small considering the possibility of numerical errors. Note that almost all values of $\hat{\alpha}_{mom}(\epsilon)$ lies outside the 95% confidence interval around the true value of α which is acceptable because the reduced model is not sensitive to small numerical errors in the parameters α and σ. Table 2.2 represents the relative absolute error of $\hat{\alpha}_{mom}(\epsilon)$ with respect to true value $\alpha = 1.8553$ as $\epsilon \to 0$. The relative absolute error in table 2.2 is given

Table 2.2: Relative absolute error of $\hat{\alpha}_{mom}(\epsilon)$ (2.80) computed using observations of the full model

ϵ in full model	1	0.5	0.3	0.1	0.05
Relative Absolute error of $\hat{\alpha}_{mom}(\epsilon)$	2%	5.8%	6.4%	4.4%	4.3%

by

$$\text{Relative Error} = \frac{|\hat{\alpha}_{mom}(\epsilon) - \alpha|}{\alpha}.$$

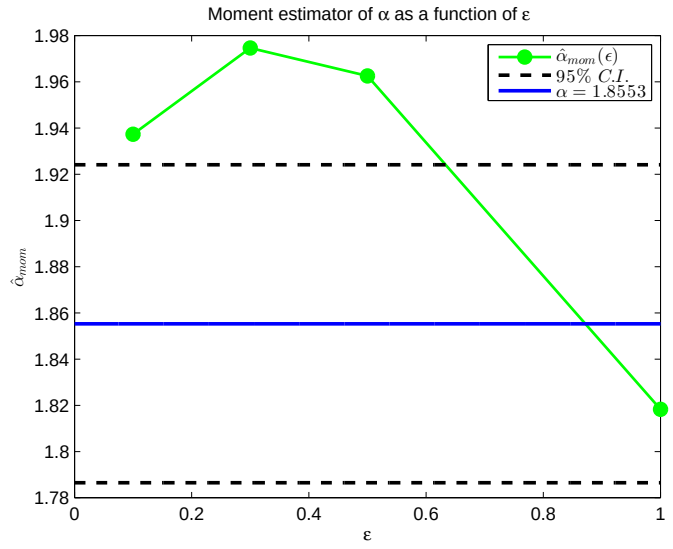

Figure 2.11: Moment estimator $\hat{\alpha}_{mom}$ based on x_k data set from the full model for $\epsilon = \{1, 0.5, 0.3, 0.1\}$. In the full model, parameters are fixed as in (4.5) and force and damping are fixed as $F = 10$ and $\gamma = 2$, respectively. Solid curve - $\hat{\alpha}_{mom}(\epsilon)$ given by (2.77) as a function of ϵ. Straight line represents the "true" value of α, given by (2.30), computed numerically using the fast subsystem. Top and bottom straight lines represent the 95% confidence interval, $(1.7865, 1.9241)$, for parameter α. Middle straight solid line - "true" value of α.

Figure 2.11 also shows that moment estimator $\hat{\alpha}_{mom}(\epsilon)$ is not too sensitive to changes in value of $\epsilon > 0$. approximate Maximum bias of $\hat{\alpha}_{mom}(\epsilon)$ is observed for $\epsilon = 0.3$ with 6.3% relative error. Considering numerical errors and a relatively large value of $\epsilon = 0.3$, 6.3% error is considered a very good agreement between the estimator and the true value of α.

Since the moment estimator $\hat{\alpha}_{mom}(\epsilon)$ depends only on one-point moments of the discrete observations of x_k, it doesn't depend on the subsampling time-step of the data. The dependence on Δ enters weekly through the use of the Euler discretization in the derivation of $\hat{\alpha}_{mom}(\epsilon)$. We analyze the dependence of $\hat{\alpha}_{mom}(\epsilon)$ on ϵ by doing regression of $\hat{\alpha}_{mom}(\epsilon)$ for various ϵ values. In figure 2.11, we observe an approximate quadratic curve of $\hat{\alpha}_{mom}(\epsilon)$ as function of ϵ. We obtain the following regression relationship

$$\hat{\alpha}_{mom}(\epsilon) = 1.9195 + 0.2440\epsilon - 0.344\epsilon.^2. \tag{2.81}$$

Using regression, we observe $\hat{\alpha}_{mom}(\epsilon) \to 1.9195$ as $\epsilon \to 0$ which is 3.2% relative error with true value of $\alpha = 1.8553$. Since, reduced model is not sensitive to small errors in parameter α and taking numerical errors in consideration, we conclude that $\hat{\alpha}_{mom}$ is a sufficiently accurate estimator.

2.6.3 Comparison between the approximate Maximum Likelihood estimator and the moment estimator using observations of the reduced model

Comparison using observations of the reduced model: We compare $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$ under Direct Observability to compare the errors due to Euler discretization. The idea is that we take discrete observations of X_k from reduced model and on the same data with fixed subsampling time-step $\Delta > 0$, we compute both estimators $\hat{\alpha}_{mle},\ \hat{\alpha}_{mom}$. Then we compare the error of estimators from true value of α. We use the Monte-Carlo approach here.

Numerical Setting: In the reduced model, we fix parameters as given in (4.5). We test estimators $\hat{\alpha}_{mle}, \hat{\alpha}_{mom}$ for two set of parameters:

$$I : F = 10, \gamma = 2$$

$$II : F = 24, \gamma = 0.1$$

In section 2.3.2, we derived the true value of $\alpha = 1.8553$ corresponding to parameters in (4.5). We used observations of the fast subsystem, (2.4) to compute α numerically. Recall that α, given by (2.30), does not change with different values of force F and damping γ. Hence for different F, γ, we consider value of true α fixed as $\alpha = 1.8553$.

Time-averaging on one trajectory: For each set of parameters, we perform simulation and obtain discrete observations of X_k in reduced model for each $k = 1 \dots K$. We fix integration time-step as $\delta t = 0.0001$ with different total time of observations T. Then we compute both estimators $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$ using given observations of X_k. Note that $\hat{\alpha}_{mle}$ is computed using subsampling time-step same as integration time-step, i.e.

$$\Delta = \delta t = 0.0001.$$

Hence, both estimators are computed on original simulated trajectory and then compared with true value of $\alpha = 1.8553$. We compute "absolute error" (Error) and "Square of error" (SE) as

$$\begin{aligned} \text{Error}_{mle} &= |\hat{\alpha}_{mle} - \alpha|, \quad \text{Error}_{mom} = |\hat{\alpha}_{mom} - \alpha|, \\ \text{SE}_{mle} &= (\hat{\alpha}_{mle} - \alpha)^2, \quad \text{SE}_{mom} = (\hat{\alpha}_{mom} - \alpha)^2. \end{aligned} \tag{2.82}$$

Monte-Carlo Approach: For each parameter setting , i.e. I and II and for fixed value of T, we run $M = 6$ trajectories with same details as explained above. For all six trajectories, absolute error and square of error is computed for estimators $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$. Then we average both type of errors and define them as the "Bias" and "Mean Squared Error (MSE)" respectively, to compare the errors accurately, i.e., we define averaged bias and MSE as

$$\text{Bias}_{mle} = E(|\hat{\alpha}_{mle} - \alpha|), \quad \text{Bias}_{mom} = E(|\hat{\alpha}_{mom} - \alpha|), \tag{2.83}$$

$$\text{MSE}_{mle} = E((\hat{\alpha}_{mle} - \alpha)^2), \quad \text{MSE}_{mom} = E((\hat{\alpha}_{mom} - \alpha)^2). \tag{2.84}$$

Comparison: In all simulations, integration time-step is fixed as $\delta t = 0.0001$. The goal is

to compare Monte-Carlo errors in the two estimators computed with the same subsampling time-step, but changing time of observations. We consider five values of T as

$$T = [2000, 5000, 10000, 50000, 100000]$$

Note that as T increases, the number of observations $N = T/\delta t + 1$ increases. Hence, for each parameters set of I and II, we have five values of T, for which we obtain six trajectories for each T.

Parameter setting I: $F = 10, \gamma = 2$

Figure 2.12 shows comparison of the mean bias and the mean squared errors for estimators $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$ as time of observation T increases. We observe that for various time T or number of observations N, Bias and MSE of $\hat{\alpha}_{mom}$ is less than corresponding errors of $\hat{\alpha}_{mle}$.

Parameter setting II: $F = 24, \gamma = 0.1$

Figure 2.13 shows comparison of the mean bias and the mean squared errors for estimators $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$ as time of observation T increases. We observe that as time T increases, Bias and MSE of $\hat{\alpha}_{mom}$ is less than corresponding errors of $\hat{\alpha}_{mle}$ under Direct Observability.

2.6.4 Conclusion on comparison of estimators $\hat{\alpha}_{mom}$ and $\hat{\alpha}_{mle}$

Given observations of X_k from the reduced model (2.29), we observe numerically that as number of discrete observations N increases, $\hat{\alpha}_{mom}$ has less bias and mean square error than $\hat{\alpha}_{mle}$. This observation implies that $\hat{\alpha}_{mom}$ estimates parameter α more accurately than $\hat{\alpha}_{mle}$ using observations of the reduced model. Also, in section 2.5, we observed that $\hat{\alpha}_{mle}(\epsilon)$ depends on auto-correlation of observations, leading to subsampling issue. To ensure that $\hat{\alpha}_{mle}(\epsilon)$ is a sufficiently accurate estimator of α, we need to determine range of an optimal subsampling time, then subsample the given data with that optimal time-step. It is not feasible to find an explicit analytical expression for the optimal subsampling time

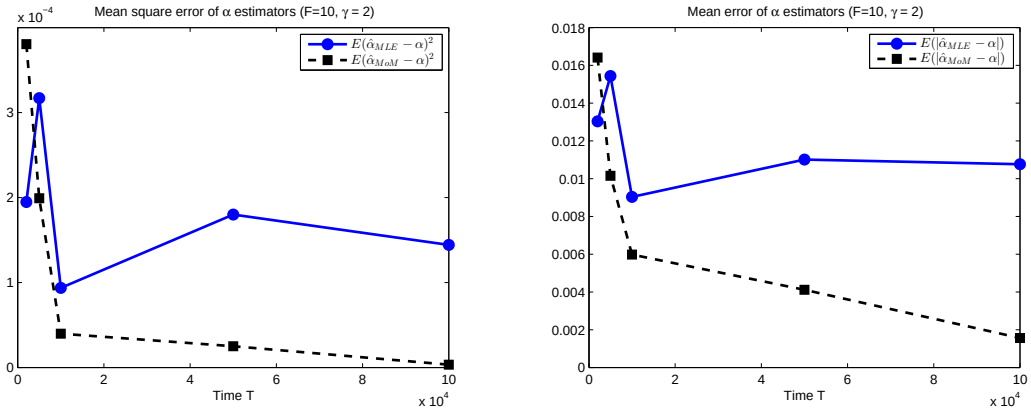

Figure 2.12: Bias and mean squared error (MSE) of estimators $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$ compared with true parameter $\alpha = 1.8553$. Estimators are based on discrete observations from reduced model (2.29) with fixed integration time-step $\delta t = 10^{-5}$ and varying total observation time $T = [2000, 5000, 10000, 50000, 100000]$. Parameters are fixed as in (4.5)and $F = 10, \gamma = 2$. For each fixed T, Bias and MSE is computed by averaging the errors from six trajectories of reduced model (with different random seed and initial value of X_k). Estimator $\hat{\alpha}_{mle}$ is computed at subsampling time-step equal to integration time-step $\Delta = \delta t$. Left part - Bias of estimators, given by (2.84). Right part - MSE of estimators, given by (2.84). Solid line - Bias and MSE of $\hat{\alpha}_{mle}$, dashed line - Bias and MSE of $\hat{\alpha}_{mom}$. As time T or number of observations N increases, $\hat{\alpha}_{mom}$ seems to have less bias and mean squared error than $\hat{\alpha}_{mle}$ making $\hat{\alpha}_{mom}$ a more robust estimator of parameter α.

range and that makes the subsampling process rather complex. On the other hand, the moment estimator $\hat{\alpha}_{mom}(\epsilon)$ is based only on one-point moments of discrete observations, therefore, we do not need any subsampling to compute the moment estimator. Overall, we conclude that $\hat{\alpha}_{mom}(\epsilon)$ is a more robust estimator than $\hat{\alpha}_{mle}(\epsilon)$ since it has no subsampling complexity.

Recall that moment estimator $\hat{\alpha}_{mom}(\epsilon)$ is valid only if the empirical mean of x_k is not too small. In practice for the $\hat{\alpha}_{mom}(\epsilon)$ to be acceptable, the mean of x_k should be O(1) or larger. In the full model (2.2), if force F is close to zero, then the stationary mean of x_k is also close to zero. Hence, for small force F in the full model, $\hat{\alpha}_{mom}$ can have numerical problems estimating α.

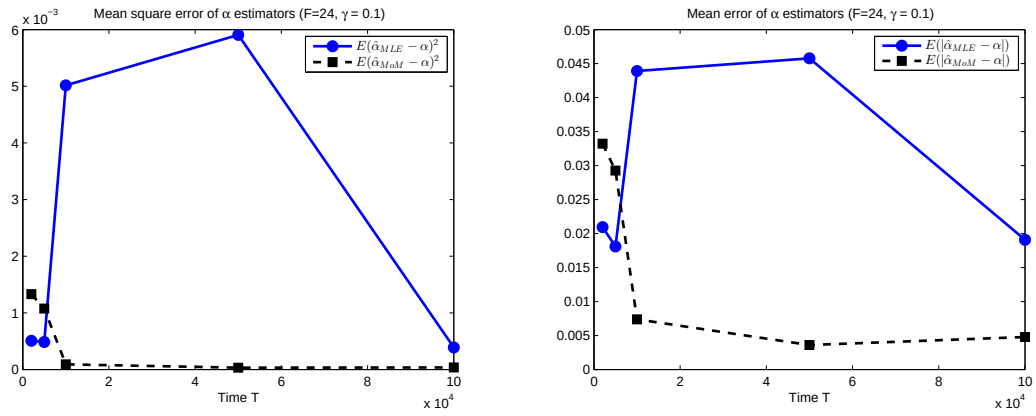

Figure 2.13: Bias and mean squared error (MSE) of estimators $\hat{\alpha}_{mle}$ and $\hat{\alpha}_{mom}$ compared with true parameter $\alpha = 1.8553$. Estimators are based on discrete observations from reduced model (2.29) with fixed integration time-step $\delta t = 10^{-5}$ and varying total observation time $T = [2000, 5000, 10000, 50000, 100000]$. Parameters are fixed as in (4.5)and $F = 24, \gamma = 0.1$. For each fixed T, Bias and MSE is computed by averaging the errors from six trajectories of reduced model (with different random seed and initial value of X_k). Estimator $\hat{\alpha}_{mle}$ is computed at subsampling time-step equal to integration time-step $\Delta = \delta t$. Left part - Bias of estimators, given by (2.84). Right part - MSE of estimators, given by (2.84). Solid line - Bias and MSE of $\hat{\alpha}_{mle}$, dashed line - Bias and MSE of $\hat{\alpha}_{mom}$. As time T or number of observations N increases, $\hat{\alpha}_{mom}$ seems to have less bias and mean squared error than $\hat{\alpha}_{mle}$ making $\hat{\alpha}_{mom}$ a more robust estimator of parameter α.

2.7 True Values of Parameters in the Reduced Model as a Function of One-point Moments of the Slow Variables

In this section, our goal is to estimate the "true" value of parameters in the reduced model without computing all correlations for the fast variables. Direct numerical simulations of the evolution of the dynamics in the full L96 model is computationally expensive, due both to the large number of small-scale variables and the necessity to choose a small discretization step in order to resolve the fast components of L96. Hence, our objective is to derive a low-dimensional stochastic model for slow variables such that behavior of slow variables in full model weakly converges to the corresponding statistics in reduced model as scale

separation parameter $\epsilon \to 0$. In section 2.3.2, homogenization procedure is used to derive the explicit equation for the reduced model and also the "true" values of parameters in reduced model. The "true" values of parameters, given by (2.30), depend on lagged cross-covariance of the fast variables which can not be computed analytically. Therefore, these "true" values are computed using one long simulation of the fast subsystem (2.4) and using empirical values of lagged cross-correlation of the fast variables. We observe numerically that lagged cross-correlations of the fast variables are approximately equal to zero, i.e.

$$\langle y_i(0) y_j(\tau) \rangle \approx 0, i \neq j, \tag{2.85}$$

hence, "true" values are computed by lagged auto-covariance of the fast variables alone. In other situations, it can be hard to resolve all the correlation of the fast variables which can make numerical computation of "true" values of parameters in reduced model very expensive. Numerical computation of all lagged cross-covariances can increase the simulation time because of the number of possible combinations of cross-correlations and summations needed to compute numerical approximation of each averaged cross-correlation. For example, in the next chapter, the "true" values are dependent on fourth-moments of the fast variables which is expensive to compute due to the amount of possible index combinations. Hence, in this section, we derive an alternate formula for "true" values of parameters which depends on mean and variance of the slow variables alone.

In section 2.3.2, homogenization procedure is used to derive the explicit equation for the reduced model and also the "true" values of parameters in the reduced model. The reduced model can be rewritten as

$$dX_k = X_{k-1}(X_{k+1} - X_{k-2})dt - (\gamma + \alpha)X_k dt + F dt + \sigma dW_k, \tag{2.86}$$

where $k = 1 \dots K$ and W_k are independent Brownian motions. There is a relationship

between parameters α and σ derived in (2.30) given by

$$\sigma^2 = \frac{k_x s^2}{J k_y c}\alpha, \tag{2.87}$$

which is proved analytically and hence considered exact in reduced model.

In theorem 2.5.3, we derived another exact relationship between parameters α and σ, given in (2.70), using Ito's formula for the reduced model. The parameters α and σ also obey the relationship

$$\sigma^2 = 2((\gamma+\alpha)r_0 - F\mu), \tag{2.88}$$

where $r_0 = \left\langle X_k^2 \right\rangle, \mu = \langle X_k \rangle$. Combining (2.87) and (2.88), we obtain another formula for the true α given by

$$\alpha_{new} = 2Jk_yc\left(\frac{\gamma r_0 - F\mu}{k_x s^2 - 2r_0 J k_y c}\right), \tag{2.89}$$

where $\mu = \langle X_k \rangle, r_0 = \left\langle X_k^2 \right\rangle$. Note that the formulas in (2.87) and (2.88) are derived analytically without any approximation and discretization, hence, expression in α in (2.89) can also be used to cmpute the "true" value of α and to avoid computing all correlations of the fast variables.

We test (2.89) on data generated by the full model with a small value of ϵ. Assume to be given the observations of the slow variables x_k in the full L96 model, i.e. $\{x_k(t), t = 0 : \Delta : T, k = 1 \cdots K\}$ where $\Delta > 0$ is subsampling time-step and $N = T/\Delta + 1$ is number of total observations of x_k. Then, the new "true" value of α can be estimated as

$$\hat{\alpha}_{new} = 2Jk_yc\left(\frac{\gamma \hat{r}_0 - F\hat{\mu}}{k_x s^2 - 2\hat{r}_0 J k_y c}\right), \tag{2.90}$$

where $\hat{r}_0$ and $\hat{\mu}$ are the standard empirical moment estimators defined as

$$\hat{r}_0 = \frac{1}{N}\sum_{t=0:\Delta:T} x_k^2, \quad \hat{\mu} = \frac{1}{N}\sum_{t=0:\Delta:T} x_k.$$

Important Observation: Note that the expression for estimating the "true" value given by $\hat{\alpha}_{new}$ in (2.90) is not under Indirect Observability framework. Indirect Observability is when we compute estimators from the data of x_k without using equations of the full L96 model. Estimator $\hat{\alpha}_{new}$ is derived by combining homogenization on the full L96 model and Ito's formula for the reduced model. Also, to compute $\hat{\alpha}_{new}$ in (2.90), we need to know constants k_y, c, and s in the equation for y variables and the coupling coefficients k_x of slow and fast variables in the full L96 model (2.2). Hence, we need to know the explicit equations of the full model to derive and compute the estimator $\hat{\alpha}_{new}$ for the "true" value of α.

2.7.1 True value $\hat{\alpha}_{new}$ as function of ϵ, using observations from full model

Given observations of the large-scale variable x_k from the full model, the true value of α in the reduced model can be estimated using mean and variance of given data alone. Note that $\hat{\alpha}_{new}$ in (2.90) is independent of the subsampling time-step Δ and hence is only a function of ϵ and the number of observations N. In this subsection, we numerically investigate the behavior of $\hat{\alpha}_{new}$ as $\epsilon \to 0$ and compare it with the "true" value of α in (2.30) computed using correlations of the fast variables.

Numerical Setting: In the full model (2.2), fix parameters as in (4.5), force $F = 10$, and damping $\gamma = 2$. For values of $\epsilon = \{1, 0.3, 0.2, 0.1\}$, we simulate the full model and obtain discrete observations of x_k. Integration time-steps δt and total time of observations T for each ϵ are specified in Table 2.1. Based on these data, we compute estimator $\hat{\alpha}_{new}$ for various values of ϵ. We compare $\hat{\alpha}_{new}$ with parameter $\alpha = 1.8553$ as $\epsilon \to 0$.

Figure 2.14 shows $\hat{\alpha}_{new}(\epsilon)$ computed using observations of x_k in the full model for several values of ϵ. For $\epsilon = 0.1$, $\hat{\alpha}_{new}$ has 1.7% relative error compared with α. Since $\alpha = 1.8553$ is computed numerically in (2.30), we consider the possibility of numerical errors in computing

Table 2.3: Relative absolute error of $\hat{\alpha}_{new}(\epsilon)$ (2.90), using observations of the full model as $\epsilon \to 0$, with respect to true value α in (2.30)

ϵ in full model	1	0.5	0.3	0.1
Relative Absolute error of $\hat{\alpha}_{mom}(\epsilon)$	9.2%	0.7%	0.45%	1.72%

α. For $\epsilon = 0.1$, $\hat{\alpha}_{new} = 1.887$ which belongs to the 95% confidence interval of true values of α. Hence, we conclude that $\hat{\alpha}_{new}$ gives an accurate estimate of true value of α for small ϵ. Table 2.3 also represents relative absolute errors of $\hat{\alpha}_{new}(\epsilon)$ with respect to true value $\alpha = 1.8553$ as $\epsilon \to 0$.

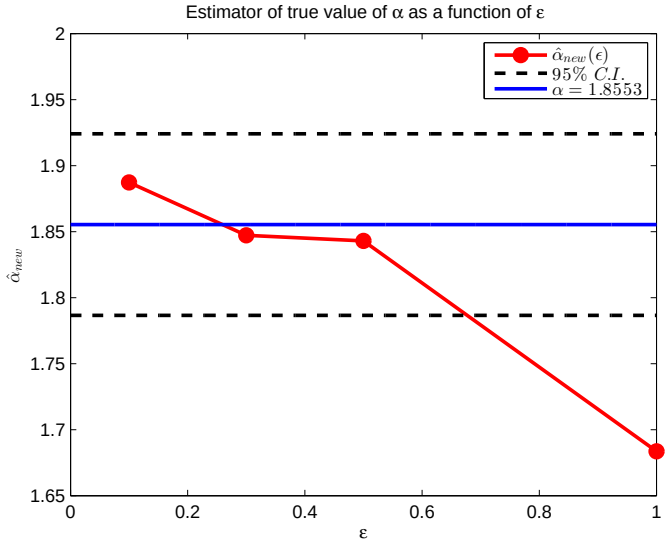

Figure 2.14: Estimator $\hat{\alpha}_{new}$ computed using the observations of x_k in the full model for $\epsilon = \{1, 0.5, 0.3, 0.1\}$. In the full model, parameters are as in (4.5) and force and damping are $F = 10$ and $\gamma = 2$, respectively. Solid curve line - $\hat{\alpha}_{new}$ given by (2.90) as a function of ϵ. Straight line represents the "true" value of α, given by (2.30), computed numerically using the fast subsystem. Top and bottom straight lines represent the 95% confidence interval, $(1.7865, 1.9241)$, for parameter α. Middle straight solid line - "true" value $\alpha = 1.8553$.

2.8 Conclusion

Our objective is to develop an efficient and accurate parametric estimation procedure for the effective reduced equation representing a limiting process which is not observable. The data used in the estimation procedure are subsampled from an observable approximating process. This objective has been explicitly illustrated using an observable process as the L96 model (2.2).

In section 2.3, the homogenization procedure is used to derive explicit equations for the reduced model, given by (2.29), and the "true" values of parameters α, σ in the reduced model, given by (2.30). The reduced equation represents the effective unobservable process. The "true" values, derived using homogenization procedure, depends on lagged covariance of fast variables in the L96 model. Due to the complexity of the equation of fast variables in the L96 model, we can not evaluate "true" values of parameters analytically and hence they are computed using one long simulation of the fast subsystem of L96. It can be hard to resolve all the correlation of the fast variables in the fast subsystem (2.4) which can make numerical computation of the "true" values of parameters very expensive. Hence, in section 2.7, we derive an alternative formula for the "true" values of parameters using only mean and variance of the slow variables. If we know the constant parameters of the fast subsystem and one-point moments of the slow variables in the full model, we can compute "true" value of α, given by (2.90), without a need to resolve all possible auto-correlations of the fast variables in the fast subsystem.

Recall that our main objective is to estimate parameters of reduced model under the Indirect Observability, i.e., using observations of the slow variables in the full model. Hence, the "true" values for the reduced model are only used to test the behavior of the estimators as scale separation parameter ϵ decreases. In section 2.5, we consider the first estimator of parameters in the reduced model as approximate Maximum Likelihood estimator. Since

the reduced model is the limiting process of the slow variables in L96 as $\epsilon \to 0$, it is desirable that the approximate Maximum Likelihood parameter estimates computed from discrete data of the slow variables in the full model converge to "true" parameter values as $\epsilon \to 0$. Approximate Maximum Likelihood estimators $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$, given by (2.52), depend on auto-correlations of the slow variables. In subsection 4.2.1, we observe a considerable difference between the curvature of the auto-correlation of x_k in the full model and X_k in the reduced model for small time lag Δ. Hence, $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$ can loose their consistency if we use the data of the slow variables of the full model at a very dense time-step, referred as *subsampling issue*. Also, $\hat{\alpha}_{mle}(\epsilon)$ and $\hat{\sigma}_{mle}(\epsilon)$ loose their consistency at very large time-steps Δ due to the use of the Euler discretization of the reduced model in the derivation of the approximate Maximum Likelihood estimators, referred as *Bias due to the Euler discretization*. Therefore, there is a need to subsample the data at an optimal time-step Δ such that the approximate Maximum Likelihood estimators approximate true parameter well for small ϵ. It is not feasible to derive an analytical expression for the optimal subsampling time-step in the approximate Maximum Likelihood estimators. Hence, the only solution is to compare the approximate Maximum Likelihood estimator with "true" values of parameter (2.30) and investigate numerically the optimal subsampling time-step. This is computationally expensive and not a feasible solution. Therefore, our next objective is to derive estimators which do not depend on subsampling time-step.

The second estimator we derive is the moment estimator $\hat{\alpha}_{mom}$ by applying "Method of Moments" on the Euler discretized reduced model driving X_t. One advantage of this estimator is its dependence on only one-point stationary moments.

Under Direct Observability, i.e. given observations of X_k from the reduced model (2.29), we observe numerically that as number of discrete observations N increases, $\hat{\alpha}_{mom}$ has less bias and mean square error than $\hat{\alpha}_{mle}$. This observation implies that $\hat{\alpha}_{mom}$ estimates

parameter α more accurately than $\hat{\alpha}_{mle}$ using observations of the reduced model. Under Indirect Observability, there is a need to subsample the observations of x_k in the full model at an optimal time-step Δ to compute consistent $\hat{\alpha}_{mle}(\epsilon)$ but there is no feasible solution to find analytical expression for the optimal time-step Δ. On the other hand, we do not need any subsampling to compute the moment estimator under Indirect Observability.

Overall, we conclude that if the mean of the slow variables in the multi-scale non-linear model is relatively large and the reduced model is not very sensitive to the change in parameters then $\hat{\alpha}_{mom}(\epsilon)$ is considered a more robust estimator compared with $\hat{\alpha}_{mle}(\epsilon)$ since it has no subsampling complexity. We need to study and analyze the statistics of the full model to find an optimal subsampling time-step for consistent value of approximate Maximum Likelihood estimator $\hat{\alpha}_{mle}(\epsilon)$. Another alternative solution is to plot both estimators and we can assume from plot of both estimators in Figure 2.10 that the value of α near intersection of both estimators will give acceptable value for α in the reduced model.

Recall that moment estimator $\hat{\alpha}_{mom}(\epsilon)$ is valid only if the empirical mean of x_k is not too small. In practice for the $\hat{\alpha}_{mom}(\epsilon)$ to be acceptable, the mean of x_k should be O(1) or larger. In the full model (2.2), if force F is close to zero, then the stationary mean of x_k is also close to zero. Hence, for small force F in the full model, $\hat{\alpha}_{mom}(\epsilon)$ can have numerical problems estimating α.

CHAPTER 3

Stochastic Mode-reduction of Multi-scale Models with Energy as a Hidden Slow Variable

3.1 Introduction

Motivation. In this chapter, we develop procedure of the stochastic mode reduction for the *parabolic-hyperbolic coupled systems*. One of the main application of these systems are in Thermo-elasticity. The system of the thermo-elasticity is the superposition of a wave-like and a heat equation. While solutions of the wave equation do reproduce the same pattern for all times, those of the heat equation are very quickly dissipated. In [54], a linear system of thermo-elasticity is considered, consisting of a wave equation coupled to a heat equation. The considered system of thermo-elasticity in [54] is given by

$$\begin{aligned} u_{tt} - c^2\Delta u + \alpha\Delta\theta = 0, \\ \theta_t - \nu\Delta\theta + \beta u_t = 0. \end{aligned} \tag{3.1}$$

with certain boundary and initial conditions. The variable u is the slow variable representing the wave vector coupled with equation of heat variable represented by fast variable vector θ. The coupling parameters α, β and the viscosity ν are assumed to be positive constants. The fast subsystem of the multi-scale model in (3.1) is the wave equation without coupling terms and hence conserving energy. The wave equation is an important second-linear partial differential equation for the description of waves. It has applications in acoustics, electromagnetics, and fluid dynamics. Since there is no energy dissipation term in wave equation, it conserves energy. There is similar model of thermo-elasticity as in (3.1), considered in [23]. In our work, we consider a prototype model as the generalization of *Additive Triad Model* having energy conserved fast subsystem, as an example of parabolic-hyperbolic coupled systems.

Outline of our work. In this chapter we consider an application of the stochastic mode-reduction to multi-scale models with energy as a hidden slow variable. In particular, we consider the situation when the stochastic terms are added to the slow variables in energy-conserving nonlinear systems. Since the noise only affect the slow variables, the fast subsystem is deterministic evolving on a sphere of constant energy. On the other hand, the radius of the sphere slowly changes due to the coupling between the slow and fast dynamics. Therefore, in order to apply the stochastic reduction techniques, one needs to consider energy of the fast subsystem as an additional hidden slow variable.

3.2 Generalized Triad Model

To illustrate our approach we first consider a prototype coupled triad model

$$\begin{aligned} dx &= \sum_{j,k=1}^{n} A_{1jk}^{xyy} y_j y_k dt - \gamma x dt + \sigma dW_t, \\ dy_j &= \sum_{k=1}^{n} A_{j1k}^{yxy} x y_k dt + \sum_{i,k=1}^{n} B_{jik}^{yyy} y_i y_k dt, \end{aligned} \tag{3.2}$$

where x is the slow variable, y_k, $k = 1, \ldots, n$ are the fast variables, A_{1jk}^{xyy}, A_{j1k}^{yxy}, B_{jik}^{yyy} are interaction coefficients obeying the relationships

$$\begin{aligned} A_{1jk}^{xyy} + A_{j1k}^{yxy} + A_{k1j}^{yxy} = 0, \\ B_{jik}^{yyy} + B_{ikj}^{yyy} + B_{kji}^{yyy} = 0. \end{aligned} \tag{3.3}$$

Note that if we ignore the damping and diffusion terms in equation for x in (3.2), then the non-linear terms will be left in the Triad model conserving energy $x^2 + \sum y_k^2$ due to the constraint on the coupling coefficients given by (3.3).

One can attempt to apply the stochastic mode-reduction strategy to the model in (3.2). In the stochastic mode-reduction procedure one needs to consider the behavior of the fast subsystem

$$\dot{y_j} = \sum_{i,k=1}^{n} B_{jik}^{yyy} y_i y_k \tag{3.4}$$

and integrated lagged correlations functions of y-variables typically enter as coefficients into the reduced model. The behavior of the fast subsystem depends drastically on the energy level, but the energy in the fast modes changes in time due to the interaction with the slow variable. Therefore, we consider energy of the fast subsystem

$$E = \sum_{k=1}^{n} y_k^2 \tag{3.5}$$

as an additional slow variable and the full rescaled model becomes

$$\begin{aligned} dx &= \frac{1}{\varepsilon}\sum_{j,k=1}^{n} A_{1jk}^{xyy} y_j y_k dt - \gamma x dt + \sigma dW_t, \\ dE &= -2\frac{1}{\varepsilon}\sum_{j,k=1}^{n} A_{1jk}^{xyy} x y_j y_k dt, \\ dy_j &= \frac{1}{\varepsilon}\sum_{k=1}^{n} A_{j1k}^{yxy} x y_k dt + \frac{1}{\varepsilon^2}\sum_{i,k=1}^{n} B_{jik}^{yyy} y_i y_k dt, \end{aligned} \tag{3.6}$$

where we have used $A_{j1k}^{yxy} + A_{k1j}^{yxy} + A_{1jk}^{xyy} = 0$.

There are certain assumptions we apply on the full model (3.6) which as given as follows

- We take number of fast variables as large, i.e. $n = 10$ to ensure the ergodicity of the fast subsystem (3.4).
- The energy conservation constraint of the coupling coefficients between x and y , i.e. first equation in (3.3) is not an essential assumption for derivation of the reduced model. It is considered to simplify certain computations in this chapter.
- The energy conservation constraint of the interaction coefficients in the fast subsystem, i.e. the second equation in (3.3) is a mandatory constraint leading to the fact that the fast subsystem conserves energy.

3.3 Stationary Distribution of the Generalized Triad Model

The stationary distribution of the generalized triad model (3.6) can be computed explicitly for any ε and it is easy to show that the stationary distribution for $x, y_1, \ldots, y_n$ is a product of Gaussian distributions with mean zero and identical variances. The stationary distribution does not depend on the small parameter ε.

Fokker-Planck equation for the invariant measure $P = P(x, y_1, .., y_n)$ for $x, y_1, \ldots, y_n$ in

(3.6) can be written as

$$0 = \left(A_0 + \frac{1}{\varepsilon} A_1 + \frac{1}{\varepsilon^2} A_2 \right) P$$

where

$$\begin{aligned} A_0 &= -\gamma x \partial_x + \frac{\sigma^2}{2} \partial_{xx}, \\ A_1 &= \sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x + \sum_{j,k} A_{j1k}^{yxy} x y_k \partial_{y_j}, \\ A_2 &= \sum_{i,j,k} B_{jik}^{yyy} y_i y_k \partial_{y_j}. \end{aligned}$$

The differential operator A_0 annihilates Gaussian density

$$\frac{\sqrt{2\gamma}}{\sqrt{2\pi}\sigma} exp\left(-\frac{\gamma}{\sigma^2} x^2 \right)$$

The operators A_1 and A_2 annihilate separately any function of the full energy $x^2 + E$ with E in (3.5) due to the conservation of energy by the non-linear terms in (3.2) using the constraints (3.3). Therefore, they will also annihilate the function

$$P = \left(\frac{\sqrt{2\gamma}}{\sqrt{2\pi}\sigma} \right)^{n+1} exp\left(-\frac{\gamma}{\sigma^2} \left[x^2 + \sum_j y_j^2 \right] \right).$$

Therefore, the invariant measure for x, $y_1, \ldots, y_n$ in (3.6) is a product measure with

$$x, y_1, \ldots, y_n \sim N\left(0, \frac{\sigma^2}{2\gamma} \right). \tag{3.7}$$

The stationary distribution for the energy E in (3.5) can be easily derived since the stationary distribution of fast variables is a product of Gaussian densities.

$$\rho_E(E) = C E^{(n-2)/2} e^{-\frac{E\gamma}{\sigma^2}}. \tag{3.8}$$

3.4 Limit of the Full Model as $\varepsilon \to 0$

We apply the stochastic mode-reduction to the model (3.6) treating the energy E of the fast subsystem as a slow variable. The Kolmogorov backward equation associated with (3.6) for

a scalar function $u = u(t, x, E, y_1, \ldots, y_n)$ is given by

$$-\partial_t u = L_0 u + \frac{1}{\epsilon} L_1 u + \frac{1}{\epsilon^2} L_2 u \tag{3.9}$$

where the operators above are

$$\begin{aligned} L_0 &= -\gamma x \partial_x + \frac{\sigma^2}{2} \partial_{xx}, \\ L_1 &= \sum_{j,k} A^{xyy}_{1jk} y_j y_k \partial_x + \sum_{j,k} A^{yxy}_{j1k} x y_k \partial_{y_j} - 2 \sum_{j,k} A^{xyy}_{1jk} x y_j y_k \partial_E, \\ L_2 &= \sum_{i,j,k} B^{yyy}_{jik} y_i y_k \partial_{y_j}. \end{aligned} \tag{3.10}$$

We introduce the projection operator

$$\mathbb{P}\, g = \int g\, \mu(\vec{y}|E) d\vec{y},$$

where g is any bounded Borel function and $\mu(\vec{y}|E)$ is the invariant measure of the fast subsystem (3.4) on the sphere of constant energy (3.5).

We follow same homogenization procedure as explained before in section 2.3.2. Considering the expansion

$$u = u_0 + \varepsilon u_1 + \varepsilon^2 u_2$$

and collecting powers of ε we obtain the same relations as in (2.14) with operators (3.11). Similar as in section 2.3.2, from the relations in (2.14), we obtain that u_0 is independent of fast variables, i.e. $u_0 = u_0(x, E)$, the compatibility condition $\mathbb{P}L_1 = 0$, and the reduced operator

$$L = -\mathbb{P} L_1 L_2^{-1} L_1.$$

We compute operator L and simplify it in later sections.

Note about the compatibility condition: The compatibility condition $\mathbb{P}L_1 = 0$ mean the following conditions on the averages of $\vec{y}$ in the fast subsystem

$$\mathbb{E}_\mu y_j = 0, \qquad \mathbb{E}_\mu y_j y_k = 0 \text{ for } j \neq k. \tag{3.11}$$

The first condition is quite plausible, it simply implies the symmetry of the flow. The second condition is a bit more problematic, one simple way ensure the second condition is to assume $\mathbb{E}_\mu y_j y_k = \mathbb{E}_\mu y_j \mathbb{E}_\mu y_k = 0$, but this is not true since the $\vec{y}$-variables are not independent in the fast subsystem (3.4). If we assume that the stationary measure of the fast subsystem is the uniform measure on the sphere, i.e. $\mu(\vec{y}|E_0) = \delta\left(\sum_{k=1}^{n} y_k^2 - E_0\right)$ (where E_0 is the particular energy level), then the marginal joint distribution for y_j, y_k converges to a product of independent Gaussian distributions as $n \to \infty$ ($n = \dim(\vec{y})$). Therefore, we expect the second condition in (3.11) to become approximately true for large n. This is similar to the condition for the mode-reduction in the TBH model considered in [62]. Therefore, the dimension n should be reasonable large, $n = O(10)$ or larger. This is to be investigated numerically.

Substituting L_1 and neglecting the derivatives ∂_{y_j} on the right (since the effective operator is applied to the function $u_0(x, E, t)$ which does not involve $\vec{y}$)

$$\begin{aligned} L &= -\int \left[\left(\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x + \sum_{j,k} A_{j1k}^{yxy} x y_k \partial_{y_j} - 2 \sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E \right) L_2^{-1}, \right. \\ &\quad \left. \left(\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x - 2 \sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E \right) \right] \mu(\vec{y}|E) d\vec{y}. \end{aligned} \tag{3.12}$$

3.4.1 Derivation of diffusion coefficients for the reduced model

Operator L in (3.12) can be partitioned into infinitesimal generator of the drift and infinitesimal generator of the diffusion process for the reduced model. We know in the Fokker-Planck equation that ∂_x^2 leads to the diffusion terms in the equation of x, similarly the diffusion terms for the reduced model is represented by ∂_x^2 and ∂_E^2 in (3.12). Thus, the diffusion part

in (3.12) is derived by neglecting the terms which involve ∂_{y_j} in L_1, i.e.

$$L_{diff} = -\int \left[\left(\sum_{j,k} A^{xyy}_{1jk} y_j y_k \partial_x - 2 \sum_{j,k} A^{xyy}_{1jk} x y_j y_k \partial_E \right) L_2^{-1}, \left(\sum_{j,k} A^{xyy}_{1jk} y_j y_k \partial_x - 2 \sum_{j,k} A^{xyy}_{1jk} x y_j y_k \partial_E \right) \right] \mu(\vec{y}|E) d\vec{y}. \quad (3.13)$$

Substituting the action of L_2^{-1} we obtain

$$L_{diff} = A \left[\partial_{xx} + 4x^2 \partial_{EE} - 2\partial_x x \partial_E - 2x \partial_E \partial_x \right], \quad (3.14)$$

where

$$A = \sum_{j,k,j',k'} A^{xyy}_{1jk} A^{xyy}_{1j'k'} Q_{j,k,j',k'}, \qquad Q_{j,k,j',k'} = \int_0^\infty \mathbb{E}_\mu \left[y_j y_k Y_{j'}(t) Y_{k'}(t) \right] dt, \quad (3.15)$$

where $Y_j(t)$ is the solution of the fast subsystem with $Y_j(0) = y_j$. The quantity $Q_{j,k,j',k'}$ is the area under fourth-order two-point moment in the fast subsystem (3.4) on the energy level E.

The infinitesimal generator of the diffusion process given by (3.14) is not in the canonical form, but can be easily rewritten as

$$L_{diff} = A \left[\partial_{xx} + 4x^2 \partial_{EE} - 2x \partial_x \partial_E - 2x \partial_E \partial_x \right] - 2A \partial_E, \quad (3.16)$$

Here we concentrate on the diffusion part of the operator, and the drift part will be considered later. The diffusion part in (3.16) can be written as

$$L_{diff} = \frac{1}{2} \sum_{i,k=1,2} (DD^T)_{ik} \partial_{z_i} \partial_{z_k} \quad (3.17)$$

where $z_1 \equiv x$, $z_2 \equiv E$,

$$DD^T = 2A \begin{pmatrix} 1 & -2x \\ -2x & 4x^2 \end{pmatrix}, \qquad D = \sqrt{2A} \begin{pmatrix} 1 & 0 \\ -2x & 0 \end{pmatrix}. \quad (3.18)$$

3.4.2 Rescaling the fast subsystem

In previous subsection 3.4.1, we defined $Q_{j,k,j',k'}$ as

$$Q_{j,k,j',k'} = \int_0^\infty \mathbb{E}_{\mu(\vec{y}|E)} \left[y_j y_k Y_{j'}(t) Y_{k'}(t) \right] dt, \tag{3.19}$$

where $Y_j(t)$ is the solution of the fast subsystem with initial condition $Y_j(0) = y_j$ and fixed energy E, $\mu(\vec{y}|E)$ is the stationary density of the fast subsystem (3.4) and $\mathbb{E}_{\mu(\vec{y}|E)}$ represents the expectation of the lagged fourth-moment with respect to the stationary invariant measure $\mu(\vec{y}|E)$ of the fast subsystem having conserved energy E. The quantity $Q_{j,k,j',k'}$ is the integrated fourth-order two-point moment in the fast subsystem (3.4) computed on the energy level E.

It is not feasible to compute and tabulate the fourth-moment above on all energy levels in order to perform simulations of the reduced model. Fortunately, the fast subsystem is invariant under the rescaling

$$y_j^{new} = \sqrt{C}\, y_j, \quad t^{new} = \frac{1}{\sqrt{C}}\, t$$

for any constant C. We use this property with $C = n/E$, i.e.

$$y_j \;=\; \sqrt{\frac{n}{E}}\, y_j, \quad t = \sqrt{\frac{E}{n}}\, t, \tag{3.20}$$

to rescale the fast subsystem to the energy level n.

Hence, the integrated lagged fourth-moment Q can be rewritten as

$$Q_{j,k,j',k'} = \left(\frac{E}{n}\right)^{3/2} \int_0^\infty \mathbb{E}_{\mu(y|E=n)} \left[y_j y_k Y_{j'}(t) Y_{k'}(t) \right] dt = \left(\frac{E}{n}\right)^{3/2} q_{j,k,j',k'}$$

where $q_{j,k,j',k'}$ is the integrated lagged fourth-moment in fast subsystem (3.4) with fixed energy $E = n$. Therefore, we can perform only one microcanonical simulation of the fast subsystem on the energy shell $E = n$ to compute $q_{j,k,j',k'}$, and quantities $Q_{j,k,j',k'}$ can be obtained by rescaling.

3.4.3 Deriving drift using Fokker-Planck equation

The drift in the effective equation can be computed directly by considering the part of the effective operator which was neglected in the computation of the diffusion or we can use the explicit knowledge of the stationary distribution for x and E to compute the drift by imposing that the Fokker-Planck operator annihilates the joint stationary density. The stationary density for x and E is the product of the Gaussian and the χ^2 densities.

We can write the form for the reduced model, using the diffusion terms obtained in (3.17), as

$$\begin{aligned} dx &= -\gamma x + \sigma dW_1 - \Gamma_1(x,E)dt + \sqrt{2A}dW_2, \\ dE &= -\Gamma_2(x,E)dt - 2x\sqrt{2A}dW_2, \end{aligned} \tag{3.21}$$

where A depends on E, i.e.

$$A = \sum_{j,k,j',k'} A^{xyy}_{1jk} A^{xyy}_{1j'k'} Q_{j,k,j',k'}, \quad Q_{j,k,j',k'} = \left(\frac{E}{n}\right)^{3/2} q_{j,k,j',k'}$$

or

$$A = \left(\frac{E}{n}\right)^{3/2} \sum_{j,k,j',k'} A^{xyy}_{1jk} A^{xyy}_{1j'k'} q_{j,k,j',k'} = \left(\frac{E}{n}\right)^{3/2} \beta, \tag{3.22}$$

where β does not depend on E and the model (3.21) can be rewritten as

$$\begin{aligned} dx &= -\gamma x + \sigma dW_1 - \Gamma_1(x,E)dt + \sqrt{2}\left(\frac{E}{n}\right)^{3/4} \beta dW_2, \\ dE &= -\Gamma_2(x,E)dt - 2\sqrt{2}x\left(\frac{E}{n}\right)^{3/4} \beta dW_2. \end{aligned} \tag{3.23}$$

The Fokker-Planck equation for the system (3.23) reads

$$\begin{aligned} &\left(-\frac{\partial}{\partial x}\Gamma_1 - \frac{\partial}{\partial E}\Gamma_2 + \frac{\partial^2}{\partial x^2}\Sigma_{11} + \frac{\partial^2}{\partial x \partial E}\Sigma_{12} + \frac{\partial^2}{\partial E \partial x}\Sigma_{21} + \frac{\partial^2}{\partial E^2}\Sigma_{22}\right)\rho(x,E), \\ = &\left(-\partial_{z_i}\Gamma_i + \sum_{i,j} \partial_{z_i}\partial_{z_j}\Sigma_{ij}\right)\rho(x,E) = 0, \end{aligned} \tag{3.24}$$

where $\Sigma = \frac{1}{2}DD^T$ with D in (3.44), i.e.

$$\Sigma = A \begin{pmatrix} 1 & -2x \\ -2x & 4x^2 \end{pmatrix} = \beta \left(\frac{E}{n}\right)^{3/2} \begin{pmatrix} 1 & -2x \\ -2x & 4x^2 \end{pmatrix}. \tag{3.25}$$

Also, we have neglected the terms in Fokker-Planck equation (3.24) which correspond to the OU terms in equation of x given by

$$dx = -\gamma x dt + \sigma dW_1,$$

since these terms will annihilate the Gaussian density separately from the rest of the Fokker-Planck equation (3.24).

The Fokker-Planck equation for the system (3.23) can be rewritten as

$$0 = \left(-\partial_{z_i}\Gamma_i(z) + \sum_{i,j} \partial_{z_i}\partial_{z_j}\Sigma_{ij} \right) \rho(x, E)$$

with Σ in (3.25). The idea is to transform the stationary distribution for x and E in the form $e^{-v(x,E)}$ and the diffusion terms in the form

$$\sum_i \partial_{z_i} \sum_j \left(\rho(z)\partial_{z_j}\Sigma_{ij} + \Sigma_{ij}\partial_{z_j}\rho(z) \right).$$

Defining $(\Sigma_{ij})_x = \partial_x \Sigma_{ij}$ and $(\Sigma_{ij})_E = \partial_E \Sigma_{ij}$, we open the summation in above equation as

$$\begin{aligned} & \sum_{i,j} \partial_{z_i}\partial_{z_j}\Sigma_{ij} = \\ & \sum_i \partial_{z_i} \sum_j (\partial_{z_j}\Sigma_{ij} + \Sigma_{ij}\partial_{z_j}) = \\ = \;& \partial_x \left((\Sigma_{11})_x + \Sigma_{11}\partial_x + (\Sigma_{12})_E + \Sigma_{12}\partial_E \right) \\ + \;& \partial_E \left((\Sigma_{21})_x + \Sigma_{21}\partial_x + (\Sigma_{22})_E + \Sigma_{22}\partial_E \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \Gamma_1\rho(x, E) &= \left[(\Sigma_{11})_x + \Sigma_{11}\partial_x + (\Sigma_{12})_E + \Sigma_{12}\partial_E \right] \rho(x, E), \\ \Gamma_2\rho(x, E) &= \left[(\Sigma_{21})_x + \Sigma_{21}\partial_x + (\Sigma_{22})_E + \Sigma_{22}\partial_E \right] \rho(x, E). \end{aligned}$$

Terms arising from differentiating Σ_{ik} become

$$\begin{aligned}(\Sigma_{11})_x + (\Sigma_{12})_E &= -3x\beta\frac{E^{1/2}}{n^{3/2}}, \\ (\Sigma_{21})_x + (\Sigma_{22})_E &= -2\beta\left(\frac{E}{n}\right)^{3/2} + 6x^2\beta\frac{E^{1/2}}{n^{3/2}}.\end{aligned} \tag{3.26}$$

The stationary distribution for x and E can be written in the following form

$$\begin{aligned}&\rho(x,E) = Ce^{-\left(\frac{\gamma}{\sigma^2}(E+x^2) - \frac{n-2}{2}\log(E)\right)} = Ce^{-v(x,E)} \\ &\text{with } \; v(x,E) = \frac{\gamma}{\sigma^2}(E+x^2) - \frac{n-2}{2}\log(E) \\ &\text{and } \; \partial_x v(x,E) = \frac{\gamma}{\sigma^2}2x, \qquad \partial_E v(x,E) = \frac{\gamma}{\sigma^2} - \frac{n-2}{2E}.\end{aligned}$$

We use $\partial_x\rho(x,E) = -\rho(x,E)v_x(x,E)$ and $\partial_E\rho(x,E) = -\rho(x,E)v_E(x,E)$ to compute

$$\begin{aligned}&\Sigma_{11}\rho_x(x,E) + \Sigma_{12}\rho_E(x,E) = \\ &\qquad -2x\beta\left(\frac{E}{n}\right)^{3/2}\frac{\gamma}{\sigma^2} + 2x\beta\left(\frac{E}{n}\right)^{3/2}\left[\frac{\gamma}{\sigma^2} - \frac{n-2}{2E}\right],\end{aligned} \tag{3.27}$$

$$\begin{aligned}&\Sigma_{21}\rho_x(x,E) + \Sigma_{22}\rho_E(x,E) = \\ &\qquad 4x^2\beta\left(\frac{E}{n}\right)^{3/2}\frac{\gamma}{\sigma^2} - 4x^2\beta\left(\frac{E}{n}\right)^{3/2}\left[\frac{\gamma}{\sigma^2} - \frac{n-2}{2E}\right].\end{aligned}$$

Combining (3.26) and (3.27) we obtain

$$\begin{aligned}\Gamma_1 &= -x\beta(n+1)\frac{E^{1/2}}{n^{3/2}}, \\ \Gamma_2 &= -2\beta\left(\frac{E}{n}\right)^{3/2} + 2x^2\beta(n+1)\frac{E^{1/2}}{n^{3/2}}.\end{aligned}$$

And the reduced model for x and E becomes

$$\begin{aligned}dx &= -\gamma x dt - (n+1)\beta x\frac{E^{1/2}}{n^{3/2}} + \sigma dW_1 + \sqrt{2\beta}\left(\frac{E}{n}\right)^{3/4}dW_2, \\ dE &= -2\beta\left(\frac{E}{n}\right)^{3/2} + 2(n+1)\beta x^2\frac{E^{1/2}}{n^{3/2}} - 2\sqrt{2\beta}x\left(\frac{E}{n}\right)^{3/4}dW_2.\end{aligned} \tag{3.28}$$

3.5 Alternative Derivation of the Reduced Model using Uniform Stationary Measure on the Sphere

In this section, we derive the reduced model (3.28) using uniform measure on the sphere as an explicit formula for the stationary distribution of the fast subsystem. The fast subsystem (3.4) is a deterministic system which conserves energy E, given by (3.5). The energy E of the fast variables changes in time due to the interaction with the slow variables and is assumed to be a hidden slow variable in (3.6). In this section, we assume the stationary distribution of the fast subsystem, for initial fixed energy as E, to be uniform on the sphere of radius $\sqrt{E}$.

3.5.1 Stationary distribution of the fast subsystem

We assume that the fast subsystem (3.4) is ergodic on the hypersphere defined by energy of the fast subsystem with respect to the uniform distribution on the sphere, i.e. the trajectory of the fast subsystem comes arbitrary close to any point on the sphere and covers the sphere randomly. As considered in [62], let $y_i(t, c)$ be the solution of the fast subsystem (3.4) at time t for the initial condition

$$y_i(0, c) = c_i,$$

with fixed energy given by

$$E_0 \;=\; \sum_{i=1}^{n} c_i^2. \tag{3.29}$$

Then the stationary distribution of the fast subsystem can be written as

$$\mu(\vec{y}|E) = S_n^{-1} E^{1-n/2} \delta\left(E - E_0\right), \tag{3.30}$$

where n is the total number of fast variables, E is the given fixed energy (3.29) in the fast subsystem, $S_n^{-1}E^{1-n/2}$ is the normalizing constant and δ is the dirac-delta function is to

specify that energy is conserved at every point in the fast subsystem.

3.5.2 Limit of the full model as $\epsilon \to 0$

We apply the stochastic mode reduction to the full model (2.2) exactly in similar way as done in section 3.4 but using the stationary measure of the fast subsystem assumed in (3.30). The Kolmogorov backward equation associated with (2.2) for a scalar function u is given by (3.9) where the operators are same as defined in (3.11). The compatibility conditions necessary for the stochastic mode reduction specified in (3.11) holds as the number of fast variables converges to infinity. Assuming the stationary measure of the fast subsystem to be the uniform measure on the sphere given by (3.30), the marginal joint distribution for y_j, y_k converges to the product of independent Gaussian as $n \to \infty$. Therefore, we expect the second condition in (3.11) to become approximately true for large n. The effective backward operator in (3.12) is rewritten as

$$\begin{aligned} L &= -\int \left[\left(\sum_{j,k} A^{xyy}_{1jk} y_j y_k \partial_x + \sum_{j,k} A^{yxy}_{j1k} x y_k \partial_{y_j} - 2 \sum_{j,k} A^{xyy}_{1jk} x y_j y_k \partial_E \right) L_2^{-1}, \right. \\ &\quad \left. \left(\sum_{j,k} A^{xyy}_{1jk} y_j y_k \partial_x - 2 \sum_{j,k} A^{xyy}_{1jk} x y_j y_k \partial_E \right) \right] \mu(\vec{y}|E) d\vec{y}, \end{aligned} \tag{3.31}$$

where the operator L_2^{-1} can be derived using Feynmann-Kac technique and is given by

$$L_2^{-1} f(y) = -\int_0^\infty E\left[(f(y_\tau | y_0 = y) \right] d\tau, \tag{3.32}$$

where y_τ is the solution of the fast subsystem (3.4) at time τ and $E\left[(f(y_\tau|y_0 = y) \right]$ is the conditional expectation with respect to y_τ if given initial value of y is y_0.

Since the stationary measure of fast variables in the fast subsystem depends on the energy E and the y variables, hence the operator L_2^{-1} also depends on both, E and y_i.

Therefore, to understand the derivatives $\partial_E L_2^{-1}$ and $\partial_{y_i} L_2^{-1}$, we split the effective operator L in (3.31) into three parts, given by

$$L = -(I_1 + I_2 + I_3), \tag{3.33}$$

where the splitting is as follows after applying the action of L_2^{-1}

$$\begin{aligned}
I_1 &= \int \mu(\vec{y}|E) \left[\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x L_2^{-1} \left(\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x - 2 \sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E \right) \right] d\vec{y}, \\
I_2 &= \int \mu(\vec{y}|E) \left[\sum_{j,k} A_{j1k}^{yxy} x y_k \partial_{y_j} L_2^{-1} \left(\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x - 2 \sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E \right) \right] d\vec{y}, \\
I_3 &= \int \mu(\vec{y}|E) \left[-2 \sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E L_2^{-1} \left(\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x - 2 \sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E \right) \right] d\vec{y}.
\end{aligned} \tag{3.34}$$

Next, we will simplify all three terms in operator L separately.

Computing I_1

Applying action of L_2^{-1} on the I_1 term in (3.34), we obtain

$$I_1 = \int_0^\infty \int \mu(\vec{y}|E) \left[\sum_{j,k} A_{1jk}^{xyy} y_j y_k \partial_x \left(\sum_{j',k'} A_{1j'k'}^{xyy} Y_{j'}(t) Y_{k'}(t) (\partial_x - 2x\partial_E) \right) \right] d\vec{y} d\tau$$

where $Y_j(t)$ is the solution of the fast subsystem with $Y_j(0) = y_j$.

Since stationary measure of fast variables in fast subsystem is independent of x, the partial derivatives can be pulled out and the I_1 expression gets simplified as

$$I_1 = A(\partial_{xx} - 2x\partial_{xE} - 2\partial_E), \tag{3.35}$$

where

$$A = \sum_{j,k,j',k'} A_{1jk}^{xyy} A_{1j'k'}^{xyy} Q(j,k,j',k')$$

and $Q_{j,k,j',k'}$ is the integrated fourth-order two-point moment, given by (3.15), in the fast subsystem (3.4) on the energy level E. The expression for I_1 in (3.35) can be further simplified after rescaling of the fast subsystem as follows

$$I_1 = \left(\frac{E}{n}\right)^{3/2} \beta(\partial_{xx} - 2x\partial_{xE} - 2\partial_E), \tag{3.36}$$

where the fast subsystem is rescaled to the system with fixed energy $E = n$, described in (3.20) and (3.22).

Computing I_2

Applying the action of L_2^{-1} (3.32) on I_2 term in (3.34), we obtain

$$\begin{aligned} I_2 &= \int_0^\infty \int \mu(\vec{y}|E) \left[\sum_{j,k} A^{yxy}_{j1k} x y_k \partial_{y_j} \left(\sum_{j',k'} A^{xyy}_{1j'k'} Y_{j'}(t) Y_{k'}(t)(\partial_x - 2x\partial_E) \right) \right] d\vec{y} d\tau, \\ &= x \sum_{j,k,j',k'} A^{yxy}_{j1k} A^{xyy}_{1j'k'} \int_0^\infty \int \mu(\vec{y}|E) \left[y_k \partial_{y_j} \left(Y_{j'}(t) Y_{k'}(t)(\partial_x - 2x\partial_E) \right) \right] d\vec{y} d\tau, \end{aligned} \tag{3.37}$$

where $Y_j(t)$ is the solution of the fast subsystem with $Y_j(0) = y_j$. The integration in the expression for I_2 (3.37) can be further simplified by integration by parts if we denote

$$u = y_k \mu(\vec{y}|E),\ dv = \partial_{y_j} \left(Y_{j'}(t) Y_{k'}(t) \left(\partial_x - 2x\partial_E\right) \right) d\vec{y},$$

where u and v are two continuously differentiable functions and dv is the partial derivative of v with respect to y_j. We obtain partial derivative of u w.r.t. y_j and integration of v as

$$du = -2y_j y_k S_n^{-1}(E)^{1-n/2} \delta'\left(E - \sum_{r=1}^n y_r^2\right),\ v = Y_{j'}(t) Y_{k'}(t) \left(\partial_x - 2x\partial_E\right),$$

where $\delta'(\cdot)$ is the distributional derivative of dirac-delta function $\delta(\cdot)$. Above integration by parts simplifies I_2 term in (3.37) as follows

$$I_2 \;=\; -2x S_n^{-1}(E)^{1-n/2} D \left(\partial_x - 2x\partial_E\right),$$

where D is defined as follows

$$D = \sum_{j,k,j',k'} A^{yxy}_{j1k} A^{xyy}_{1j'k'} \int_0^\infty \int y_j y_k Y_{j'}(t) Y_{k'}(t) \delta'\left(E - \sum_{r=1}^n y_r^2\right) d\vec{y} d\tau. \tag{3.38}$$

Computing I_3

Applying the action of L_2^{-1} (3.32) on I_3 term in (3.34), we obtain

$$\begin{aligned} I_3 &= -2\int_0^\infty \int \mu(\vec{y}|E)\left[\sum_{j,k} A_{1jk}^{xyy} x y_j y_k \partial_E \left(\sum_{j',k'} A_{1jk}^{xyy} Y_{j'}(t) Y_{k'}(t)(\partial_x - 2x\partial_E)\right)\right] d\vec{y}d\tau, \\ &= -2x \sum_{j,k,j',k'} A_{1jk}^{xyy} A_{1j'k'}^{xyy} \int_0^\infty \int \mu(\vec{y}|E)\left[y_j y_k \partial_E \left(Y_{j'}(t)Y_{k'}(t)(\partial_x - 2x\partial_E)\right)\right] d\vec{y}d\tau. \end{aligned}$$

Since the solution of the fast subsystem at time t, i.e., $Y_{j'}(t)$ depends on the fixed energy E of the fast subsystem, the only way to simplify the partial derivative w.r.t. E in I_3 term is by using product rule as

$$\begin{aligned} I_3 &= -2x \sum_{j,k,j',k'} A_{1jk}^{xyy} A_{1j'k'}^{xyy} \int_0^\infty \int \mu(\vec{y}|E)\left[y_j y_k \partial_E \left(Y_{j'}(t)Y_{k'}(t)\right)\right] d\vec{y}d\tau \ (\partial_x - 2x\partial_E) \\ &\quad -2xA(\partial_{Ex} - 2x\partial_{EE}), \end{aligned}$$

where A is the function of integrated fourth-order two-point moment, given by (3.15), in the fast subsystem (3.4) on the energy level E.

Let the first term left for integration in I_3 be denoted by I_3' which can be simplified by tedious but otherwise completely straightforward algebraic manipulation. Hence, I_3 term can be rewritten as

$$I_3 = -2x \sum_{j,k,j',k'} A_{1jk}^{xyy} A_{1j'k'}^{xyy} I_3' \ (\partial_x - 2x\partial_E) - 2Ax(\partial_{Ex} - 2x\partial_{EE}), \tag{3.39}$$

where I_3' is given by

$$I_3' = \int_0^\infty \int \mu(\vec{y}|E) y_j y_k \partial_E \left[Y_{j'}(t)Y_{k'}(t)\right] d\vec{y}d\tau. \tag{3.40}$$

Note that we can't pull partial derivative w.r.t. E out of integration in expression of I_3' (3.40) since the fast variables and their density $\mu(\vec{y}|E)$ are functions of energy E in the fast subsystem (3.4). Hence, we will consider separately the situation of having ∂_E outside the integral and manipulate the considered expression to simplify the required I_3' (3.40).

The expression similar to I_3' but with ∂_E outside the integration is given below with the simplification of considered expression using the 'Fundamental Theorem of Calculus',

$$\partial_E \left[\int_0^\infty \int \mu(\vec{y}|E) y_j y_k Y_{j'}(t) Y_{k'}(t) d\vec{y} d\tau \right] = \int_0^\infty \int \mu(\vec{y}|E) \partial_E \left[y_j y_k Y_{j'}(t) Y_{k'}(t) \right] d\vec{y} d\tau.$$

The partial derivative w.r.t. energy E in above expression can be split into the fourth lagged moment of y variables and the measure $\mu(\vec{y}|E)$ using the product rule as

$$\partial_E \left[\int_0^\infty \int \mu(\vec{y}|E) y_j y_k Y_{j'}(t) Y_{k'}(t) d\vec{y} d\tau \right] = \int_0^\infty \int y_j y_k Y_{j'}(t) Y_{k'}(t) \partial_E \left[\mu(\vec{y}|E) \right] d\vec{y} d\tau + I_3',$$

where I_3' is given in (3.40). Therefore, the required I_3' can be rewritten as

$$I_3' = \partial_E \left[Q(j,k,j',k') \right] - \int_0^\infty \int y_j y_k Y_{j'}(t) Y_{k'}(t) \partial_E \left[\mu(\vec{y}|E) \right] d\vec{y} d\tau, \tag{3.41}$$

where $Q(j,k,j',k')$ is the integrated fourth-order two-point moment, given by (3.19), in the fast subsystem (3.4) on the energy level E. Since the stationary measure of fast variables $\mu(\vec{y}|E)$ given by (3.30), is a function of E, we can simplify the second term in I_3' using chain rule as

$$\partial_E \left[\mu(\vec{y}|E) \right] = (1 - n/2) E^{-1} \mu(\vec{y}|E) + S_n^{-1} E^{1-n/2} \delta' \left(E - \sum_{i=1}^n y_k^2 \right),$$

where $\delta'(\cdot)$ represent the distributional derivative of dirac-delta function $\delta(\cdot)$. Substituting the above specified chain rule in (3.41) simplifies I_3' as

$$\begin{aligned} I_3' &= \partial_E \left[Q(j,k,j',k') \right] - (1 - N/2) E^{-1} Q(j,k,j',k') \\ &\quad - S_n^{-1} E^{1-n/2} \int_0^\infty \int y_j y_k Y_{j'}(t) Y_{k'}(t) \delta' \left(E - \sum_{i=1}^n y_k^2 \right) d\vec{y} d\tau, \end{aligned}$$

where $Q(j,k,j',k')$ is the function of the energy E and the fast variables, given by (3.19). Note that rescaling of the fast subsystem, described in (3.20) factorizes $Q(j,k,j',k')$ into two separate functions of the energy E and the fast variables, respectively. Applying the rescaling technique of the fast subsystem, given by (3.20), in first two terms of I_3' and

partially differentiating the factorization of $Q(j,k,j',k')$ (3.21) w.r.t. energy E, simplifies I_3' as follows

$$\begin{aligned} I_3' &= \frac{3}{2n^{3/2}}E^{1/2}q(j,k,j',k') - \frac{1-n/2}{n^{3/2}}E^{1/2}q(j,k,j',k') \\ &\quad -S_n^{-1}E^{1-n/2}\int_0^\infty \int y_j y_k Y_{j'}(t)Y_{k'}(t)\delta'\left(E-\sum_{i=1}^n y_k^2\right)d\vec{y}d\tau, \\ &= \frac{n+1}{2n^{3/2}}E^{1/2}q(j,k,j',k') - S_n^{-1}E^{1-n/2}\int_0^\infty \int y_j y_k Y_{j'}(t)Y_{k'}(t)\delta'\left(E-\sum_{i=1}^n y_k^2\right)d\vec{y}d\tau, \end{aligned}$$

Substituting I_3' back in the main equation of I_3 (3.39) and substituting $Q(j,k,j',k')$ with rescaled version given by (3.21), we obtain

$$\begin{aligned} I_3 &= -x\frac{n+1}{n^{3/2}}E^{1/2}\beta(\partial_x - 2x\partial_E) - 2x\frac{E^{3/2}}{n^{3/2}}\beta(\partial_{Ex}-2x\partial_{EE}) \\ &\quad +2xS_n^{-1}E^{1-n/2}D(\partial_x - 2x\partial_E), \end{aligned} \tag{3.42}$$

where β is summations of the area under fourth-order two-point moment, given by (3.22), in the rescaled fast subsystem (3.4) on the energy level n and D is given by (3.38).

Recall that the effective L operator is given by (3.31) which is splitted into I_1, I_2 and I_3 terms as in (3.33). Substituting I_1, I_2 and I_3 from (3.36), (3.38) and (3.42), respectively, effective L operator can be rewritten as

$$\begin{aligned} L &= \frac{1}{n^{3/2}}E^{3/2}\beta(\partial_{xx}-2x\partial_{xE}-2\partial_E) - 2xS_n^{-1}E^{1-n/2}D\left(\partial_x - 2x\partial_E\right) \\ &\quad -\frac{n+1}{n^{3/2}}xE^{1/2}\beta(\partial_x-2x\partial_E) - \frac{2}{n^{3/2}}xE^{3/2}\beta(\partial_{Ex}-2x\partial_{EE}) \\ &\quad +2xS_n^{-1}E^{1-n/2}D(\partial_x-2x\partial_E), \\ &= -\frac{E^{1/2}}{n^{3/2}}\beta\left((n+1)x\partial_x + 2E\partial_E - 2x^2(n+1)\partial_E\right) \\ &\quad +\frac{E^{3/2}}{n^{3/2}}\beta\left(\partial_{xx}-2x\partial_{xE}-2x\partial_{Ex}+4x^2\partial_{EE}\right), \end{aligned}$$

where β is summations of the area under fourth-order two-point moment, given by (3.22), in the rescaled fast subsystem (3.4) on the energy level n and D is given by (3.38). Let

L_{diff} represent the diffusion part and L_{drift} part of the operator L in (3.43). The diffusion part of L can be rewritten as

$$L_{diff} = \frac{1}{2} \sum_{i,k=1,2} (DD^T)_{ik} \partial_{z_i} \partial_{z_k} \tag{3.43}$$

where $z_1 \equiv x$, $z_2 \equiv E$,

$$DD^T = 2\beta \left(\frac{E}{n}\right)^{3/2} \begin{pmatrix} 1 & -2x \\ -2x & 4x^2 \end{pmatrix}, \qquad D = \sqrt{2\beta} \left(\frac{E}{n}\right)^{3/4} \begin{pmatrix} 1 & 0 \\ -2x & 0 \end{pmatrix}. \tag{3.44}$$

Using the effective L operator in (3.43) and diffusion in (3.43), we obtain same reduced model for x and E as in the first derivation (3.28), given by

$$\begin{aligned} dx &= -\gamma x dt - (n+1)\beta x \frac{E^{1/2}}{n^{3/2}} + \sigma dW_1 + \sqrt{2\beta} \left(\frac{E}{n}\right)^{3/4} dW_2, \\ dE &= -2\beta \left(\frac{E}{n}\right)^{3/2} + 2(n+1)\beta x^2 \frac{E^{1/2}}{n^{3/2}} - 2\sqrt{2\beta} x \left(\frac{E}{n}\right)^{3/4} dW_2. \end{aligned} \tag{3.45}$$

where β is the function of the integrated fourth-order two-point moment, given by (3.22), in the rescaled fast subsystem (3.4) on the energy level n.

3.6 Numerical Simulations

We will fix the parameters in full model (3.2) as

$$\gamma = 1, \quad \sigma = 2.236, \quad \dim(\vec{y}) = n = 10. \tag{3.46}$$

$$\text{Number of } xyy \text{ triads } = 10, \qquad \text{Number of } yyy \text{ triads } = 19$$

Coefficients are given in Tables 3.2 and 3.1. We perform simulations with coefficients $A = A/4$ where A is given in Table 3.1, the coefficients are divided by four to achieve better scale separation between x and y variables.

Numerical Method

We use a split-step method to integrate the full model (3.2). We perform a high-order deterministic step for the deterministic part of the model and then use Euler discretization to add a Gaussian random variable which approximates the increment in the Brownian motion. The full model is sensitive to the deterministic integrator, thus we use RK5 formula to integrate numerically the deterministic part of the system (3.2).

We use a time-averaging combined with Monte-Carlo approach to accelerate computations. In particular, we perform M runs of the full model with $T = 40,000$ with different initial conditions and then average over the M realizations. We choose $M = 3$ for $\varepsilon = 1$, 0.5 and $M = 10$ for $\varepsilon = 0.25$, 0.1 because smaller values of ε are more sensitive and $\varepsilon = 0.25$, 0.1 are used to compare with the simulations of the reduced model. We use integration time-step as $\delta t = 10^{-6}$ and $\delta t = 2 \times 10^{-5}$ for $\varepsilon = 0.1$ and $\varepsilon = 0.25$, respectively.

The reduced model is integrated using the same method with the RK5 deterministic integrator and Euler discretization for the noise. We also perform a hybrid approach combining time-averaging and Monte-Carlo simulations. We run $M = 10$ trajectories each with $T = 100,000$, $\Delta t = 0.00001$, and different initial conditions, and then take the empirical average of statistical quantities computed from all M trajectories. The coefficient β in reduced model (3.28) is computed from the fast subsystem numerically and is

$$\beta = 1.2759.$$

Convergence of full model as $\varepsilon \to 0$. We perform simulations with

$$\varepsilon = 1,\ 0.5,\ 0.25,\ 0.1$$

too illustrate the convergence of the full model as $\varepsilon \to 0$. The stationary density for x and E is reproduced quite accurately for any ε and we only depict the convergence of correlation functions in Figures 3.1 and 3.2.

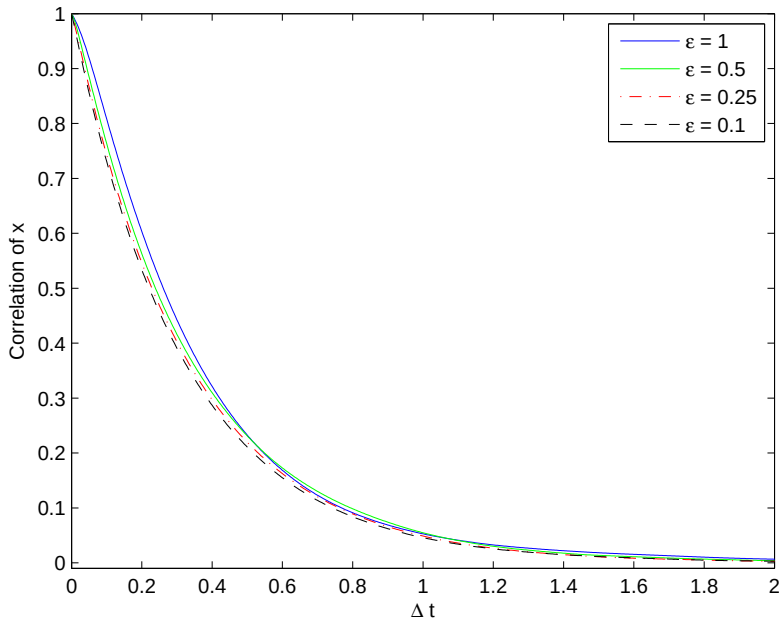

Figure 3.1: Correlation function of x in full model (3.2) for $\varepsilon = 1,\ 0.5,\ 0.25,\ 0.1$.

Comparison with the Reduced Model.

We show the weak convergence of full model (3.2) to the reduced model (2.29) as $\epsilon \to 0$ by comparing the auto-correlation, density and Kurtosis of slow variables in the full and reduced model. We perform simulations of full model with

$$\varepsilon = 0.25,\ 0.1$$

to illustrate the convergence of the full model to the reduced model as $\epsilon \to 0$.

Convergence of auto-correlation. In figure 3.3, the left and right part shows the convergence of auto-correlation of x and E, respectively, in the full model to the corresponding auto-correlations in the reduced model as $\epsilon \to 0$.

Convergence of density. In figure 3.4, the stationary density for x and E is reproduced quite accurately for any ϵ by the reduced model. Note that stationary density of

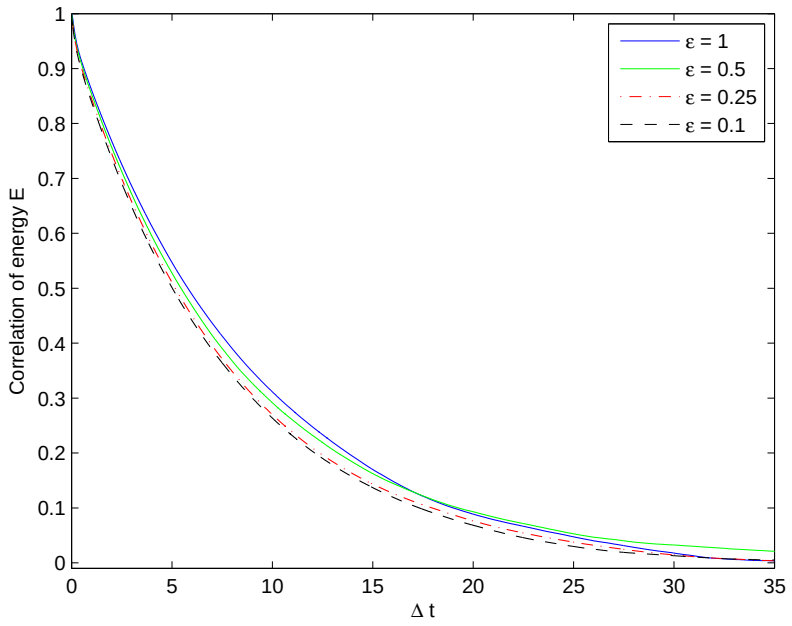

Figure 3.2: Correlation function of E in full model (3.2) for $\varepsilon = 1,\ 0.5,\ 0.25,\ 0.1$.

energy E in right part of figure 3.4 is same as the density of the chi-square distribution with nine degrees of freedom.

Convergence of kurtosis. Figure 3.5 shows the convergence of kurtosis of x and E in the full model to the corresponding kurtosis in the reduced model as $\epsilon \to 0$. Left part of figure 3.5 shows that at time lag $\tau = 0$, x is Gaussian but has non-Gaussian behavior as time lag τ increases and the figure shows that the reduced model is able to reproduce the non-Gaussian behavior very well as $\epsilon \to 0$.

Kurtosis of x at time lag τ is given by

$$Kurt(\tau) = \frac{\mathrm{Cov}_{x^2}(\tau)}{\mathrm{Cov}_x^2(0) + 2\mathrm{Cov}_x^2(\tau)}, \tag{3.47}$$

where the lagged covariance is as follows

$$\mathrm{Cov}_x(\tau) = \mathbb{E}\left[x(t)x(t+\tau)\right].$$

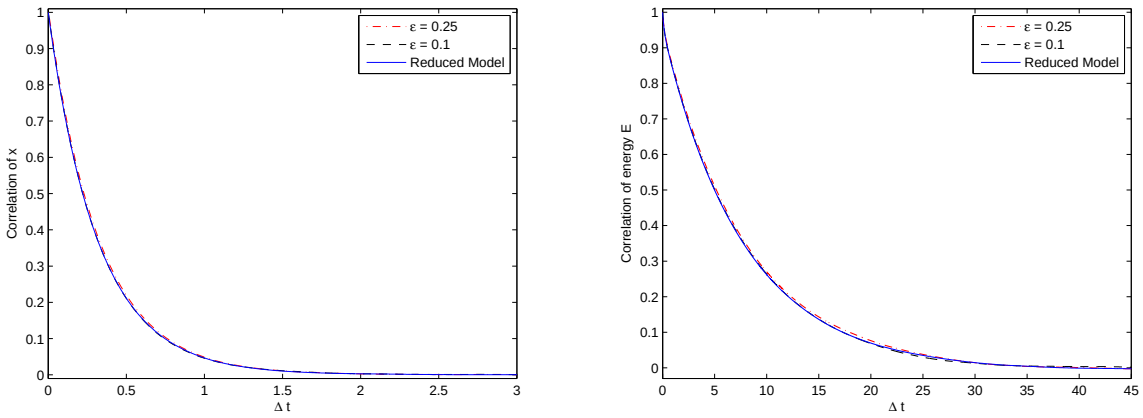

Figure 3.3: Correlation function of x and E in the full model (3.2) with $\varepsilon = 0.25$, 0.1 and in the reduced model (3.28).

3.7 Conclusion Remarks

In this chapter, we considered the system (3.6) where stochastic terms are added to the slow variables in energy-conserving non-linear systems. Since the noise only affect the slow variables, the fast subsystem is deterministic evolving on a sphere of constant energy. On the other hand, the radius of the sphere changes in full model due to the coupling between the slow and fast dynamics, we consider the energy of the fast subsystem as an additional hidden slow variable. The main goal is to derive a closed form stochastic model for the slow modes alone, i.e., x and hidden slow variable as energy E.

The reduced model for slow variables in full model (3.6) is derived using two similar ways in this chapter. In section 3.3, it is shown that x and y variables follow independent Normal distribution with identical mean and variance, their product measure is given by (3.7). Also, hidden slow variable E in full model follow chi-square distribution, given by (3.8). In 3.4, we use the explicit knowledge of the stationary distribution for x and E to

Coefficients	y variables	A_1	A_2	A_3
$A_{1,1,2}$	y_1, y_2	4.8	-2.2	-2.6
$A_{1,10,11}$	y_{10}, y_{11}	2.1	1	-3.1
$A_{1,4,12}$	y_4, y_{12}	5.4	-2.9	-2.5
$A_{1,7,8}$	y_7, y_8	4.5	-2	-2.5
$A_{1,3,9}$	y_3, y_9	5.4	-2.9	-2.5
$A_{1,1,12}$	y_1, y_{12}	2.1	1	-3.1
$A_{1,2,4}$	y_2, y_4	4.8	-2.2	-2.6
$A_{1,5,8}$	y_5, y_8	4.5	-2	-2.5
$A_{1,7,9}$	y_7, y_9	3.5	-1.2	-2.3
$A_{1,3,6}$	y_3, y_6	5.0	-2.5	-2.5

Table 3.1: Coefficients $A^{xyy}, A^{yxy}, A^{yyx}$ used in coupling of x and y variables in full model (3.2).

Coefficients	y variables	B_1	B_2	B_3
$B_{1,2,3}$	y_1, y_2, y_3	2	2.5	-4.5
$B_{1,2,4}$	y_1, y_2, y_4	4.2426	2.8284	-7.071
$B_{1,2,9}$	y_1, y_2, y_9	-1.2247	2.9393	-1.7146
$B_{1,2,10}$	y_1, y_2, y_{10}	2.1166	2.9103	-5.0269
$B_{1,3,4}$	y_1, y_3, y_4	1.7321	2.5981	-4.3302
$B_{1,5,6}$	y_1, y_5, y_6	3.8013	4.9193	-8.7206
$B_{1,9,10}$	y_1, y_9, y_{10}	3.9598	-2.2627	-1.6971
$B_{2,3,4}$	y_2, y_3, y_4	-2	4	-2
$B_{2,5,6}$	y_2, y_5, y_6	-4.5	2.1	2.4
$B_{2,9,10}$	y_2, y_9, y_{10}	1.7393	1.4230	-3.1623
$B_{3,7,8}$	y_3, y_7, y_8	1.1608	2.3217	-3.4825
$B_{4,7,8}$	y_4, y_7, y_8	-1.7321	-2.0785	3.8106
$B_{5,6,7}$	y_5, y_6, y_7	2.9566	2.0912	-5.0478
$B_{5,6,8}$	y_5, y_6, y_8	-2.6192	-1.4966	4.1158
$B_{5,7,8}$	y_5, y_7, y_8	4.6476	2.7111	-7.3587
$B_{5,6,9}$	y_5, y_6, y_9	-3	-1.8	4.8
$B_{5,6,10}$	y_5, y_6, y_{10}	1.8554	2.2677	-4.1231
$B_{6,7,8}$	y_6, y_7, y_8	4.6669	2.9698	-7.6367
$B_{8,9,10}$	y_8, y_9, y_{10}	3.923	2.3974	-6.3204

Table 3.2: Coefficients B^{yyy} used in coupling of y variables in full model (3.2).

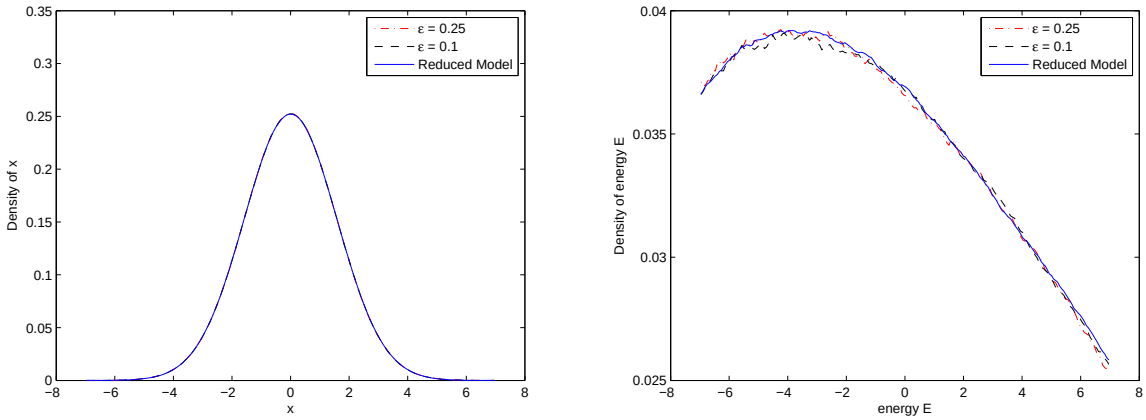

Figure 3.4: Marginal density of slow variables x and E in the full model with $\varepsilon = 0.25$, 0.1 and in the reduced model.

derive the reduced model by imposing that the Fokker-Planck operator annihilates the joint stationary density.

In section 3.5, we derive the reduced model in an alternative way using the stationary distribution of y variables in fast subsystem, given by (3.30). The fast subsystem (3.4) is deterministic, hence, we assume y variables to be uniformly distributed over the sphere of radius as initial energy of the system. The alternative method provides the same reduced model as derived using Fokker-Planck equation in section 3.4.

The reduced model consisting of variables x and E is given by (3.28) and the analytical expressions for parameters of reduced model is dependent on the fourth-order two-point moments of y variables, denoted as $Q(j,k,j',k')$ which on other hand, is dependent on initial energy of the fast subsystem. Since the radius of the sphere changes stochastically in full model, we would need to simulate the fast subsystem with all possible values of E as initial energy and compute $Q(j,k,j',k')$ which is clearly not feasible. Fortunately, the fast subsystem (3.4) is invariant under the rescaling given by (3.20) which factorizes $Q(j,k,j',k')$ into function of E and fourth-order two-point moment of fast subsystem with

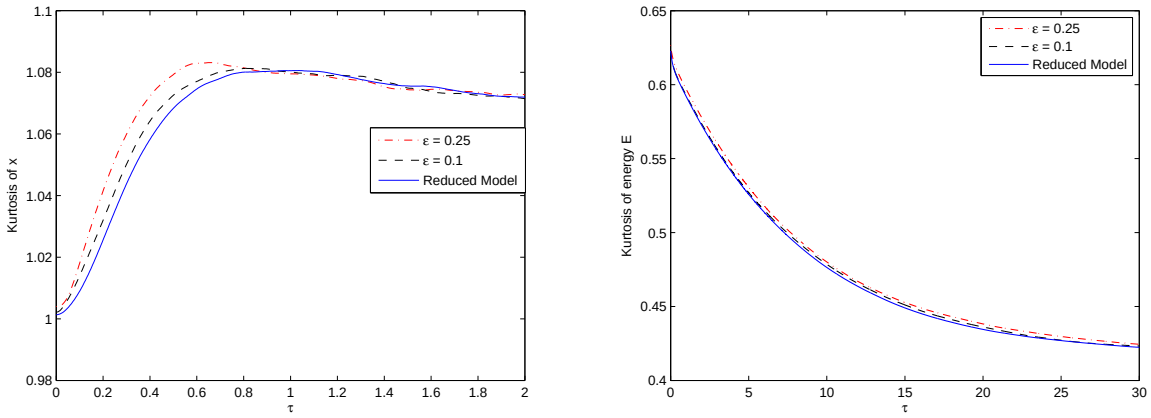

Figure 3.5: kurtosis (3.47) of slow variables x and E in full model with $\varepsilon = 0.25$, 0.1 and in the reduced model. Note: variable x is only weakly non-Gaussian; the vertical scale for the kurtosis of x is rather small; errors between the full and reduced model for the kurtosis of x are approximately 1%.

fixed energy n, separately. The consequence of rescaling leads to simulation of the fast subsystem only once with fixed initial energy $E = n$ (n is number of y variables) and hence, simplification of parameters in reduced model.

In section 3.6, fixing the coupling, drift and diffusion coefficients in full model (3.6), we verify numerically the convergence of full model to reduced model as $\epsilon \to 0$. The statistics that we have used to verify the numerical convergence is auto-correlation, density and kurtosis of x and E variables. The stochastic mode reduction technique developed in this chapter works on certain assumptions on the full model which are as follows

- The number of fast variables needs to be large to ensure the ergodicity of the fast subsystem.

- The fast subsystem needs to be chaotic to ensure the decay of the lagged fourth two-point moment in finite time, i.e. fourth-order two-point moment used to compute Q(j,k,j',k') in (3.15) should decay in finite time else it will be not feasible to compute

accurate value of Q(j,k,j',k').

- Coupling coefficients of the fast subsystem needs to be energy conservative (second equation in (3.3)) to ensure that the fast subsystem conserves energy.

- Energy E given by (3.5) is required to be slow variable.

CHAPTER 4

Parametric Estimation for Fast-oscillating Potential Model under Indirect Observability

In this chapter, we study the adequate data subsampling for consistent parametric estimation of unobservable stochastic differential equations (SDEs) under Indirect Observability, similar to the study done for Lorenz-96 model in chapter 2. Unlike non-linear reduced model (2.29) in chapter 2, here, we consider the multi-scale model having linear reduced model. In two recent papers [11, 10], authors have provided a rigorous foundation for the parameters estimation of linear stochastic model under Indirect Observability. The authors in [11, 10] considered the asymptotic behavior of the approximate Maximum Likelihood estimators for the unknown parameters of the stochastic model, using the observations of multi-scale approximating process as scale separation parameter $\epsilon \to 0$. In particular, they demonstrated that for consistent estimation of the diffusion parameters the underlying dataset has to be subsampled with time-steps constrained by specific subsampling criteria, depending on the

value of the multi-scale parameter ϵ. Otherwise, if these subsampling criteria are violated, the estimated underlying diffusion model will not reproduce the statistical features of the data and the corresponding parameter estimators will be biased even in the limit $\epsilon \to 0$.

We extend the results in [11, 10] on the model with the fast-oscillating potential to illustrate the subsampling problem. First, the numerical investigation of the subsampling criteria derived in [11, 10] in the context of homogenized models is performed.

Another important aspect discussed in the chapter is estimation of an effective model from a dataset generated with a fixed but unknown value of the scale separation parameter ϵ. This issue is important in practical situations, since there has been a considerable effort to efficiently parametrize a stochastic model for the large-scale structures from numerical simulations of various geophysical models. In [9], authors introduced a regression approach for constructing bias-corrected estimators from a single dataset generated by a multi-scale approximate dynamics with a fixed, but unknown value of the parameter ϵ. We extend the regression approach introduced in [9] to the multi-scale fast-oscillating potential model and verify it numerically.

4.1 Multi-scale Model with Fast-oscillating Potential

We consider the multi-scale model given by

$$dx_t = -gV'(x_t)dt - \frac{1}{\epsilon}p'\left(\frac{x_t}{\epsilon}\right)dt + \sqrt{2s}\,dB_t, \tag{4.1}$$

where B_t is the Brownian motion, the large-scale potential is $V(x) = \frac{1}{2}x^2$, and the fast-oscillating part of the potential is $p(x) = \cos(x)$. The resulting SDE becomes

$$dx_t = -gx_t dt + \frac{1}{\varepsilon}\sin\left(\frac{x_t}{\varepsilon}\right)dt + \sqrt{2s}dB_t. \tag{4.2}$$

The model in (4.2) has been considered previously in [68, 67] in which the authors derived the reduced model for (4.2) and tested the subsampling issue for estimating the

parameters of the reduced model when the data is generated by the model (4.2). In this chapter, we perform a more detailed numerical study of parametric estimation under Indirect Observability and state the explicit conditions for the consistency of the estimators. Second, we describe the construction of new bias-corrected estimators from datasets generated by the multi-scale model (4.2) with an unknown fixed parameter ϵ.

The invariant density of the model (4.2) can be computed explicitly using the Fokker-Planck equation and is equal to

$$\rho_{inv}(x) = Ce^{-v(x)}, \quad \text{where} \quad v(x) = \frac{g}{2s}x^2 + \frac{1}{s}\cos\left(\frac{x}{\varepsilon}\right), \tag{4.3}$$

and C is the normalization constant. Therefore, the invariant density has a fast-oscillating component, but converges to the Gaussian density weakly in the sense of test-functions.

4.1.1 Homogenization for the potential model

The homogenization for parabolic equations has been shown in [69] and the references therein. As $\epsilon \to 0$, the process x_t in (4.2) converges to the Ornstein-Uhlenbeck process

$$dX_t = -\gamma X_t + \sigma dW_t, \tag{4.4}$$

with parameters γ and σ given by

$$\gamma = \alpha\frac{L^2}{Z\tilde{Z}}, \quad \sigma = \sqrt{2s\frac{L^2}{Z\tilde{Z}}}, \tag{4.5}$$

where L is the period of function $p(x)$ and hence equal to 2π and

$$Z = \int_0^L e^{\cos(y)/s}\,dy, \quad \tilde{Z} = \int_0^L e^{-\cos(y)/s}\,dy.$$

The convergence of $x_t \to X_t$ as $\epsilon \to 0$ is a weak convergence of generators, as proved in [69] using homogenization technique for parabolic equations. Note that $Z = \tilde{Z}$ in the particular case of model (4.2) due to the symmetry of the full model.

In later sections, we fix the damping and diffusion coefficients of the full model (4.2) as

$$g = 0.5, \quad s = 1, \tag{4.6}$$

and thus obtain the corresponding homogenized coefficients using (4.5) as

$$\gamma = 0.3119, \quad \sigma = 1.117. \tag{4.7}$$

4.2 Parameter Estimation for the Model with the Fast-oscillating Potential

The solution of X_t in the OU SDE (4.4) is well known having the covariances

$$r_0 = \mathbb{E}[X_t^2] = \frac{\sigma^2}{2\gamma}, \quad r_1 = E\big[X_t X_{t+\Delta}\big] = r_0 \exp(-\gamma\Delta), \tag{4.8}$$

where $\mathbb{E}[\cdot]$ denotes the expected value with respect to the invariant distribution. Therefore, the parameters γ and σ in the OU SDE (4.4) can be expressed through the lagged covariances using the relationships in (4.8), as

$$\gamma = g(r_0, r_1) = -\frac{1}{\Delta}\log\left(\frac{r_1}{r_0}\right), \quad \sigma^2 = h(r_0, r_1) = -\frac{2r_0}{\Delta}\log\left(\frac{r_1}{r_0}\right) = 2r_0\gamma. \tag{4.9}$$

The main goal is to estimate the parameters γ and σ efficiently in reduced model (4.4) using the observations of x_t in full model as $\epsilon \to 0$. Assume to be given the discrete observations of x_t in full model (4.2) for fixed $\epsilon > 0$, subsampled at time $\Delta > 0$, $\{x_t, t = 0 : \Delta : T\}$ where T is the total time of the discrete observations and $N = T/\Delta + 1$ is the total number of observations. We define the estimators $\hat{\gamma}_\epsilon, \hat{\sigma}_\epsilon^2$ in reduced model (4.4) under Indirect Observability as

$$\hat{\gamma}_\epsilon = \hat{\gamma}_\epsilon(N, \Delta, \epsilon) = g\left(\hat{r}_0^\epsilon, \hat{r}_1^\epsilon\right), \quad \hat{\sigma}_\epsilon^2 = \hat{\sigma}_\epsilon^2(N, \Delta, \epsilon) = h\left(\hat{r}_0^\epsilon, \hat{r}_1^\epsilon\right), \tag{4.10}$$

where the covariance estimators are the standard empirical covariance estimators $\hat{r}_k^\epsilon$ given by

$$\hat{r}_k^\epsilon = \hat{r}_k^\epsilon(N, \Delta, \epsilon) = \frac{1}{N} \sum_{t=0:\Delta:T} x_t x_{t+\Delta}, \quad \text{for } k = 0, 1.$$

Estimators in (4.10) are asymptotically equivalent to the approximate Maximum Likelihood estimators for the OU SDE (4.4).

The main objective is to determine the necessary and sufficient conditions for the consistency of parametric estimation under Indirect Observability, i.e. conditions to ensure that estimators $\left(\hat{\gamma}_\epsilon, \hat{\sigma}_\epsilon^2\right) \to (\gamma, \sigma^2)$ as $\epsilon \to 0$ when the estimators in (4.10) are computed using the observations of x_t in full model (4.2)..

4.2.1 Subsampling strategy

In [10] authors considered the full model as Smoothed Ornstein-Uhlenbeck process having reduced model as OU model and derived the necessary and sufficient conditions for the consistency of the approximate Maximum Likelihood estimators for parameters in OU model. In [9], the similar procedure of finding the necessary and sufficient conditions for the consistency of the approximate Maximum Likelihood estimators for parameters in OU model is considered but for the full model as *Additive Triad model*. In [10, 9], the procedure of deriving necessary and sufficient conditions for consistency of approximate Maximum Likelihood estimators under Indirect Observability is based on the comparison of correlation of slow variables in full and reduced model. In this subsection, we show numerically that the error between correlation of x_t in potential model (4.2) and X_t in reduced model (4.4) is same as the error shown for Triad model in [9]. Therefore, the conditions of consistency of approximate Maximum Likelihood estimators is same as derived for Triad model.

The bias of the estimator $\hat{\gamma}_\epsilon$ in (4.10) compared to true value γ in (4.5) is dependent on the difference between correlation of x_t in full model (4.2) and correlation of X_t in (4.4) as

$\epsilon \to 0$. Fix the constant parameters g, s in full model same as specified in (4.6). Figure 4.1 shows the difference between the correlation function computed for the process x_t in (4.2) and the correlation function for the X_t in Ornstein-Uhlenbeck process with parameters (4.7) as $\epsilon \to 0$ for two different time lags $\Delta = 0.2$ (left part) and $\Delta = 0.5$ (right part). Figure 4.1 indicates that the correlation function for the process x_t in (4.2) converges with the same speed as the correlation function of the slow variable in the triad model in [9]. Therefore, the errors between the correlation functions of x_t in full model and X_t in reduced model follows the ϵ^2 power law decay as $\epsilon \to 0$.

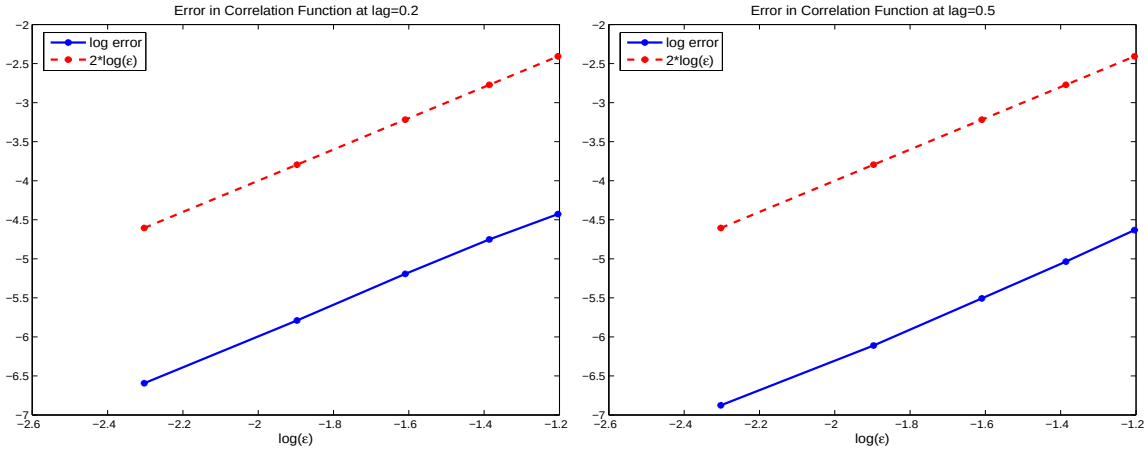

Figure 4.1: Log-log plot for the decay of the error between correlation function $|CF_{x_t}(\Delta) - CF_{X_t}(\Delta)|$ for $\varepsilon = 0.3, 0.25, 0.2, 0.15, 0.1$ computed at a particular lag $\Delta = 0.2$ (left part) and $\Delta = 0.5$ (right part) where x_t is the process (4.2) and X_t is the Ornstein-Uhlenbeck process with parameters (4.7).

Parameter estimate $\hat{\sigma}_\epsilon^2$ also depends on the estimator of second-moment $\hat{r}_0^\epsilon$ which also needs to be analyzed as $\epsilon \to 0$. We show explicitly the convergence of the variance, i.e. $Var\{x_t\} \to Var\{X_t\}$ as $\epsilon \to 0$ in Appendix A. Using the proof sketched in the appendix

A, one can derive that an upper bound for the speed of convergence of $Var\{x_t\}$ as

$$|Var\{x_t\} - Var\{X_t\}| \leq C\epsilon^2.$$

Thus, similar to the Triad model in [9], the convergence of the estimators $\hat{\gamma}_\epsilon$ and $\hat{\sigma}^2_\epsilon$ in (4.10) computed using the data of x_t in (4.2) is determined by the behavior of the correlation function $\hat{r}^\epsilon_1/\hat{r}^\epsilon_0$. Therefore, we conjecture using the results of Triad model in [9] that the necessary conditions for consistency of the subsampling strategy for the potential model (4.2) should be given by

$$\Delta = \epsilon^{2\alpha}, \quad \alpha \in (0,1), \quad N = \epsilon^{-2\beta}, \quad \alpha < \beta. \tag{4.11}$$

and the bias for the parametric estimation under Indirect Observability is proportional to $\gamma\epsilon^2/\Delta$, i.e.

$$\hat{\gamma}_\epsilon - \gamma \sim C\frac{\gamma\epsilon^2}{\Delta} \quad \text{for } \frac{N\epsilon^4}{\Delta} >> 1. \tag{4.12}$$

Thus, the bias $\hat{\gamma}_\epsilon - \gamma$ should be constant with respect to ϵ if $\hat{\gamma}_\epsilon$ is computed by subsampling the observations of x_t in the full model (4.2) with $\Delta = \epsilon^2$.

To support the above conjecture, we compare several subsampling strategies similar to the ones done on Triad model in [9] numerically. In figure 4.2, we consider the numerical simulations of x_t in full model (4.2) subsampled at various time-steps Δ, the total number of sample points N is much larger than Δ/ϵ^4 for all the simulations. For the largest considered $\Delta = 0.3$, the number of sample points is $N = 2 \times 10^6$, thus, $N\varepsilon^4/\Delta \approx 600 \gg 1$ for the smallest value of $\varepsilon = 0.1$ considered in the simulations.

Left part of figure 4.2 displays the behavior of the parameter estimator $\hat{\gamma}_\epsilon$ in (4.10) computed using the discrete observations of x_t in full model (4.2) for four distinct subsampling strategies $\Delta = \epsilon$, $\Delta = \epsilon^2$, $\Delta = 4\epsilon^2$, and $\Delta = \epsilon^3$. Right part of figure 4.2 displays the subsampling strategies with $\Delta = \epsilon^2, \epsilon^{1.75}$, $\epsilon^{1.5}$, $\epsilon^{1.25}, \epsilon$. Both left and right parts of

figure 4.2 show the identical behavior of the estimator $\hat{\gamma}_\epsilon$ computed using the data of the fast-oscillating potential (4.2) as the corresponding behavior of $\hat{\gamma}_\epsilon$ using the data of Triad model in [9]. As conjectured, the error in the estimation $|\hat{\gamma}_\epsilon(\Delta = \epsilon^2) - \gamma|$ remains constant to ϵ as $\epsilon \to 0$ and becomes four times smaller if $\Delta = 4\epsilon^2$. Similar to the triad case, estimation errors decay to zero for subsampling strategies $\Delta = \epsilon^p$ with $p < 2$ and the bias grows unboundedly for subsampling with $\Delta = \epsilon^3$. In particular, the relative errors for the scaling $\Delta = \varepsilon$ follow a linear relationship and the intercept of this line computed by the linear regression is approximately 0.57 which corresponds to the estimated 0.57% relative error as $\epsilon \to 0$.

Therefore, calculations presented in this section confirm that the behavior of the bias for the estimators $\hat{\gamma}_\epsilon$ computed from the data of x_t in the model with the fast-oscillating potential (4.2) is as given by (4.12), similar to the behavior of the bias for the estimators in Triad model in [9].

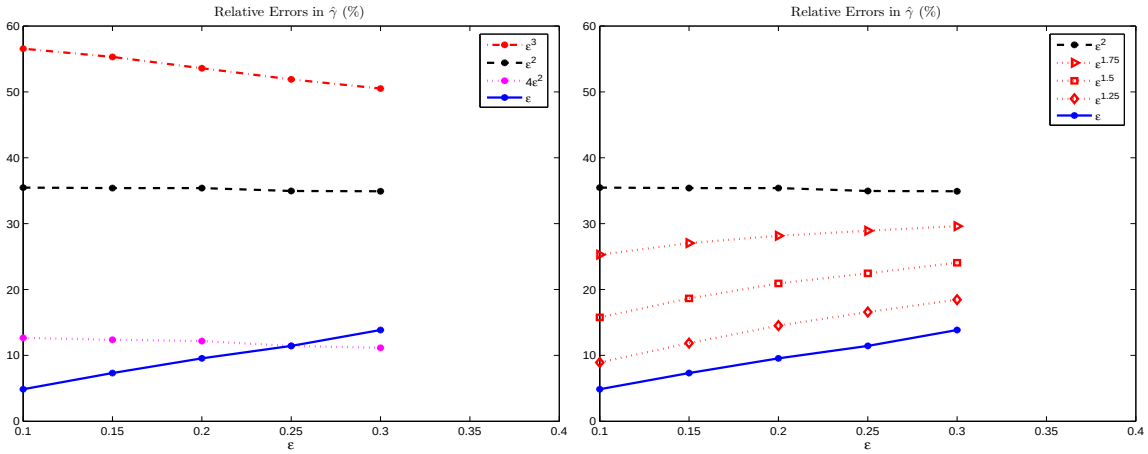

Figure 4.2: Relative errors (percent) in the estimator $\hat{\gamma}$ computed from the data of x_t generated by the model with the fast-oscillating potential (4.2) subsampled with several different strategies. Left part: subsampling with $\Delta = \varepsilon^3$, $\Delta = \varepsilon^2$, $\Delta = 4\varepsilon^2$, $\Delta = \varepsilon$. Right part: subsampling with $\Delta = \varepsilon^2$, $\Delta = \varepsilon^{1.75}$, $\Delta = \varepsilon^{1.5}$, $\Delta = \varepsilon^{1.25}$, $\Delta = \varepsilon$.

4.2.2 Analysis of the data generated by the model with the fast-oscillating potential for fixed but unknown ε

In the previous section, we used datasets of x_t subsampled at various time-steps Δ in the potential model (4.2) with known values of ϵ to validate numerically the subsampling strategy (4.11) which guarantees the convergence of bias to zero as $\epsilon \to 0$. In this subsection, we use numerical data generated by the full model (4.2) with a fixed particular value of ϵ to demonstrate how to model a specific dataset when the scale parameter ϵ is fixed but unknown. The goal is to determine from the data of x_t (with fixed but unknown ϵ) alone the correct subsampling regime corresponding to consistent estimation of the parameters γ and σ. To this end, we analyze the behavior of $\hat{\gamma}_\epsilon(\Delta)$ computed using the data of x_t subsampled at several distinct values of Δ in the full model with fixed ϵ.

In [9], authors develop an approach for constructing the bias-corrected estimators when the data are generated from a trajectory of slow variables in full model with a fixed, but unknown value of the multi-scale parameter ϵ. Authors show that the curve $\hat{\gamma}_\epsilon(\Delta)$ vs Δ clearly identifies the correct subsampling regime. Moreover, the bias-corrected estimators can then be easily computed by linear regression of $\hat{\gamma}_\epsilon(\Delta)$ Δ versus Δ. We apply the similar approach shown in [9] to our full model (4.2) and analyze the curve $\hat{\gamma}_\epsilon(\Delta)$ vs Δ for fixed value of ϵ in potential model.

We use the discrete observations generated by the x_t in the fast-oscillating potential model (4.2) with $\varepsilon = 0.15$ to investigate the behavior of the estimator $\hat{\gamma}_\epsilon(\Delta)$ as a function of Δ. We assume that the value of ε is fixed but unknown and apply regression technique specified in [9] to recover the correct value of the parameters. The left part of figure 4.3 illustrates the behavior of the estimator $\hat{\gamma}_\epsilon(\Delta)$ as a function of Δ. The behavior of $\hat{\gamma}_\epsilon(\Delta)$ vs Δ exhibits a hyperbolic profile consistent with the estimate for the bias in (4.12). The "true" homogenized value of the damping parameter is $\gamma = 0.3119$ and values of the estimator

$\hat{\gamma}_\epsilon(\Delta)$ for $\Delta \in [0.002, \ldots, 0.22]$ are in the range $[0.49, \ldots, 0.326]$. This is the consequence of the fact that the correlation function for x_t in the potential model (4.2) decays faster than the homogenized correlation function $e^{-\gamma\tau}$ for small lags. Note that the critical scaling threshold for Δ is in this case $\epsilon^2 = 0.0225$, but we point out that the estimator $\hat{\gamma}_\epsilon$ is biased for any value of Δ, including $\Delta > 0.0225$ when computed using observations of x_t in full model for $\epsilon = 0.15$. The comparison of $\hat{\gamma}_\epsilon(\Delta)$ versus Δ for a given dataset of x_t in full model (4.2) for a fixed but unknown value of ϵ does not provide direct quantitative information about the unknown value of ϵ.

Similar to Triad model in [9], we consider the graph of the function $\hat{\gamma}_\epsilon(\Delta)\Delta$ versus Δ to deduce the value of the scale parameter ϵ. Right part of figure 4.3 shows the behavior of $\hat{\gamma}_\epsilon(\Delta)\Delta$ vs Δ using the data set of x_t in full model (4.2) for fixed $\epsilon = 0.15$. The graph $\hat{\gamma}_\epsilon(\Delta)\Delta$ vs Δ becomes approximately a straight line between $\Delta = 0.02$ and $\Delta = 0.04$ providing an approximate indication for the range of the multi-scale effects in the data ($\varepsilon^2 = 0.0225$ in this case). Therefore, we can also use the regression estimator to estimate the effective damping parameter from the slope of the line $\hat{\gamma}_\epsilon(\Delta)\Delta$ vs Δ where $\hat{\gamma}_\epsilon(\Delta)$ is computed from the dataset generated by the potential model (4.2) with $\varepsilon = 0.15$. It is clearly visible from the graph of $\hat{\gamma}_\epsilon(\Delta)\Delta$ vs Δ that the first two points for $\Delta = 0.002,\ 0.01$ do not follow the linear relationship and should be neglected in the regression estimator. This also can be quantified numerically by computing the goodness of fit for the linear regression of the graph $\hat{\gamma}_\epsilon(\Delta)\Delta$ vs Δ on different intervals for the subsampling parameter Δ. Thus, the regression estimator computed using the data $\hat{\gamma}_\epsilon(\Delta)$ with $0.02 < \Delta \leq 0.22$ becomes

$$\hat{\gamma}_{regression}(\varepsilon = 0.15) \approx 0.3139$$

which is in a very good agreement (only 0.6% relative error) with the analytical prediction

for the homogenized coefficient $\gamma = 0.3119$. We would like to point out that if a straightforward estimation is used, the estimator $\hat{\gamma}_\epsilon(\Delta = 0.22) = 0.326$ which amounts to the 4.5% relative error. Therefore, the regression estimator significantly outperforms the straightforward estimation in the Indirect Observability context. Moreover, the regression estimator computed on the interval $0.02 < \Delta \leq 0.22$ is much more accurate than the standard estimator $\hat{\gamma}_\epsilon(\Delta = 0.22)$, but the regression estimator is computed using the same number of observational point, N, as the standard estimator $\hat{\gamma}_\epsilon(\Delta = 0.22)$.

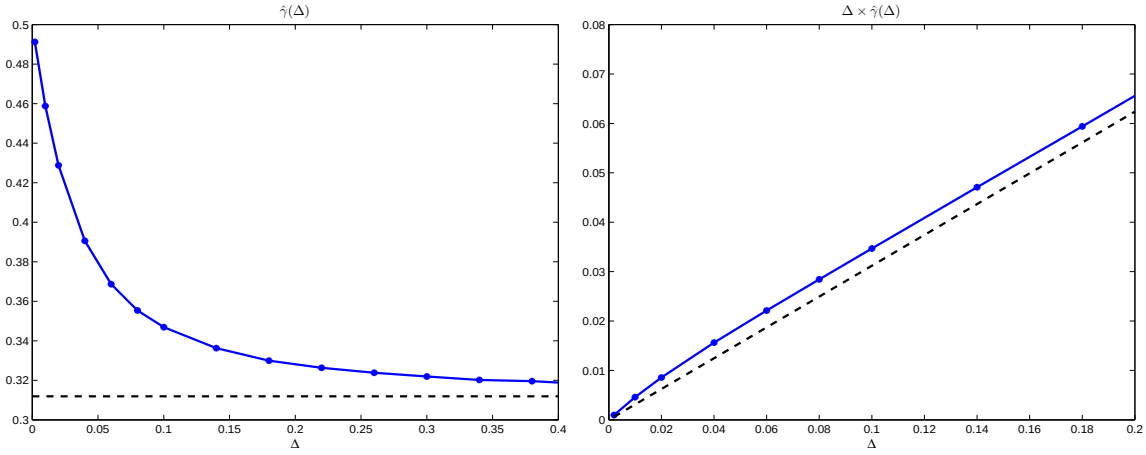

Figure 4.3: Left part: estimator $\hat{\gamma}_\epsilon(\Delta)$ for different values of Δ computed from the data generated by the SDE with the fast-oscillating potential (4.2) with $\varepsilon = 0.15$. Solid line: $\hat{\gamma}_\epsilon(\Delta)$, dashed line: analytical asymptotic value in (4.5) computed from (4.7). Right part: behavior of $\hat{\gamma}_\epsilon(\Delta)\Delta$ with $\hat{\gamma}_\epsilon$ computed for different values of Δ from the data generated by the SDE with the fast-oscillating potential (4.2) with $\varepsilon = 0.15$. Solid line: $\hat{\gamma}_\epsilon(\Delta)\Delta$, Dashed line - straight line with the slope $\gamma = 0.3119$ given by the analytical formula in (4.5).

4.3 Conclusion

In this chapter we studied the parametric estimation of an effective SDE for X_t using the subsampled observations of x_t in fast-oscillating potential model (4.2). In section 4.1.1,

we refer to [68] and show that the x_t process in full model (4.2) converges weakly to the Ornstein-Uhlenbeck process (4.4) as scale separation parameter $\epsilon \to 0$. The homogenization procedure also provides the explicit values of parameters γ and σ as given in (4.5). The main goal is to test the estimation of parameters γ and σ under Indirect Observability, i.e. estimators computed using the observations of x_t in the full model (4.2). Therefore, the homogenized coefficients γ and σ in (4.5) are used to test the estimators under Indirect Observability.

We consider the estimators for the OU reduced model (4.4) as $\hat{\gamma}_\epsilon$ and $\hat{\sigma}_\epsilon$ in (4.10) under Indirect Observability which are asymptotically equivalent to the approximate Maximum Likelihood estimators for the OU SDE. The main objective is to determine the necessary and sufficient conditions for the consistency of parametric estimation of γ and σ under Indirect Observability, i.e. conditions to ensure that the estimators $(\hat{\gamma}_\epsilon, \hat{\sigma}^2_\epsilon) \to (\gamma, \sigma^2)$ as $\epsilon \to 0$ when the estimators in (4.10) are computed using the observations of x_t in full model (4.2).

In [9], authors derived the necessary and sufficient conditions for the parametric estimation of OU SDE under Indirect Observability using the approximate data of Triad model. We extend the results in [9] to the fast-oscillating potential model (4.2) and analyze them numerically. In section 4.2.1 we show that similar to the Triad model in [9], the convergence of the estimators $\hat{\gamma}_\epsilon$ and $\hat{\sigma}^2_\epsilon$ in (4.10) computed using the data of x_t in (4.2) is determined by the behavior of the correlation function $\hat{r}^\epsilon_1/\hat{r}^\epsilon_0$. Therefore, we conjecture using the results of Triad model in [9] that the necessary conditions for consistency of the subsampling strategy for the potential model (4.2) should be

$$\Delta = \epsilon^{2\alpha}, \quad \alpha \in (0,1), \quad N = \epsilon^{-2\beta}, \quad \alpha < \beta. \tag{4.13}$$

and the bias for the parametric estimation under Indirect Observability is proportional to

$\gamma\epsilon^2/\Delta$, i.e.

$$\hat{\gamma}_\epsilon - \gamma \sim C\frac{\gamma\epsilon^2}{\Delta} \quad \text{for } \frac{N\epsilon^4}{\Delta} >> 1. \tag{4.14}$$

We verify the above conjecture numerically in section 4.2.1. Thus, the bias $\hat{\gamma}_\epsilon - \gamma$ is constant with respect to ϵ if $\hat{\gamma}_\epsilon$ is computed by subsampling the observations of x_t in the full model (4.2) with $\Delta = \epsilon^2$.

The next important practical question discussed in this chapter is to construct the bias-corrected estimators computed using the observations of x_t in full model (4.2) for a fixed but unknown value of the multi-scale parameter ϵ. The goal is to determine from the data of x_t (with a fixed but unknown ϵ) alone the correct subsampling regime Δ which can lead to consistency of estimators $\hat{\gamma}_\epsilon$ and $\hat{\sigma}_\epsilon$ given by (4.10). To this end, we analyze the behavior of $\hat{\gamma}_\epsilon(\Delta)$ computed using the data of x_t subsampled at several distinct values of Δ in the full model with fixed ϵ. In section 4.2.2, we extend the results of Triad model in [9] and show numerically that the curve $\hat{\gamma}_\epsilon(\Delta)$ vs Δ identifies the correct subsampling regime. Moreover, the bias-corrected estimators can then be easily computed by linear regression of $\hat{\gamma}_\epsilon(\Delta)$ Δ versus Δ.

APPENDIX A

Computing Variance in the Model with the Fast-oscillating Potential

Consider function $f(x) > 0$ which is integrable on R and having the property $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$. We also assume that $f'(x)$ and $f''(x)$ exist, are continuous and integrable on R. Also, consider a bounded continuous function $g(v) > 0$ on R.

The goal is to compute the behavior of the fast-oscillating integral

$$J(\varepsilon) = \int_{-\infty}^{\infty} f(x) g\left(\cos\left(\frac{x}{\varepsilon}\right)\right)\, dx,$$

where the small parameter $\varepsilon > 0$ tends to 0.

Let us consider a partition $I_k = [\, 2\pi k\varepsilon,\, 2\pi(k+1)\varepsilon\,]$, $k = -\infty, \ldots, \infty$. Then the integral $J(\varepsilon)$ is equal to the infinite sum of integrals over elementary intervals I_k, i.e. $J(\varepsilon) = \sum_k J_k(\varepsilon)$ where

$$J_k(\varepsilon) = \int_{I_k} f(x) g\left(\cos\left(\frac{x}{\varepsilon}\right)\right)\, dx.$$

Since the function $f(x)$ is slowly varying, we use the Taylor expansion to obtain a quadratic

approximation for $f(x)$ on each elementary interval I_k. Then an approximation for $J_k(\varepsilon)$ can be computed as

$$J_k(\varepsilon) = \int_{I_k} [\, f(2\pi k\varepsilon) + (x - 2\pi k\varepsilon) f'(2\pi k\varepsilon) + \frac{1}{2}(x - 2\pi k\varepsilon)^2 f''(z_k)\,] g\left(\cos\left(\frac{x}{\varepsilon}\right)\right)\, dx. \quad \text{(A.1)}$$

with $z_k \in I_k$.

Next, we can integrate explicitly the fast-oscillating function over each elementary interval I_k in (A.1). In particular, if we define the following constants

$$\int_{I_k} g\left(\cos\left(\frac{x}{\varepsilon}\right)\right)\, dx = \varepsilon \int_0^{2\pi} g\left(\cos(y)\right)\, dy = \varepsilon Z_0,$$
$$\int_{I_k} (x - 2\pi k\varepsilon) g\left(\cos\left(\frac{x}{\varepsilon}\right)\right)\, dx = \varepsilon^2 \int_0^{2\pi} y g\left(\cos(y)\right)\, dy = \varepsilon^2 Z_1,$$
$$\int_{I_k} (x - 2\pi k\varepsilon)^2 g\left(\cos\left(\frac{x}{\varepsilon}\right)\right)\, dx = \varepsilon^3 \int_0^{2\pi} y^2 g\left(\cos(y)\right)\, dy = \varepsilon^3 Z_2,$$

then the expression (A.1) for $J_k(\varepsilon)$ becomes

$$J_k(\varepsilon) = Z_0 \varepsilon f(2\pi k\varepsilon) + Z_1 \varepsilon^2 f'(2\pi k\varepsilon) + \frac{1}{2} Z_2 \varepsilon^3 f''(z_k). \quad \text{(A.2)}$$

Substituting (A.2) into the summation for $J(\varepsilon)$ we obtain an approximate expression for $J(\varepsilon)$

$$2\pi J(\varepsilon) = Z_0 \sum_k f(2\pi k\varepsilon) 2\pi\varepsilon + \varepsilon Z_1 \sum_k f'(2\pi k\varepsilon) 2\pi\varepsilon + \frac{\varepsilon^2}{2} Z_2 \sum_k f''(z_k) 2\pi\varepsilon,$$

where $z_k \in I_k$ and we also multiplied both sides by 2π. The final step is to treat the infinite summations in the above expression as Riemann sums for the corresponding integrals and obtain the approximation

$$2\pi J(\varepsilon) = Z_0 \int_R f(x)\, dx + \varepsilon Z_1 \int_R f'(x)\, dx + O(\varepsilon^2)$$

where $O(\varepsilon^2)$ terms arise due to converting the Riemann sum into the integral and, also, from estimating the remainder term with the second derivative. Since $\int_R f'(x)\, dx = f(+\infty) -$

$f(-\infty) = 0$, we obtain the second-order expansion

$$2\pi J(\varepsilon) = Z_0 \int_R f(x)\, dx + O(\varepsilon^2). \tag{A.3}$$

The second-order expansion (A.3) for the integral $J(\varepsilon)$ can be used to both, prove the convergence of the variance for the process x_t in the model (4.2) to the variance of the Ornstein-Uhlenbeck process X_t with parameters in (4.5) and, also, to estimate the speed of convergence. Clearly, since the linear term in ε is not present in (A.3), the speed of convergence is at least as ε^2.

Bibliography

[1] R. V. Abramov. Approximate linear response for slow variables of dynamics with explicit time-scale seperation. *Journal of Computational Physics*, 229:7739–7746, 2010.

[2] R. V. Abramov. A simple closure approximation for slow dynamics of a multi-scale system: non-linear and multiplicative coupling. *Multi-scale Modelling and Simulation (To appear)*, 2012.

[3] R. V. Abramov. A simple linear response closure approximation for slow dynamics of a multi-scale system with linear coupling. *Multi-scale Model Simulation*, 10:28–47, 2012.

[4] R. V. Abramov and A. J. Majda. Blended response algorithms for linear fluctuation-dissipation for complex non-linear dynamical systems. *Nonlinearity*, 20:2793–2821, 2007.

[5] R. V. Abramov and A. J. Majda. New approximations and tests of linear fluctuation-response for chaotic non-linear forced-dissipative dynamical systems. *Journal of Nonlinear Science*, 18:303–341, 2008.

[6] U. Achatz and G. Branstator. A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability. *Journal of Atmospheric Sciences.*, 56:3140–3160, 1999.

[7] T. Alperovich and A. Sopasakis. Stochastic description of traffic flow. *Journal of Statistical Physics*, 133(6):1083–1105, 2008.

[8] L. Arnold, P. Imkeller, and Y. Wu. Reduction of deterministic coupled atmosphere-ocean models to stochastic ocean models: a numerical case study of the Lorenz-Maas system. *Dynamical Systems*, 18(4):295–350, 2003.

[9] R. Azencott, A. Beri, A. Jain, and I. Timofeyev. Subsampling and parametric estimation for multi-scale dynamics. *Communications in Mathematical Sciences (To appear)*, 2012.

[10] R. Azencott, A. Beri, and I. Timofeyev. Adaptive subsampling for parametric estimation of Gaussian diffusions. *Journal of Statistical Physics*, 139(6):1066–1089, 2010.

[11] R. Azencott, A. Beri, and I. Timofeyev. Parametric estimation of stationary stochastic processes under indirect observability. *Journal of Statistical Physics*, 144(1):150–170, 2011.

[12] J. Berner. Linking non-linearity and non-Gaussianity of planetary wave behavior by the Fokker-Planck equation. *Journal of the Atmospheric Sciences*, 62:2098–2117, 2005.

[13] G. Branstator and S. E. Haupt. An empirical model of barotropic atmospheric dynamics and its response to tropical forcing. *Journal of Climate*, 11:2645–2667, 1995.

[14] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics, New York, second edition, 1991.

[15] R. Buizza, M. Miller, and T. Palmer. Stochastic representation of model uncertainty in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Metereological Society*, 125:2887–2908, 1999.

[16] A. Chertock, A. Kurganov, and A. Polizzi. Multi-class traffic flow model with lookahead dynamics (in preparation).

[17] A. Chertock, A. Kurganov, A. Polizzi, and I. Timofeyev. Pedestrian flow models with slowdown interactions. *Mathematical Models and Methods in Applied Sciences (To appear)*.

[18] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences*, 97:2968–2973, 2000.

[19] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction with memory. *Journal of Physica D*, 166:239–257, 2002.

[20] A. J. Chorin, A. P. Kast, and R. Kupferman. Optimal prediction of under-resolved dynamics. *Proceedings of the National Academy of Sciences.*, 95:4094–4098, 1998.

[21] A. J. Chorin, A. P. Kast, and R. Kupferman. Unresolved computation and optimal prediction. *Communications on Pure and Applied Mathematics*, 52:1231–1254, 1999.

[22] A. J. Chorin, A. P. Kast, and R. Kupferman. Unresolved computation and optimal prediction. *Journal of Physica D*, 24(2):99–112, 1999.

[23] I. Chueshov. Invariant manifolds and non-linear master-slave synchronization in coupled systems. *Applicable Analysis*, 86:269–286, 2006.

[24] D. Crommelin and E. Vanden-Eijnden. Sub-grid scale parameterization with conditional Markov chains. *Journal of Atmospheric Sciences*, 65:2661–2675, 2008.

[25] J. Culina, S. Kravtsov, and A. H. Monahan. Stochastic parameterization schemes for use in realistic climate models. *Journal of Atmospheric Sciences*, 68:284–299, 2011.

[26] T. Del-Sole. A fundamental limitation of Markov models. *Journal of the Atmospheric Sciences*, 57:2158–2168, 2000.

[27] T. Del-Sole and B. F. Farrel. Quasi-linear equilibration of a thermally maintained, stochastically excited jet in a quasigeostrophic model. *Journal of Atmospheric Sciences*, 53:1781–1797, 1996.

[28] P. Deuflhard and C. Schütte. Molecular conformation dynamics and computational drug desgin. *In Applied mathematics entering the 21st century: invited talks from the ICIAM 2003 Congress, page 91. Society for Industrial Mathematics*, 2004.

[29] S. I. Dolaptchiev, U. Achatz, and I. Timofeyev. Stochastic closure for local averages in the finite-difference discretization of the forced Burgers equation. *Theoretical and Computational Fluid Dynamics (To appear)*, 2012.

[30] S. I. Dolaptchiev, U. Achatz, and I. Timofeyev. Subgrid-scale closure for the inviscid Burgers-Hopf equation. *Communications in Mathematical Sciences (To appear)*, 2012.

[31] J. Dorrestijn, D. T. Crommelin, A.P. Siebesma, and H.J.J. Jonker. Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theoretical and Computational Fluid Dynamics (To appear)*, 2012.

[32] N. Dundon and A. Sopasakis. *Stochastic modeling and simulation of multi-lane traffic*, pages 661–689. Transportation and Traffic Theory (Elsevier), New York, 2007.

[33] J. Egger. Master equations for climate parameter sets. *Climate Dynamics*, 18:169–177, 2001.

[34] J. Egger and K.-P. Hoinka. Covariance analysis of the global atmospheric axial angular momentum budget. *Monthly Weather Review, AMS Journal*, 130:1063–1070, 2002.

[35] R.S. Ellis and M.A. Pinsky. The first and second fluid approximation to the linearized Boltzmann equation. *Journal of Pure and Applied Mathematics*, 54(9):125–156, 1975.

[36] I. Fatkullin and E. Vanden-Eijnden. A computational strategy for multi-scale systems with applications to Lorenz-96 model. *Journal of Computational Physics*, 200:605–638, 2004.

[37] C. Franzke, A. Majda, and E. Vanden-Eijnden. Low-order stochastic model reduction for a realistic barotropic model climate. *Journal of Atmospheric Sciences*, 62:1722–1745, 2005.

[38] C. Franzke and A. J. Majda. Low-order stochastic mode reduction for a prototype atmospheric GCM. *Journal of Atmospheric Sciences*, 63:457–479, 2006.

[39] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17:R55–R127, 2004.

[40] K. Hasselmann. Stochastic climate models, Part I: Theory. *Tellus*, 28:473–485, 1976.

[41] C. Hauck, Y. Sun, and I. Timofeyev. On cellular automata models of traffic flow with look-ahead potential. *Stochastics and Dynamics (submitted)*.

[42] I. Horenko, E. Dittmer, A. Fischer, and C. Schutte. Automated model reduction for complex systems exhibiting metastability. *SIAM Multi-scale Modeling and Simulation*, 5(3):802–827, 2007.

[43] I. Horenko, C. Hartmann, C. Schutte, and F. Noe. Data-based parameter estimation of generalized multidimensional Langevin processes. *Physical Review E*, 76(1):016706, 2007.

[44] I. Horenko and C. Schutte. Likelihood-based estimation of multidimensional Langevin models and its application to biomolecular dynamics. *Multi-scale Modeling Simulation*, 7(2):731–773, 2008.

[45] G. Hummer. Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics*, 7(1):34, 2005.

[46] E. L. Kang and J. Harlim. Filtering non-linear spatio-temporal chaos with autoregressive linear stochastic models. *Journal of Physica D*, 241(12):1099–1113, 2012.

[47] M. Katsoulakis, A. Majda, and A. Sopasakis. Multi-scale couplings in prototype hybrid deterministic/stochastic systems: Part 1, deterministic closures. *Communications in Mathematical Sciences*, 2:255–294, 2004.

[48] M. Katsoulakis, A. Majda, and A. Sopasakis. Multi-scale couplings in prototype hybrid deterministic/stochastic systems: Part 2, stochastic closures. *Communications of Mathematical Sciences*, 3:453–478, 2005.

[49] M. Katsoulakis, A. Majda, and A. Sopasakis. Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems. *Nonlinearity*, 19(5):1021–1047, 2006.

[50] R. Z. Khasminsky. A limit theorem for the solutions of differential equations with random right-hand sides. *Theory Probability and its Applications*, 11:390–406, 1966.

[51] R. Z. Khasminsky. On stochastic processes defined differential equations with a small parameter. *Theory of Probability and its Applications*, 11:211–228, 1966.

[52] T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *Journal of Functional Analysis*, 12:55–67, 1973.

[53] T. G. Kurtz. Semigroups of conditioned shifts and approximations of Markov processes. *Annals of Probability*, 3:618–642, 1975.

[54] G. Lebeau and E. Zuazua. Null controllability of a system of linear thermo-elasticity. *Archive for Rational Mechanics and Analysis*, 141:297–329, 1998.

[55] E. N. Lorenz. *Predictability: A problem partly solved*, pages 40–58. Cambridge University Press, New York, 2006.

[56] E. N. Lorenz and K. Emanuel. Optimal sites for supplementary weather observations. *Journal of Atmospheric Science*, 55:399–414, 1998.

[57] A. Majda, R. Abramov, and M. Grote. *Information Theory and Stochastic for Multiscale Nonlinear Systems (CRM Monograph Series)*, volume 25. American Mathematical Society, New York, 2005.

[58] A. Majda, I. Timofeyev, and E. Vanden-Eijnden. Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences.*, 96:14687–14691, 1999.

[59] A. Majda, I. Timofeyev, and E. Vanden-Eijnden. A mathematical framework for stochastic climate models. *Communications in Pure and Applied Mathematics*, 54:891–974, 2001.

[60] A. Majda, I. Timofeyev, and E. Vanden-Eijnden. A priori tests of a stochastic mode reduction strategy. *Journal of Physica D*, 170:206–252, 2002.

[61] A. Majda, I. Timofeyev, and E. Vanden-Eijnden. Systematic strategies for stochastic mode reduction in climate. *Journal of Atmospheric Sciences*, 60:1705–1722, 2003.

[62] A. Majda, I. Timofeyev, and E. Vanden-Eijnden. Stochastic models for selected slow variables in large deterministic systems. *Nonlinearity*, 19:769–794, 2006.

[63] K. Nimsaila and I. Timofeyev. Markov chain stochastic parameterizations of essential variables. *SIAM Multi-scale Modeling Simulation*, 8(5):2079–2096, 2010.

[64] T. Palmer. A non-linear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Meteorological Society*, 127:279–304, 2001.

[65] G. Papanicolaou. *Introduction to the asymptotic analysis of stochastic equations (Modern modeling of continuum phenomena)*, volume 16, pages 109–149. The American Mathematical Society, 1975.

[66] G. Papanicolaou. Some probabilistic problems and methods in singular perturbations. *Rocky Mountain Journal of Mathematics*, 6:653–673, 1976.

[67] A. Papavasiliou, G. A. Pavliotis, and A. M. Stuart. Maximum likelihood drift estimation for multi-scale diffusions. *Stochastic Processes and their Applications*, 119:3173–3210, 2009.

[68] G. A. Pavliotis and A. M. Stuart. Parameter estimation for multi-scale diffusions. *Journal of Statistical Physics*, 127(4):741–781, 2007.

[69] G. A. Pavliotis and A. M. Stuart. *Multi-scale methods: Averaging and homogenization.* Springer, New York, 2008.

[70] P. D. Sardeshmukh and J. A. Whitaker. A linear theory of extratropical synoptic eddy statistics. *Journal of Atmospheric Sciences*, 55:237–258, 1998.

[71] C. Schutte, J. Walter, C. Hartmann, and W. Huisinga. An averaging principle for fast degrees of freedom exhibiting long-term correlations. *Multi-scale Modeling and Simulation*, 2(3):501–526, 2004.

[72] C. Schutte, J. Walter, C. Hartmann, and W. Huisinga. An averaging principle for fast degrees of freedom exhibiting long-term correlations. *Multi-scale Modeling and Simulation.*, 2(3):501–526, 2004.

[73] A. Sopasakis and M. A. Katsoulakis. Stochastic modeling and simulation of traffic flow: Asymmetric single exclusion process with Arrhenius look-ahead dynamics. *SIAM Journal of Applied Mathematics*, 66(3):921–944, 2006.

[74] R. Tsay. *Analysis of financial time series.* Wiley Series in Probability and Statistics, USA, 2005.

[75] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Communications in Mathematical Sciences*, 1:385–391, 2003.

[76] V. Volosov. Averaging in systems of ordinary differential equations. *Russian Mathematical Surveys*, 17:1–126, 1962.

[77] D. S. Wilks. Effects of stochastic parameterizations in the Lorenz-96 system. *Quarterly Journal of the Meteorological Society*, 131:389–407, 2005.