PREDICTABILITY AND INFORMATION LOSS

IN COMPLEX 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

Mirza Nofil Barlas

August 2009

PREDICTABILITY AND INFORMATION LOSS

IN COMPLEX SYSTEMS

Mirza Nofil Barlas

APPROVED:

Dr. Ilya Timofeyev
Department of Mathematics
Committee Chairman

Dr. Krešimir Josić
Department of Mathematics

Dr. William Ott
Department of Mathematics

Dr. Gregor Kovacic
Rensselaer Polytechnic Institute

Dr. John L. Bear
Dean, College of Natural Sciences
and Mathematics

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Ilya Timofeyev, for his guidance and constant support through the past several years. He has been a patient teacher, a great mentor, and most importantly a good friend. I thank him for sharing his knowledge and wisdom and exposing me to a variety of topics in mathematics. He taught me how to become a good mathematician and a better person. Lastly, I am particularly grateful for his help in my job search. It was a grueling process but his supervision and advice made me feel at ease. For everything you have done for me Dr. Timofeyev, I thank you. I would also like to thank my dissertation committee members Dr. Krešimir Josić, Dr. William Ott, and Dr. Gregor Kovacic, for taking the time to read my dissertation and giving me helpful suggestions. In particular, I would like to thank Dr. Josić for giving me the opportunity to work with him and helping me with my research when I first started. His commitment to high standards inspired and motivated me. I would also like to thank Dr. Sergei Lapin for emphasizing minor details that mattered most and encouraging me to ask questions that I was hesitant to ask.

I would like to specially thank our department chair, Dr. Jeff Morgan, for his assistance and guidance in getting my graduate career started on the right foot. His supervision during my undergraduate years has made it possible for me to be here. He has helped me at times of great difficulty and has always believed in me in all my endeavors. His friendship combined with his accessibility has been a constant reminder of his support. His jokes were meticulously placed in conversations to lighten up an otherwise bad day.

I would also like to thank my friends, Ankita, Ojas, and Matt, for their much

needed humor and entertainment in what could have been a somewhat stressful graduate life. Our weekly sit downs for "Seinfeld" and "24" were exactly the distractions that I needed. Furthermore, I would like to thank the Department of Mathematics at University of Houston and all the faculty members who were a part of my success. Notably, I would like to thank Dr. Michael Friedberg, Dr. Ronald Hoppe, Dr. Charles Peters, Dr. John Hardy, Dr. Alexandre Caboussat, Dr. Henry Decell, Dr. Matthew O'Malley, Dr. Richard Sanders, and Mrs. Leigh Hollyer, for their kindness and generosity.

Finally, I would like to thank my family, especially my parents, for having faith in me and who so lovingly and unselfishly sacrificed for me. Your continued encouragement was a blessing and without your emotional and financial support all of this would be a mere dream. Ammi and Abbu, I love you and I will always be in debt to you. You are my inspiration and my reason for success.

PREDICTABILITY AND INFORMATION LOSS

IN COMPLEX 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

Mirza Nofil Barlas

August 2009

ABSTRACT

Coupled ocean-atmosphere models are complex due to the large number of modes. To understand the predictability and dynamics of such large systems, lower dimensional projection on a subset of variables can be used to gain insight about the full model. Using certain simplifications, explicit analysis can be performed to explain the behavior of the reduced dynamics, where the neglected degrees of freedom are represented stochastically. Several techniques can be utilized to reduce the complexity of the models; in particular, the stochastic mode-reduction approach has been considered. Furthermore, due to the oscillatory behavior of the systems, mechanisms for the oscillations in predictability has been investigated and the approach of the non-equilibrium behavior to the equilibrium state is studied.

In the first half of this thesis, we examine loss of predictability in two-dimensional stochastic systems that have oscillatory behavior. We show that the information provided by an initial distribution about the state of the system decays to zero non-uniformly in time. In particular, the oscillatory behavior of the systems is responsible for the non-uniformities in predictability. Furthermore, the system as a whole will loose information, but on a subset of variables information can be gained. This return of information will lead to the notion of "return of skill". Marginal distributions will be used to study this increase in information.

In the second half of this thesis, we apply the stochastic mode-reduction strategy to a particular class of prototype coupled ocean-atmosphere models, where self-interactions of the slow variables are given by a rotationally invariant gradient system. The problem addresses the interaction of coherent structures with noise, where the diffusion/drift term in the reduced system contains information about the full dy-

namics of the system. The stochastic mode-reduction strategy is utilized to derive stochastic reduced models, which gives a simple description of the phenomena that occurs from breaking the original rotational symmetry. The direction of the symmetry breaking can be predicted apriori without any information about the statistical behavior of the fast modes. Furthermore, we show a connection of the full and the mode-reduced system using the notion of predictability from the first part of thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Background

1.1 Motivation

Most atmosphere-ocean system are coupled system where changes in the dynamics of one part of the system can lead to large changes in other parts of the system. Furthermore, due to lack of knowledge about the original state of the system and limited spatial and temporal resolution in observational data, there is some uncertainty involved in the current state of the system. Hence, prediction of the future state of the atmosphere can be awfully difficult and mostly impossible for times past two weeks [20].

Significant amount of research has been done in understanding the dynamics of the state of the atmosphere. There are two popular methods used in forecasting the dynamics of the atmosphere: deterministic and probabilistic. Deterministic technique relies on integrating differential equations that yield solutions which are dependent on initial conditions. Assumptions about the initial conditions are then used to predict

the future evolution of the atmosphere. In probabilistic forecasting, observational data about the current and past state of the system is used. Then minimizing the error in prediction, one determines the coefficient of the equation which is modeled by a stochastic process. Using ensemble forecasting, several cluster of initial conditions are generated with a certain distribution. According to how the system evolves for each initial conditions, one can compute the actual conditional probability of a specific event. Hence, with certain confidence, one can predict the future state of the system.

Weather forecasting can also be thought from an information theoretic point of view. By efficiently capturing all the information about the current state of the system, one can make a successful weather prediction. Entropy is a measure of uncertainty in the transfer of this information. Hence, entropy measures the uncertainty over the true state of the system. One can argue that the true state of the system might be unknown. Extensive numerical and experimental data suggest that the true state of the atmosphere is close to the equilibrium distribution (i.e. stationary in time). Hence, the notion of weather predictability is the degree to which a correct prediction has been made. The Fokker-Planck Equation (FPE) explains how distributions evolve in time. Once the system is observed, the future evolution of the weather system can be evaluated using the FPE.

Climate prediction requires coupled ocean-atmosphere models. Coupled ocean-atmosphere models have wide range of timescales; ocean evolves on a slow time scale (i.e. years or decades) while atmosphere evolves on a fast time scale (i.e. few weeks). Fluctuation in the ocean dynamics drive variations in the climate through the interaction coefficients [8]. The difficulty in predicting the state of the ocean using the coupled model requires replacing the atmosphere dynamics by a systematic tech-

nique so that the reduced dynamics are efficiently captured. Hence a rigorous method is needed to mimic the interaction of the fast dynamics whereby still retaining the prominent features of the system [42]. The reduced models are numerically efficiently while still preserving the important components of the climate.

Several techniques can be used to reduce the number of degrees of freedom and still be able to preserve the dynamics of the system. One such technique is called Proper Orthogonal Decomposition (POD) [36]. POD is also sometimes referred to as principal component analysis. The method uses empirical data to find the "optimal" subspace to approximate the given data. Low-dimensional projections onto this subspace yields the reduced model. In particular, the reduced system is highly efficient numerically while still retaining the original properties of the original system. Another efficient and similar method used frequently in weather forecasting is Empirical Orthogonal Function (EOF). In particular, EOF captures the temporal and spatial patterns unlike POD. EOF decomposes the data in terms of orthogonal basis functions where the basis functions are calculated using the eigenvectors of the covariance matrix of the data. Each orthogonal basis function is responsible for the variance in the model and using the leading orthogonal basis function, one can maximize the variability in the model. Hence the essential orthogonal basis functions represent the subsystem.

1.2 Coherent Structures and Stochastic Mode Reduction

Coherent structures have been used to explain many physical phenomena. Coherent structures include stable periodic orbits, circle of equilibria, homoclinic orbits, heteroclinic connections, etc. These structures are evident is various low-dimensional systems, in particular linear oscillator and the Duffing equation which will studied in detail in Chapter 2. Coherent structures are dynamical structures visible in more-or-less the same region of the phase space, are present for long periods of time and maintain the same shape and size. Numerous examples of coherent structures in turbulent flows have been studied. Interaction of coherent structures with noise have also been studied extensively and are important in our applications. In Chapter 3 we study models that have an underlying coherent structure. These coherent structures when coupled with other degrees of freedom have a strong signature present in the full system. Using a particular model reduction technique, we achieve the reduced dynamics where the coherent structure remains and the neglected variables are replaced by noise. The interaction of coherent structures with noise becomes important since this structure in the reduced dynamics is preserved under various conditions.

Many ideas from stochastic modeling have been used successfully to resolve the unresolved degrees of freedom. Ideas from dynamical systems and bifurcation theory have also been considered to truncate the non-essential modes. Identifying coherent structures in the phase space of reduced climate variables becomes essential in understanding the large dynamical behavior. The retained modes can contain multiple equilibria due to the non-linear interactions of the other variables [34]. Other coherent

structures due to non-linear instability include Hopf bifurcations and periodic flows [28]. Furthermore, limit cycles [26] can also be present due to the oscillatory behavior of the system.

The presence of coherent structures in large deterministic system when projected onto the slow dynamics have been discussed before. Due to the coupling effect o the slow and fast modes, certain isolated regions occur that are not present in the original coupled model. In [42], two large isolated peaks in the joint distribution occur that have a weak signature of the original stable periodic orbit. Hence observations in the reduced atmosphere-ocean model can be used to search for the signature of these "ghost" present in the full coupled climate model [43, 44, 27].

We will employ the stochastic mode-reduction strategy to replace the non-essential variables using stochastic terms. Stochastic mode reduction relies on the splitting the degrees of freedom into the essential modes (slow modes), whose dynamics are of interest and the non-essential modes (fast modes), whose interactions will be effectively replaced by the noise term. The stochastic mode reduction technique is rigorous in the limit that the fast modes evolves much faster than the slow modes. The non-linear interactions between the fast modes are then replaced by stochastic terms. This simplification of the original dynamics of the unresolved modes is extremely useful. Various techniques used in model reduction involve replacing the non-essential degrees of freedom by some stochastic terms which may or may not mimic the features of the full system. Furthermore, stochastic mode reduction can also be used to predict apriori the emergence of previously discussed "ghost" behaviors. We will see this in more detail in Chapter 3.

1.3 Measures of Predictability

Work by Lorenz in [35] suggested that weather predictions are sensitive to initial conditions, where small changes in initial conditions can lead to significantly different states of the system. Hence accurate weather prediction are not possible beyond 2 weeks. Due to this uncertainty in the atmosphere, statistical methods that use Monte-Carlo simulations of an ensemble of initial conditions are used often in weather predictions. The spread of the ensemble is used as a measure to quantify the reliability in prediction. Hence, the variability in the system can be used to measure predictability. Several measures of predictability have been studied. In this section, we will introduce several measures of predictability and their pitfalls to suggest the need for the natural notion of distance in information theory, relative entropy as a measure of predictability.

"Potential predictability" (PP) can be quantified as a ratio

$$PP = \frac{\sigma_E^2 - \sigma_{En}^2}{\sigma_E^2}, \tag{1.1}$$

where σ_{En}^2 and σ_E^2 are ensemble and equilibrium variances, respectively. Since PP compares variances of equilibrium and ensemble distributions, it measure the relative spread of the ensemble to the equilibrium spread. Initially, the ensemble of the distribution has small variance compared to the equilibrium distribution, hence PP is close to 1. As the ensemble distribution spreads towards equilibrium distribution, PP decays to 0. Notice that PP apriori only takes into account the dispersion of the equilibrium distribution. To see the pitfalls, assume that the equilibrium distribution has zero mean and non-zero variance. Consider an initial ensemble distribution with

a unit mean and a non-zero equilibrium variance. Under this particular situation, PP will be zero and hence no significant information can be recovered from this measure even though the ensemble distribution has large departures from equilibrium mean. Hence, the ensemble still contains significant information.

Another measure of particular usefulness is the Root Mean Squared Error ($RMSE$) [10]. $RMSE$ can be defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)}, \tag{1.2}$$

where F_i is forecast and O_i is observation for the ith data. Notice that $RMSE$ calculates the average magnitude of the forecast errors. If the prediction forecast is the same as the observational data, then $RMSE$ will be a perfect score of 0. $RMSE$ puts more influence on large errors as compared to small errors and hence can encourage conservative forecasting. Other variants of $RMSE$ include Mean Error (ME), Mean Absolute Error (MAE) and Mean Square Error (MSE).

Another common measure used is forecast prediction is the Anomaly Correlation Coefficient (ACC) [6]. ACC is defined as [51]

$$ACC = \frac{\mathrm{cov}(O - E)(M - E)}{[\mathrm{var}(O - E)\mathrm{var}(M - E)]^{1/2}} \tag{1.3}$$

where var is variance, cov is covariance, O is observed values, M is modeled values, and E is equilibrium values at a discretized point in time. ACC calculates correlation in errors of observed values and modeled values with respect to the equilibrium values. Intuitively, ACC measures deviation of the modeled and observed anomalies from

the equilibrium. The measure in particular only detects patterns rather calculate absolute values.

Notice that in order to use an efficient measure of predictability, we need to consider the *total* distribution which can include mean, variance, and higher moments. Autoregressive approach has been used as a measure of potential predictability of a system [4]. Several measures of predictability have been introduced in dynamical systems theory including Lyapunov exponents [5, 60] and various notions of entropy [47, 25].

1.4 Relative Entropy for SDEs

In this section[1], we give an introduction to a different measure of predictability using information theoretic concepts. This measure is particularly well suited to quantify the predictability of a stochastic dynamical system. The measure is referred to as the *Relative Entropy* (also called *Kullback-Leibler divergence*) and was first introduced by Kleeman [31] in the atmospheric context.

Before a formal definition of relative entropy can be given, we would like to introduce the notion of entropy. Consider an event X, which occurs with a probability $p(x)$. Then the measure of information h of the event X is

$$h(X = x) = \ln \frac{1}{p(x)}. \tag{1.4}$$

Hence if an event occurs with a very high probability, h will be very small. Therefore,

knowledge that the event occurred gives us little to no new information. So the average information content weighted by the probability of occurrence of the event is called entropy and is given in [16] by

$$H(X) = \sum_x p(x) \; h(p(x)) = -\sum_x p(x) \ln p(x), \qquad (1.5)$$

a natural measure of uncertainty.

Relative entropy calculates the difference between two distributions, p and q. In our applications, these distribution are denoted as transient (non-equilibrium) distribution $p(x, t)$ and equilibrium distribution $q(x)$. Furthermore, we assume that $q(x)$ exists, is unique and does not evolve in time. As transient distribution converges asymptotically to equilibrium distribution, and the long term behavior of the system reaches equilibrium state, relative entropy R on a discrete set of states is denoted as

$$R(p(x, t), q(x)) = \sum_x p(x, t) \ln \left(\frac{p(x, t)}{q(x)} \right), \qquad (1.6)$$

where the multi-dimensional continuous analogue of (1.6) can be given by

$$R(p(\vec{x}, t), q(\vec{x})) = R(t) = \int p(\vec{x}, t) \log \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) d\vec{x}. \qquad (1.7)$$

Relative entropy measures "how far" $p(\vec{x}, t)$ is away from $q(\vec{x})$. Hence the relative entropy can be thought of as a measure of "distance" between the distributions $p(\vec{x}, t)$ and $q(\vec{x})$. Notice that R is not exactly a distance since in general $R(p, q)$ is not symmetric, that is $R(p, q) \neq R(q, p)$, nor does (1.7) satisfy the triangle inequality.

Relative entropy is the extra amount of information available from $p(\vec{x}, t)$ over
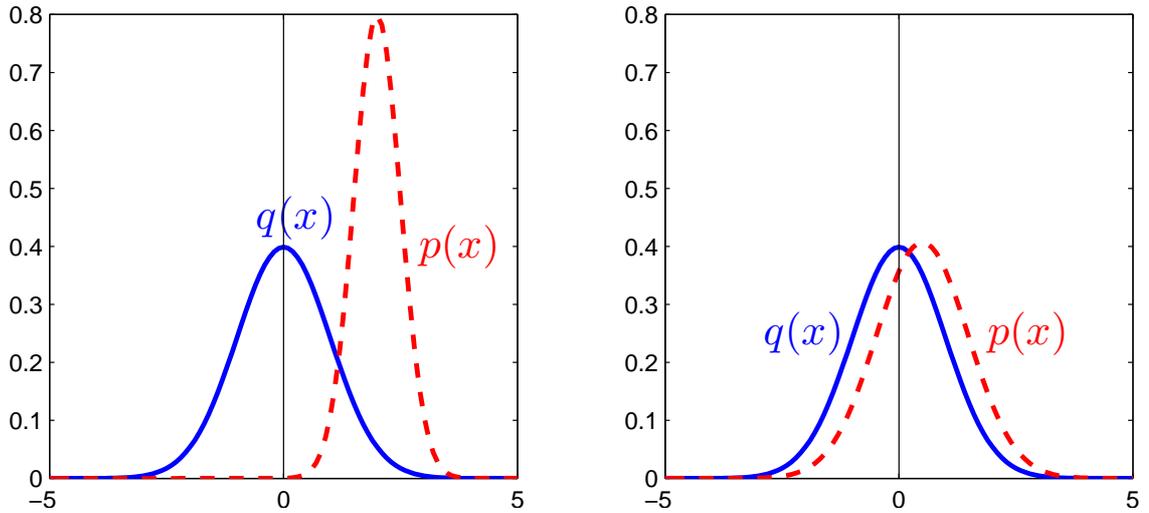
Figure 1.1: Illustration of the relative entropy for two normal Gaussian distributions where $p(x)$ is prediction distribution and $q(x)$ is climatological distribution. Comparison of high relative entropy (Left) versus low relative entropy (Right).

the equilibrium $q(\vec{x})$. More precisely, $R(t)$ corresponds to the amount of information that the distribution $p(\vec{x}, t)$ provides about the state of the system in excess of that given by the equilibrium distribution $q(\vec{x})$ [2]. It is therefore natural to interpret $R(t)$ as a measure of the utility of the prediction provided by an ensemble of particular realizations. An illustration of relative entropy is given in Figure 1.1.

As discussed before, shortfalls in measures of predictability is due to measures not taking advantage of all the statistical behavior of the equilibrium and non-equolibrium behavior. Hence the inclusion of mean, variance and higher moments become significant in capturing the total distribution Unlike some other measures of utility, in particular (1.1), relative entropy reflects differences in all moments, including the mean and variance, of the transient and equilibrium distributions. An increase in the utility of a prediction may be due to the narrow spread of the ensemble (reflected in a

difference between the variances of p and q), or the fact that this ensemble indicates a large departure from normal conditions (reflected in a difference between the means of p and q). In addition, relative entropy satisfies several important mathematical properties which make it a relatively unique measure of predictability.

Typically, the predictability properties of a given system are characterized by the behavior of the relative entropy averaged over the equilibrium distribution of the system obtained by Monte-Carlo simulations with an ensemble of ensembles. Each individual ensemble of ensembles in the simulation describes the decay of predictability for an initial state. The mean of each initial state is chosen at random from the equilibrium distribution, and their variances reflect uncertainties due to imperfect measurements.

Stochastic systems with Markov property, is one in which knowing the present state, the future of the system is independent of the past. Formally, a stochastic process $X(t)$ is called a Markov process if

$$Pr(X(t+h) = y | X(s) = x(s), \forall s \leq t] = Pr[X(t+h) = y | X(t) = x(t)], \quad \forall h > 0.$$

For Markov processes, relative entropy R simplifies significantly. If the dynamical system being modeled is Markov and $q(\vec{x})$ is the stationary distribution, then R always decreases monotonically with time. One of the basic laws of physics, the second law of thermodynamics, states that the entropy of an isolated system is non-decreasing. Hence the second law of thermodynamics is a generalization of the notion of entropy. In our case, due to ergodicity, for sufficiently long lag times the relative entropy always decreases monotonically to zero as the transient distribution approaches the

equilibrium behavior. Hence, relative entropy satisfies three important mathematical properties:

1. It is invariant under well behaved non-linear transformations of state variables [38],

2. It is non–negative [13] and,

3. It declines monotonically in time for Markov processes [13].

Decrease in relative entropy over time can be interpreted as a decline in the utility of a prediction or decrease in the predictive nature of the system. Notice that the word *utility of prediction* will be used interchangeably with *skill*.

In order to calculate the "total" or "overall" predictability of a model, relative entropy is calculated by averaging over all initial states that are generated by ensemble of ensembles.

1.4.1   Relative Entropy for Gaussian Distributions

Suppose that $q(\vec{x})$ and $p(\vec{x}, t)$ are n-dimensional multivariate Gaussian distributions where the mean and variance are known. Continuous form of relative entropy in this case becomes

$$R = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \ldots \int_{-\infty}^{\infty} dx_n \ p(\vec{x}, t) \log\left(\frac{p(\vec{x}, t)}{q(\vec{x})}\right). \tag{1.8}$$

For multidimensional case of relative entropy, μ_q and μ_p are vector-valued means and σ_q, σ_p are variance-covariance matrices. Using the multivariate normal density function of $q(\vec{x})$ and $p(\vec{x}, t)$ in (1.8), we get the simplified analytical expression of

relative entropy [31]

$$R = \frac{1}{2}\left(\log\left(\frac{\det(\sigma_q^2)}{\det(\sigma_p^2)}\right) + \text{Tr}(\sigma_p^2(\sigma_q^2)^{-1}) - n\right) \quad \text{Dispersion}$$

$$+ \quad \frac{1}{2}\left(\mu_p - \mu_q\right)^T (\sigma_q^2)^{-1}\left(\mu_p - \mu_q\right) \quad \text{Signal.}$$

(1.9)

Notice that the two expressions were intentionally separated into the "Dispersion" and the "Signal". The dispersion component reflects the difference in the variances of $q(\vec{x})$ and $p(\vec{x}, t)$ with a normalizing coefficient n. Hence dispersion is only affected by the "spread" of the distributions. The "Signal" term is usually relatively small in applications and in most cases $\mu_q = 0$. Nevertheless, the signal term can be thought of as the sum of the squares of the differences in means normalized by the variance of $q(\vec{x})$.

In (1.9), both the mean and variance of the equilibrium and transient distributions are utilized. Hence, in essence the effects of PP are incorporated in the relative entropy for Gaussian distributions.

In many applications, the overall dynamics of a stochastic differential equation is extremely complex. We know that Gaussian distributions are completely determined by their mean and variance. Hence, (1.9) captures all the predictive power of Gaussian distribution since it takes into account both, the mean and variance.

For both $p(x, t)$ and $q(x)$ Gaussian distribution, one advantage of R is that it can be written in Signal-Dispersion decomposition. In [31], Kleeman showed the practical utility of the signal component in prediction of stochastic and deterministic models for the El-Nino. Also, in [32], Kleeman, Majda and Timofeyev showed the importance

of the signal component in determining predictive utility for the truncated Burgers-Hopf (TBH) models, where the assumption on Gaussianity of initial conditions is needed. Finally, in [38] the Signal-Dispersion decomposition is generalized to suitable non-Gaussian climate distribution.

1.4.2 Relative Entropy as a Lyapunov Functional for the Fokker-Planck Equation

The theory of continuous time Markov processes is developed from the point of view of the corresponding Fokker–Planck equation (FPE), which gives the time evolution of probability density function for the system. Multivariate FPE can be defined as

$$\frac{\partial p(\vec{x}, t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} \left[A_i(\vec{x}) p(\vec{x}, t) \right] + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left[(BB^T)_{ij}(\vec{x}) p(\vec{x}, t) \right], \qquad (1.10)$$

where $A(\vec{x}, t)$ is the drift vector, $B(\vec{x}, t)$ is the diffusion matrix, and the distribution $p(\vec{x}, 0)$ provides the initial data for the Fokker–Planck equation. We assume that the system under consideration has a unique equilibrium solution $q(\vec{x})$. The stochastic differential equation (SDE) described by a conditional probability satisfying the FPE is equivalent to the Itô stochastic differential equation

$$d\vec{x}(t) = A(\vec{x}) \, dt + B(\vec{x}) \, d\vec{W}(t), \qquad (1.11)$$

where $\vec{x}(t)$ is the state and $\vec{W}(t)$ is a standard multi-dimensional Weiner process. All the stochastic differential equations in our case will be in the Itô's sense, but can be readily converted to an equivalent Stratonovich SDE and back again. Hence, all the

ideas are still valid for an equivalent Stratonovich SDE.

If the system is ergodic and as $t \to \infty$, the transient distribution $p(x, t)$ approach the stationary distribution q. Under what conditions does this happen and how is the relative entropy involved? One way to answer this question is to show that relative entropy defined in (1.7) is a Lyapunov functional of the Fokker-Planck equation [24].

The relative entropy $R(p(\vec{x}, t), q(\vec{x})) = R(t)$ is dependent on the initial data $p(\vec{x}, 0)$, corresponding to the distribution of initial conditions of an ensemble in a Monte-Carlo simulation. We will show in Chapter 2 that the decay of relative entropy can vary markedly for different choices of initial ensembles.

We know from Section 1.4 that $R \geq 0$. A proof of non–negativity of R is presented in [24], as well as [13]. A direct calculation can show that the relative entropy decays monotonically in time. To show this, consider the definition of relative entropy given in (1.7). Differentiating (1.7) we obtain

$$\frac{dR}{dt} = \int d\vec{x} \left[\frac{\partial p(\vec{x}, t)}{\partial t} \left(\log \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) + 1 \right) - \frac{\partial q(\vec{x})}{\partial t} \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) \right], \qquad (1.12)$$

which can be further written in an abbreviated form

$$\frac{dR}{dt} = \int d\vec{x} \left[\frac{\partial p(\vec{x}, t)}{\partial t} (\log p(\vec{x}, t) + 1 - \log q(\vec{x})) - \frac{\partial q(\vec{x})}{\partial t} \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) \right]. \qquad (1.13)$$

Let us also assume that $q(\vec{x})$ is zero at infinity, where it and its first derivatives vanish. The contributions to dR/dt from the drift ($\left(\frac{dR}{dt} \right)_{\text{drift}}$) and diffusion ($\left(\frac{dR}{dt} \right)_{\text{diff}}$) terms in the Fokker-Planck equation can also be obtained by the same calculation:

$$\left(\frac{dR}{dt} \right)_{\text{drift}} = \sum_i \int d\vec{x} \frac{\partial}{\partial x_i} \left[-A_i p(\vec{x}, t) \log \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) \right], \qquad (1.14)$$

15

$$\left(\frac{dR}{dt}\right)_{\text{diff}} = -\frac{1}{2}\sum_{i,j}\int d\vec{x}\, p(\vec{x},t) B_{ij}\left[\frac{\partial}{\partial x_i}\log\frac{p(\vec{x},t)}{q(\vec{x})}\right]\left[\frac{\partial}{\partial x_j}\log\frac{p(\vec{x},t)}{q(\vec{x})}\right]. \qquad (1.15)$$

Under the given assumptions on $q(\vec{x})$, it can be shown [24] that

$$\left(\frac{dR}{dt}\right)_{\text{drift}} = 0 \qquad \text{and} \qquad \left(\frac{dR}{dt}\right)_{\text{diff}} \leq 0. \qquad (1.16)$$

Hence, it follows that the decreases in relative entropy are due only to the diffusion term, since the drift term in that case is zero. Even though the results sound surprising, they are not. From an information theoretic point of view, it is the diffusion term that correspond to the stochastic component of the equation that lead to information loss. The diffusion term lead to the spread of the system and hence are responsible for the decay in the relative entropy. The faster the spread of initial conditions, the faster the loss in utility of prediction. The most immediate effect of diffusion term is to increase the spread in the ensemble forecast, but the diffusion terms interact in some nontrivial fashion with the drift term which determines the rate of this increase. A detailed study of this idea is presented in Chapter 2.

1.5 Return of Skill

The term, "return of skill", has no specific definition present in atmospheric science books. The notion is part of many discussion and is well understood in the scientific community. A heuristic definition of "return of skill" warrants the need for a formal definition of "skill". A weather forecast is said to have "skill" (i.e. skillful) if it validates an observation. The comparative observation contains the dynamics of the real atmosphere. Hence, skillful weather forecast captures some information that an

16

observation may contain. If a forecast has no "skill", it is said to unskillful.

From an information theoretic point of view, skillful weather forecast captures the correct information content of an observation. Hence, skill can be used as a criteria of evaluation for a prediction process [54].

The notion of predictability discussed before is intuitively tied to the term forecast skill. The idea that good forecast skill leads to better weather prediction clarifies this connection. Our measure of predictability, the relative entropy, has good prediction power for high values. Hence, for large values of relative entropy the forecast is said to be skillful. Several statistical methods can be used to verify that the skill in a forecast is not due to happenstance [1].

The measure of predictability used in our application decays in time. Once the connection between forecast skill and predictability has been established, our notion of predictability suggest that forecast skill would decrease as time evolves. Hence, "return of skill" is said to be the return of information in a system. Since relative entropy of joint variables decays in time, the return of skill is not possible for full relative entropy. As we will see in Chapter 2, the return of skill will not be present in the full relative entropy, but will be visible for the marginal relative entropies. Collectively, the system will loose information, but on a subset the system may gain information. Hence confidence in our prediction can gain over time on a subset of the system. We will see in the next chapter that the oscillatory behavior of our system leads to the "return of skill". As the main mass of the transient distribution moves away from the equilibrium mass, the return of skill is observed. Transient return to equilibrium behavior leads to oscillations in marginal entropy and hence the return of skill.

Chapter 2

Non-Uniform Decay of Predictability in Stochastic Systems

2.1 Introduction

Stochastic differential equations are frequently used to represent time evolution of the atmosphere. The driving noise term in stochastic differential equation represents the uncertainty present in the atmosphere. Due to this atmospheric uncertainty, weather and climate forecasting requires using Monte-Carlo simulations (MC) for an ensemble of trajectories. Most forecasting models have no explicit solutions, hence numerical schemes like Milsten 2.0 or Euler scheme are used to calculate various averaged statistical quantities.

Consider the system $y_t = f(y)$, where t represents time. In this case, Monte-Carlo

algorithm can be explained as follows:

Step 1: Generate random ensemble of initial conditions $y_0 = X = (x_0, x_1, \ldots, x_m)$ using a distribution.

Step 2: Plug in $x_i(0)$ in the above equation y_t to find $x_i(h)$, where h is the step-size in time.

Step 3: Do step 2 for $i = 0$ to m,

Step 4: Use the results to generate mean, variance, confidence intervals, etc. for $t = h$.

Now repeat steps 2, 3 and 4 for $t = 2h, 3h, \ldots$.

As discussed in Section 1.4 and proved in [31], utility of ensemble predictions for Markov processes declines monotonically to zero in time. In this chapter[1], we show that for oscillatory systems, and in particular Markov Processes, the mechanisms that leads to the loss of predictability have surprising and counterintuitive results. While the utility of prediction decays exponentially [31, 58] for ensembles, it can behave significantly differently for each individual trajectory. In fact, Palmer [48] showed that for a Lorenz system the predictability depends strongly on initial conditions. This implies that the predictability of a model can be considered as a functional dependent on the initial state used in the prediction. Averaging over all such initial conditions may lead to loss of information. Pioneering work by Lorenz [35] suggest that even slight change in initial conditions can lead to large changes in the state of

---

[1]Recently published work: Barlas, N., Josić, K., Lapin, S. and Timofeyev, I. (2007). Non-Uniform Decay of Predictability and Return of Skill in Stochastic Oscillatory Models. *Physica D* **232(2)** 116-127.

the system. Hence weather predictability are sensitive to inaccurate choice of initial conditions. Furthermore, the "memory" of the initial conditions are significantly important in weather predictability [52].

For a certain class of models, averaged predictability might not decrease at an exponential rate. In particular, in this chapter we show that for certain models as well as for certain initial states, there are large intervals of time during which the utility of the prediction remains nearly constant. Therefore, the predictability of a particular ensemble is very different from from the exponentially decaying averaged predictability of the system. The main reason for this behavior is due to the oscillatory transport of the initial ensemble towards and away from the main mass of the equilibrium distribution. During the times of plateaus in predictability, the ensemble mean is in areas where the mass of the equilibrium distribution is small. As the main mass of the non-equilibrium distribution returns to the ensemble mean, predictability is lost significantly. Hence ensemble spread of the transient distribution leads to loss in averaged predictability.

Markov processes depend on the present state of the system rather than the past states. Hence, Markov processes are not functionally dependent on their initial conditions. Using information theoretic lingual, in Markov processes, information that is lost cannot be regained. Therefore, the utility of prediction cannot increase in time for Markov models. This leads to the so called no *return of skill*. As we will see in this chapter, this result might be valid for the predictability of the full phase space of the system being modeled, it certainly does not apply on a subset of the phase space. Indeed, we show that the marginal relative entropies and in some cases conditional relative entropies of *all* variables in the model may increase in time, while the relative

entropy of their joint distribution will always decay. Therefore, while information about the state of the system will be lost over time, information about each individual attributes may be regained. Furthermore, we illustrate that the mechanism leading to the return of skill on a subset of phase space and the plateaus in relative entropy on certain time intervals are closely related.

It may be difficult to make sense of the flow of information between the variables defining a Markov model since marginal entropies can change together or independently. Hence, it is more natural to consider the flow of information between the conditional and marginal distributions. We also show that the phenomena described above can be understood in terms of such a flow of information.

We consider three noisy oscilaltory models to explain the mechanism behind the non-uniform decay of relative entropy and return of skill. These low-dimensional stochastic models exhibit three main types of oscillatory behavior: damped, self-sustained, and heteroclinic. The first model we consider is the stochastically perturbed linear oscillator. The choice of the model is due its simplicity and its analytical solution. We also consider the stochastic perturbation of a nonlinear oscillator (obtained from the normal form of a Hopf bifurcation) as well as the Duffing equation (homoclinic cycle). The three models considered cover a wide range of dynamical behaviors.

2.2   Stochastic Linear Oscillator

Several low-dimensional equations of various complexity have been used to describe El Niño/Southern Oscillation (ENSO). ENSO is a coupled ocean-atmosphere phe-

nomenon that occurs globally. ENSO has far reaching effects with signatures in the Pacific, Atlantic and Indian Oceans. Furthermore, it causes variable weather patterns and is the main source of inter-annual variability in weather and climate around the world.

Numerous models have been used to describe the observed variability in ENSO. One particular model that has been successfully used is the 2-dimensional stochastic linear oscillator [30]. Kestin et al used a stochastic linear oscillator to show that the ENSO is a well-behaved stochastic system. They showed that the ENSO can be approximated well by an AR(3) model (or the ARMA(3,1) model of [59]) which is simply a linear oscillator driven by stochastic forcing. Furthermore, Penland and Sardeshmukh [49] found that the ENSO can be approximated well by a linear system forced by white noise. Hence the above results indicate that ENSO system can be approximated well by a stochastic linear oscillator.

The idea that ENSO variability is due to stochastic forcing has been gaining significant popularity. Models of different level of difficulties, specifically, simple ENSO stochastic models, intermediate, stochastic physical models [46], and a stochastically forced hybrid coupled model [18] have been used. Despite the varying complexity of these models, one common feature among them is the oscillatory behavior of solutions and the use of stochastic forcing. Therefore, as a prototype behavior we consider the simple two-dimensional stochastic linear oscillator similar to the one described in [30]. Even in this simple model, relative entropy decays non-monotonically, and the marginal relative entropies can oscillate. The results of this section were obtained using analytical expressions for the relative entropy. The model is given by the following
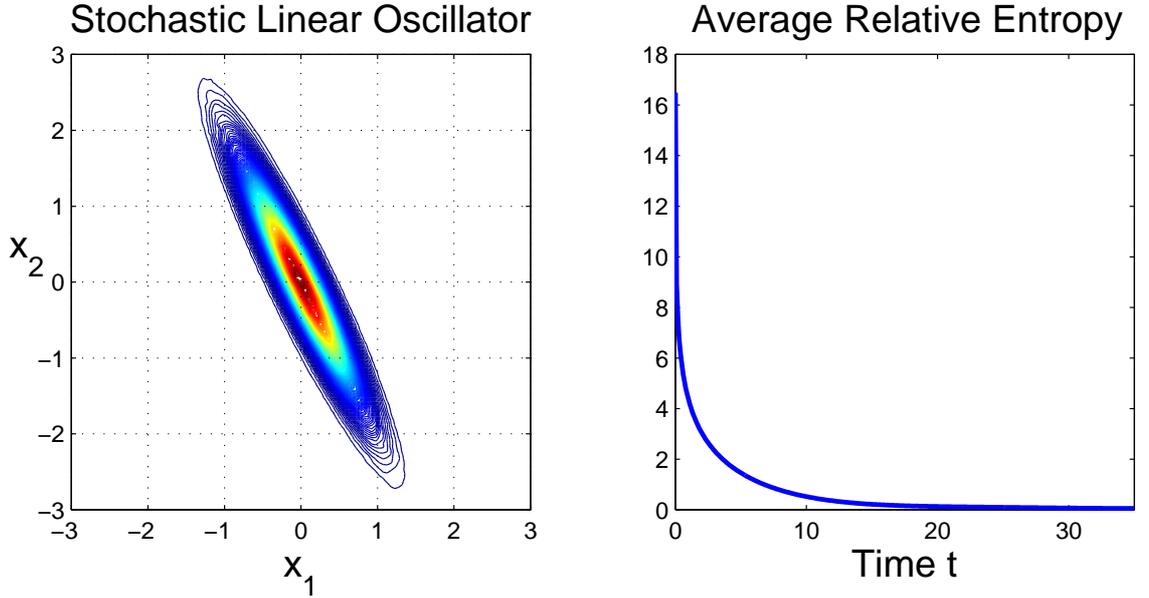
Figure 2.1: Left: The equilibrium distribution $q(\vec{x})$ for the linear oscillator (2.1). Right: Average relative entropy in the case of the stochastic linear oscillator. Parameters: $\alpha = 0.45661, \beta = 0.2439, \gamma = -1.08679, \delta = -0.512161$ and $\varepsilon = 0.1$.

two-dimensional stochastic differential equation:

$$dx_1 \ = \ \alpha x_1 \ dt + \beta x_2 \ dt,$$

$$(2.1)$$

$$dx_2 \ = \ \gamma x_1 \ dt + \delta x_2 \ dt + \varepsilon \ dW_2,$$

where W_1 and W_2 are independent Wiener processes for x_1 and x_2 with noise level 0 and ε respectively, and the remaining parameters are chosen so that with $\varepsilon = 0$ the system exhibits damped oscillations. Figure 2.1 (Left) shows the contour plot of the equilibrium distribution in one specific case. Although we use a specific set

23

of parameters for these simulations, we show below that our results hold under very general conditions.

Notice that equation (2.1) can be written in an abbreviated form

$$d\mathbf{x} = A\mathbf{x}\, dt + B\, d\mathbf{W}, \tag{2.2}$$

where

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \tag{2.3}$$

and

$$B = \begin{pmatrix} 0 & 0 \\ 0 & \epsilon \end{pmatrix}, \tag{2.4}$$

The Fokker-Planck equation describing the evolution of an initial density is given by equation (1.10), where A and B are given in (2.3) and (2.4) respectively. This equation can be solved explicitly assuming deterministic initial condition $p(\vec{x}, 0) = \delta_{\vec{x}^0}(\vec{x})$ where $\vec{x}^0 = (x_1^0, x_2^0)$. The solution of (2.1) at time t [24] is

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = e^{At} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} + \int_0^t \begin{pmatrix} dW_1(t') \\ dW_2(t') \end{pmatrix} e^{A(t-t')} \begin{pmatrix} 0 & 0 \\ 0 & \epsilon \end{pmatrix}. \tag{2.5}$$

where W_1 and W_2 are independent Weiner processes. The solution of (2.5) at time t is Gaussian with mean

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = e^{At} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}, \tag{2.6}$$

24

and covariance matrix

$$\left[\begin{array}{cc} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle \end{array} \right] = \int_0^t dt' e^{A(t-t')} \left(\begin{array}{cc} 0 & 0 \\ 0 & \varepsilon^2 \end{array} \right) e^{A^T(t-t')}. \tag{2.7}$$

Hence we know the solution, mean and variance of the system given in (2.1) explicitly. Since the equilibrium distribution $q(\vec{x})$ can also be computed explicitly [24], (1.9) can be used to obtain the relative entropy analytically in this case. Hence we have a closed form for the relative entropy as well.

2.2.1 The Decay of Relative Entropy

The full relative entropy of a model is determined by the average rate of decay of the relative entropy, where the average is taken over all initial conditions weighted by the stationary distribution. This requires taking Monte-Carlo simulations of ensemble of ensembles [31]. Each ensemble is Gaussian distributed where the mean is randomly sampled the equilibrium distribution. Hence, averaging over all such ensembles provides a measure of the average predictability of a given model. The average relative entropy for the linear oscillator (2.1) averaged over initial ensembles corresponding to densities $\delta_{\vec{x}^0}(\vec{x})$ is shown in Figure 2.1 (Right).

Due to the monotonic decay of the relative entropy in time, information is lost. Stochastic forcing is responsible for this loss of information as each ensemble of initial conditions moves forward in time. Numerical simulations indicate that diffusion is initially responsible for this decay in relative entropy. Around $t = 0$, relative entropy decays like $-\log(t)$, after which there is a long interval during which relative entropy decays exponentially. This exponential decay is shown in Figure 2.1 (Right). Hence
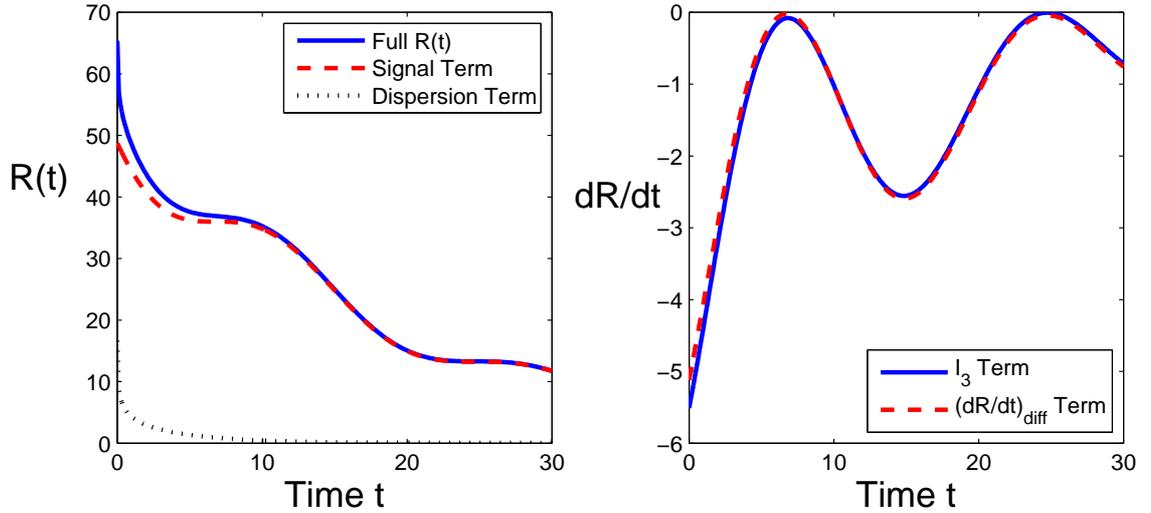
Figure 2.2: Left: Relative entropy for a particular initial condition $(x_1^0, x_2^0) = (1, 1)$ (solid line) and the contribution to the relative entropy due to the signal term (dashed line) and dispersion term (dotted line) in equation (1.9). Right: Behavior of $\left(\dfrac{dR}{dt}\right)_{\text{diff}}$ (solid line), and I_3 term of (2.12) (dashed line) in time for an initial condition $(x_1^0, x_2^0) = (1, 1)$. Parameters as in Figure 2.1.

the system loses information on average and there is no increase in the information content of the system as a whole. Notice that this might not be the case with any particular initial ensemble whose mean is far from the mean of the equilibrium distribution. Instead of decaying exponentially, relative entropy has intervals during which it remains nearly constant. This is illustrated in Figure 2.2 (Left), using the relative entropy for an ensemble of trajectories with the initial condition $(x_1^0, x_2^0) = (1, 1)$ so that $p(\vec{x}, 0) = \delta_{(1,1)}(\vec{x})$. The ensemble is generated using independent realizations of the Wiener process.

In particular, around time intervals $[5, \ldots, 10]$ and $[22, \ldots, 27]$ (see Figure 2.2 (Left)), the decay rate of relative entropy is nearly zero. During these intervals,

26

relative entropy remains nearly constant, and our confidence in our prediction does not decrease. Furthermore, Figure 2.2 (Right) shows that during the time intervals where rate of decay of relative entropy is nearly constant, the derivative of $R(t)$ w.r.t t is almost close to zero. Compared to averaged relative entropy given in Figure 2.1 (Right), relative entropy of individual ensemble can be lower or higher depending on the initial conditions. Due to this dependence on ensemble choice, prediction utility of individual ensemble decays slowly compared to averaged prediction.

Relative entropy decays in a similar manner for any initial distribution whose mean differs sufficiently from the equilibrium distribution. However, the plateaus in relative entropy do not occur at the same time and hence are not due to the differences in mean of the equilibrium and non-equilibrium distribution. We next show that the plateaus in relative entropy depend on the change of the mean of the initial ensemble. Therefore, the average over different initial ensembles provides a somewhat misleading picture: compared to the rate of decay of relative entropy for a *particular ensemble*, the rate corresponding to the average relative entropy is much larger at plateaus and much smaller between them. Furthermore, we show that this effect is due to the fact that after a short change, the value of relative entropy is mainly determined by the location of the mean of an ensemble, and the means for each initial condition can oscillate in- or out of phase.

2.2.2 Analysis of the Rate of Decay of $R(t)$

We can explain the non–uniform decay of the relative entropy by considering (1.9). As mentioned before the dispersion term is responsible for the spread of the initial conditions and hence is the main factor in the change in initial variance of the system.

After a sudden initial spread of the initial conditions, the variance of the system increases significantly. This quick increase in the variance of the transient distribution $p(\vec{x}, t)$ results in a logarithmic singularity of $R(t)$ and $\left(\dfrac{dR}{dt}\right)_{\text{diff}} \sim -1/t$. Further changes in the variance occur on a slow timescale compared to that of the oscillations (see the left column of Figure 2.4 or Figure 2.5). Hence after the initial effect, the variance plays no significant role in the behavior of the system and the dispersion terms is negligible thereafter. The effect of the signal term given in expression (1.9) can be expected to dominate once the dispersion is effectively zero. This is indeed the case as illustrated in Figure 2.2 (Left). The dispersion term after $t = 10$ has no affect on the relative entropy of the system. Furthermore, the behavior of the full relative entropy and the signal term is exactly the same after $t = 10$. Notice that even at times past $t = 10$, where the relative entropy has plateaus, there is no change in the behavior of the dispersion and signal terms.

We illustrate the analysis in a particular case where matrix A has the form

$$A = \begin{pmatrix} -\frac{1}{k} & \frac{1}{b} \\ -b & -\frac{1}{k} \end{pmatrix}, \tag{2.8}$$

and B is the same as in (2.4). Notice that this is a particular case of matrix A but the results are still valid in the general case, but the analysis is more tedious. The mean of the solution of (1.10) with the initial data x_1^0 and x_2^0 using (2.6) is,

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \begin{pmatrix} x_1^0 e^{-t/k} \cos t - \dfrac{x_2^0 e^{-t/k} \sin t}{b} \\ e^{-t/k} x_2^0 \cos t + b x_1^0 e^{-t/k} \sin t \end{pmatrix}. \tag{2.9}$$

For presentation, we use the simplification $x_2^0 = 0$. Hence (2.9) becomes,

$$\bar{x}_1(t) = x_1^0 e^{-t/k} \cos t, \qquad \bar{x}_2(t) = b x_1^0 e^{-t/k} \sin t, \qquad (2.10)$$

so that k is a damping coefficient, and b determines how the solutions are stretched in the y direction. Notice that as $t \to \infty$, for fixed values of x_1^0, k and b, the mean of the solution decays to zero. Furthermore, for fixed values of k and b of the same order, $k \gg 1$, and $b \gg 1$, small noise and $x_1(0)$ sufficiently large, the signal term initially dominates all other terms in (1.9) (for example, a particular case with $k = 10$, $b = 10$, $\varepsilon = 0.1$, and $x_1^0 = 3$ is shown in Figure 2.3). In Figure 2.3 (Bottom Right), the difference in the relative entropy and the signal term is zero after the initial spread of the ensemble. Using (2.10), and evaluating matrix A given in (2.8) and matrix B given in (2.4) into (2.7), signal term in (1.9) has the form

$$R_{\text{signal}}(t) = e^{-\frac{2t}{k}} \frac{2 \, b^2 \, (x_1^0)^2 \, (1 + k^2 + \cos(2\,t) + k \, \sin(2\,t))}{\varepsilon \, k^3}.$$

In the parameter regime of interest, the term proportional to $e^{-\frac{2t}{k}} \sin(2\,t)$ determines the non-uniformities in the decay of relative entropy.

The plateaus in relative entropy therefore occur at the times in which $\sin(2\,t)$ is increasing. Those intervals correspond to the time during which $|x_2(t)|$ increases from 0 to b, and the mean of the non-equilibrium distribution in x_2 moves away from the mean of the stationary distribution (see the Figure 2.3). Similarly, information is lost rapidly during the times at which $|x_2(t)|$ decreases from b to 0.

Intuitively, this is a consequence of the fact that information is gained in x_2 as the means of the transient and equilibrium distribution move apart and away from
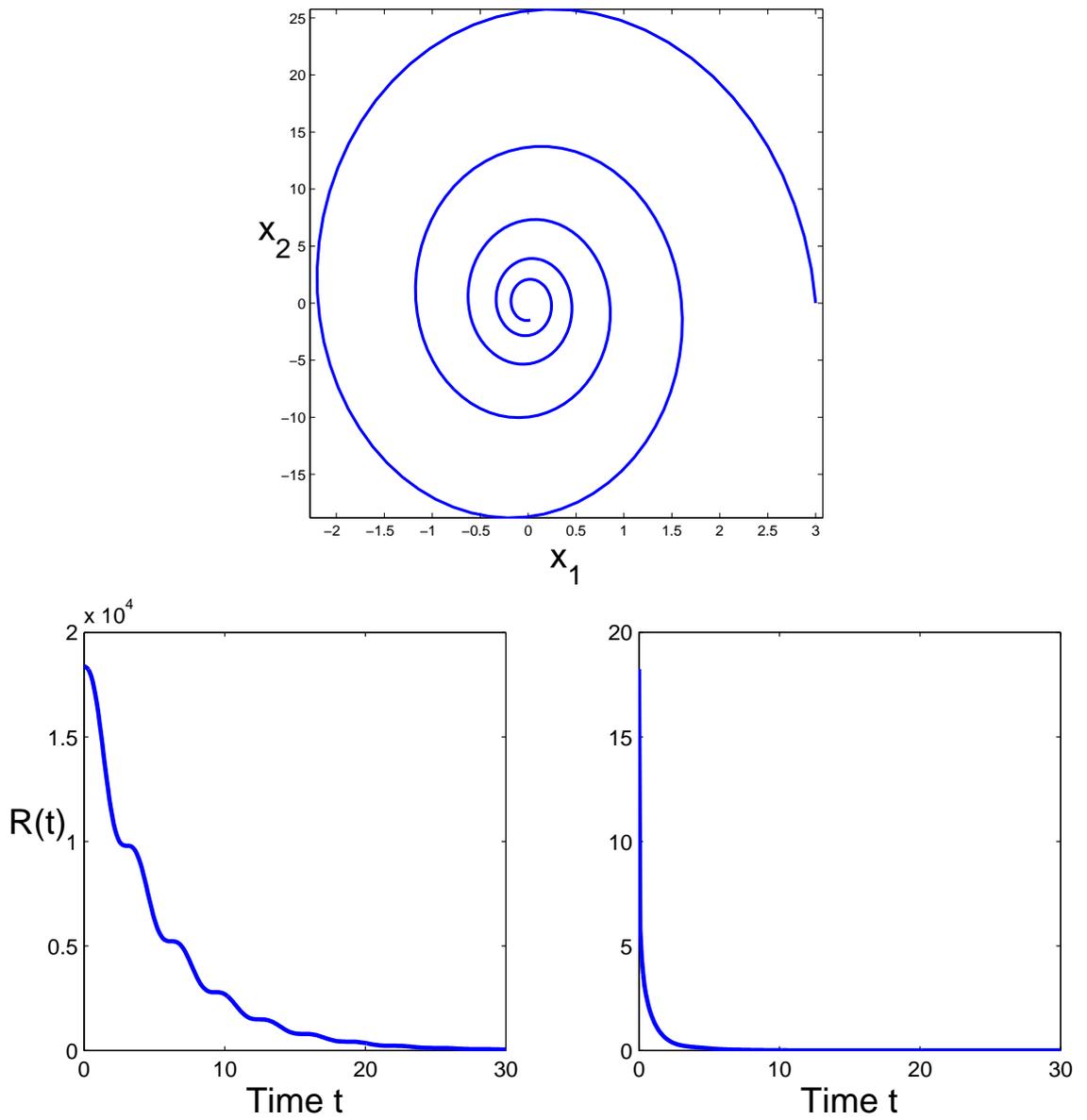
Figure 2.3: Case with matrix A defined by (2.8) for an initial condition $(x_1^0, x_2^0) = (3, 0)$. Top: Trajectory of the mean of the distribution. Bottom Left: Relative entropy behavior. Bottom Right: Difference between the full relative entropy and the signal term.

each other. This gain in mean is balanced by the the loss of information due to the increase in the variance of the transient distribution through the diffusion term. The mean of the two distribution move together during times when $\sin(2\,t)$ is decreasing, hence information is lost due to changes in both mean and variance. These intervals correspond to the sharp drop in relative entropy and hence a rapid loss of information following the plateaus. The interval between the first two panels in the left column of Figure 2.4 (or Figure 2.5) corresponds to a plateau in relative entropy, while the interval between the last two panels corresponds to the sharp drop in relative entropy following a plateau. Therefore, as the transient distribution moves around the loop in the equilibrium distribution, during times of plateaus the mean and variance of the distribution balance the gain and loss of information and hence we see the plateaus. This corresponds to the increase in $\sin(2\,t)$. At times when $\sin(2\,t)$ decays, the mean and variance of the distributions both lead to the loss in information content of the system.

According to the discussion in Section 1.4, the decay of relative entropy is entirely due to diffusion and the spread of the ensemble. Therefore the fact that the decay of relative entropy is dominated by the behavior of the mean which is completely determined by the drift term appears somewhat surprising. Additionally, the mean of the ensemble depends on the initial conditions chosen and therefore the choice of initial conditions also in turn affect the relative entropy of the ensemble. To explain the contradictory behavior of the mean, we need to look at the analysis a little differently.

For the stochastic linear oscillator in (2.1) the diffusion part of the corresponding

Fokker-Planck equation (1.15) reduces to

$$\left(\frac{dR}{dt}\right)_{\text{diff}} = -\frac{\varepsilon}{2}\int d\vec{x}\, p(\vec{x},t)\left[\frac{\partial}{\partial x_2}(\log(p(\vec{x},t)/q(\vec{x})))\right]^2. \qquad (2.11)$$

This can be evaluated using a straightforward, but lengthy calculation. The expression (2.11) can be rewritten as

$$\left(\frac{dR}{dt}\right)_{\text{diff}} = -\frac{\varepsilon}{2}\left[\underbrace{\int d\vec{x}\, p(\vec{x},t)\left(\frac{\partial \log p(\vec{x},t)}{\partial x_2}\right)^2}_{I_1} - \underbrace{2\int d\vec{x}\, p(\vec{x},t)\frac{\partial \log p(\vec{x},t)}{\partial x_2}\frac{\partial \log q(\vec{x})}{\partial x_2}}_{I_2}\right.$$

$$\qquad (2.12)$$

$$\left. + \underbrace{\int d\vec{x}\, p(\vec{x},t)\left(\frac{\partial \log q(\vec{x})}{\partial x_2}\right)^2}_{I_3}\right] \equiv I_1 + I_2 + I_3.$$

A direct computation shows that only the third integral (I_3) depends on the mean of the transient distribution μ_p, while the first two integrals I_1 and I_2 depend only on the variances σ_p^2 and σ_q^2. Moreover, this integral is exactly the time-derivative of the signal part of the relative entropy, *i.e.* $(I_3) = \frac{d}{dt}[(\mu_p)^T(\sigma_q^2)^{-1}(\mu_p)]$. Thus, since the relative entropy is dominated by the signal term, the behavior of $\left(\frac{dR}{dt}\right)_{\text{diff}}$ almost equal to I_3, as depicted in Figure 2.2 (Right). Therefore, although the decay of the relative entropy is entirely due to diffusive terms in the equation, the magnitude of $\left(\frac{dR}{dt}\right)_{\text{diff}}$ is almost completely determined by the mean of the ensemble. We showed that the relative entropy of the full distribution decreases monotonically to zero in time, but at a non-uniform rate. As we will see next, the situation is quite different for the relative entropies of the marginal distributions $p(x_1,t)$ and $p(x_2,t)$, which may increase in time.
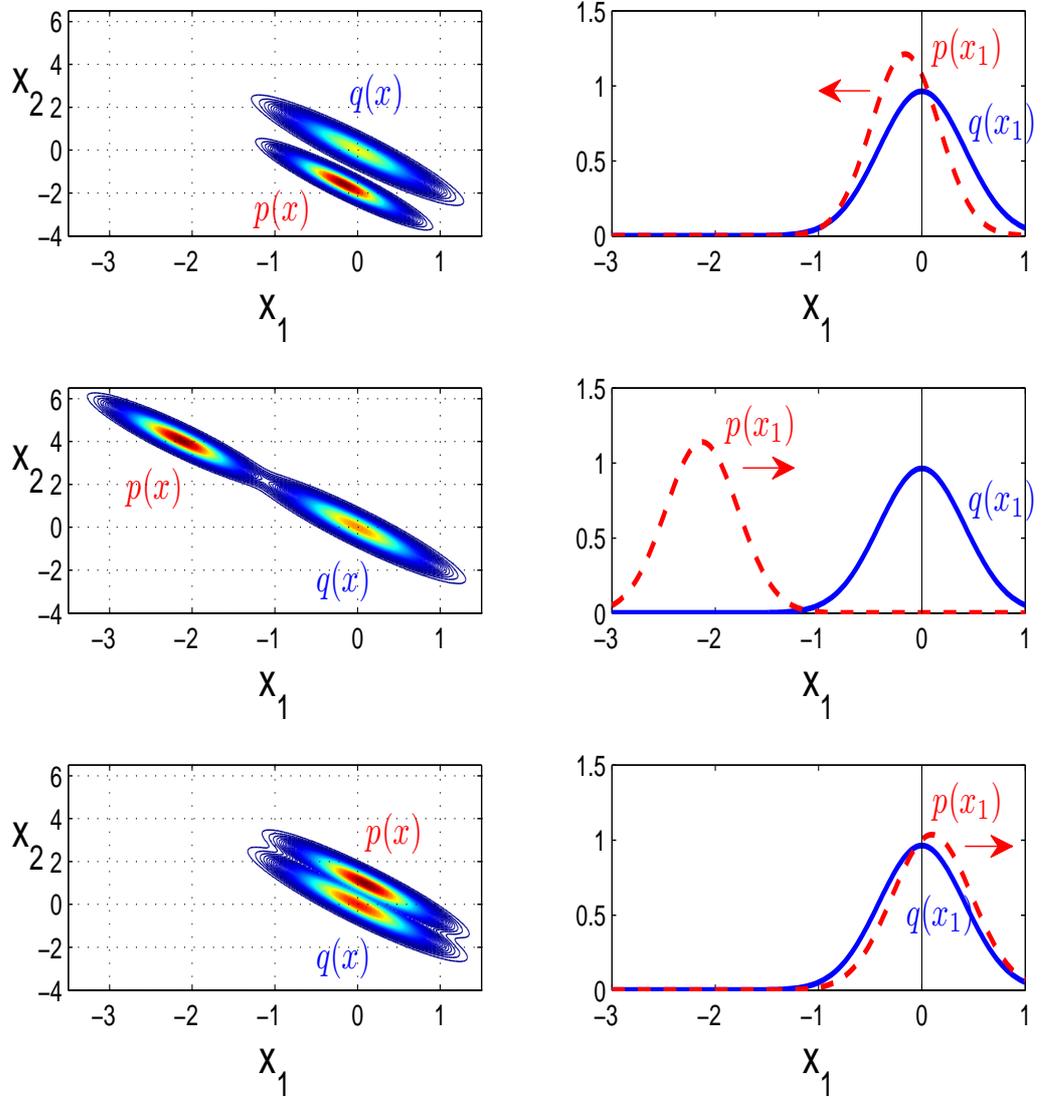
Figure 2.4: Probability density function of x_1 for the transient distribution (dashed line) and the equilibrium distribution (solid line) for times $t = 17.5$ (top), $t = 25$ (middle) and $t = 35$ (bottom). The arrows indicate the direction in which the transient distribution moves at the time of the snapshot.
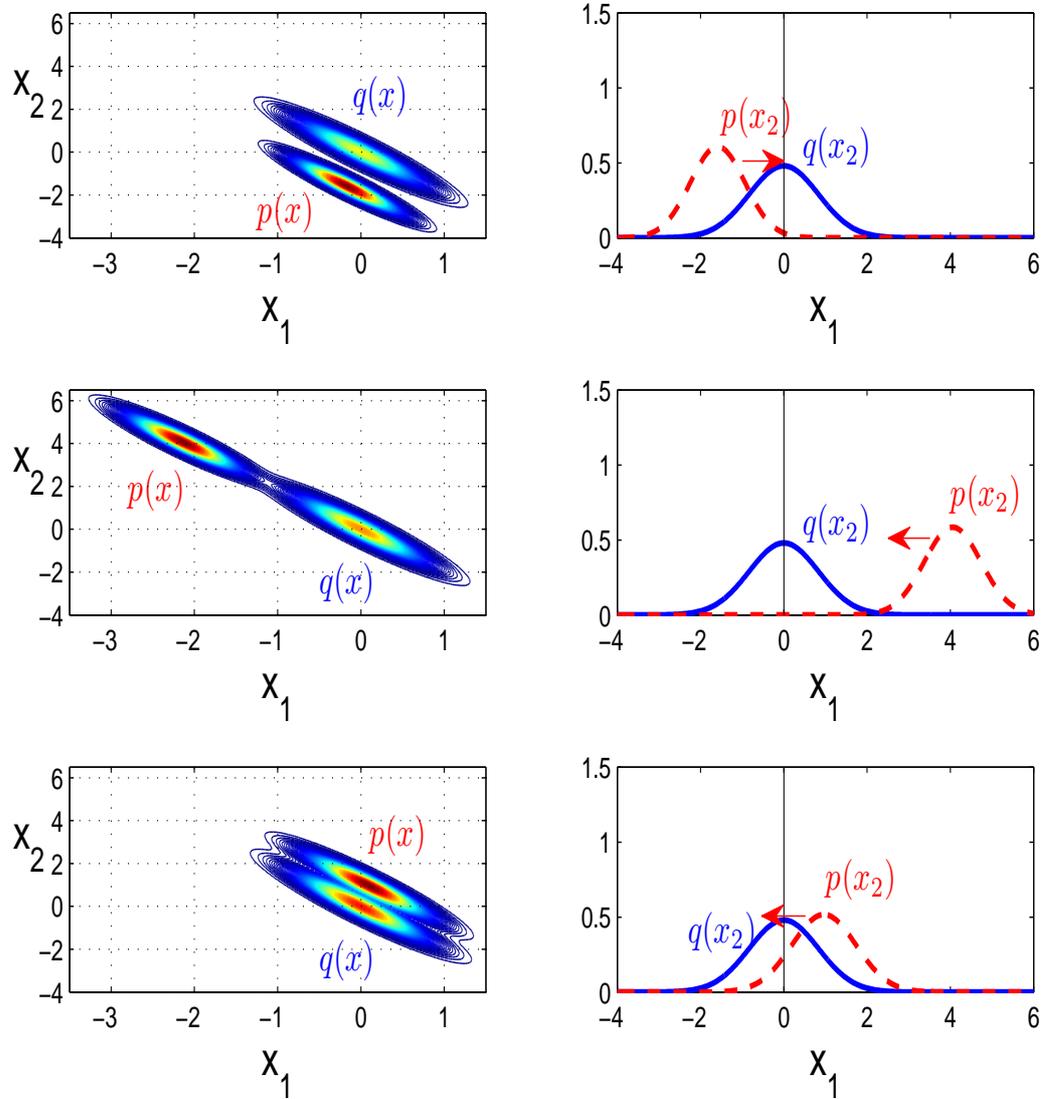
Figure 2.5: Probability density function of x_2 for the transient distribution (dashed line) and the equilibrium distribution (solid line) for times $t = 17.5$ (top), $t = 25$ (middle) and $t = 35$ (bottom). The arrows indicate the direction in which the transient distribution moves at the time of the snapshot.

2.2.3 Return of Skill for Marginal Entropies

We next consider the marginal entropies of the equilibrium and non-equilibrium distributions. The relative entropy $R_{x_1}(t)$ and $R_{x_2}(t)$ for the two marginal distributions are defined using (1.7), and can be interpreted as the amount of information that the marginal distribution $p(x_1, t)$ provides about the state of the variable x_1 at time t, in excess of the information provided by the marginal stationary distribution $q(x_1)$. Notice that a change in coordinates has no effect on the full relative entropy of the system. Hence, we emphasize the fact that marginal entropies are not invariant under coordinate changes, so the results of this section are highly coordinate dependent. Note that as compared to the full relative entropy, the marginal relative entropies do not necessarily decay in time. Relative entropy and marginal relative entropy are related through *conditional relative entropies* using the following equation

$$
\begin{aligned}
R_{x_2|x_1}(t) &= R(p(x_2|x_1, t), q(x_2|x_1)) \\
&= \int p(x_1, t) \int p(x_2|x_1, t) \log \frac{p(x_2|x_1, t)}{q(x_2|x_1)} dx_1 dx_2.
\end{aligned}
\tag{2.13}
$$

Here $p(x_2|x_1, t)$ denotes the conditional distribution of x_2 at time t given x_1, and $R_{x_2|x_1}(t)$ is the excess information provided by the marginal distribution $p(x_2|x_1, t)$ over $q(x_2|x_1)$. Another way to think about $R_{x_2|x_1}(t)$ is, how much information does x_2 contain about the system given that there is no change in the information content of x_1. Hence one can think of $R_{x_2|x_1}(t)$ as the effect in information content of x_2 given that x_1 has no effects.

The *chain rule for relative entropy* [24] relates the full, marginal, and conditional

relative entropy

$$R(t) = R_{x_2|x_1}(t) + R_{x_1}(t). \tag{2.14}$$

Relative entropy of the system can be thought of as the sum of information available from x_1 alone and the information attained from x_2 given that x_1 remains constant (i.e. information content of x_1 has no bearing on the information of x_2).

We start by calculating the relative entropy of the marginal distributions which can be obtained analytically using (1.9). The evolution of the marginal relative entropies is shown in Figure 2.6 (Right). For the initial condition $(x_1^0, x_2^0) = (1, 1)$, the oscillations are significant. The graphs of $R_{x_1}(t)$ and $R_{x_2}(t)$ in Figure 2.6 (Right) have very similar behavior. $R_{x_1}(t)$ and $R_{x_2}(t)$ are exactly the same after the initial spread of the ensemble, except for the horizontal shift. The x_1 and x_2 variables in (2.1) have a very similar structure except for the diffusion term present in x_2 variable. The drift term holds together the ensemble while the diffusion term spreads the ensemble. This constant interaction between the two terms lead one term to dominate at time intervals when the other term have a weak effect. Also, this observation implies that the information about the variables x_1 and x_2 taken separately can increase in time, while information about their joint distribution must always decrease.

The left and right columns of Figure 2.4 as well as Figure 2.5 compare the evolution of the full and marginal distributions. The increases in marginal relative entropy correspond to the times at which the mean of the marginal distribution moves away from the mean of the stationary distribution, *i.e.* the plateaus in the full relative entropy. Therefore, when the mean of the marginal distributions move towards the mean of the stationary distribution, the system as a whole loses information since the marginal relative entropy decays. Furthermore, the conditional relative entropy of the
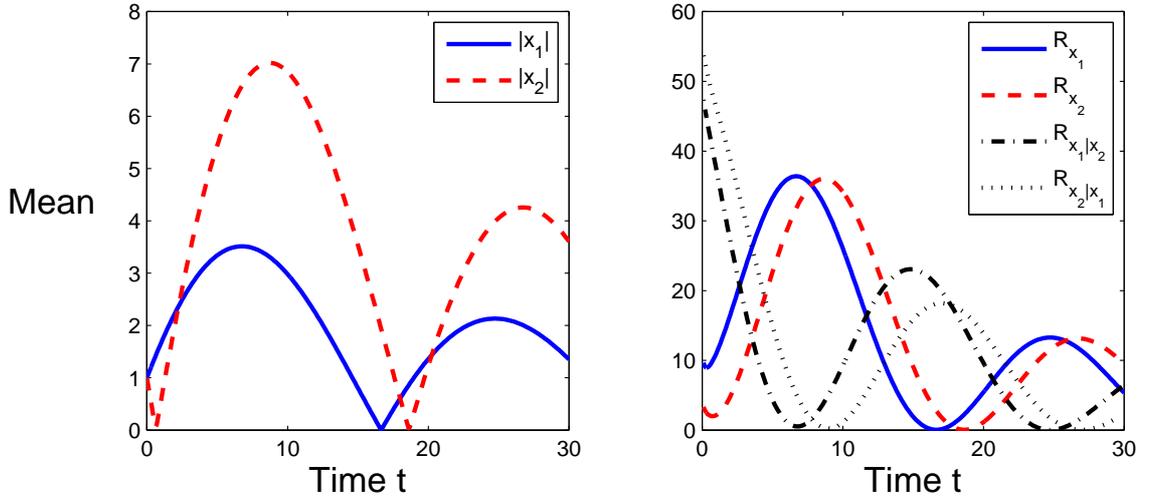
Figure 2.6: Left: Means of the variables $|x_1|$ (solid line) and $|x_2|$ (dashed line). Right: Marginal relative entropies $R_{x_1}(t)$ (solid line) and $R_{x_2}(t)$ (dashed line) and conditional relative entropies $R_{x_1|x_2}(t)$ (solid-dot line) and $R_{x_2|x_1}(t)$ (dotted line) for an initial condition $(x_1^0, x_2^0) = (1, 1)$.

system also decays during the time intervals of decrease in $R(t)$ which when combined with the marginal relative entropy leads to a sharp fall and sudden loss of information. Another way to think about this is using Equation (2.14). Notice that $R(t) \geq R_{x_1}(t)$ and $R(t) \geq R_{x_2|x_1}(t)$ and the oscillatory behavior of $R_{x_1}(t)$ and $R_{x_2|x_1}(t)$ is apparent from Figure 2.6 (Right). If the magnitude of the oscillations in $R_{x_2|x_1}(t)$ dominate $R_{x_1}(t)$, $R(t)$ mimics the behavior of $R_{x_2|x_1}(t)$ and vice versa. Opposing oscillations in $R_{x_2|x_1}(t)$ and $R_{x_1}(t)$ of the same order, causes $R(t)$ to plateau. The same ideas are valid in Equation (2.14), if x_1 and x_2 are exchanged. We can further explain the main factors contributing to this behavior as identified in the previous section, and here we provide another equivalent intuitive explanation using the mean and variance.

The right column of Figure 2.4 shows that the variance of the distribution $p(x_1)$

37

remains nearly constant during one oscillation. However, during the time between the first two panels the mean of the distribution moves away from 0, which leads to an increase in $R_{x_1}(t)$. Similarly, the movement of the transient to the stationary marginal distribution during the period between the last two panels, leads to a decrease in $R_{x_1}(t)$. The fact that both marginal relative entropies $R_{x_1}(t)$ and $R_{x_2}(t)$ increase at the same time, is because of the fact that for the solution of the corresponding deterministic system both $\bar{x}_1(t)$ and $\bar{x}_2(t)$ can increase at the same time. Note that this would not be true in a different coordinate system since marginal entropies are dependent on the choice of coordinates. In particular, for the matrix A in (2.8), $\bar{x}_1(t)$ and $\bar{x}_2(t)$ and the marginal relative entropies oscillate out of phase.

One question that is quite natural to ask is how the information contained in the marginal distributions of x_1 and x_2 is generated and is there a flow of information between the two variables. Again Equation (2.14) provides the answer: With an increase in information about the marginal Rx_1, comes a decrease in information about the conditional distribution $R_{x_2|x_1}$, that is, a decrease in the excess of information that a knowledge of x_1 provides about x_2 over that provided by the stationary distribution $q(x_2|x_1)$ (see Figure 2.6 (Right)). Similar argument applies for R_{x_2} and $R_{x_1|x_2}$.

We also note that there is no direct "flow of information" between the variables x_1 and x_2. However, one can intuitively think of a flow of information between the marginal and conditional distributions when the full relative entropy is approximately constant, since during that time the sum of the two is approximately constant as well. This is exactly the situation when the oscillations in the conditional and marginal are opposing and have the same magnitude.

2.3   Stochastic Non-Linear Oscillator

We saw in previous section that relative entropy of a stochastic linear oscillator depends on its initial conditions. Contrary to the idea that relative entropy for the full system decays monotonically, we showed that only averaged relative entropy decays monotonically, while marginal and conditional entropy can oscillate in time. These results extend to much more general stochastic systems as well. The non-uniform decay of relative entropy occurs whenever the main mass of the distribution $p(\vec{x}, t)$ approaches, and then diverges from the main mass of the stationary distribution $q(\vec{x})$. This is irrespective of the type of distribution chosen for the initial conditions. Oscillations in the marginal relative entropies occur when such divergence occurs in the marginal distributions. The following example shows that such behavior can be also expected for a non-linear stochastic oscillator, when the mass of the stationary distribution is distributed non-uniformly around the limit cycle.

2.3.1   Decay of Relative Information for Nonlinear Oscillators

We would like to show that the behavior of relative entropy described in the previous section can be observed in the case of the stochastic non-linear oscillator as well. In particular, we consider a planar, stochastic system with a limit cycle arising from a supercritical *Hopf bifurcation*. Supercritical Hopf bifurcation have been successfully used to describe various climate phenomena. Jin and Ghil [28] showed that Hopf bifurcation can be used to explain the dynamics of intraseasonal oscillations in the Northern Hemisphere extratropics. Furthermore, Korobeinikov and Mcnabb [33] showed that glacial oscillations can be associated with supercritical Hopf bifurcation

in the global climate system.

Due to the realistic applications of systems with Hopf bifurcation, we consider:

$$dx_1 = \mu x_1 dt - c\omega x_2 dt + \Theta x_1(x_1^2 + c^2 x_2^2)dt + \varepsilon dW_1,$$

$$\text{(2.15)}$$

$$dx_2 = \frac{1}{c}\left(\omega x_1 + c\mu x_2 + c\Theta x_2(x_1^2 + c^2 x_2^2)\right)dt + \varepsilon dW_2,$$

where $W_{1,2}$ are independent Wiener processes. Notice that for this particular system Fokker-Planck solution cannot be solved explicitly and hence we examine the system numerically with the following parameters values

$$\mu = 0.5, \quad \omega = 1.0,$$
$$c = 0.6 \text{ and } \Theta = -1.0.$$

$$\text{(2.16)}$$

Without the diffusion term, the system in (2.15) has a stable periodic orbit which has the form

$$x_1(t) = \sqrt{-\frac{\mu}{\Theta}}\cos(\omega t + \phi_0), \qquad x_2(t) = \frac{1}{c}\sqrt{-\frac{\mu}{\Theta}}\sin(\omega t + \phi_0), \qquad \text{(2.17)}$$

with period $T_{per} = 2\pi/\omega$. For very small noise, the invariant measure is concentrated sharply around the vertical extrema of the unperturbed orbit. As in the case of a linear oscillator, the invariant measure is stretched in the x_2 direction to better explain the non-uniform decay of relative entropy. Note that the speed at which a trajectory moves around the attracting periodic orbit of the deterministic system is at a minimum at the top and bottom where the extrema of the orbit lies. These are

therefore the places at which the equilibrium distribution will have local maxima, and high probability regions will occur. Similar behavior can be observed for other values of c for which the equilibrium measure is distributed non-uniformly along the limit cycle.

As mentioned before, since the Fokker-Planck eqution of the system is not explicitly solvable, the relative entropy is evaluated numerically by discretizing the phase space into a uniform mesh. Stochastic Euler method was used to integrate the equation. The equilibrium distribution $q(\vec{x})$ is estimated utilizing bin-counting from a single long realization and then calculating the distribution. Equilibrium distribution of the system given in (2.17) is shown in Figure 2.7. To calculate the initial non-
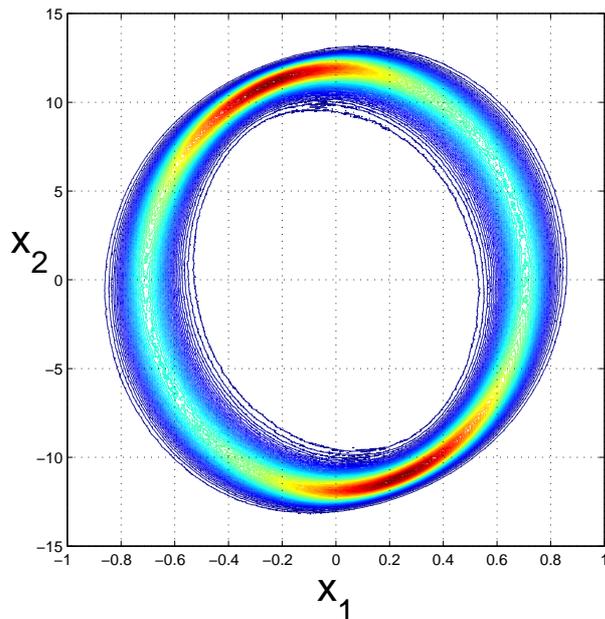
Figure 2.7: Equilibrium distribution of system given in (2.17) with parameters from (2.16)

equilibrium distribution, $250,000$ points are generated from the uniform distribution with width 0.3×0.3 centered at $(x_1, x_2) = (0.5, 0)$, away from the mean of the equilibrium distribution. Numerical estimates for relative entropy $R(t)$ and marginals relative entropies $R_{x_1}(t), R_{x_2}(t)$ are shown in Figure 2.8. The non-uniform decay of relative entropy and oscillations in marginal relative entropies are clearly visible after
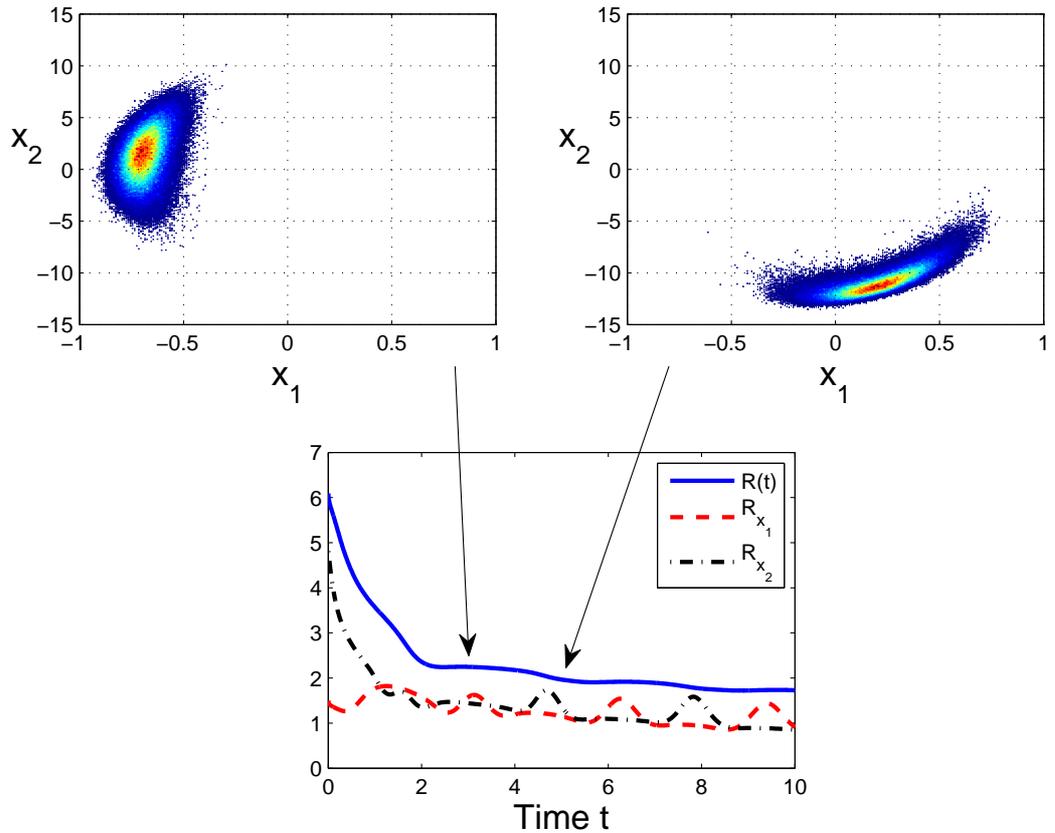
Figure 2.8: Top: Probability density function for $t = 3$ (left) and $t = 5$ (right). Bottom: Full relative entropy $R(t)$ (solid line) and marginal relative entropies $R_{x_1}(t)$ (dashed line) and $R_{x_2}(t)$ (dot-dash line) in simulations of (2.15) with $\epsilon = 0.1$ and initial ensemble centered at $(x_1^0, x_2^0) = (0.5, 0)$.

42

a short transient period. These results are similar to the one considered in Section 2.2. Since the periodic orbit given in (2.17) is stable, the transient period is due to the fast initial transition of the initial ensemble towards the orbit. Hence initially, the ensemble spreads and the relative entropy drops significantly.

After this initial phase, the relative entropy decays more slowly. As in the case of the damped stochastic linear oscillator, the variance of the transient distribution increases slowly compared to the time of the oscillations. Figure 2.8 shows that the plateaus in relative entropy correspond to the times during which the mass of the transient distribution moves away from a peak in the mass of the stationary distribution. Therefore the plateaus occur and the nonuniform decay of relative entropy is exactly due to the mechanism discussed in the previous section.

2.3.2 Return of Skill for Marginal Entropies

Marginal entropies depicted in Figure 2.8 exhibit strong out of phase oscillatory behavior with frequency 2ω, the same frequency we see in the plateaus of the full relative entropy. The marginal stationary distribution $q(x_1)$ is approximately unimodal. As in the case of the linear oscillator, the minima of the marginal entropy $R_{x_1}(t)$ occur at the times at which the mean value of the transient marginal distribution coincides with the mean value of the equilibrium distribution.

Since the marginal distribution $q(x_2)$ is strongly bimodal, the situation is little different. The minima of $R_{x_2}(t)$ occur at the times when the mean of the transient distribution $p(x_2)$ is between the two peaks in the stationary distribution $q(x_2)$. Since this occurs exactly when the distribution $p(x_2)$ is at its farthest away from $q(x_2)$, the marginal relative entropies $R_{x_1}(t)$ and $R_{x_2}(t)$ oscillate out of phase. This out of phase

oscillations are seen in the conditional relative entropies as well. Clearly, the behavior of $R(t)$ is very similar to the case of a linear oscillator and given the explanation for marginal entropies, it is not surprising that the conditional entropies $R_{x_1|x_2}(t)$ and $R_{x_2|x_1}(t)$ also oscillate out of phase using (2.14). This is visible in Figure 2.9 (Right) for time past the initial phase of the system. Notice in Figure 2.9 (Left), the system
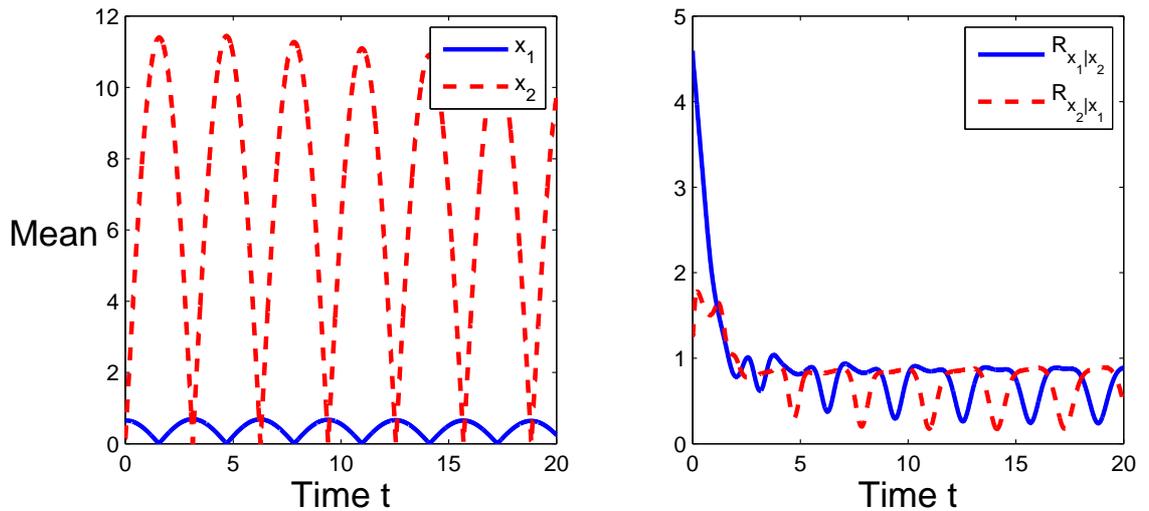
Figure 2.9: Left: Means of the variables $|x_1|$ (solid line) and $|x_2|$ (dashed line) for the system given in (2.17). Right: Conditional relative entropies $R_{x_1|x_2}(t)$ (solid line) and $R_{x_2|x_1}(t)$ (dashed line) for the system given in (2.17) with initial condition $(x_1^0, x_2^0) = (0.5, 0)$.

oscillates with a much higher frequency for x_2 than x_1, which is not the case in Figure 2.6 (Left). Also notice that the mean of the non-linear oscillator for x_1 and x_2 are completely out of phase as compared to the one for the linear oscillator. Even though the damping in 2.9 (Right) for x_1 is not quite visible, one can see a slight decay in mean of x_2. The x_1 variable also decays but with much slower rate compared to x_2. This damping behavior is very similar to the linear oscillator.

44

2.4 Stochastically Perturbed Duffing Equation

Coherence resonance is a dynamical systems phenomenon where addition of certain amount of noise leads to coherent oscillations. We next consider a system of nonlinear stochastic differential equations exhibiting coherence resonance [50, 17]. Although the deterministic dynamical behavior of the new system is very different from both previous examples, stochastic perturbations lead to intervals of extended predictability and the return of skill for marginal distributions [14].

We consider the Duffing equation driven by white noise [57]

$$dx_1 = x_2\ dt + \varepsilon\ dW_1,$$

$$(2.18)$$

$$dx_2 = (x_1 - x_1^3 - \gamma x_2 + \beta x_1^2 x_2)\ dt + \varepsilon\ dW_2,$$

where $W_{1,2}$ are independent Wiener processes, γ and β are parameters and ε is the noise strength. For $\varepsilon = 0$ and parameters $\gamma = 0.4$ and $\beta = 0.497$, Equation (2.19) yields a non-linear system possessing an attracting double homoclinic cycle (Figure 2.10 (Left)) with a saddle point at the origin $(0,0)$. The signature of the homoclinic loop is clearly visible in the joint distribution of the stochastic system in (2.19) shown in Figure 2.10 (Right).

To demonstrate the non-uniform decay of predictability and the return of skill for marginal distributions, we chose a particular $250,000$-member initial ensemble centered at $x_1 = 0.25$, $x_2 = 0.25$, and calculated the relative entropy numerically as in the previous example. The choice of the initial conditions are again relevant to the behavior of the relative entropy. The distributions $p(\vec{x}, t)$ are computed utilizing
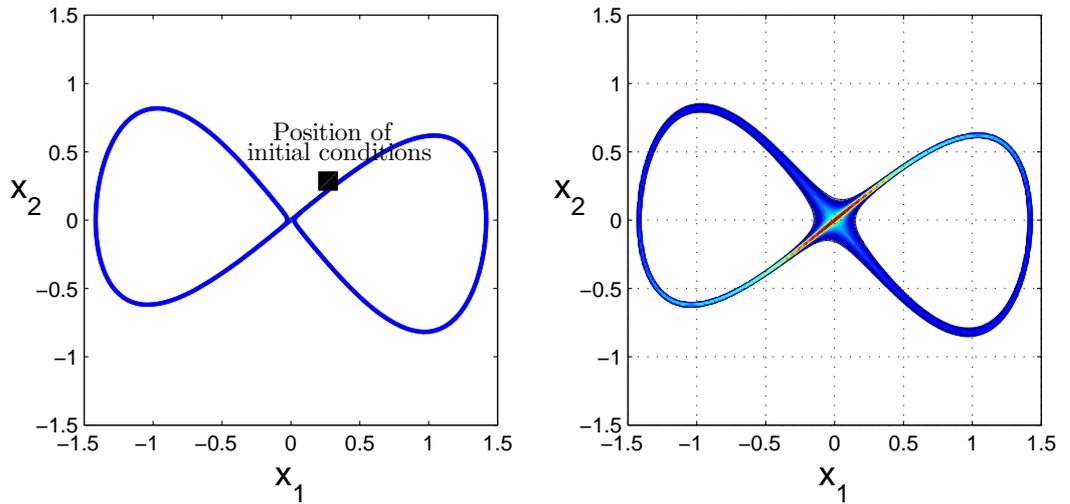
Figure 2.10: Left: Homoclinic loop for the Duffing equation in (2.19) with $\varepsilon = 0$. Right: Contour plot of probability density function for $\varepsilon = 0.01$.

the Monte-Carlo simulations with the initial ensemble generated from the uniform distribution on a square $[0.3] \times [0.3]$ (see Figure 2.12 (Left)). The choice of initial ensemble may not necessarily be uniform. The computed relative entropy for $\varepsilon = 0.01$ is presented in the bottom panel of Figure 2.11. As in the linear oscillator case, the relative entropy is almost constant over several time intervals. There are four plateaus in the graph of relative entropy, although the reason for the first plateau (at times $[4, \ldots, 8]$) is somewhat different from the subsequent ones. Furthermore, the first plateau seems to occur for a much longer time period than others.

Recall that the decay in relative entropy is only due to the diffusion in equation
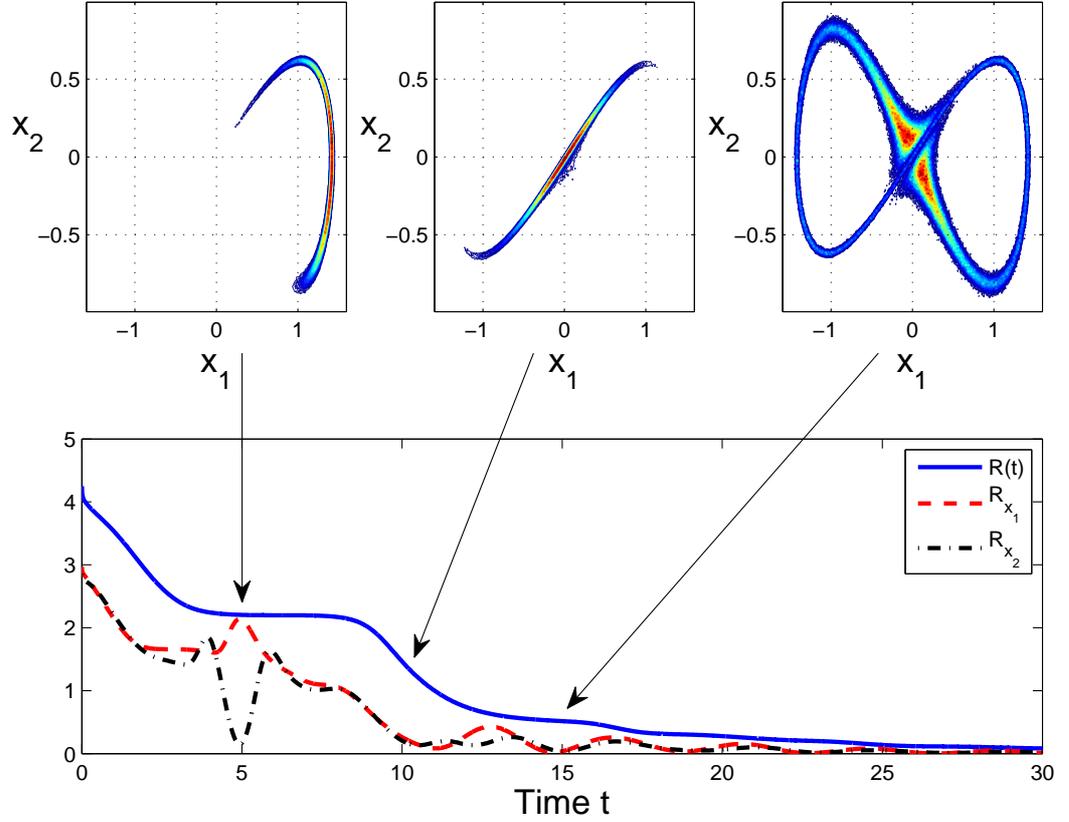
Figure 2.11: Top: Transient probability density function for $t = 5$ (left), $t = 10$ (middle) and $t = 15$ (right). Bottom: Full relative entropy $R(t)$ (solid), Marginal relative entropies $R_{x_1}(t)$ (solid) and $R_{x_2}(t)$ (dashed) in simulations of (2.19) with initial ensemble centered at $(x_1^0, x_2^0) = (0.25, 0.25)$ and $\varepsilon = 0.01$.

(1.15). For the model (2.19) the diffusion term becomes

$$\left(\frac{dR}{dt}\right)_{\text{diff}} = -\frac{\varepsilon}{2} \int dx_1 dx_2 \; p(x_1, x_2) \sum_{i=1,2} \left[\frac{\partial}{\partial x_i} \left(\log \frac{p(x_1, x_2)}{q(x_1, x_2)} \right) \right]^2. \qquad (2.19)$$

For the stochastic Duffing equation the behavior of the $\left(\dfrac{dR}{dt}\right)_{\text{diff}}$ is more complicated

47

than in the case of the linear oscillator and the non-linear oscillator. Namely, the value of $\left(\dfrac{dR}{dt}\right)_{\text{diff}}$ depends not only on the means, but also on all terms involving variances of the equilibrium and non-equilibirum distributions.

Since the initial ensemble is chosen on one side of the heteroclinic loop, different trajectories do not separate during the first passage along the heteroclinic loop (see the first two top panels in Figure 2.11). As the cluster of initial conditions moves away from the origin we observe the long plateau in the graph of relative entropy, since the non-equilibrium distribution moves away from the origin where the main mass of the equilibrium distribution is concentrated. The transient distribution spreads away from the main mass located at the origin resulting in the sharp decline in relative entropy initially. Indeed, Figure 2.12 illustrates that the mean of transient distribution is largest at times $[4, \ldots, 8]$, coinciding with the first plateau in relative entropy $R(t)$. After the first transition, individual realizations return close to the origin, but separate following the two different sides of the homoclinic loop. Therefore, the mean of the ensemble is approximately zero (see Figure 2.12 (Left)). Due to coherence resonance [17, 50, 57], there exist a mean exit time as most of the mass of the transient distribution is ejected from the vicinity of the origin around the same time. The second plateau in the graph of relative entropy occurs when the two main portions of the transient distribution are at their farthest distance from the main mass of the equilibrium distribution at times $[15, \ldots, 17]$. The bimodal behavior of the transient distribution during this time implies that the oscillatory behavior is manifested strongly through the variance of the ensemble (see Figure 2.12 (Right)). Hence the oscillatory behavior of the mean is responsible for the plateau initially while the variance becomes important for the plateaus that occur later. During the

oscillatory behavior in mean/variance, the other term variance/mean is approximately zero.

Although the details are somewhat different from the previous examples, the non-uniform decay in relative entropy is again due to the fact that the stationary distribution is concentrated in one area of the phase space, and oscillations in the system that take the transient distribution recurrently close to the main mass of the stationary distribution.

2.4.1 Return of Skill for Marginal and Conditional Entropies

Marginal relative entropies for x_1 and x_2 are shown in the bottom panel of Figure 2.11. The explanation for the oscillations in both marginal entropies is similar to
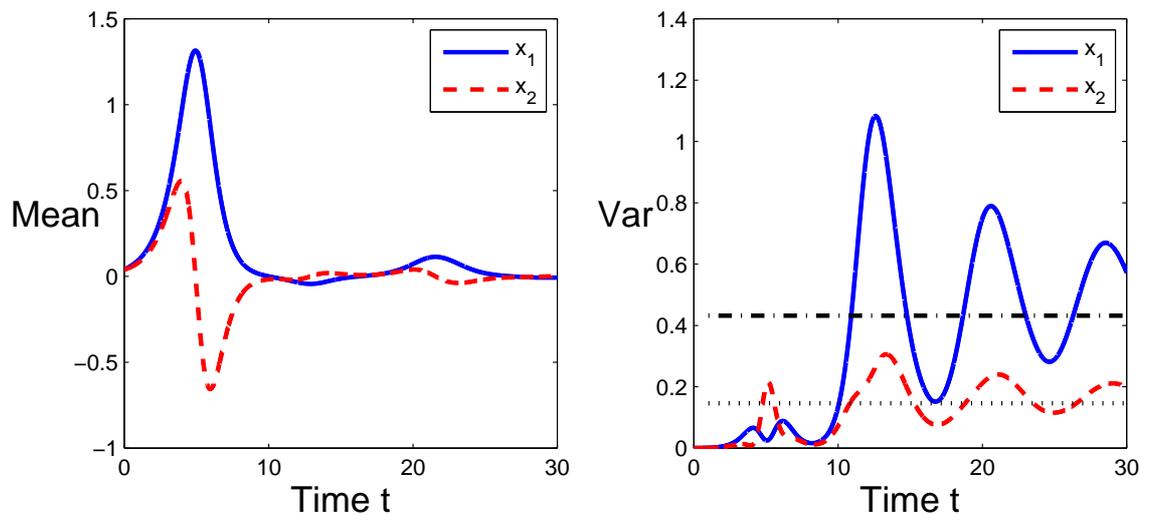
Figure 2.12: Left: Mean in x_1 (solid line), mean in x_2 (dashed line). Right: Variance of x_1 (solid line) and variance of x_2 (dashed line) in simulations of the stochastic Duffing equation in (2.19) with $\varepsilon = 0.01$. (Horizontal lines show equilibrium variances).

the one described in the preceding examples. Figure 2.11 shows that the marginal entropies are at a maximum at the times during which the main mass of the marginal transient distribution diverges maximally away from the marginal of the stationary distribution.

The first increase in x_1 marginal and x_2 marginal (around $t = 5$ for x_1 and $t = 4, 6$ for x_2) corresponds to the situation when most of the mass is at the furthest distance from the fixed point in x_1 direction or x_2 direction, respectively (Figure 2.11 (Top Left)). Furthermore, in Figure 2.12 the increases in marginals of relative entropy occurs when the mean of the distribution deviates away from the fixed point (equilibrium mean) of zero. Therefore, the first oscillation in marginal relative entropies can be explained by observing changes in the mean of x_1 and x_2. Note that after an initial increase, the mean decays approximately to zero and reaches the equilibrium mean. Hence mean does not explain the subsequent oscillations.

Oscillations in marginal relative entropies are also observed for later times $t = 14, 17, 24, 28$. These oscillations can be attributed to the change in the variance of the corresponding dynamical variable. Note that in Figure 2.11 (Bottom) there are a total of four peaks in the marginal relative entropies after $t = 10$. On the other hand there are only two peaks in the variance. The reason for this is that the variances of the marginals are larger than the variances of the stationary marginal distribution at times around $t = 15$ and $t = 24$, but are smaller than the variances of the stationary marginal distribution around $t = 17$ and $t = 27$. The overshoots and undershoots from the equilibrium variances are a consequence of coherence resonance. The fact that the system oscillates and resonates at the same time is the reason for these oscillations.
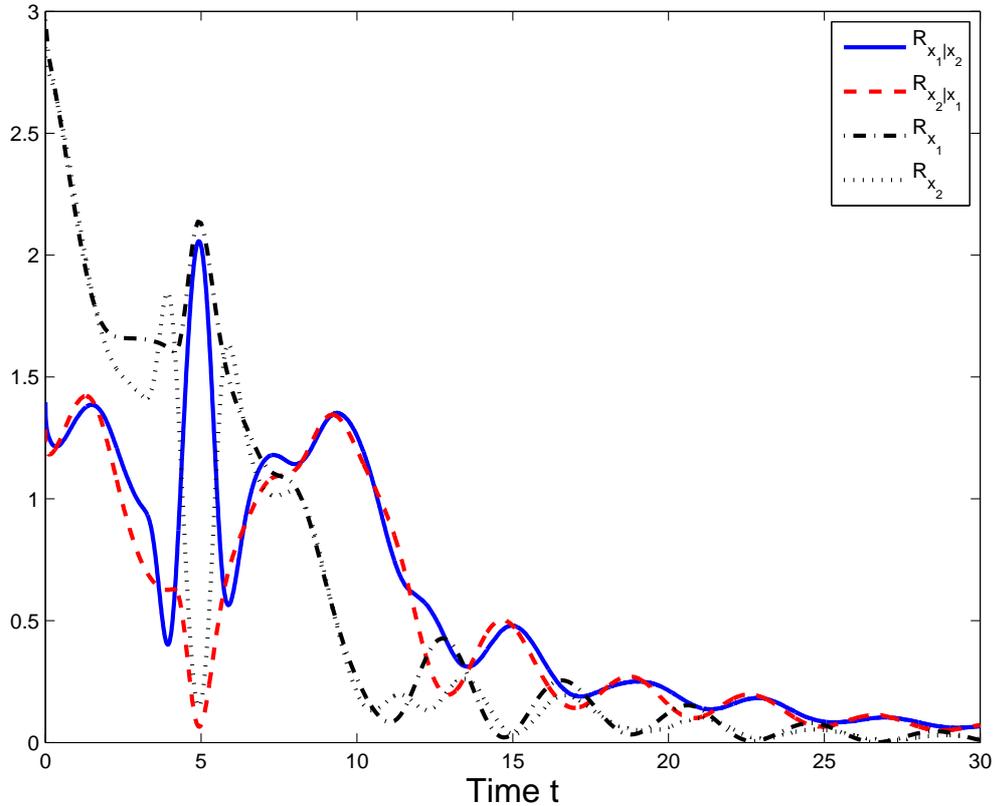
Figure 2.13: Conditional relative entropies $R_{x_1|x_2}(t)$ (solid line) and $R_{x_2|x_1}(t)$ (dashed line) with marginal relative entropies R_{x_1} (dash-dot line) and R_{x_2} (dotted line) in simulations of (2.19) with initial ensemble centered at $(x_1^0, x_2^0) = (0.25, 0.25)$ and $\varepsilon = 0.01$.

Oscillations in conditional entropies can also be explained similarly. This fact is much easier to explain once marginal entropies have been considered. In Figure 2.13, the first oscillation in the conditional entropies $R_{x_1|x_2}$ and $R_{x_2|x_1}$ occur at around $t = 1.5$. This oscillation can be attributed to the dramatic change in the decay rate of the marginal relative entropy which occur around the same time. In particular, we discuss explicitly the case of R_{x_2} and $R_{x_1|x_2}$. During time interval $[4, \ldots, 8]$, when

relative entropy remains constant, troughs occur in $R_{x_1|x_2}$ at around $t = 4, 6$ exactly at times where marginal of x_2 peaks. Furthermore, trough in marginal of x_2 at time $t = 5$ coincides with peak in $R_{x_1|x_2}$. The sharp decay in relative entropy in the time interval $[9, \ldots, 13]$ contributes to the peak in conditional entropy $R_{x_1|x_2}$ at around $t = 9$. The rate of decay in relative entropy is slower compared to the marginal entropy at time interval $[8, \ldots, 9]$, hence this significant difference in the decay rates lead to the peak in the conditional entropy. After $t = 13$ relative entropy is approximately constant; the peaks in conditional entropy $R_{x_1|x_2}$ occur at times where the marginal entropy of x_2 have troughs. The similar explanation can be applied to $R_{x_2|x_1}$ using marginal of x_1.

The return of skill (maxima in the graph of marginal relative entropies) occurs when the transient variance deviates from the equilibrium variance. On the other hand, minima in oscillations of the marginal relative entropies correspond to times when the variance of the ensemble is nearly identical to the equilibrium value. Again, the idea that the mean is responsible for the initial oscillations and the variance is the main factor in the subsequent oscillations, seems to be the theme. Hence the mean and variance can explain the plateaus in the full relative entropy as well as the oscillations in the marginal entropies.

2.5   Conclusions

We considered the predictability of three models that encapsulate a wide range of systems with oscillatory behavior. The particular emphasis was on the non-uniform decay of the utility of predictions and return of skill (oscillations in marginals and con-

ditional entropies) for dynamic variables. These models were constructed as stochastic perturbations of linear oscillator, non-linear oscillator (Hopf normal form) and homoclinic cycle (Duffing equation), and are good representatives of a wide class of stochastic oscillators.

Relative entropy is utilized to characterize the predictability properties of these prototype systems. Relative entropy was used as a measure of predictability due to its particular mathematical properties. For Markov process, relative entropy simplifies significantly and was particularly well-suited for the models that were considered. The average (with respect to many initial ensembles) predictability of all three systems decays exponentially in time. Nevertheless, relative entropy of particular initial conditions can have time intervals where there might be plateaus. In particular, the two related phenomena that emerge in the behavior of the relative entropy functional and marginal entropies for each particular ensemble simulation were (i) the full relative entropy decays at a non-uniform rate, and (ii) there is return of skill (oscillatory behavior) for the marginal entropies for all three systems. Notice that both the phenomena can be explained by the oscillatory behavior of the systems considered.

Interestingly, we can also think of the return of skill as a flow of information from the conditional to the marginal non-equilibrium distribution. Both of these phenomena are driven by oscillations of the mean of the non-equilibrium ensemble, and an increase in the variance of the non-equilibrium ensemble. Furthermore, variance of the transient distribution is responsible for the sharp initial decay in relative entropy.

The main idea in this chapter is the transport of the non-equilibrium distribution in phase space by the underlying oscillatory dynamics. This results in a slower rate of decay for the relative entropy when the mean of the non-equilibrium ensemble is

moving away from an area in which the equilibrium distribution is concentrated. The same mechanism causes oscillations in marginal distributions and return of skill in each dynamic variable. Hence, the plateaus seen in the relative entropy were due to the non-equilibrium mean moving away from the equilibrium mean.

While the exact details differ between the oscillatory systems considered, the non-uniform decay rate of the relative entropy functional is similar in all three cases. The oscillatory behavior is due to the initial ensemble concentrated in the tails of the equilibrium distribution, but can also be detected for other initial data. This suggests that similar behavior can be detected in more complex systems, especially when the initial ensembles are concentrated around rare events.

Chapter 3

Apriori Prediction of Symmetry Breaking in Stochastic Systems

3.1 Introduction

Recently, stochastic modeling has been given increased attention as a tool to understand and simplify multi-scale systems. One such example is atmosphere/ocean modeling. Most models in the atmosphere-ocean have multiple time scales, hence straightforward numerical calculations are very lengthy or even infeasible. Stochastic-Mode Reduction (see Appendix B) is one particular approach that was be considered; in this approach the non-essential degrees of freedom are eliminated and substituted by stochastic terms.

Another important area of research in atmosphere-ocean modeling is the understanding of low-dimensional coherent structures. In the atmosphere-ocean applications, stable low-dimensional structures such as multiple equilibria, periodic orbits,

and homoclinic/heteroclinic connections in the low-dimensional projected phase space have been utilized to explain many physical phenomena. Large coupled models are hard to study, so to get an insight into the full behavior of the system, low-dimensional projections were used. In [9], Charney and DeVore used low-dimensional projections with external forcing to find multiple equilibrium states in a barotropic channel model. The choice of external forcing was important since different results were obtained for different type of forcing. Charney and DeVore suggested significant application of the results in atmosphere phenomena since equilibrium states correspond to a system that will remain there for a long period of time. Furthermore, constant transitioning between different states can explain the variability and hence, lack of predictability. While Charney and DeVore focused on equilibrium states, Crommelin in [14] suggest the importance of homoclinic dynamics in atmospheric ultralow-frequency variability (i.e. timescale in years to decades). The subsystem derived from the large system has strong signature of the homoclinic dynamics. Crommelin suggested that homoclinic orbits are responsible for the long-timescale variability in large systems. Hence, the importance of coherent structure has been suggested in many physical applications.

Numerous techniques are utilized to replace neglected variables with appropriate terms. Effect of random noise is one method employed to model interactions with the neglected variables. Berner and Branstator in [7], study an atmospheric circulation model and investigate low dimensional projections of the system on the phase space. They observe clear linear and non-linear signatures in the mean dynamics of a subset of the variables. The linear signature was effectively replaced by least squares operator driven by Gaussian white noise while the non-linear signature was replaced by two linear functions. Berner and Branstator argue the importance of reduced dynamics

in capturing the full behavior, where random fluctuations were used to replace the non-essential variables.

In this chapter, coupled systems are analyzed which are designed to address the interaction between coherent structures and noise. In particular, systems considered here address the effect of coupling and/or perturbations of systems with a stable periodic orbits. In addition, systems presented here help explain the significance of the recently developed mode-reduction strategy [40, 41] for the high-dimensional dynamical systems with separation of time scales. The mode-reduction strategy was designed to reduce the dimension of the problem by effectively replacing the fast variables by appropriate stochastic terms. In [41], it was shown that the stochastic-mode reduction works extremely well and gives rise to stochastic terms with multiplicative and additive noise. This was shown using two examples, in particular the triad model and the coupled Truncated Burgers–Hopf (TBH) system. Mode-Reduction was effectively applied to other low-dimensional systems [39, 42] and more realistic atmospheric models [21, 22].

In this work, we consider low-dimensional models coupled with additional degrees of freedom where the coupling is "additive", i.e. coupling terms are replaced by damping and additive white noise using the stochastic mode-reduction strategy. It is shown that the coupling with additional degrees of freedom destroys the original rotational symmetry of the truncated low-dimensional system. Additionally, the direction of the symmetry breaking can be predicted apriori without any knowledge of the statistical dynamics of the fast modes. The nature of the symmetry breaking is evident in the statistical behavior of the slow variables. These ideas are first applied to a simple gradient system coupled with fast modes. In this particular case, a sim-

ple analysis yields an explanation for the maxima in the joint probability density of the slow modes. Essentially, the eigenvectors of the effective damping matrix play a crucial role in explaining the statistical distribution of the dynamic variables. The role of the diffusion matrix is also crucial in the analysis. Furthermore, analysis simplifies, since the diffusion matrix can be written as a constant multiple of the drift matrix. Hence, the eigenvectors of both the drift and the diffusion matrices are the same, leading to the explanation of statistical properties of the slow variables. We will further elaborate if the drift and diffusion matrix do not have similar eigenvalue decomposition.

Next, we consider a two-dimensional system with a stable circle of equilibria coupled with two additional variables. Although this system cannot be recast as a gradient system and no potential function exists, the signature of the periodic orbit is strong in the coupled model and two peaks occur in the joint density of the slow variables. We demonstrate that the same mechanism as the one for gradient systems is responsible for occurrence of these peaks. In particular, only two points on the stable periodic orbit "survive" the perturbation and these two points can be predicted apriori from the structure of the effective damping.

Effect of coupling strength between the slow and fast variables is also addressed. We show that due to an increase in strength of coupling, mode-reduction strategy fails. The slow-fast variables have no clear separation of time-scales which is evident by looking at correlation functions. Nevertheless, the reduced system effectively predicts the orientation of the joint pdf in the slow variables.

3.2 Gradient Systems

To explain symmetry breaking, we consider a particular case of rotationally invariant gradient system. In particular, interactions of the slow variables can be represented by a gradient system. General gradient systems have the form

$$\frac{d}{dt}\mathbf{x} = -\nabla V, \tag{3.1}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \mathbf{V} = \mathbf{V}(|\mathbf{x}|^2). \tag{3.2}$$

For ease and simplicity of presentation, a two-dimensional case is studied here. It is important to point out that this approach can be generalized to higher dimensions. We consider the potential

$$V(x_1, x_2) = -\frac{1}{2}\mu \left(1 - \frac{\alpha_0}{2}|\mathbf{x}|^2\right)|\mathbf{x}|^2, \tag{3.3}$$

and the system of equations in (3.1) becomes

$$\dot{x}_1 = \mu(1 - \alpha_0|\mathbf{x}|^2)x_1,$$

$$\tag{3.4}$$

$$\dot{x}_2 = \mu(1 - \alpha_0|\mathbf{x}|^2)x_2,$$

where $|\mathbf{x}|^2 = x_1^2 + x_2^2 = \alpha_0^{-1}$ is the stable circle of equilibria. The circle of equilibria is stable when $\mu > 0$. Graphical representation of V is given in Figure 3.1. Fast
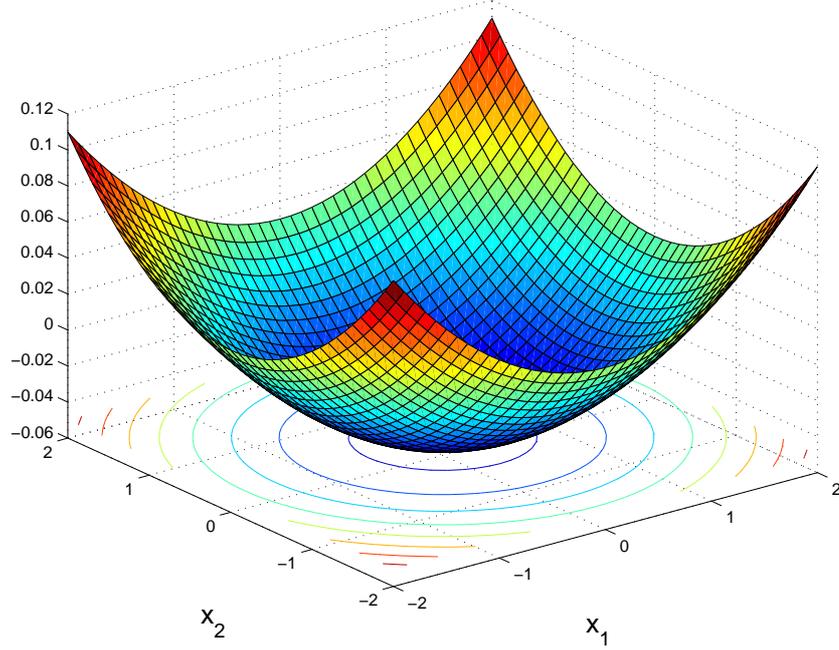
Figure 3.1: 3D Plot of function $V(x_1, x_2)$ given in (3.3) with $\alpha_0 = 0.8$ and $\mu = 0.1$

variables, u_1 and u_2, are utilized to break the original symmetry. We couple the fast variables, u_1 and u_2 with the slow variables, x_1 and x_2

$$
\begin{aligned}
\dot{x}_1 &= (-\nabla V)_1 + \lambda A_1 u_1 u_2, \\
\dot{x}_2 &= (-\nabla V)_2 + \lambda B_1 u_1 u_2, \\
\dot{u}_1 &= \lambda A_2 x_1 u_2 + \lambda B_2 x_2 u_2 - \gamma_1 u_1 + \sigma_1 \dot{W}_1, \\
\dot{u}_2 &= \lambda A_3 x_1 u_1 + \lambda B_3 x_2 u_1 - \gamma_2 u_2 + \sigma_2 \dot{W}_2,
\end{aligned}
\tag{3.5}
$$

where $(\nabla V)_i$ is the ith component of the vector-valued function ∇V, W_1 and W_2 are independent Wiener processes, A_j and B_j for $j = 1, 2, 3$ are known interaction coefficients and λ controls the strength of coupling between the slow and fast modes.

60

In this system, the energy-conserving interactions between the slow, $\mathbf{x}$, and the fast, $\mathbf{u}$, variables can be recast using the interaction coefficients as follows

$$A_1 + A_2 + A_3 = 0, \qquad B_1 + B_2 + B_3 = 0. \tag{3.6}$$

Under the assumption that the slow variables decorrelate more slowly than the fast variables, the stochastic-mode reduction strategy (see Appendix B) can be applied to the system given in (3.5). The reduced model for $\mathbf{x} = (x_1, x_2)^T$ is given by the stochastic differential equation

$$d\mathbf{x} = -\nabla V \, dt - A\mathbf{x} \, dt + \Sigma \, d\mathbf{W}, \tag{3.7}$$

where $\mathbf{W} = (W_1, W_2)^T$ is a vector of independent Wiener processes and A and Σ are 2×2 damping and diffusion matrices, respectively. In particular, the matrices are

$$A = \frac{\lambda^2}{\gamma_1 + \gamma_2} \begin{pmatrix} A_1 A_2 \frac{\sigma_2^2}{2\gamma_2} + A_1 A_3 \frac{\sigma_1^2}{2\gamma_1} & A_1 B_2 \frac{\sigma_2^2}{2\gamma_2} + A_1 B_3 \frac{\sigma_1^2}{2\gamma_1} \\[2ex] A_2 B_1 \frac{\sigma_2^2}{2\gamma_2} + A_3 B_1 \frac{\sigma_1^2}{2\gamma_1} & B_1 B_2 \frac{\sigma_2^2}{2\gamma_2} + B_1 B_3 \frac{\sigma_1^2}{2\gamma_1} \end{pmatrix}, \tag{3.8}$$

$$\Sigma = \frac{\lambda \sigma_1 \sigma_2}{\sqrt{2\gamma_1 \gamma_2}} \frac{1}{\sqrt{\gamma_1 + \gamma_2}} \begin{pmatrix} \frac{A_1^2}{\sqrt{A_1^2 + B_1^2}} & \frac{A_1 B_1}{\sqrt{A_1^2 + B_1^2}} \\[2ex] \frac{B_1 A_1}{\sqrt{A_1^2 + B_1^2}} & \frac{B_1^2}{\sqrt{A_1^2 + B_1^2}} \end{pmatrix} \tag{3.9}$$

61

Using (3.6), $\sigma_1 = \sigma_2 \equiv \sigma$ and $\gamma_1 = \gamma_2 \equiv \gamma$, (3.8) and (3.9) become

$$A = \frac{\lambda^2 \sigma^2}{4\gamma^2} \begin{pmatrix} A_1^2 & A_1 B_1 \\ A_1 B_1 & B_1^2 \end{pmatrix},$$

(3.10)

$$\Sigma = \frac{\lambda \sigma^2}{2\gamma^{3/2}} \frac{1}{\sqrt{A_1^2 + B_1^2}} \begin{pmatrix} A_1^2 & A_1 B_1 \\ B_1 A_1 & B_1^2 \end{pmatrix}.$$

(3.11)

3.2.1 Symmetry Breaking in Gradient Systems

Matrix A and Σ are symmetric due to the energy conserving interactions. Hence

$$A^T = A, \quad \Sigma^T = \Sigma,$$

(3.12)

with the property

$$\Sigma = \text{const } A.$$

(3.13)

Moreover, the damping and diffusion matrices have several other additional properties. Firstly, each matrix has only one non-zero eigenvalue while the other eigenvalue is zero. This is a direct consequence of the energy–conserving coupling between the slow and fast variables. It is easy to evaluate the eigenvalues and in particular the non-zero eigenvalue of A is positive. In a simple case like (3.10), we can see that the eigenvalues are 0 and $A_1^2 + B_1^2$, where the latter is clearly positive. Secondly, damping and diffusion matrices are similar matrices up to a constant, hence we have the same eigenvectors, and thus, the same eigenvalue decomposition which follows from the

62

property (3.13). Therefore, matrices A and Σ can be diagonalized as follows

$$A = R^{-1}\tilde{A}R, \quad \Sigma = R^{-1}\tilde{\Sigma}R, \quad \tilde{\Sigma} = c\tilde{A}, \tag{3.14}$$

where $\tilde{A}$ and $\tilde{\Sigma}$ are diagonal matrices of eigenvalues and R is the matrix of eigenvectors. Moreover, R is unitary due to the property (3.12), which implies that the eigenvectors are perpendicular and R can be recasted as rotation matrix, i.e.

$$R = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}. \tag{3.15}$$

A particular value of ψ depends only on the choice of interaction coefficients A_j and B_j, $j = 1, 2, 3$, and can be computed apriori without any knowledge about the behavior of the fast variables. For (3.10), the eigenvectors are $\left(-\dfrac{B_1}{A_1}, 1\right)^T$ and $\left(\dfrac{A_1}{B_1}, 1\right)^T$. It is easy to see that the eigenvectors are orthogonal, more precisely, taking advantage of the symmetric matrix A, the eigenvectors form an orthonormal basis.

As shown in (3.8), leading order statistics of the fast variables (i.e. γ_i and σ_i, $i = 1, 2$) enters as a constant in front of matrices composed of interaction coefficients in the expressions for A and Σ. Therefore, statistical behavior of the fast variables affects only the magnitude of eigenvalues and has no affect on the eigenvectors. We would like to emphasize that properties in (3.13), (3.14), and (3.15) are general; they follow from the properties of the mode-reduction strategy and energy-conserving choice of coupling.

To explain the symmetry breaking we consider a change of variables

$$\mathbf{y} = \mathbf{R}\mathbf{x},$$

where $\mathbf{R}$ is the rotation given in (3.14). This allows the system in (3.7) to be rewritten in a simple form with diagonal damping and forcing

$$d\mathbf{y} = \mu\left(1 - \alpha_0|\mathbf{y}|^2\right)\mathbf{y}dt - \lambda^2\tilde{A}\mathbf{y}dt + \lambda\tilde{\Sigma}d\tilde{\mathbf{W}}, \tag{3.16}$$

where $\tilde{A}$ and $\tilde{\Sigma}$ are diagonal matrices from (3.14), and $\tilde{\mathbf{W}} = R\mathbf{W}$ is a two-dimensional vector of independent Wiener processes. Using the rotation invariance property of Wiener processes and the fact that $|R| = 1$, we see that $\tilde{\mathbf{W}}$ is also a Wiener process. For further justification and proof, see [24]. Simultaneous diagonalization of both the damping term and the diffusion term is possible due to the properties of A and Σ in (3.14). Furthermore, taking into account the above mentioned fact that one eigenvalue of both, $\tilde{A}$ and $\tilde{\Sigma}$ is zero, the system in (3.16) can be rewritten to emphasize the one-dimensional structure of the stochastic perturbation

$$dy_1 = \mu\left(1 - \alpha_0|\mathbf{y}|^2\right)y_1 dt,$$

$$\tag{3.17}$$

$$dy_2 = \mu\left(1 - \alpha_0|\mathbf{y}|^2\right)y_2 dt - \lambda^2\tilde{a}y_2 dt + \lambda\tilde{\sigma}d\tilde{W}_2,$$

where $\tilde{a}$ and $\tilde{\sigma}$ are the non-zero eigenvalues of A and Σ, respectively. Since A and Σ have the same eigenvalue decomposition, in particular $\tilde{a} = \tilde{\sigma} = A_1^2 + B_1^2$. The system in (3.17) can be recast as a stochastic perturbation of a gradient system with a potential which is no longer rotation invariant. The system in (3.7) can also be

64

recast as a stochastic perturbation of a gradient system

$$d\mathbf{x} = -\nabla U dt + \lambda \Sigma d\mathbf{W}, \tag{3.18}$$

with the potential

$$U = -\frac{1}{2}\mu \left(1 - \frac{\alpha_0}{2}|\mathbf{x}|^2\right)|\mathbf{x}|^2 + \frac{A^{11}}{2}x_1^2 + \frac{A^{22}}{2}x_2^2 + A^{12}x_1x_2, \tag{3.19}$$

where A^{ij} are the ij^{th} entry of the damping matrix A. The minima of the potential
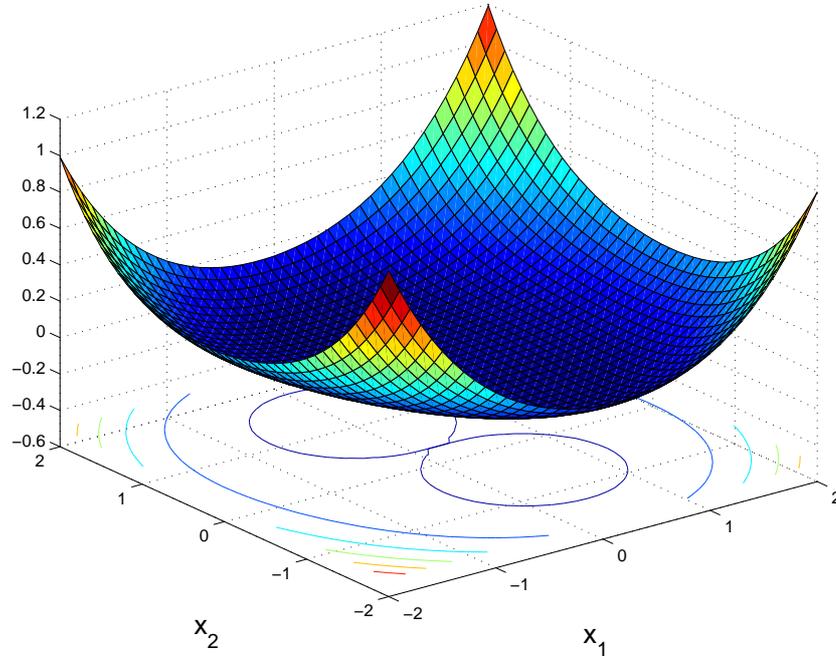
Figure 3.2: 3D Plot of function U given in (3.19) with parameters from (3.20)

in (3.19) coincide with the intersection of the circle $|\mathbf{x}|^2 = \alpha_0^{-1}$ and the neutral direction of the damping matrix, A. The direction for the zero eigenvalue is, in

65

turn, perpendicular to the direction of the damping. Therefore, coupling induces a symmetry breaking and a single preferred direction emerges from the interaction between the deterministic dynamics and stochastic terms.

3.2.2 Effect of Coupling on Gradient Systems

Stochastic mode reduction has been designed to reproduce statistical behavior of complex models with time-scale separation. The time-scale separation is controlled by the coupling strength, λ. For small values of λ, $0 \leq \lambda \leq 1$, behavior of $u_{1,2}$ in (3.5) is dominated by stochastic terms. This is the exactly the regime where the mode-elimination strategy works accurately. The system has clear scale separation and the variables, $x_{1,2}$ evolve slowly compared to $u_{1,2}$. To see this, we consider correlation functions which are computed as time averages; correlation function of $f(t)$ is given by $CF(\tau) = \left\langle (f(t) - \bar{f})(f(t + \tau) - \bar{f}) \right\rangle_t$, where $\bar{f}$ is the mean of f, and $\langle \cdot \rangle_t$ denotes temporal average. For our system, we consider correlation functions that are normalized by the variance, so that $NCF(0) = 1$, where $NCF(\tau) = \dfrac{CF(\tau)}{CF(0)}$. For $\lambda = 0.5$, the coupling is small and there is clear scale separation. This is visible in Figure 3.3, where $x_{1,2}$ evolve on a longer time-scale while the fast variables, $u_{1,2}$ evolve on a faster time-scale. In Figure 3.4 we justify that the mode-elimination strategy works and the slow variables $x_{1,2}$ in the full coupled system and reduced gradient system have normalized correlation functions that have good agreement.
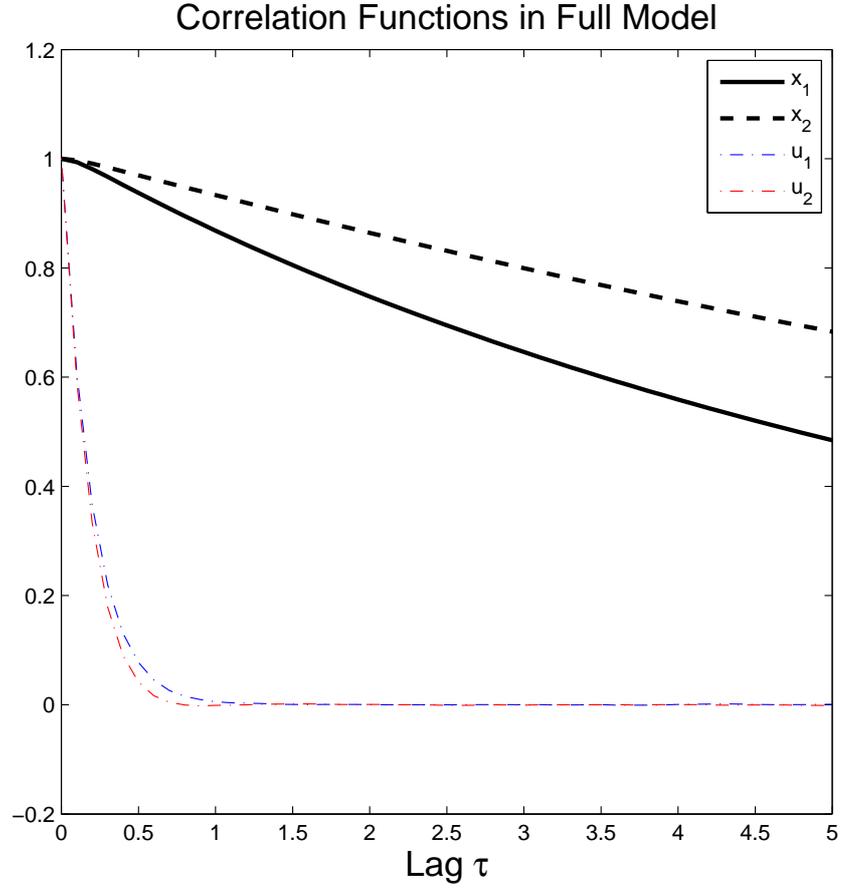
Figure 3.3: Comparison of Normalized Correlation Functions of x_1 (solid line), x_2 (dashed line) with correlation functions of u_1, u_2 (dash-dot line) for simulations of the coupled system in (3.5) with parameters in (3.20)

The parameters used in the simulations are

$$\mu = 0.1, \quad \alpha_0 = 0.8, \quad \lambda = 0.5,$$

$$A_{1,2,3} = -2, \ -2.5, \ 4.5, \quad B_{1,2,3} = -0.5, \ -0.5, \ 1, \tag{3.20}$$

$$\gamma_1 = \gamma_2 \equiv \gamma = 5, \quad \sigma_1 = \sigma_2 \equiv \sigma = 3.1622.$$
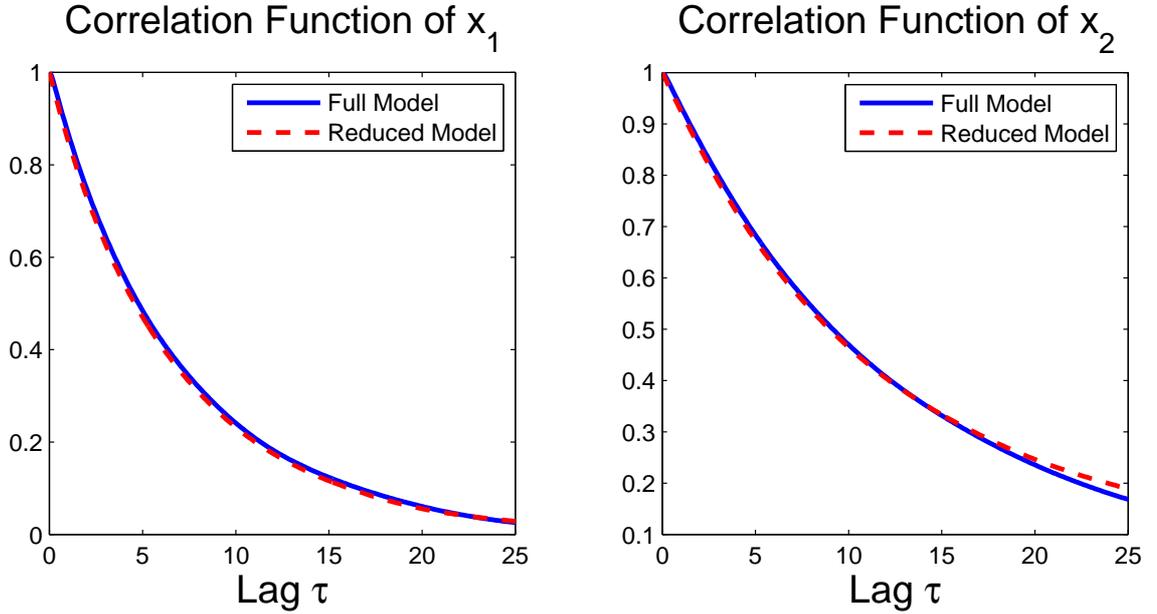
Figure 3.4: Comparison of Normalized Correlation Functions of x_1 and x_2 in the simulations of the full coupled system in (3.5) (solid line) in the regime (3.20) and the corresponding reduced equation in (3.7) (dashed line).

Since the mode-elimination strategy reproduces the statistical properties of the slow modes in the full system using the reduced system, other statistical quantities, in particular, probability density function (pdf) of x_1 and x_2 are also computed. With this in mind, marginal density functions were calculated for the slow variables and compared. In Figure 3.5, the close agreement between the pdf of the slow variables in the full and corresponding reduced system is evident. Further simulations for the case where $\lambda = 2$ were also performed. This value of λ is an upper bound for our case where the scale separation still exists and the mode reduction works well. The figure for the case where $\lambda = 2$ is not presented here due to the monotony of the presentation. To verify the mode-reduction strategy, other statistical quantities of the slow variables
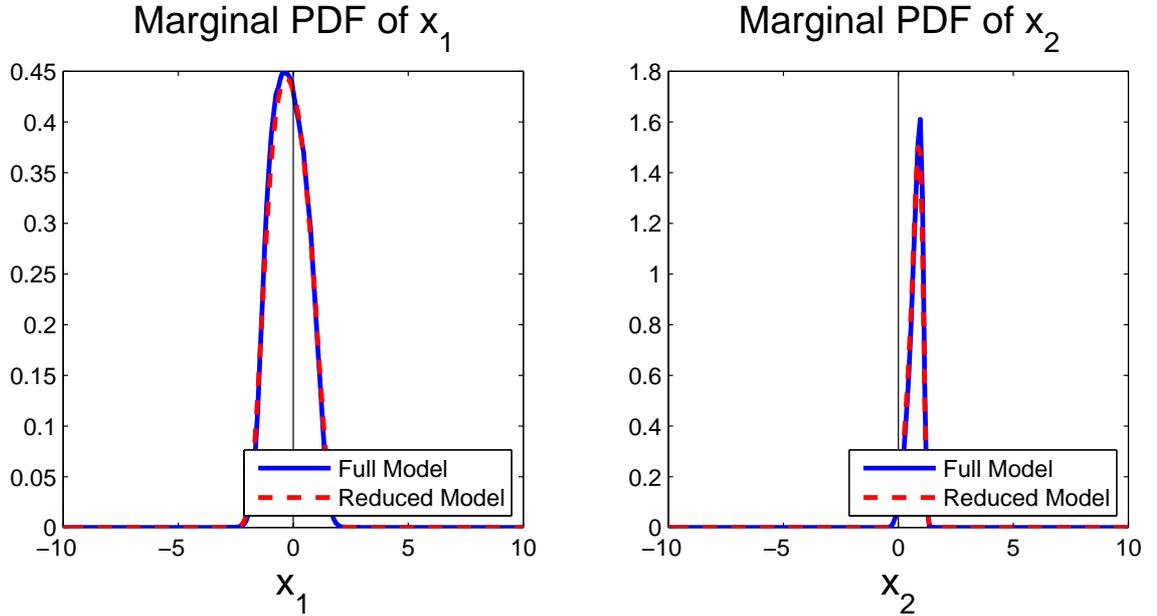
Figure 3.5: Comparison of marginal probability density functions of x_1 and x_2 of the full coupled system in (3.5) (solid line) in the regime (3.20) and the corresponding reduced equation in (3.7) (dashed line).

$x_{1,2}$ were computed. The agreement of the slow dynamics were clearly evident. Joint probability distribution (jpdf) of the slow variables were also considered. The choice of jpdf is due to its effectiveness in explaining the symmetry breaking phenomena. We will illustrate the idea first using low coupling strength where λ lies in the range of $[0, \ldots, 2]$. In Figure 3.6, jpdf of the slow variables are presented. In particular, we used $\lambda = 0.5$ to show the joint distribution of the slow modes in the full coupled system and the mode-reduced system. Initial conditions in both simulations were chosen above the eigendirection corresponding to the non-zero eigenvalue. It is important to note that if the initial conditions are chosen below the eigenvector corresponding to the non-zero eigenvalue, we will not see the same behavior. In particular, we will
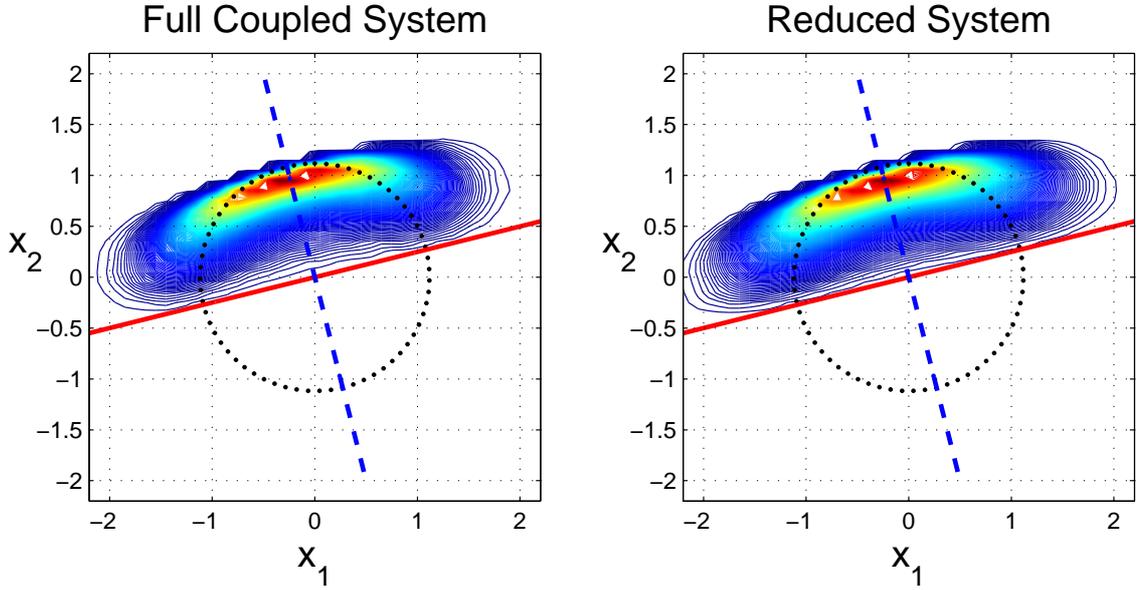
Figure 3.6: Contour plot of the joint probability density function of x_1, x_2 for the coupled system in (3.5) (left) and the reduced model in (3.7) (right) with parameters in (3.20). The eigenvectors of the matrix A in (3.8) are shown in blue dashed (zero eigenvector) and red solid (non-zero eigenvector). Circle $x_1^2 + x_2^2 = \alpha_0^{-1}$ is plotted as dotted line.

see a similar jpdf below the eigendirection associated with zero eigenvalue. We would like to comment that the reduced system is not ergodic, since trajectories cannot cross the line corresponding to the non-zero eigenvector. Therefore, a symmetric part with respect to the direction of the non-zero eigenvalue would also emerge in the joint distribution of x_1, x_2 for ensemble simulations of both systems. Notice that in our case, the joint distribution lies strictly above the solid line. Furthermore, there is very good agreement in the contour plots of the joint distribution of x_1, x_2. Hence the mode-elimination strategy worked well under the set of parameters given in (3.20). The orientation of the joint distribution is almost parallel to the eigenvector

for the non-zero eigenvalue. This is due to the strength of noise moving parallel to the eigendirection of the non-zero eigenvalue. Hence it is possible to predict the orientation of the joint distribution using the eigenvector of the damping matrix, A. As discussed before, we can also use the diffusion matrix, Σ, to reach a similar conclusion.

In Figure 3.6, notice that the intersection of the circle of equilibrium $x_1^2 + x_2^2 = \alpha_0^{-1}$ with the eigenvector associated with the zero eigenvalue lies exactly at the local basin of attraction. This local minimum of the potential in (3.19) corresponds to peak in the joint probability function. Hence, the position of the peak can be predicted apriori as the point of intersection of the stable circle of equilibria with the eigendirection of the zero eigenvalue.

The fact that we can use the damping matrix in the reduced system as a tool to predict the joint distribution of x_1 and x_2 in the full system is quite remarkable. Though this seems trivial, the consequences of this result are far reaching. For example, systems with many degrees of freedom that have significantly small number of slow modes as compared to the fast modes can be reduced and simply using the damping/drift matrix we can predict the preferred direction of the joint distribution as well as the local basin of attraction.

We will now explain the same phenomena, using different set of parameters. In particular, the effect of coupling strength on the full system given in (3.5) and its effective model in (3.7) will be considered. An increase in $\gamma_{1,2}$ and $\sigma_{1,2}$ in system (3.5) is the regime where the mode-reduction strategy works accurately. The applicability of the result in a regime where the mode-reduction fails is of significant interest. With this in mind, parameters are chosen in a regime with no scale separation between

the slow and the fast modes. The strength of coupling λ and its effect on the joint probability distribution of x_1, x_2 in the full and the mode-reduced model are considered. Under conditions where mode reduction fails, the joint probability distribution is affected tremendously. For this reason, λ is increased from 0.5 in (3.20) to 5. Now since $\lambda \geq 2$, the system is dominated by λ on the right side of (3.5) when $\gamma_{1,2}$ and $\sigma_{1,2}$ are fixed. Hence $u_{1,2}$ are dominated by λ, the case where there is no scale separation. Parameters used in the simulations are

$$\mu = 0.1, \quad \alpha_0 = 1.0, \quad \lambda = 5.0,$$
$$A_{1,2,3} = -2, \ -2.5, \ 4.5, \quad B_{1,2,3} = -0.5, \ -0.5, \ 1, \tag{3.21}$$
$$\gamma_1 = \gamma_2 \equiv \gamma = 2.5, \quad \sigma_1 = \sigma_2 \equiv \sigma = 2.2361.$$

Notice again that $\dfrac{\sigma^2}{2\gamma} = 1$. To demonstrate lack of scale separation, we draw the correlation functions of $x_{1,2}$ as well as $u_{1,2}$. Figure 3.7 (Left) shows clearly that there is no scale separation between the fast and slow modes. Notice how the slow variables $x_{1,2}$ evolve on almost the same time scale as the fast variables $u_{1,2}$. In Figure 3.7 (Right) the disagreement in the correlation function of x_1 in the coupled full model and the effective reduced model is shown; another justification that the mode-reduction strategy has failed. Notice the difference in Figure 3.7 and Figure 3.4. Similar behavior is manifested in the correlation function of x_2. The discrepancy is also visible in the joint probability distribution of $x_{1,2}$ in the full and the reduced model. Joint distribution of $x_{1,2}$ are drawn in Figure 3.8, where parameters are taken from (3.21). Irrespective of this disagreement, notice that the preferred direction of the joint distribution can be predicted using the eigendirection of the non-zero eigenvector of the drift matrix A. As mentioned before, in Figure 3.8 (Left) it is
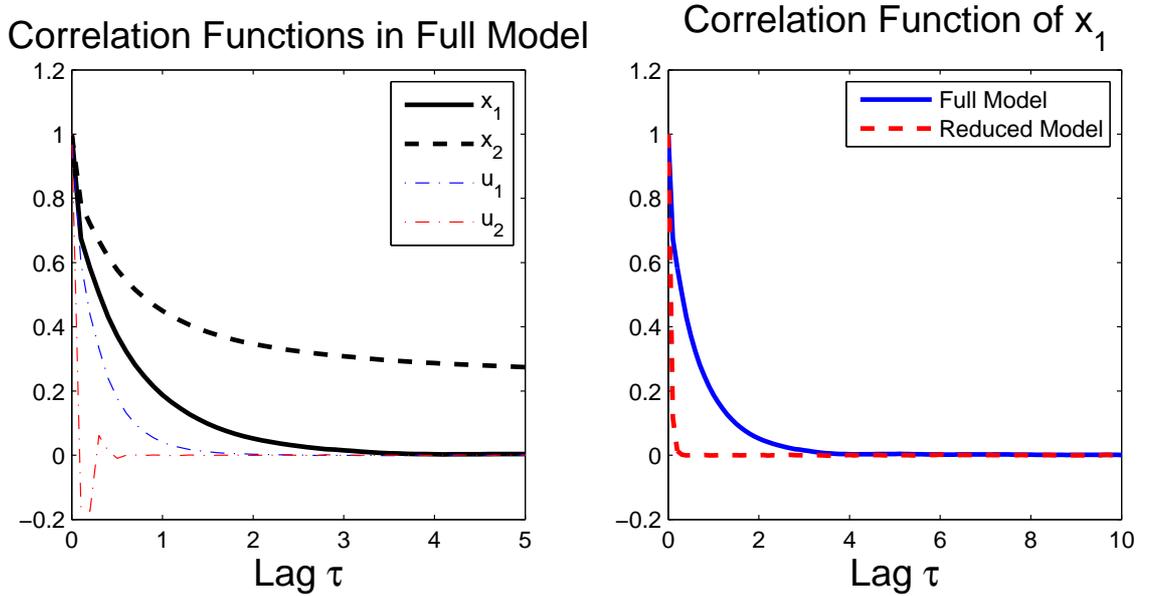
Figure 3.7: Left: Comparison of correlation functions of x_1 (solid line), x_2 (dashed line) with correlation functions of u_1, u_2 (dash-dot line) for simulations of the coupled system in (3.5) with parameters in (3.21). Right: Comparison of the Correlation Function of x_1 in the simulations of the full coupled system in (3.5) (solid line) in the regime (3.21) and the corresponding reduced equation in (3.7) (dashed line).

visible that the joint distribution lies above the eigenvector associated with the non-zero eigenvalue of matrix A, but for the reduced system such results are not possible since stochastic mode-reduction strategy does not work. Hence, it is possible to predict apriori the orientation of the joint distribution of the slow dynamics using the mode-reduced system. Therefore, the reduced system still retains some information about the full system in the essential modes.

Figure 3.8: Contour plot of the joint probability density function of x_1, x_2 for the coupled system in (3.5) (left) and the reduced model in (3.7) (right) with parameters in (3.21). The eigenvectors of the matrix A in (3.8) are shown in blue dashed (zero eigenvector) and red solid (non-zero eigenvector).

3.3 Low-Dimensional Systems with Stable Periodic Orbit

Although systems with stable periodic orbit cannot be recast as gradient systems, the argument from Section 3.2 regarding the modification of the potential can be extended to this case as well. To illustrate the symmetry breaking of systems with

74

stable periodic orbit, we consider the following low-dimensional system

$$
\begin{aligned}
\dot{x}_1 &= \mu(1 - \alpha_0|\mathbf{x}|^2)x_1 - (\alpha + \beta|\mathbf{x}|^2)x_2 + \lambda A_1 u_1 u_2, \\
\dot{x}_2 &= \mu(1 - \alpha_0|\mathbf{x}|^2)x_2 + (\alpha + \beta|\mathbf{x}|^2)x_1 + \lambda B_1 u_1 u_2, \\
\dot{u}_1 &= \lambda A_2 x_1 u_2 + \lambda B_2 x_2 u_2 - \gamma_1 u_1 + \sigma_1 \dot{W}_1, \\
\dot{u}_2 &= \lambda A_3 x_1 u_1 + \lambda B_3 x_2 u_1 - \gamma_2 u_2 + \sigma_2 \dot{W}_2.
\end{aligned}
\tag{3.22}
$$

Notice that the system (3.23) is very similar to the one considered in (3.5) except for the terms $-(\alpha+\beta|\mathbf{x}|^2)x_2$ and $(\alpha+\beta|\mathbf{x}|^2)x_1$ on the right side. The additional terms on the right side add rotation into the system, hence at times the system given in (3.23) will be referred to as the "system with rotation". For fixed values of α_0, increase in α and β increases the effect of rotation in the system. Clearly, for $\alpha = \beta \equiv 0$ the system in (3.23) is the same as (3.5). It is easy to show that the system in (3.23) projected onto x_1, x_2,

$$
\begin{aligned}
\dot{x}_1 &= \mu(1 - \alpha_0|\mathbf{x}|^2)x_1 - (\alpha + \beta|\mathbf{x}|^2)x_2, \\
\dot{x}_2 &= \mu(1 - \alpha_0|\mathbf{x}|^2)x_2 + (\alpha + \beta|\mathbf{x}|^2)x_1,
\end{aligned}
\tag{3.23}
$$

possesses a stable periodic orbit

$$
\mathbf{x}(t) = \alpha_0^{-1/2}(\cos \omega t, \sin \omega t), \quad \text{with } \omega = \alpha + \beta \alpha_0^{-1}.
\tag{3.24}
$$

Derivation of the reduced system is similar to the case of the gradient system in Section 3.3.1. A general outline of the derivation of reduced system is given in Appendix

75

B. The reduced system for (3.23) is given by

$$\dot{x}_1 = \mu(1 - \alpha_0|\mathbf{x}|^2)x_1 - (\alpha + \beta|\mathbf{x}|^2)x_2 - A^{11}x_1 - A^{12}x_2 + \Sigma_1\dot{W}_1,$$

$$(3.25)$$

$$\dot{x}_2 = \mu(1 - \alpha_0|\mathbf{x}|^2)x_2 + (\alpha + \beta|\mathbf{x}|^2)x_1 - A^{12}x_1 - A^{22}x_2 + \Sigma_2\dot{W}_2,$$

where A^{ij} are entries of the damping matrix in (3.8). Therefore, the effective damping is identical to the case of the gradient system and can be diagonalized by the rotation transformation in (3.15). Similar to the example discussed in the previous section, the diffusion matrix can also be diagonalized at the same time.

3.3.1 Symmetry Breaking in System with Stable Periodic Orbit

Although the system in (3.24) cannot be recast as a gradient system, parts of the right-hand side of the coupled system and the reduced equations are identical to the gradient systems in (3.5) and (3.7), respectively. Therefore, we expect that the peaks in the joint density of x_1, x_2 will occur near the intersection of the circle $x_1^2 + x_2^2 = \alpha_0^{-1}$ and the eigenvector corresponding to the zero eigenvalue of the damping matrix. This is confirmed by numerical simulations with surprising accuracy. Joint probability density for the coupled system and the reduced system is depicted in Figure 3.9. Parameters in the simulation were chosen to be similar to the parameters in Section
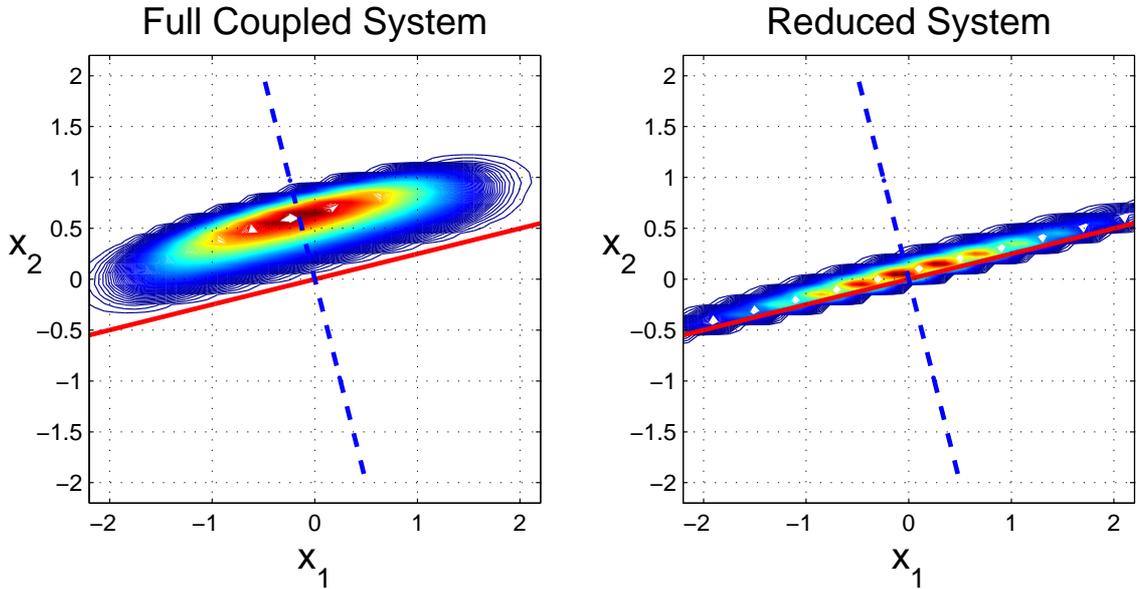
Figure 3.9: Joint probability density of x_1, x_2 for the coupled system in (3.23) (left) and the reduced model in (3.25) (right) with parameters in (3.26). The eigenvectors of the matrix A in (3.8) are shown in blue dashed (zero eigenvector) and red solid (non-zero eigenvector). Circle $x_1^2 + x_2^2 = \alpha_0^{-1}$ is plotted as dotted line.

3.2

$$\mu = 0.1, \quad \alpha_0 = 0.8, \quad \lambda = 0.5, \quad \alpha = 0.06, \quad \beta = 0.05,$$

$$A_{1,2,3} = -2, \; -1.5, \; 3.5, \quad B_{1,2,3} = -1.25, \; -1.2, \; 2.45, \tag{3.26}$$

$$\gamma = 5, \quad \sigma = 3.1622.$$

The stochastic mode-reduction was designed to reproduce the statistical features of complex models with separation of time-scales. For the coupling strength $\lambda = 0.5$, the

	$E\{x_1\}$	$E\{x_2\}$	$Var\{x_1\}$	$Var\{x_2\}$	$Kurt\{x_1\}$	$Kurt\{x_2\}$
Coupled Model	-0.00224	0.00339	0.57894	0.63812	2.1703	2.0556
Effective Model	-0.00482	0.01155	0.58613	0.65066	2.2636	2.1143

Table 3.1: One-point statistics of x_1 and x_2 for the coupled system in (3.23) and reduced system in (3.25) in the regime with $\lambda = 0.5$; $Kurt\{y\} = \langle (y-\bar{y})^4 \rangle / \langle (y-\bar{y})^2 \rangle^2$.
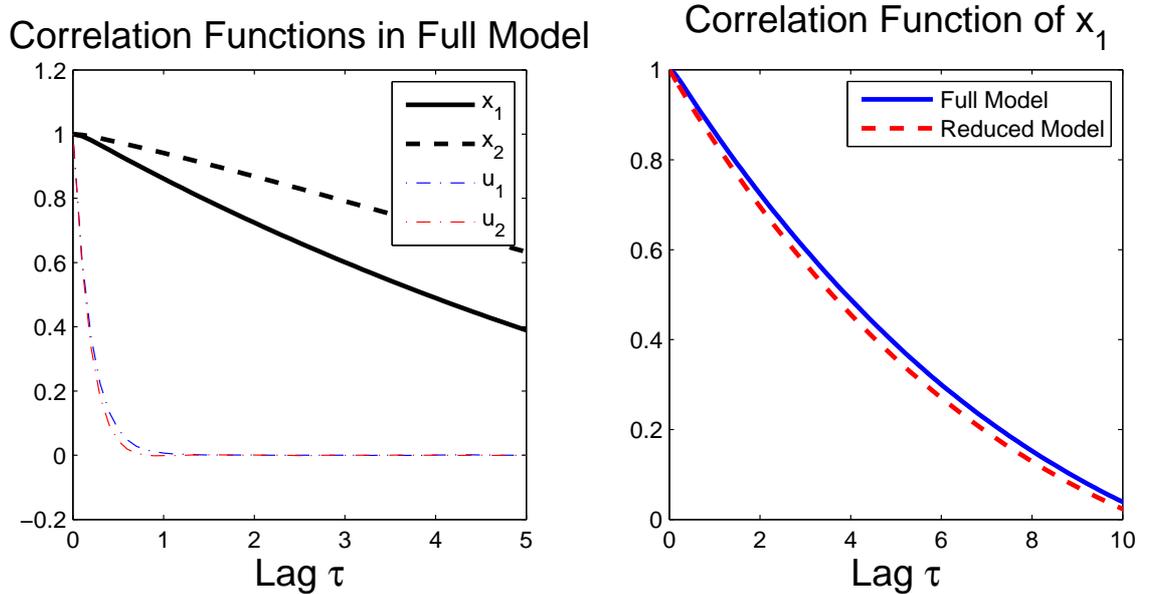
Figure 3.10: Left: Comparison of Normalized Correlation Functions of x_1 (solid line), x_2 (dashed line) with correlation functions of u_1, u_2 (dash-dot line) for simulations of the coupled system in (3.23) with parameters in (3.26). Right: Comparison of Normalized Correlation Functions of x_1 in the simulations of the full coupled system in (3.23) (solid line) in the regime (3.26) and the corresponding reduced equation in (3.25) (dashed line).

stochastic mode-reduction strategy is utilized in the correct regime and reproduces statistical features of x_1 and x_2 extremely well. Normalized correlation functions of $x_{1,2}$ and $u_{1,2}$ depicted in Figure 3.10 (Left) demonstrate that the time-scale of $x_{1,2}$ is much slower than for $u_{1,2}$. Statistical agreement of the coupled and reduced model is presented in Table 3.1 and correlation function of x_1 for two models is depicted in Figure 3.10 (Right). Correlation function of x_2 also agree between the system with rotation given in (3.23) and the effective reduced system in (3.25). The results are not shown here due to redundancy and are very similar to gradient system. From Section 3.2.2, an increase in coupling strength leads to a regime where the mode-reduction
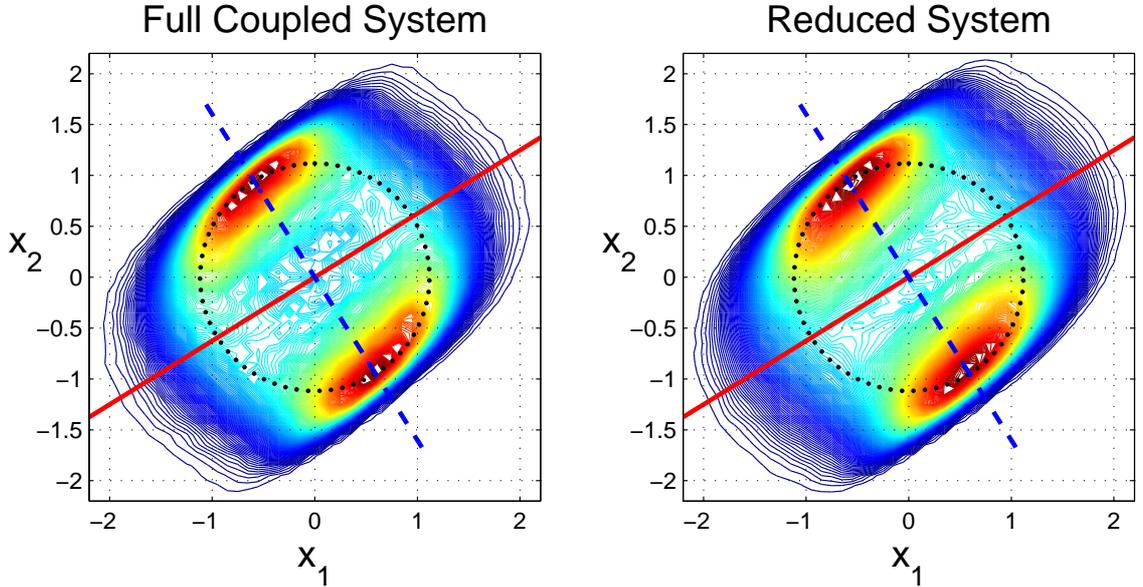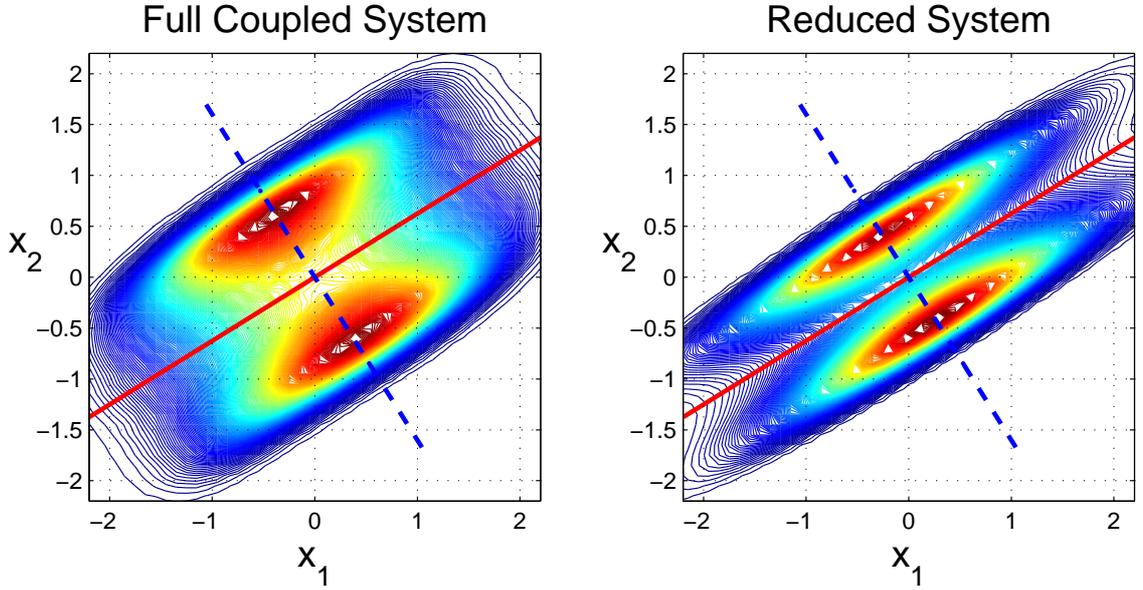
Figure 3.11: Joint probability density function of x_1, x_2 for the coupled system in (3.23) (left) and the reduced model in (3.25) (right) with parameters in (3.26) where the coupling strength $\lambda = 10$. The eigenvectors of the matrix A in (3.8) are shown in blue dashed (zero eigenvector) and red solid (non-zero eigenvector).

fails. Hence, the statistical agreement between the full and effective reduced system is weak and the stochastic mode-reduction strategy is not supposed to be utilized to explain the statistical behavior of x_1 and x_2. This was confirmed by choosing coupling strength $\lambda = 10$ and observing a large discrepancy between the correlation functions of x_1 and x_2 in the coupled and reduced models.

For large values of λ, there is no scale separation between $x_{1,2}$ and $u_{1,2}$. This behavior was verified by checking the correlation functions of $x_{1,2}$ and $u_{1,2}$. It was demonstrated that x_1 and x_2 cannot be treated as slow variables since initial decay rates of correlation functions for x_1 and u_1 are comparable. Therefore, $\lambda = 10$ is the regime which is not appropriate for the application of the stochastic mode-reduction strategy. Nevertheless, as in the case of gradient systems the reduced model in (3.25)

still can be utilized to explain the symmetry breaking phenomena qualitatively. Joint probability density for x_1 and x_2 is presented in Figure 3.11. Clearly, there is a large discrepancy between the joint distribution of these variables, but the stochastic mode-reduction predicts the location of the peaks in the joint distribution with good accuracy. This example demonstrates that in some cases the predictive power of the stochastic mode-reduction extends beyond regimes with scale-separation.

3.3.2 Effect of Rotation on Symmetry Breaking

As mentioned before, α and β in system (3.23) affect the rotation frequency of the system. Figure 3.9 (left) has peaks in the left upper half of the plane and in the right bottom half of the plane. These peaks in the joint distribution of x_1 and x_2 correspond to high probability regions in the system. An increase in the rotational terms, pushes the peaks along the circle $x_1^2 + x_2^2 = \alpha_0^{-1}$. Notice that an increase in α and β for fixed values of λ, α_0, μ, moves the peak in quadrant II down the circular path while the peak in quadrant IV is pushed up. This can be seen easily by looking at the terms $-(\alpha + \beta|\mathbf{x}|^2)x_2$ and $(\alpha + \beta|\mathbf{x}|^2)x_1$ in the right side of (3.23). For example, in quadrant II, x_1 is negative while x_2 is positive. For $\alpha \geq 0$ and $\beta \geq 0$, since $x_1 \leq 0$ and $x_2 \geq 0$ in quadrant II, $-(\alpha + \beta|\mathbf{x}|^2)x_2$ forces x_1 to become negative along the circle $x_1^2 + x_2^2 = \alpha_0^{-1}$ and $(\alpha + \beta|\mathbf{x}|^2)x_1$ forces x_2 to be negative as well. The same idea applies to quadrant IV and both x_1 and x_2 are forced to become positive and hence move anti-clockwise along the circle of equilibria $x_1^2 + x_2^2 = \alpha_0^{-1}$. The other terms on the right side of (3.23) counter these forces and the forces are constantly competing. If μ is comparatively greater than λ, α and β, then the first term will dominate the dynamics of the system. We would like to observe the behavior of the joint

probability distribution of x_1 and x_2 when rotation is increased for a fixed coupling strength. Notice that not only the regime where the mode-reduction strategy works is of interest but also when mode-reduction fails. Particularly, the effect of rotation on symmetry breaking and the apriori prediction of the peaks are of interest.

The values of α and β are positive in our simulations. It might be of interest to consider other possible combination of values for α and β. We would like to mention that all the other possible cases can be easily mimicked just by using $\alpha > 0$ and $\beta > 0$. Consider the stable periodic orbit

$$\mathbf{x}(t) = \alpha_0^{-1/2} \left(\cos \omega t, \sin \omega t \right), \quad \text{with } \omega = \alpha + \beta \alpha_0^{-1},$$

where ω depends on α an β. Hence, we consider considering only positive values of α and β. Furthermore, we only need to increase the α or β parameter in (3.23) to see the increase in rotation. There is no significant difference in increasing both the parameters as compared to increasing one parameter at a time, α or β. Hence, to study the rotational effect on symmetry, we consider possible values of α, where α increases from 0.06 to 0.3 while other parameters are fixed. The results are shown in Figure 3.12. Hence the conclusions of the previous section are valid for a wide range of parameters. In particular, the direction of symmetry breaking is related to the eigenvectors of the damping matrix and rotation terms only stretch the regions of high probability.
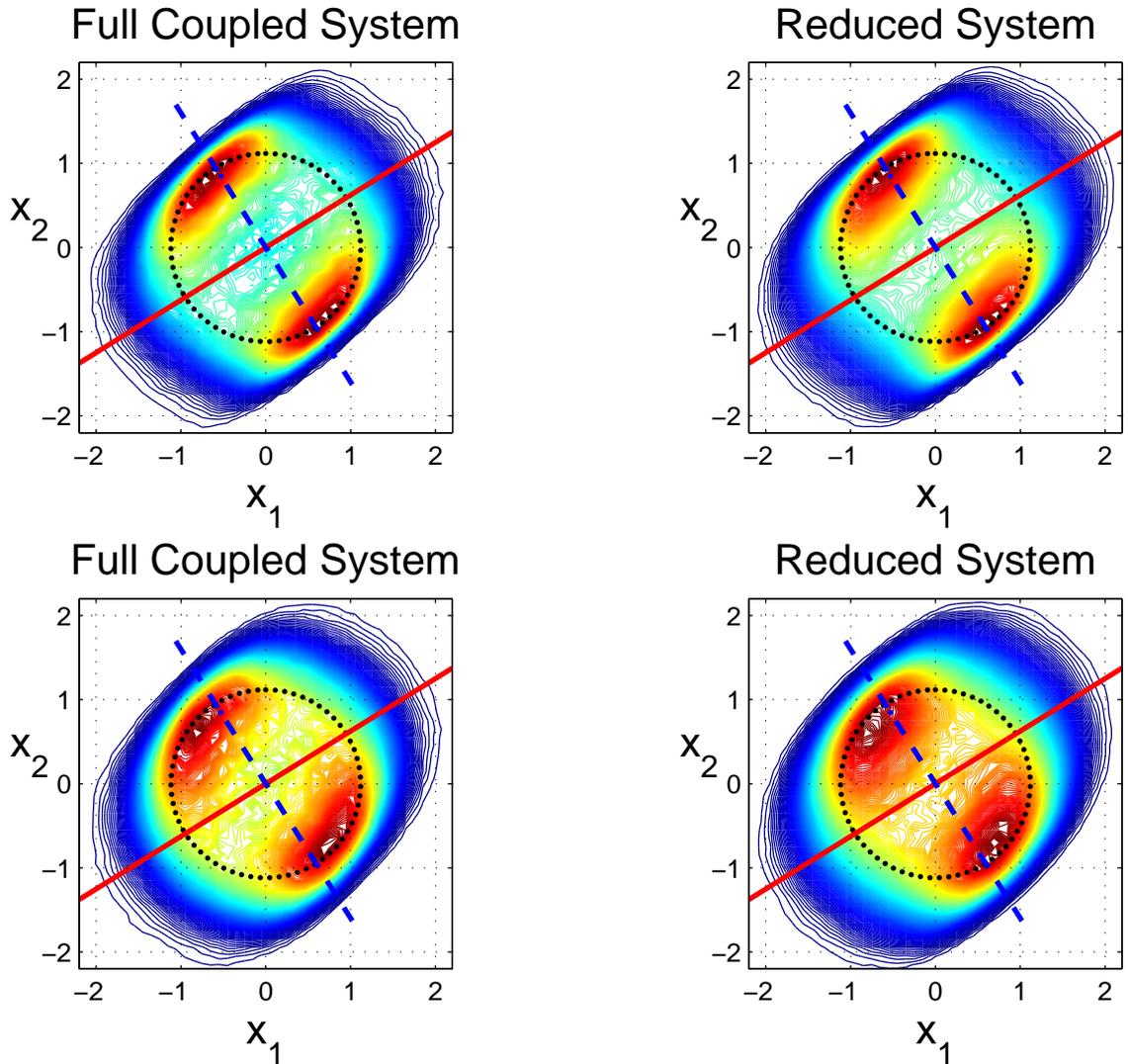
Figure 3.12: Top: Joint probability density function of x_1, x_2 for the coupled system in (3.23) (left) and the reduced model in (3.25) (right) with parameters in (3.26) except for the change in β to 0.1. The eigenvectors of the matrix A in (3.8) are shown in blue dashed (zero eigenvector) and red solid (non-zero eigenvector). Bottom: Joint probability density function of x_1, x_2 for the coupled system in (3.23) (left) and the reduced model in (3.25) (right) with parameters in (3.26) except for the change in β to 0.3. The eigenvectors of the matrix A in (3.8) are shown in blue dashed (zero eigenvector) and red solid (non-zero eigenvector).

3.4   Conclusion

Two coupled systems are presented here as prototype examples elucidating interactions between various deterministic coherent structures and noise. Coupled systems are constructed by coupling a low-dimensional dynamical system with a particular structure with additional variables. It is demonstrated that the coupling breaks the original symmetry of the underlying dynamics and a preferred direction emerges in the two-dimensional projection of the full higher-dimensional distribution. Although not all low-dimensional projections considered here can be recast as gradient systems, rotational symmetry of the gradient part of the right-hand side plays a crucial role in describing this symmetry breaking mechanism. The gradient system considered here, has a stable circle of equilibria and can be used to show analytically the symmetry breaking phenomena. Furthermore, it is demonstrated that coupling terms can be successfully replaced by stochastic terms utilizing the stochastic mode-reduction technique. Under assumptions of ergodicity, the stochastic mode reduction gives closed-form stochastic models for the slow variables in the limit of infinite separation of timescales. The stochastic models reproduce exactly the statistical behavior of the slow modes, when the fast modes are artificially accelerated to enforce the separation of timescales.

Symmetry breaking can be explained easily in the context of stochastic reduced models since the preferred direction is directly linked to the eigenvectors of the stochastic perturbation. Since the diffusion and damping matrices are similar, eigenvectors of the two matrices are the same. Hence either the eigenvectors of the damping or the diffusion matrix can be used to explain this symmetry breaking. Moreover, these eigenvectors can be computed apriori, without running any computer simu-

lations, thus, giving a powerful insight into the statistical behavior of the coupled problem. In addition, it is also demonstrated that the stochastic mode-reduction strategy can predict the preferred direction for systems without scale separation and, thus, far outside regimes of the intended applicability of this technique. Therefore, the stochastic mode-reduction acts as an effective linearization of the full coupled dynamics where the coupling terms are effectively replaced by stochastic terms. We expect that the applicability of the stochastic mode-reduction strategy to problems with symmetry breaking can be extended to more complex systems without scale separation.

Chapter 4

Predictability in Full and Reduced Systems

4.1 Introduction

Stochastic mode reduction in the regime of scale separation yields reduced systems which have good statistical agreement with the full dynamics. Statistical quantities, such as in mean, variance, higher moments, correlation functions, etc of the two systems have almost exactly identical behavior. Moreover, certain statistical properties of full systems can be inferred from the reduced system even in the regime without the scale-separation. Hence, reduced system contains significant information about the full system when the mode reduction technique is not effective. This was detailed in Chapter 3, where the reduced model was used to correctly predict the orientation of the joint pdf of x_1, x_2 in the full model. In this chapter, we investigate the non-equilibrium behavior of reduced systems. In particular, we analyze the agreement in

relative entropy of the original and the mode eliminated system. Furthermore, the conditions necessary under which such agreement occurs will also be established. We choose parameters where the discrepancy in the decay rate of relative entropy between systems occur on a longer time scale as well as on a shorter time scale, nevertheless the systems will behave similarly. A wide range of parameters need to be considered since we have to verify that the results are not due to happenstance. We show that the behavior is not due to specific values but is a general occurrence for a large set of parameters.

4.2   Measure of Difference in Decay Rates

Relative entropy for most system decays exponential in time. This is not true for oscillatory models. In this chapter, we consider systems with decay rate "close" to an exponential function. The word "close" is considered in the least square sense. In particular, we rely on the idea that decay rate of relative entropy can approximated by exponential functions. In our case, the exponential function as a measure of decay rates of relative entropy works well. The decay rate is calculated using log of the exponential fit function. Formally, given $(\mathbf{Y}, \mathbf{X})$ and an exponential function $\mathbf{g}$, we determine the unknown parameters $\boldsymbol{\beta}$ such that

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{e}, \tag{4.1}$$

where $\mathbf{X}$ is the time discretization t_i for $i = 1, .., N$, $\mathbf{Y}$ is $R(t_i)$ and $\mathbf{e}$ is the error in the approximation. In general, $\mathbf{X}$ are the independent variables while $\mathbf{Y}$ are the

dependent variables. In our application (4.1) becomes

$$Y_i = \beta_0 e^{\beta_1 X_i} + e_i, \tag{4.2}$$

where the residual $e_i = Y_i - \widehat{Y_i}$ and $\widehat{Y_i}$ is the estimate of Y_i. Method of ordinary least squares is used to minimize the function

$$RSE = \sum_{i=1}^{N} e_i^2. \tag{4.3}$$

The exponential fit function is denoted by $\widehat{Y_i}$, hence

$$\log(\widehat{Y_i}) = \log(\beta_0) + \beta_1 X_i. \tag{4.4}$$

Decay of the exponential fit function is determined from the quantity β_1. It is the slope of the log of the exponential function fitted to $R(t_i)$. We use the logarithm of the absolute error of the quantity β_1 to see the comparison in the full and reduced system. If β_{f1} is β_1 for the full system and β_{r1} is β_1 for the reduced system, then the absolute error $\delta\beta_1$

$$\delta\beta_1 = |\beta_{f1} - \beta_{r1}|, \tag{4.5}$$

$\log(\delta\beta_1)$ will be the measure that we will use to compare the difference in decay rates of the full and reduced system for $R(t)$.

Generally, the full system has far more variables as compared to its reduced version (see Appendix B). Hence, it is particularly important to determine and compare the correct set of modes. The modes that are particularly suited for this comparison

are the slow variables. Therefore, relative entropy of the full system is the marginal relative entropy of the slow modes. Consider $(\mathbf{x}, \mathbf{u}) = (slow, fast)$, the slow-fast modes in the full system. Relative entropy of the slow modes in the full system projected onto the $\mathbf{x}$-plane is denoted as $Rf_{\mathbf{x}}(t)$, while relative entropy of the reduced system is denoted as $Rr(t)$.

Under the assumption that the slow variables decorrelate much slowly than the fast variables, the stochastic-mode reduction strategy can be applied and the correlation function and the marginal/joint probability distributions of the full and the reduced system have a near perfect agreement in the slow variables, $\boldsymbol{x}$. Under no scale separation, the difference in the correlation function of x_1 was evident from Figure 3.7. In this chapter, we show that the difference in the correlation functions of the slow variables for the full and reduced system is almost the same difference that we will see in the relative entropies. Hence, establishing a clear connection between the notion of predictability and correlation function. A simplified intuitive definition of correlation relation is how fast the system forgets the initial conditions. Hence, as the system reaches equilibrium state the correlation function decays to zero and the system has no knowledge of the initial conditions. The same explanation can be used for the notion of predictability. Once the system is in equilibrium state, the predictive strength of the system is lost as a whole. Hence one cannot predict with any confidence where the system will be at a future point in time. Therefore, correlation function can somehow be used to explain the predictive nature of the system. These ideas will first be applied to a triad model and its mode reduced analogue, the Ornstein-Uhlenbeck process. This prototype model is very useful, since the OU process has explicit closed forms for correlation function and relative entropy. The ideas

are further applied to the low-dimensional system considered in Section 3.3. Notice that the system studied in (3.5) and the reduced system (3.7) were not studied since the system in (3.5) is not ergodic and hence the results are not applicable.

4.3 Comparison of Correlation Function and Relative Entropy

4.3.1 Triad System and Ornstein-Uhlenbeck Process

We investigate the triad model and the mode reduced Ornstein-Uhlenbeck (OU) process. The triad model (i.e. Full System) is

$$
\begin{aligned}
\dot{x} &= A_1 u_1 u_2, \\
\dot{u}_1 &= A_2 x u_2 - \gamma_1 u_1 + \sigma_1 \dot{W}_1, \\
\dot{u}_2 &= A_3 x u_1 - \gamma_2 u_2 + \sigma_2 \dot{W}_2,
\end{aligned}
\tag{4.6}
$$

where W_1, W_2 are independent Wiener processes, A_j for $j = 1, 2, 3$ are known interaction coefficients and γ_i and σ_i, $i = 1, 2$ are strength of drift and diffusion terms respectively. Here we consider the energy-conserving interactions between the slow variable x and fast variables $\{u_1, u_2\}$ by using the interaction coefficients

$$
A_1 + A_2 + A_3 = 0.
$$

Under the assumption that x decorrelates slowly than the fast modes $\{u_1, u_2\}$, the stochastic mode-reduction strategy (see Appendix B) can be applied to the system

in (4.6) and the reduced model for x becomes

$$\dot{x} = -\gamma x + \sigma \dot{W}, \tag{4.7}$$

where

$$\gamma = \frac{-1}{\gamma_1 + \gamma_2} \left(A_1 A_2 \frac{\sigma_2^2}{2\gamma_2} + A_1 A_3 \frac{\sigma_1^2}{2\gamma_1} \right),$$

$$\sigma = A_1 \frac{\sigma_1 \sigma_2}{\sqrt{2\gamma_1 \gamma_2}} \frac{1}{\sqrt{\gamma_1 + \gamma_2}}, \tag{4.8}$$

where W is the Weiner process. Since the system in (4.7) is a simple, explicitly representable process, the time correlation function is

$$\langle x(t+\tau), x(t) \rangle_t = \frac{\sigma^2}{2\gamma} e^{-\gamma\tau}, \tag{4.9}$$

where $\tau \geq 0$, $\langle \cdot, \cdot \rangle_t$ is averaging w.r.t t and γ, σ are defined explicitly in (4.8). Correlation function can be calculated using a single realization and averaging over two points separated by a distance τ in time.

Relative entropy requires Monte-Carlo simulation with an ensemble of initial conditions, as discussed in Section 1.4. Given Gaussian distributed initial conditions, $x(t)$ in (4.7) has mean and variance

$$\langle x(t) \rangle = \langle x(0) \rangle e^{-\gamma t},$$

$$var\{x(t)\} = [var\{x(0)\} - \frac{\sigma^2}{2\gamma}]e^{-2\gamma t} + \frac{\sigma^2}{2\gamma}, \tag{4.10}$$

90

where $\langle x(0) \rangle$ is the mean and $var\{x(0)\}$ is the variance of the initial ensemble. From (4.10), the limiting/stationary ($t \to \infty$) distribution is normally distributed with mean 0 and variance $\dfrac{\sigma^2}{2\gamma}$. From [31], when the equilibrium distribution q and prediction distribution p have standard form of Gaussian distribution [24] of finite dimension n, then the closed analytical expression for Relative Entropy, R is

$$R = \frac{1}{2}\left\{ \ln\left[\frac{det(\sigma_q^2)}{det(\sigma_p^2)}\right] + tr[\sigma_p^2(\sigma_q^2)^{-1}] + (\vec{\mu_p} - \vec{\mu_q})^T(\sigma_q^2)^{-1}(\vec{\mu_p} - \vec{\mu_q}) - n \right\}, \quad (4.11)$$

where det is determinant and tr is trace. For the OU process, $\vec{\mu_q}$ and $\vec{\mu_p}$ are vector-valued means and σ_q^2 and σ_p^2 are variance-covariance matrix. The equilibrium mean and equilibrium variance are $\vec{\mu_q}$ and $(\sigma_q^2)_{ii}$ $i = 1, 2$, respectively. Hence the relative entropy for the OU process can be calculated explicitly. For the correlation function and relative entropy of the triad model given in (4.6), Direct Numerical Simulations (DNS) are employed to achieve the results.

Explicit formula for relative entropy $R(t)$ under Gaussian initial conditions is known, hence we use an ensemble of normally distributed initial conditions. Different mean and variance for the set of initial conditions can have different relative entropy results. To circumvent the problem, we first calculate the stationary distribution of the two systems given in (4.6) and (4.7). Hence the equilibrium mean and variance are known for both systems for all the modes. For our first set of numerical simulations, the mean and variance of the initial ensemble is chosen to be the equilibrium mean and the equilibrium variance for the fast modes $\mathbf{u}$. For the slow mode x, we used the mean of the ensemble to be the equilibrium mean but the variance of the ensemble was 1/10 of the equilibrium variance.

Large variance of the ensemble will lead to little or no "spread" in the initial conditions, hence the decay rate of the relative entropy would not be significant. If the decay rate of relative entropy is not profound, full and reduced system will not have an accurate comparison. It is important to emphasize that the variance of the initial conditions in the slow mode x should be chosen "away" from the equilibrium variance. Hence the system will reach equilibrium on a very long time scale and the rate of decay in relative entropy would become significant. In the other set of simulations, the mean of the slow variables x will be further away from the equilibrium mean, however in this case the results will remain unaffected. For this set of simulations, the mean and variance of the initial ensemble of the fast modes were kept as before. Different set of simulations with different means of initial ensemble is to understand if the mean of the ensemble has any influence on the non-equilibrium behavior of the model.

4.3.1.1   Mean of the Ensemble as Equilibrium Mean

Parameters used in the simulations are

$$
\begin{aligned}
A_{1,2,3} &= -2, \ -1.5, \ 3.5, \\
\gamma &= \gamma_1 = \gamma_2, \quad \sigma = \sigma_1 = \sigma_2.
\end{aligned}
\tag{4.12}
$$

To simplify the presentation we consider γ and σ such that

$$
\frac{\sigma^2}{2\gamma} = 1,
\tag{4.13}
$$

but the argument still applies when this simplification is not used. In Figures 4.1, 4.2 and 4.3 we show a comparison of the correlation function of x and the relative

Figure 4.1: Correlation function of x (left) and relative entropy (right) of the triad model in (4.6) and the mode reduced Ornstein-Uhlenbeck process in (4.7) with parameters in (4.12). $\gamma = 1$ (top) and $\gamma = 2.5$ (bottom) with the simplification given in (4.13). Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution is the equilibrium mean.

entropy of the full triad model and the reduced OU Process for different values of γ. The result shows that for increasing value of γ, mode-reduction works and the

Figure 4.2: Correlation function of x (left) and relative entropy (right) of the triad model in (4.6) and the mode reduced Ornstein-Uhlenbeck process in (4.7) with parameters in (4.12). $\gamma = 5.0$ (top) and $\gamma = 7.5$ (bottom) with the simplification given in (4.13). Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution is the equilibrium mean.

relative entropy as well as the correlation function have a good agreement in the full and the reduced system. Furthermore, it is evident that the difference in the relative
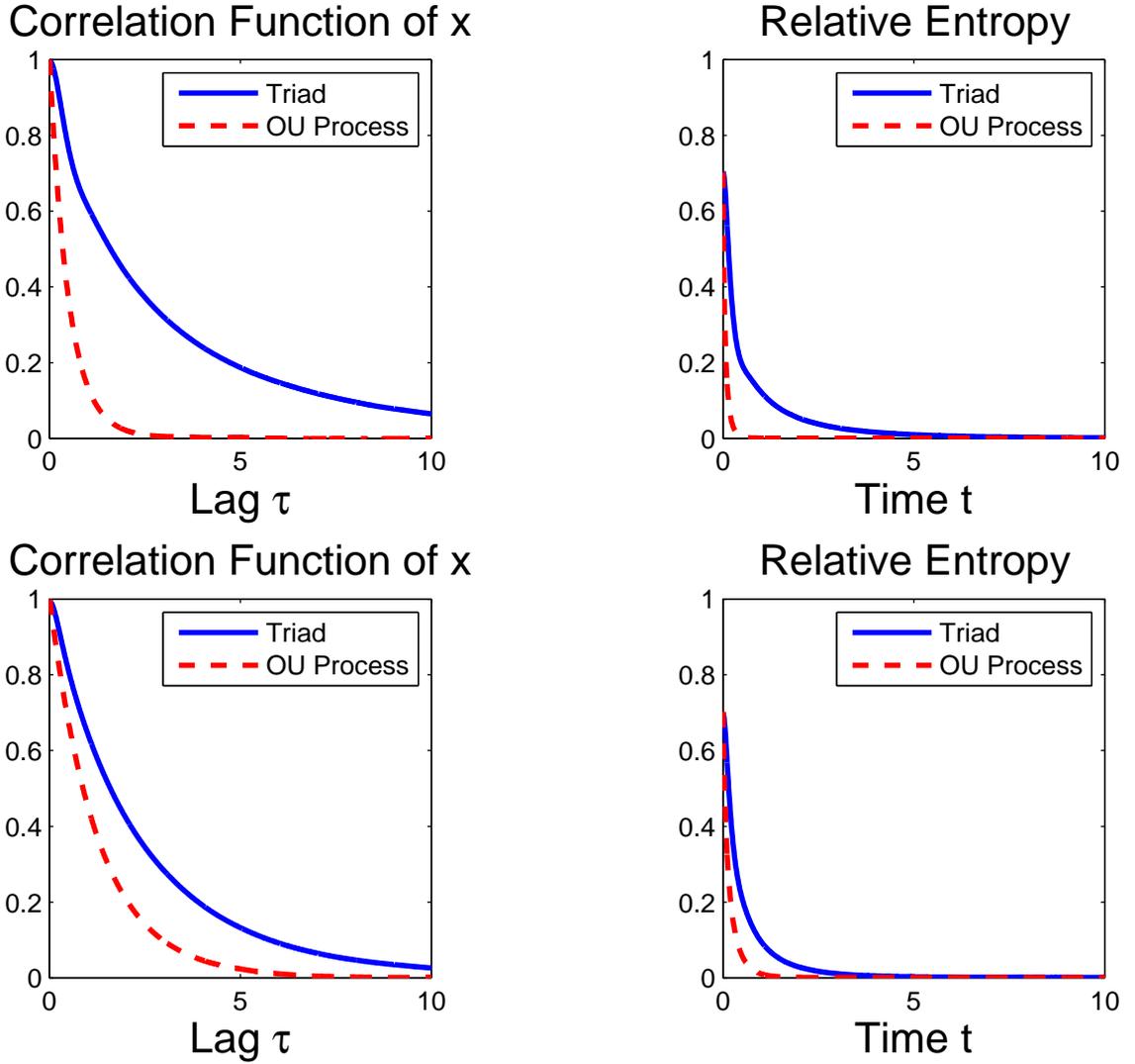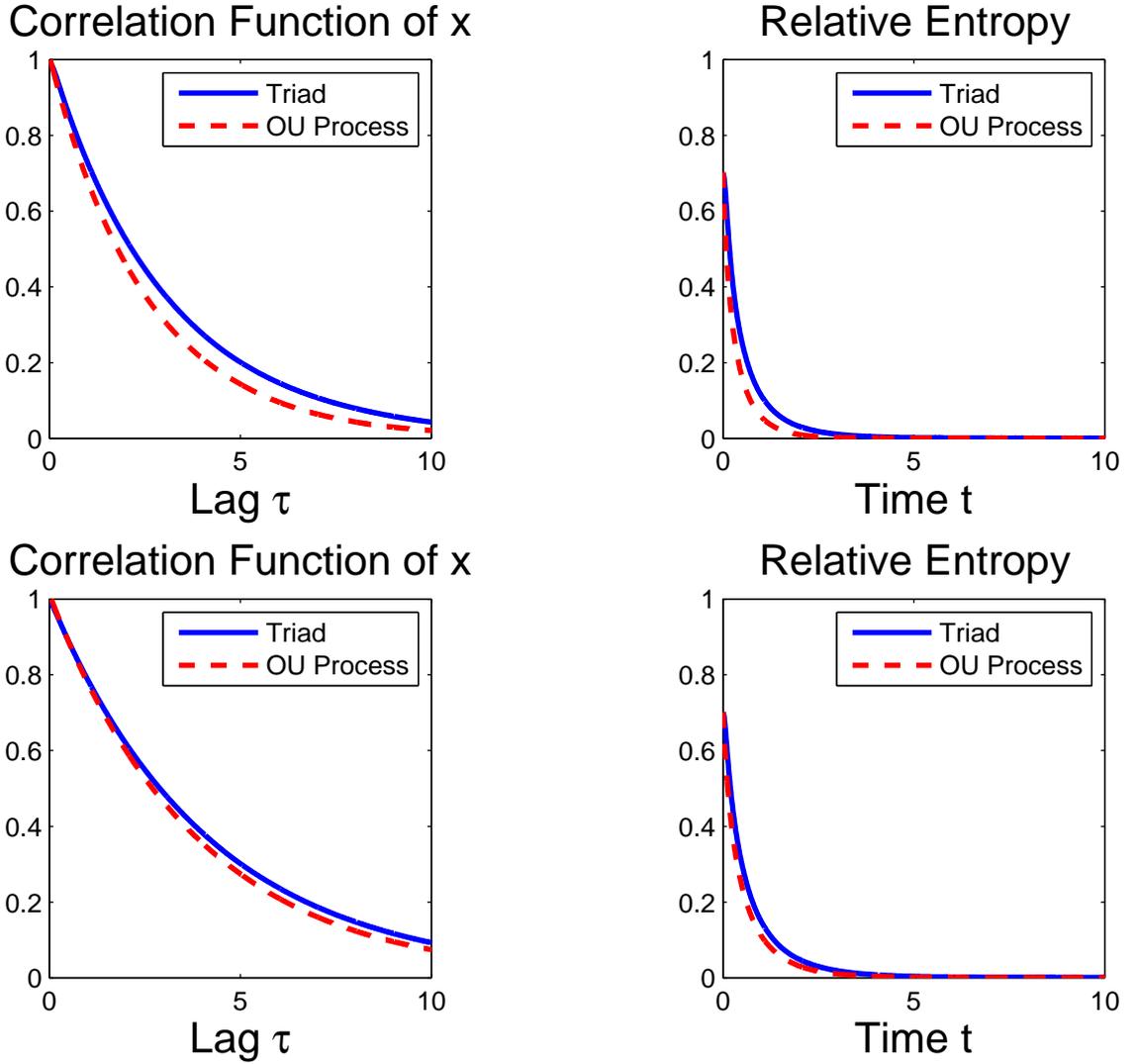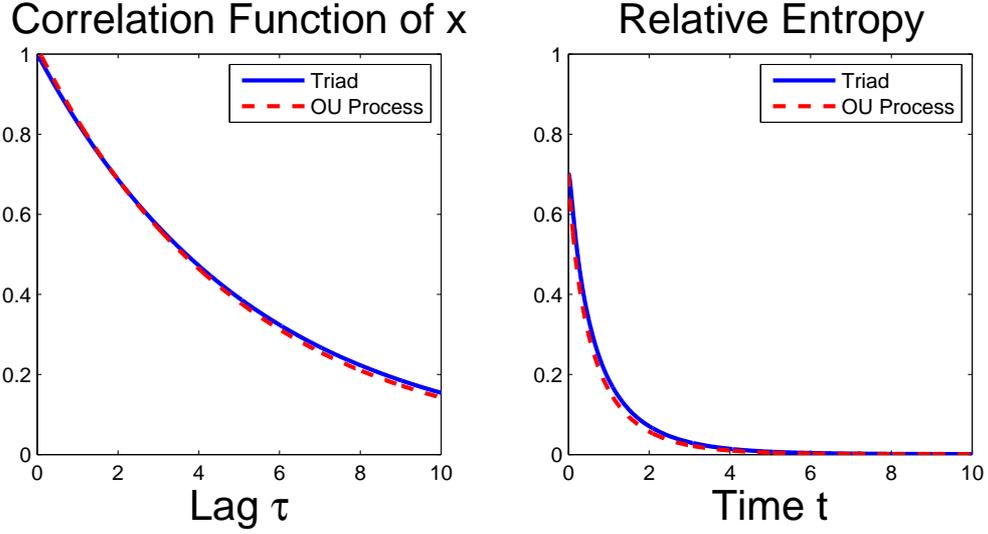
Figure 4.3: Correlation function of x (left) and relative entropy (right) of the triad model in (4.6) and the mode reduced Ornstein-Uhlenbeck process in (4.7) with parameters in (4.12). $\gamma = 10$ with the simplification given in (4.13). Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution is the equilibrium mean.

entropy for the full and reduced system is almost the same difference that we see in the correlation function of x. For the relative entropy calculation, the ensemble of initial conditions for the fast variables was normally distributed with equilibrium mean and equilibrium variance. For the slow mode, the mean of the initial condition was the same as the equilibrium mean while the variance was one-tenth of the equilibrium variance.

Now using the least squares approach in (4.2), the value of β_1 for different values of γ is calculated. Notice that that there are four different values of β_1; β_1 for the relative entropy of the full and reduced system is β_{f1-re} and β_{r1-re} respectively, while β_1 for the correlation function of x in the full and reduced system is β_{f1-cf} and β_{r1-cf} respectively. Table 4.1 show these values for β_1 for different values of γ. It is not

obvious from the table whether the difference in decay rates are the same for the correlation function of x and the relative entropy, but using the measure in (4.5), $\delta\beta_1$, and calculating the logarithm of this quantity one can quantify the difference in decay rates. In Figure 4.4, value of $\delta\beta_1$ is calculated for different values of γ. In particular, values of γ in Table 4.1 were used. The graphs for the correlation function of x and the relative entropy both have the same decay rate. Also, the graphs are almost a straight line and hence have the same slope. The difference in the logarithm of the absolute error of the correlation function and the relative entropy is a non-zero constant. It is easy to see from the graph that in this particular case, the constant is 2. Next, we demonstrate that the mean of the initial condition of the slow variable, x has no bearing on the result.

4.3.1.2 Mean of the Ensemble Different from Equilibrium Mean

To show the validity of the results, different cases of the mean of the initial conditions of the slow variable were simulated. In particular, the mean was 0.5 unit, 1 unit and 2 units away from the equilibrium mean.

To show the consistency in the results, simulations were performed where the mean

Table 4.1: β_1 for correlation function of x (**cf**) and relative entropy (**re**) for the systems given in (4.6) and (4.7) with different values of γ.

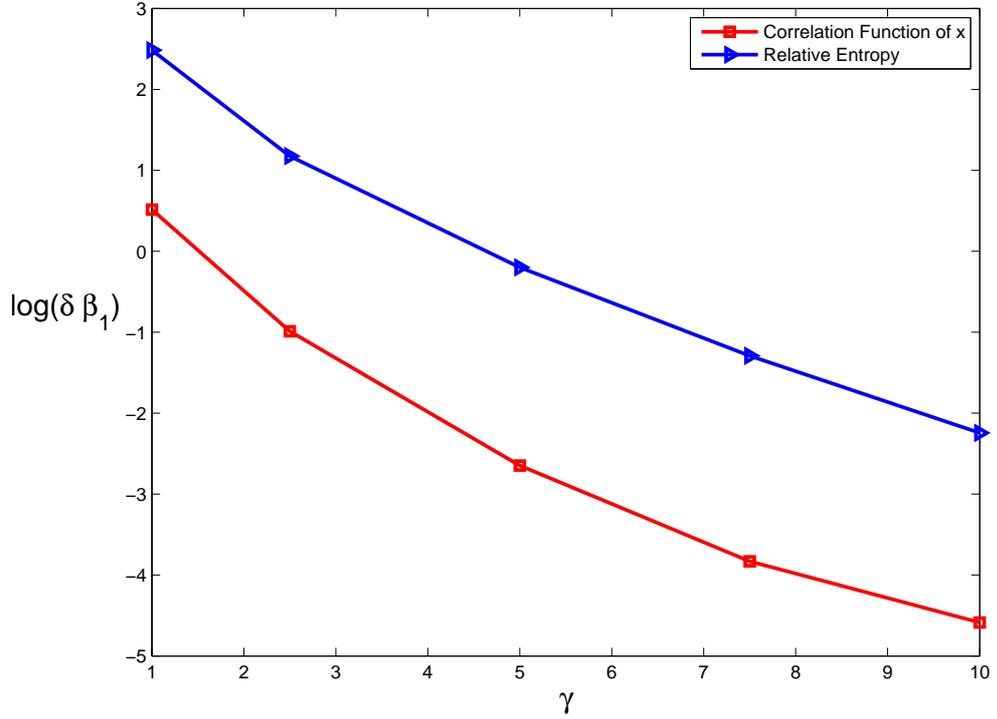|		Values of γ				
		1	**2.5**	**5**	**7.5**	**10**
Triad Model	β_{f1-re}	-1.77	-2.134	-1.849	-1.499	-1.223
	β_{f1-cf}	-0.3077	-0.4097	-0.3198	-0.2377	-0.1861
OU Process	β_{r1-re}	-13.76	-5.373	-2.668	-1.774	-1.329
	β_{r1-cf}	-1.982	-0.7816	-0.3907	-0.216	-0.1963

Figure 4.4: Logarithm of $\delta\beta_1$ for the correlation function of x (line with square) and relative entropy (line with right-pointing triangle). For the correlation function of x, logarithm of $|\beta_{f1-cf} - \beta_{r1-cf}|$ is calculated and for the relative entropy the logarithm of $|\beta_{f1-re} - \beta_{r1-re}|$ is calculated. The mean of the initial conditions of the slow variable x is the same as the equilibrium mean.

of initial conditions is the same for the fast variables, u, as in Figure 4.4 but the mean of the initial condition for the slow variable, x, is 2 units away from the equilibrium mean. We present results only for the case where the mean of the initial conditions is 2 units away the equilibrium mean. The results were valid for other values of the mean. Notice again in Figures 4.5, 4.6, and 4.7 that for larger values of γ the difference in the correlation function and relative entropy is almost the same. Furthermore, Figure 4.8 shows similar results as in Section 4.3.1.1. As a word of caution, notice that the difference in the logarithm of the absolute error of the correlation function and the

Figure 4.5: Correlation function of x (left) and relative entropy (right) of the triad model in (4.6) and the mode reduced Ornstein-Uhlenbeck process in (4.7) with parameters in (4.12). $\gamma = 1.0$ (top) and $\gamma = 2.5$ (bottom) with the simplification given in (4.13). Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution is 2 units away from the equilibrium mean.

relative entropy is still a non-zero constant but in this case the constant is different from the one in Section 4.3.1.1. Due to redundancy, the results where the mean of

Figure 4.6: Correlation function of x (left) and relative entropy (right) of the triad model in (4.6) and the mode reduced Ornstein-Uhlenbeck process in (4.7) with parameters in (4.12). $\gamma = 5.0$ (top) and $\gamma = 7.5$ (bottom) with the simplification given in (4.13). Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution is 2 units away from the equilibrium mean.

the IC of x is 0.5 unit and 1 unit away from the equilibrium mean are not presented here, but they were consistent with this presentation. Hence, mean of the ensemble
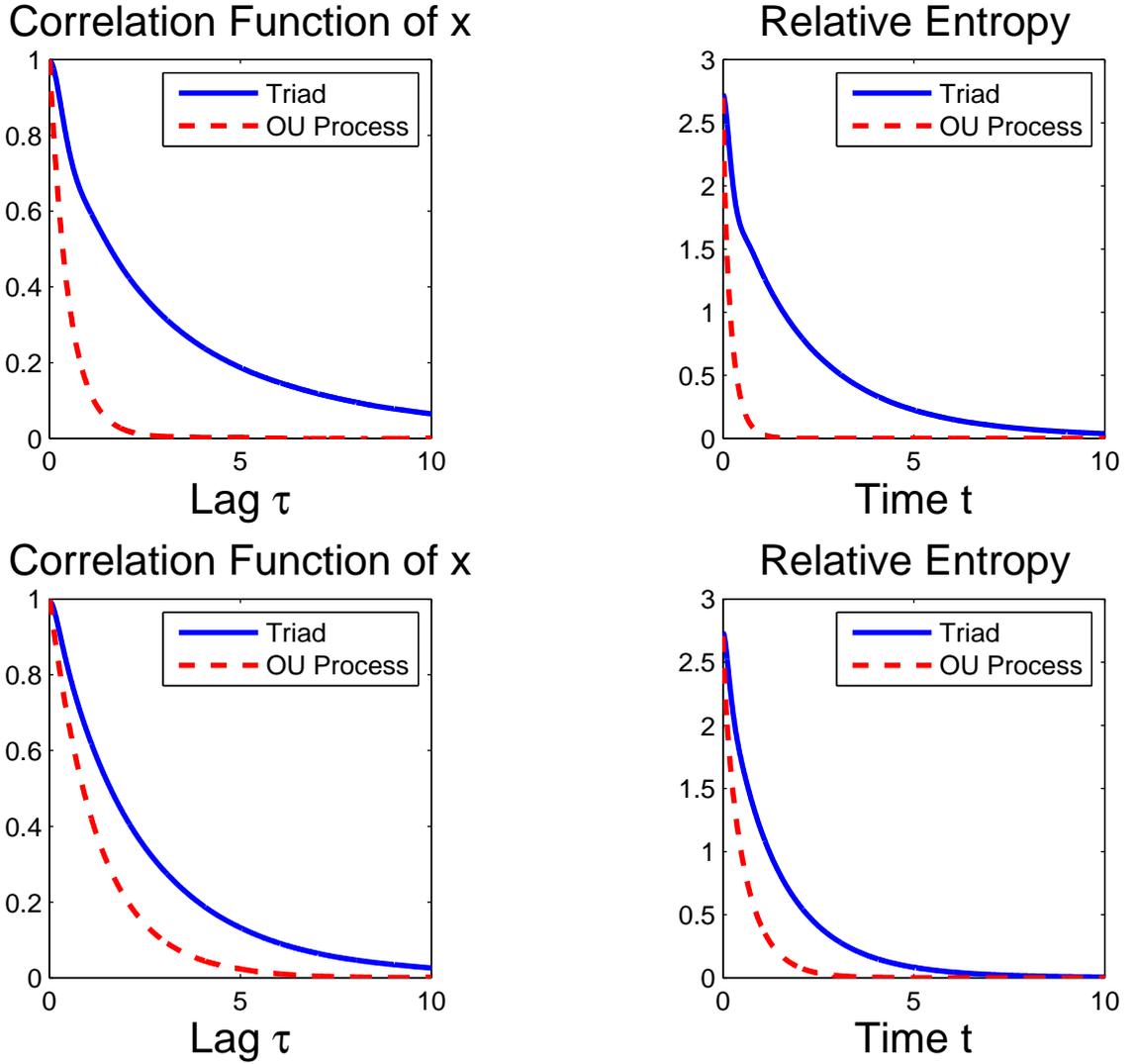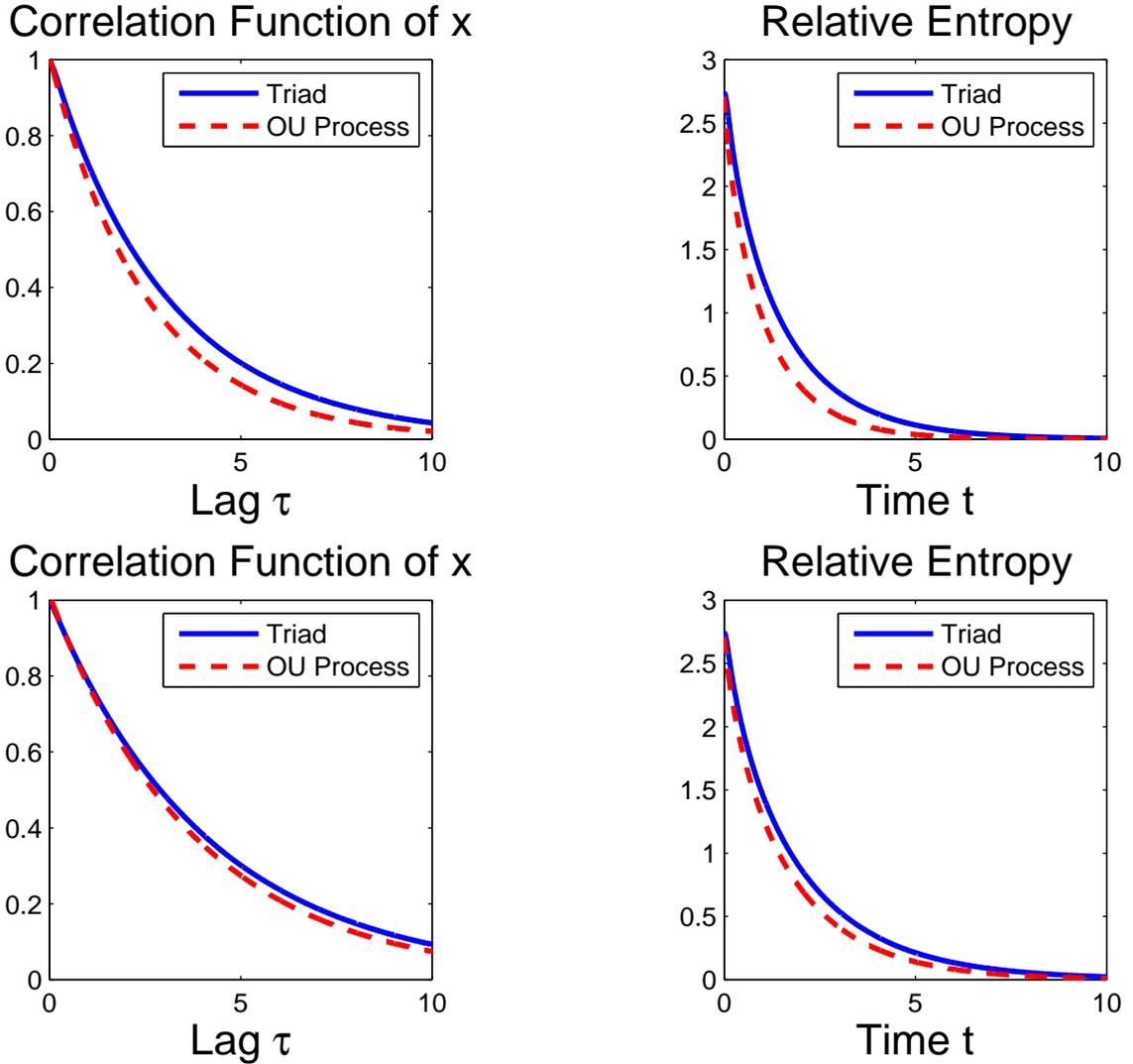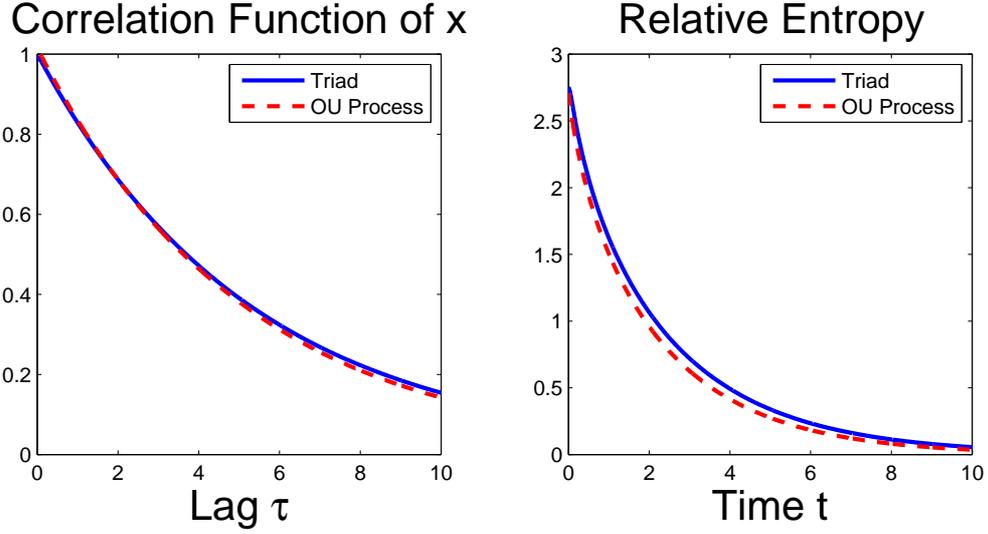
Figure 4.7: Correlation function of x (left) and relative entropy (right) of the triad model in (4.6) and the mode reduced Ornstein-Uhlenbeck process in (4.7) with parameters in (4.12). $\gamma = 10$ with the simplification given in (4.13). Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution is 2 units away from the equilibrium mean.

has no influence on the non-equilibrium behavior of the model.

4.3.2 System with Stable Periodic Orbit

We considered a low-dimensional system with a stable periodic orbit in (3.3.1). It is natural to ask, if results from Section 4.3.1 can still be valid on the system (3.23) and its mode-reduced analogue (3.24). Recall that the low-dimensional system with stable periodic orbit has a rotation term in the slow variables, $\boldsymbol{x} = (x_1, x_2)^T$. As discussed in Section 4.1, the rotation term is of great importance. The system without the rotation term is non-ergodic and the results from this section are not applicable for that particular system. Recall, the full system with stable periodic orbit has two slow variables, x_1 and x_2 and two fast variables u_1 and u_2, unlike the system in (4.6) where
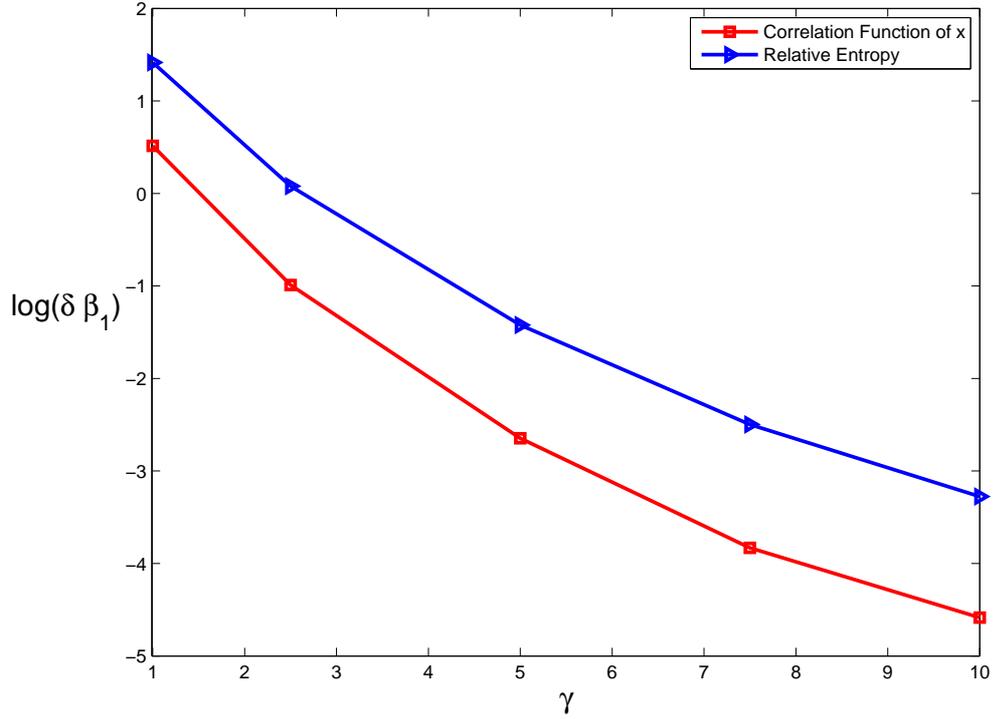
Figure 4.8: Logarithm of $\delta\beta_1$ for the correlation function of x (line with square) and relative entropy (line with right-pointing triangle). For the correlation function of x, logarithm of $|\beta_{f1-cf} - \beta_{r1-cf}|$ is calculated and for the relative entropy the logarithm of $|\beta_{f1-re} - \beta_{r1-re}|$ is calculated. The mean of the initial conditions of the slow variable x is 2 units away from the equilibrium mean.

there was only one slow variable. Hence, correlation function of the slow variables maybe either be for x_1 or x_2. The results are applicable in either case. In addition, the difference in the decay rates of correlation function in x_1 as well as in x_2 are comparable to the difference in decay rates of the relative entropy. The full system in this case is four-dimensional, but since slow variables are under consideration, relative entropy of the full system implies marginal relative entropy in the x_1, x_2 variables. The reduced system will only mimic the slow dynamics, and hence the relative entropy of the reduced system is just the full entropy of the system.

The Ornstein-Uhlenbeck Process in (4.7) is a system which has explicit expression for the correlation function of x, mean, variance, and in particular the relative entropy. As you can see from (4.9), (4.10), and (4.11), the correlation function and relative entropy of an OU Process can be approximated with great accuracy by an exponential function. Hence in the case for the triad and the OU process the results are extremely convincing. The system with the stable periodic orbit and its mode-reduced version, explicit expressions are not present. Furthermore, the correlation function of the slow variables and relative entropy are not necessarily exponential. Hence an approximation of the quantities by an exponential function will not necessarily yield very good results. Nevertheless, for high values of γ where the exponential function approximates very well, the result is still achievable. In Figure 4.9 the correlation function of x_1 and x_2 with relative entropy for $\gamma = 1$ is given. The exponential function is not a good estimate for the relative entropy as well as the correlation functions. Compared to Figures 4.10, 4.11, and 4.12 where the function can easily be approximated by an exponential function. As before, to calculate relative entropy, Monte-Carlo simulations were used where the initial ensemble has normal distribution. For the first set of simulations, the mean and variance of the initial ensemble in the slow variables x_1 and x_2 has equilibrium mean but the variance of the ensemble is one-tenth of the equilibrium variance. The reason for this choice is the same as discussed for the triad model in (4.6). The other sets of simulations were performed where the mean of the initial ensemble was 0.5 units, 1 unit, and 2 units away from the equilibrium mean. For the system with rotation, the mean of the initial ensemble for the slow variables was changed simultaneously for both x_1 and x_2. The result was

102
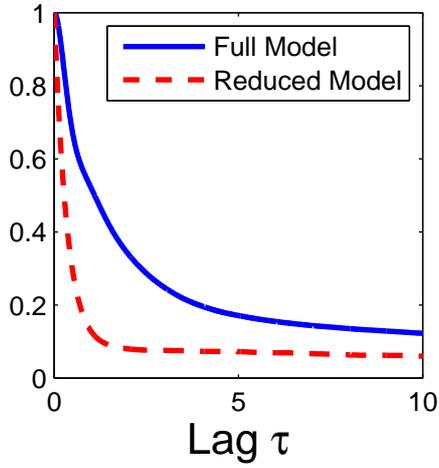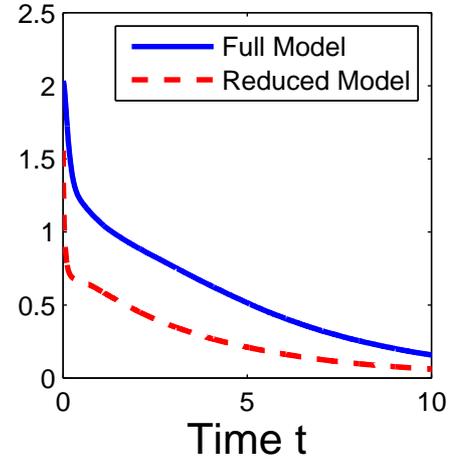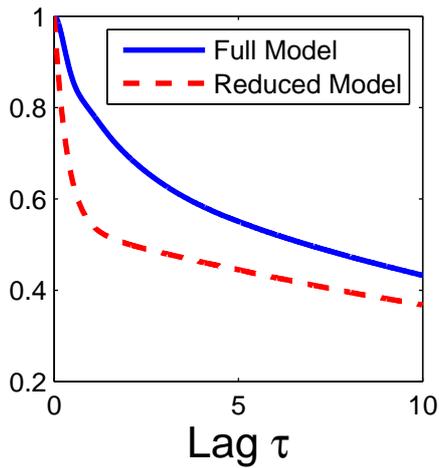
Figure 4.9: Correlation function of x_1 (top left) and x_2 (bottom left) for $\gamma = 1$. Relative entropy (top right and bottom right) for $\gamma = 1$. Solid line represents the full model in (3.23) and the dashed line is the reduced model in (3.24).

valid in this situation. Parameters used in the simulations

$$\mu = 0.1, \quad \alpha_0 = 0.8, \quad \lambda = 1, \quad \alpha = 0.06, \quad \beta = 0.05,$$

$$A_{1,2,3} = -2, \; -1.5, \; 3.5, \quad B_{1,2,3} = -1.25, \; -1.2, \; 2.45, \tag{4.14}$$

$$\gamma = \gamma_1 = \gamma_2, \quad \sigma = \sigma_1 = \sigma_2,$$

Figure 4.10: Correlation function of x_1 (top left) and x_2 (bottom left) for $\gamma = 1$. Relative entropy (top right and bottom right) for $\gamma = 5.0$. Solid line represents the full model in (3.23) and the dashed line is the reduced model in (3.24).

with the simplification that $\dfrac{\sigma^2}{2\gamma} = 1$. Notice that the coupling strength, λ, is set to 1 and hence coupling has no affect on the result. Also notice that increase in α and β for some fixed α_0 increases rotation in (3.23), but the result here are independent

Figure 4.11: Correlation function of x_1 (top left) and x_2 (bottom left) for $\gamma = 1$. Relative entropy (top right and bottom right) for $\gamma = 7.5$. Solid line represents the full model in (3.23) and the dashed line is the reduced model in (3.24).

of such changes. Furthermore, $A_{1,2,3}$ and $B_{1,2,3}$ were changed independently as well as simultaneously and there was no change. In Figure 4.13 the result is not valid for small values of γ since the best fit exponential function has high approximation
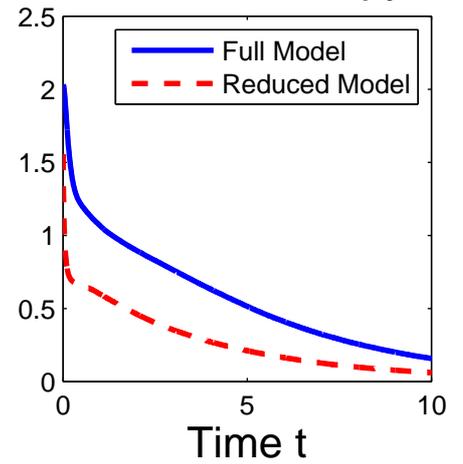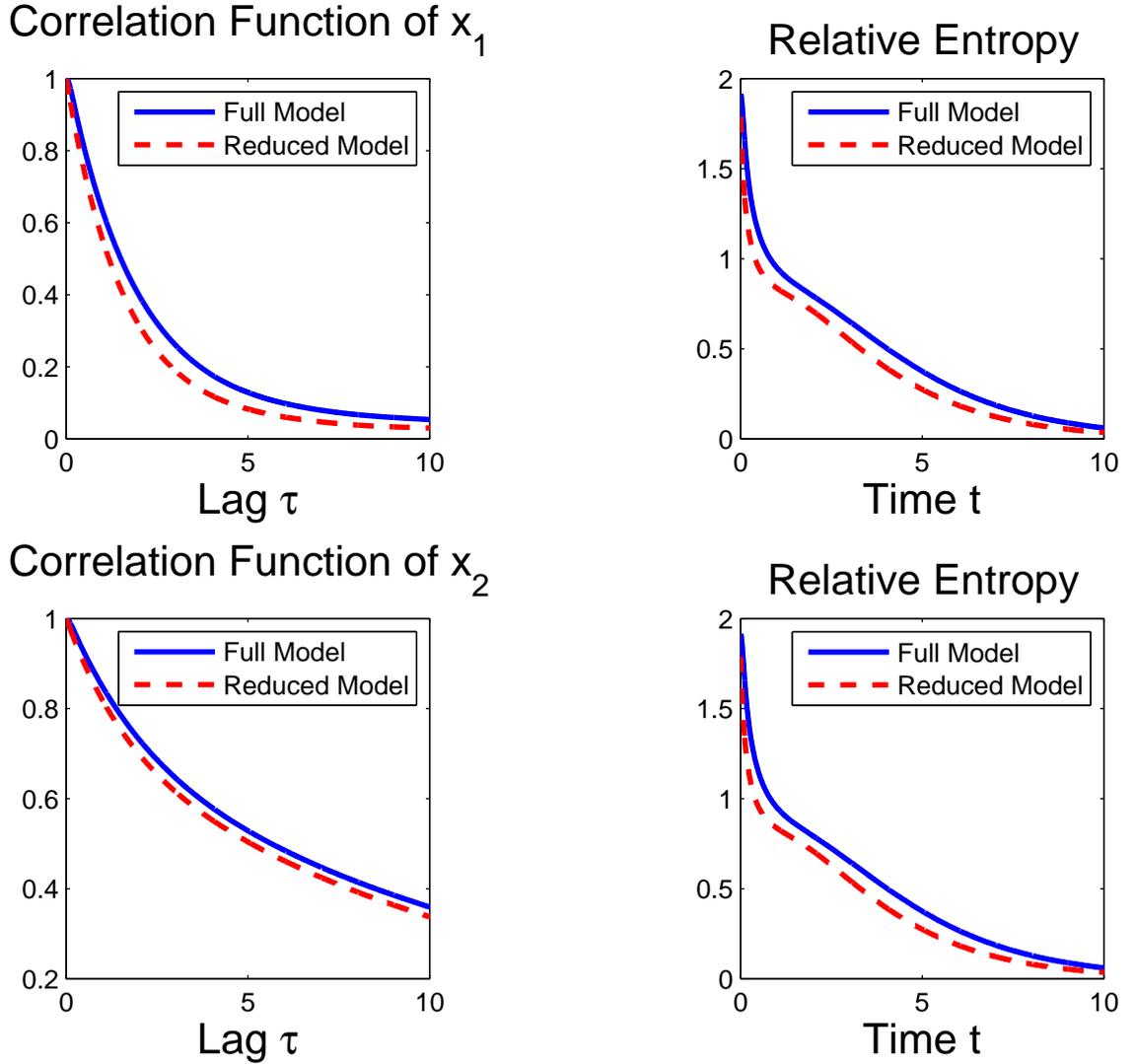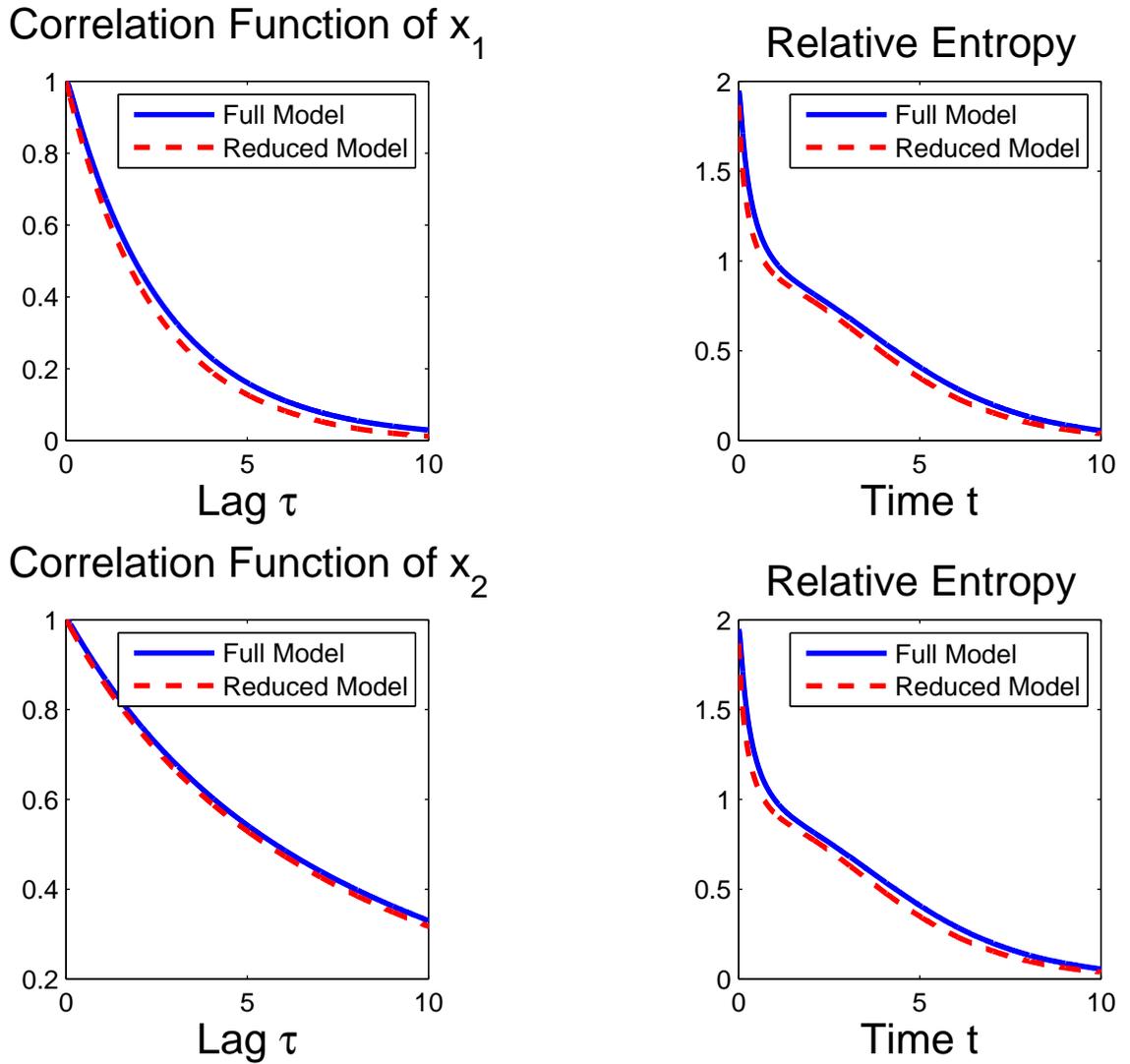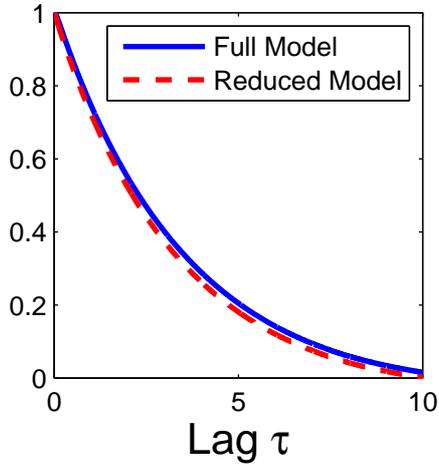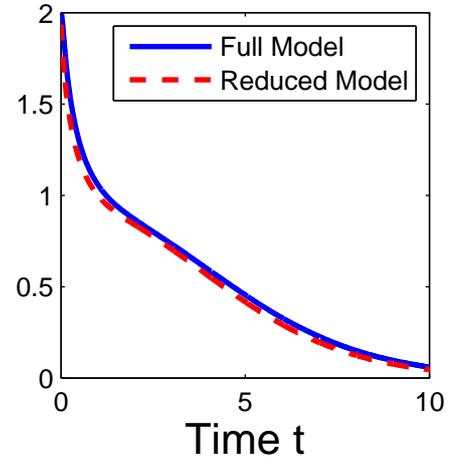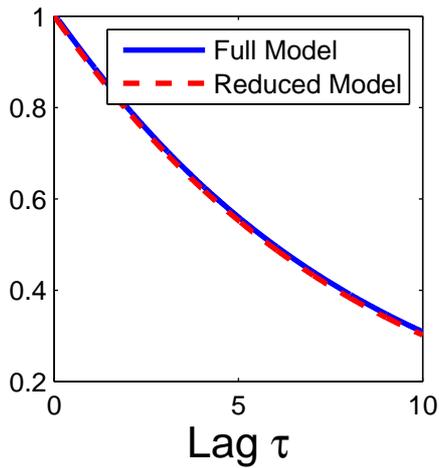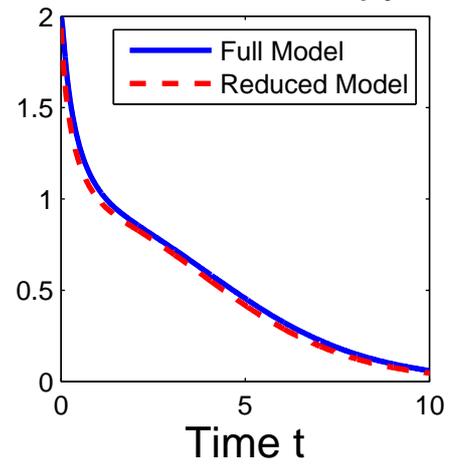
Figure 4.12: Correlation function of x_1 (top left) and x_2 (bottom left) for $\gamma = 1$. Relative entropy (top right and bottom right) for $\gamma = 10.0$. Solid line represents the full model in (3.23) and the dashed line is the reduced model in (3.24).

errors. For large values of γ, where the exponential function approximates with good accuracy, the result is valid.
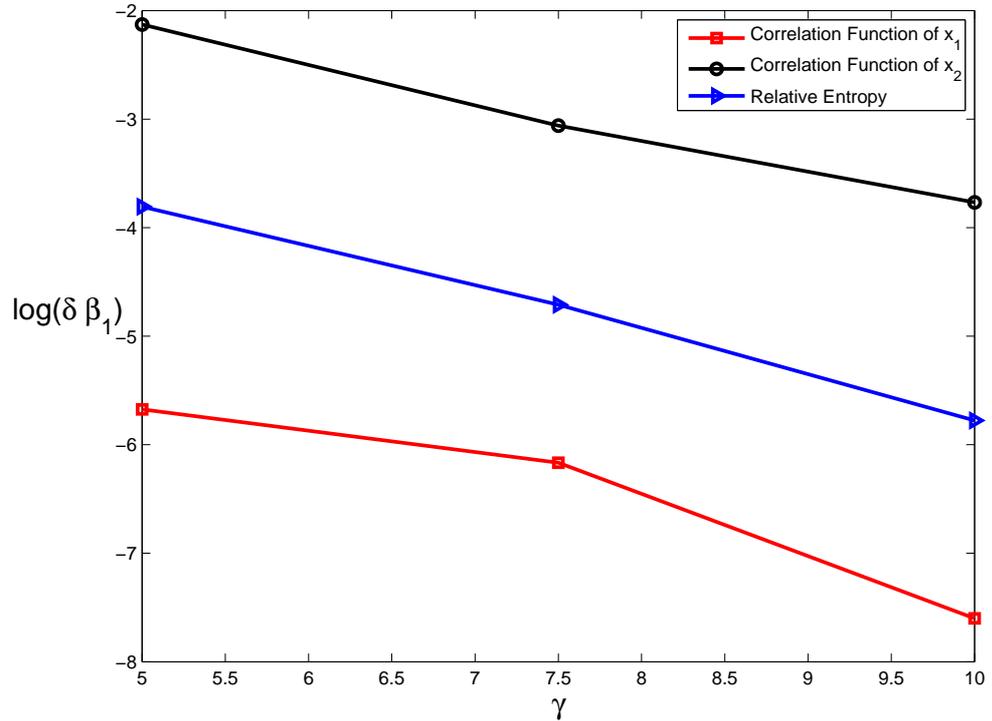
Figure 4.13: Logarithm of $\delta\beta_1$ for the correlation function of x_1 (line with square) and x_2 (line with circle) and relative entropy (line with right-pointing triangle). The variance of the initial conditions of x_1 and x_2 is one-tenth the equilibrium variance. Monte-Carlo simulation used 250,000 initial conditions normally distributed mean and variance where the mean of the distribution of x_1 and x_2 is 0.5 unit away from the equilibrium mean.

4.4 Conclusion

We demonstrated that there is a connection between the decay rate of correlation function and the idea of predictability. Moreover, correlation functions can be used as an indicator to explain how the decay rate of the relative entropy differs in the full and reduced system. Numerically, correlation functions are easy to calculate as they require a single realization, while relative entropy is calculated using Monte-Carlo simulations which requires multiple, possibly hundred and thousands of realizations.

Hence using a single realization we are able to understand how the predictability of the system behaves; moreover, we can explain how the decay rates in relative entropy differs in the full and reduced system. Numerical calculation of relative entropy requires apriori knowledge of the equilibrium distribution. Hence, in order to fully calculate relative entropy, simulations are needed to calculate the equilibrium distribution q as well as the predictive distribution p. These Direct Numerical Simulations (DNS) are time consuming and require tremendous computing power. For our problem, the calculation of "total" relative entropy requires calculating over all a subset of variables, in particular the slow variables $\boldsymbol{x}$. If the problem contains three or more slow variables, it becomes nearly impossible to calculate the relative entropy since calculating such a quantity is unrealistic. In these situations, it is easy to rely on correlation function as a measure of the predictive nature of the system rather than relative entropy itself. We would like to mention that this result has far reaching applications and is not confined to correlation functions and relative entropy. It is possible under certain conditions that the behavior is also evident in other statistical quantities, in particular the mean and variance of the slow variables. The problem with these quantities is the approximation using exponential function. In the case of the triad model and the OU process the result can still be applied using the approximation of exponential function since the mean and variance of the OU Process are exponential functions. The result is definitely not applicable for the system with stable periodic orbit and the mode-reduced analogue, since decay of the statistical quantities toward equilibrium cannot be well-approximated by exponential functions. Nevertheless, comparison of correlation functions in the full and reduced models provide overall indication about the non-equilibrium behavior of the reduced system.

Bibliography

[1] Anderson, J. and van den Dool, H. (1994). Skill and Return of Skill in Dynamic Extended-Range Forecasts. *Monthly Weather Review* **122** 507–516.

[2] Barlas, N., Josić, K., Lapin, S. and Timofeyev, I. (2007). Non-Uniform Decay of Predictability and Return of Skill in Stochastic Oscillatory Models. *Physica D* **232(2)** 116-127.

[3] Barlas, N. and Timofeyev, I. (2009). Application of the Stochastic Mode-Reduction Strategy and A-priori Prediction of Symmetry Breaking in Stochastic Systems with Underlying Symmetry. *Communications in Mathematical Sciences* (Accepted).

[4] Boer, G. (2000). A Study of Atmosphere-Ocean Predictability on Long Time Scales. *Climate Dynamics* **16** 469–472.

[5] Boffetta, G., Cencini, M., Falcioni, M. and Vulpiani, A. (2002). Predictability: A Way to Characterize Complexity. *Physics Reports* **356** 367–474.

[6] Branstator, G. (1986). The Variability in Skill of 72-hour Global-Scale NMC Forecasts. *Monthly Weather Review* **114(12)** 2628–2639.

[7] Branstator, G. and Berner, J. (2005). Linear and Nonlinear Signatures in the Planetary Wave Dynamics of an AGCM: Phase Space Tendencies. *Journal of the Atmospheric Sciences* **62** 1792–1811.

[8] Browning, K. A. and Gurney, R. J. (1999). *Global Energy and Water Cycles.* Cambridge University Press.

[9] Charney, J. G. and DeVore, J. G. (1979). Mutiple Flow Equilibria in the Atmosphere and Blocking. *Journal of the Atmospheric Sciences* **36** 1205–1216.

[10] Chu, P. C., Ivanov, L. M., Kantha, L.H., Melnichenko, O. V. and Poberezhny (2002). Power Law Decay in Model Predictability Skill. *Geophysical Research Letters* **29(15)** 1748–1753.

[11] Collins, M. (2002). Climate Predictability on Interannual to Decadal Time Scales: The Initial Value Problem. *Climate Dynamics* **19** 671–692.

[12] Collins, M. and Allen, M. (2002). Assessing the Relative Roles of Initial and Boundary Conditions in Interannual to Decadal Climate Predictability. *Journal of Climate* **15** 3104–3109.

[13] Cover, T. M. and Thomas, J. (1991). *Elements of Information Theory.* John Wiley and Sons.

[14] Crommelin, D. T. (2002). Homoclinic Dynamics: A Scenario for Atmospheric Ultra-Low-Frequency Variability. *Journal of the Atmospheric Sciences* **59** 1533–1549.

[15] Crommelin, D. T. (2003). Regime Transitions and Heteroclinic Connections in a Barotropic Atmosphere. *Journal of the Atmospheric Sciences* **60** 229–246.

[16] DelSole, T. (2004). Predictability and Information Theory. Part I: Measures of Predictability. *Journal of the Atmospheric Sciences* **61** 2425–2440.

[17] DeVille, R. E. L., Vanden-Eijnden, E. and Muratov, C. B. (2005). Two Distinct Mechanisms of Coherence in Randomly Perturbed Dynamical Systems. *Physical Review E* **72** 031105.

[18] Eckert, C. and Latif, M. (1997). Predictability of a Stochastically Forced Hybrid Coupled Model of El Niño. *Journal of Climate* **10** 1488–1504.

[19] Eckmann J. P. and Ruelle. (1985). Ergodic Theory of Chaos and Strange Attractors. *Reviews of Modern Physics* **57** 617–656.

[20] Fisher, G., Herbert, F., Madden, R. A., Schlegel, M. and Speth, P. (2002). *Thermodynamical and Dynamical Structures of the Global Atmosphere*. Springer-Verlag.

[21] Franzke, C. and Majda, A. J. (2006). Low-Order Stochastic Mode Reduction for a Prototype Atmospheric GCM. *Journal of the Atmospheric Sciences* **63** 457–479.

[22] Franzke, C., Majda, A. J. and Vanden-Eijnden, E. (2005). Low-Order Stochastic Mode Reduction for a Realistic Barotropic Model Climate. *Journal of the Atmospheric Sciences* **62** 1722–1745.

[23] Galanti, E. and Tziperman, E. (2000). ENSO's Phase Locking to the Seasonal Cycle in the Fast-SST, Fast-Wave, and Mixed-Mode Regimes. *Journal of the Atmospheric Sciences* **57** 2936–2950.

[24] Gardiner, C. W. (1985). *Handbook of Stochastic Methods*. Springer-Verlag.

[25] Gaspard, P. (1998). *Chaos, Scattering and Statistical Mechanics*, volume 9 of *Cambridge Nonlinear Science Series*. Cambridge University Press.

[26] Ghil, M. and Robertson, A. (2002). Waves vs. Particles in the Atmospheres Phases Space: A Pathway to Long-Range Forecasting? *Proceedings of the National Academy of Sciences of the United States of America* **99** 2493-2500.

[27] Itoh, H., and Kimoto, M. (1996). Multiple Attractors and Chaotic Itinerancy in a Quasigeostrophic Model with Realistic Topogaraphy: Implications for Weather Regimes and Low-frequency Variability. *Journal of Atmospheric Sciences* **53** 2217–2231.

[28] Jin, F.-F. and Ghil, M. (1990). Intraseasonal Oscillations in the Extratropics: Hopf Bifurcation and Topographic Instabilities. *Journal of Atmospheric Sciences* **47** 3007–3022.

[29] Jolliffe, I.T. and Stephenson, D.B. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley.

[30] Kestin, T., Karoly, D., Yano, J.-I. and Rayner, N. A. (1998). Time Frequency Variability of ENSO and Stochastic Simulations. *Journal of Climate* **11** 2258–2272.

[31] Kleeman, R. (2002). Measuring Dynamical Prediction Utility using Relative Entropy. *Journal of the Atmospheric Sciences* **59** 2057–2072.

[32] Kleeman, R., Majda, A. J. and Timofeyev, I. (2002). Quantifying Predictability in a Model with Statistical Features of the Atmosphere. *Proceedings of the National Academy of Sciences of the United States of America* **99(24)** 15291–15296.

[33] Korobeinikov, A. and Mcnabb, A. (2001). Long-Term Global Climate Dynamics: A Hopf Bifurcation Causing Recurrent Ice Ages. *Journal of Applied Mathematics and Decision Sciences* **5(4)** 201–214.

[34] Legras, B., and Ghil, M. (1985). Persistent Anomalies, Blocking and Variations in Atmospheric Predictability. *Journal of the Atmospheric Sciences* **42** 433-471.

[35] Lorenz, E. N. (1963). Deterministic Non-Periodic Flow. *Journal of the Atmospheric Sciences* **20** 130-141.

[36] Lumley, J. L. (1967). The Structure of Inhomogeneous Turbulence. *Atmospheric Turbulence and Wave Propagation* 166-178.

[37] Majda, A. J. and Harlim, J. (2007). Information Flow Between Subspaces of Complex Dynamical Systems. *Proceedings of the National Academy of Sciences of the United States of America* **104** 9558–9563.

[38] Majda, A. J., Kleeman, R. and Cai, D. (2002). A Mathematical Framework for Quantifying Predictability through Relative Entropy. *Methods and Applications of Analysis* **9** 425–444.

[39] Majda, A. J. and Timofeyev, I. (2004). Low-Dimensional Chaotic Dynamics versus Intrinsic Stochastic Noise: A Paradigm Model. *Physica D* **199** 339–368.

[40] Majda, A. J., Timofeyev, I. and Vanden-Eijnden, E. (2001). A Mathematical Framework for Stochastic Climate Models. *Communications on Pure and Applied Mathematics* **54** 891–974.

[41] Majda, A. J., Timofeyev, I. and Vanden-Eijnden, E. (2002). A priori Tests of a Stochastic Mode Reduction Strategy. *Physica D* **170** 206–252.

[42] Majda, A. J., Timofeyev, I. and Vanden-Eijnden, E. (2003). Systematic Strategies for Stochastic Mode Reduction in Climate. *Journal of the Atmospheric Sciences* **60(14)** 1705–1722.

[43] Marcus, S. L., Ghil, M. and Dickey, J. O. (1994). The Extratropical 40-day Oscillation in the UCLA General Circulation Model. Part I: Atmospheric Angular Momentum. *Journal of the Atmospheric Sciences* **51** 1431–1446

[44] Marcus, S. L., Ghil, M. and Dickey, J. O. (1996). The Extratropical 40-day Oscillation in the UCLA General Circulation Model. Part II: Spatial Structure. *Journal of the Atmospheric Sciences* **53** 1993–2014

[45] Mayakoda, K., Hembree, G. D., Strickler, R. F. and Shulman, I. (1972). Cumulative Results of Extended Forecast Experiments. Part I: Model Performance for Winter Cases. *Monthly Weather Review* **100** 836–855.

[46] Moore, A. M. and Kleeman, R. (1999). Stochastic Forcing of ENSO by the Intraseasonal Oscillation. *Journal of Climate* **12** 1199–1220.

[47] Ott, E. (2002). *Chaos in Dynamical Systems.* Cambridge University Press.

[48] Palmer, T. (1993). Extended-Range Atmospheric Prediction and the Lorenz Model. *Bulletin of the American Meteorological Society* **74(1)** 49–65.

[49] Penland, C. and Sardeshmukh, P. D (1995). The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate* **8** 1999–2024.

[50] Pikovsky, A. S. and Kurths, J. (1997). Coherence Resonance in a Noise-driven Excitable System. *Physical Review Letters* **78(5)** 775–778.

[51] Radok, U. and Brown, T. (1933). Anomaly Correlation and an Alternative: Partial Correlation. *Notes and Correspondence* **121** 1269–1271.

[52] Reichler, T. and Roads, J. O. (2003). The Role of Boundary and Initial Conditions for Dynamical Seasonal Predictability. *Nonlinear Processes in Geophysics* **10** 211–232.

[53] Ruelle, D. and Takens, R. (1971). On the Nature of Turbulence. *Communications in Mathematical Physics* **20** 167–192.

[54] Sarewitz, D. R., Pielke, R. A. and Byerly, R. (2000). *Prediction: Science, Decision Making, and the Future of Nature*. Island Press.

[55] Saynisch, J., Kurths, J. and Maraun, D. (2006). A Conceptual ENSO Model Under Realistic Noise Forcing. *Nonlinear Processes in Geophysics* **13** 275–285.

[56] Schreiber, T. (2000). Measuring Information Transfer. *Physical Review Letters* **85** 461–464.

[57] Stone, E. and Holmes, P. (1990). Random Perturbations of Heteroclinic Attractors. *SIAM Journal of Applied Mathematics* **50(3)** 726–743.

[58] Tribbia, J. J. and Baumhefner, D. P. (2004). Scale Interactions and Atmospheric Predictability: An Updated Perspective. *Monthly Weather Review* **132** 703–713.

[59] Trenberth, K. E. and Hoar, T. J. (1996). The 1990-1995 El Niño–Southern Oscillation Event: Longest on Record. *Geophysical Research Letters* **23** 57–60.

[60] Ziehmann, C., Smith, L. A. and Kurths, J. (2000). Localize Lyapunov Exponent and the Prediction of Predictability. *Physics Letters A* **271(4)** 237–251.

APPENDICES

APPENDIX A

Consider the general SDE

$$d\vec{X} = \mathbf{A}(\vec{X})dt + \mathbf{B}(\vec{X})d\vec{\mathbf{W}}, \tag{A.1}$$

where $\mathbf{A}$ is a vector, $\mathbf{B}$ is a matrix and $\vec{\mathbf{W}}$ is multi-dimensional linearly independent Wiener process. Notice in our applications, we can assume that $\mathbf{B}$ in diagonal. Given initial conditions $\vec{X_0}$, the backward equation associated with (A.1)

$$-\partial_t u = \sum_i A_i \partial_i u + \frac{1}{2} \sum_{ij} (B_{ij})^2 \partial_{ij}^2 u, \tag{A.2}$$

where $u(\vec{X_0}, t) = \mathbf{E_w}(\vec{X_t}|\vec{X_0})$ and $\mathbf{E_w}$ is expectation with respect to the statistics of the Wiener process $\vec{\mathbf{W}}$. The adjoint of the backward equation is the forward equation also known as the Fokker-Planck Equation. Given the initial state $(\vec{X'}, t')$ with initial condition $Pr(\vec{X}, t'|\vec{X'}, t') = \delta(\vec{X} - \vec{X'})$, the forward equation for (A.1)

$$\partial_t p = -\sum_i \partial_i [A_i p] + \frac{1}{2} \sum_{ij} \partial_{ij}^2 [(B_{ij})^2 p], \tag{A.3}$$

where $p = Pr(\vec{X}, t|\vec{X'}, t')$. Hence we define the operator associated with the forward equation as

$$L = -\sum_i \partial_i [A_i \ ] + \frac{1}{2} \sum_{ij} \partial_{ij}^2 [(B_{ij})^2 \ ],$$

and the operator associated with the backward equation, the adjoint of L as

$$L^* = \sum_i A_i \partial_i + \frac{1}{2} \sum_{ij} (B_{ij})^2 \partial_{ij}^2.$$

If we define ρ as the stationary distribution, then notice that

$$L^*1 \;=\; 0,$$

(A.4)

$$L\rho \;=\; 0.$$

Hence, 0 is a simple eigenvalue for L and L^* but the eigenvectors for the two operators are different. Notice that 0 is the simple eigenvalue, if the SDE is ergodic. Using the first equation in (A.4) we can find the stationary distribution, ρ.

APPENDIX B

B.1 General Case

We consider systems with slow-fast decomposition $\vec{X} = (\vec{x}, \vec{y})$ where the equation in time t has the form

$$
\begin{aligned}
d\vec{x} &= F_1(\vec{x}) + F_2(\vec{x}, \vec{y}), \\
d\vec{y} &= G_1(\vec{x}, \vec{y}) - \gamma\vec{y} + \beta d\vec{\mathbf{W}},
\end{aligned}
\tag{B.1}
$$

where γ is a vector and β is a matrix. We assume that the OU term in (B.1) simulate the $\vec{y}$ behavior. Introducing ϵ in (B.1), we get the equation

$$
\begin{aligned}
d\vec{x} &= F_1(\vec{x}) + F_2(\vec{x}, \vec{y}), \\
d\vec{y} &= G_1(\vec{x}, \vec{y}) - \frac{\gamma}{\epsilon}\vec{y} + \frac{\beta}{\sqrt{\epsilon}}d\vec{\mathbf{W}}.
\end{aligned}
\tag{B.2}
$$

If we coarse-grain (B.2) on a longer time scale, $\tau = \epsilon t$, to measure the slowly evolving climate variables $\vec{x}$, we derive the re-scaled stochastic climate mode in time τ

$$
\begin{aligned}
d\vec{x} &= F_1(\vec{x}) + \frac{1}{\epsilon}F_2(\vec{x}, \vec{y}), \\
d\vec{y} &= \frac{1}{\epsilon}G_1(\vec{x}, \vec{y}) - \frac{\gamma}{\epsilon^2}\vec{y} + \frac{\beta}{\epsilon}d\vec{\mathbf{W}}.
\end{aligned}
\tag{B.3}
$$

In the end we would like to compare the behavior of $\vec{X}$ on time t and τ. The the backward equation associated with the (B.3) becomes

$$-\partial_t u = Lu = L_0 u + \frac{1}{\epsilon} L_1 u + \frac{1}{\epsilon^2} L_2 u. \tag{B.4}$$

Let u be formally represented as a power series

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \ldots \tag{B.5}$$

where $u = u(\vec{x}, \vec{y}, t)$. Plugging in (B.5) in (B.4) and collecting terms for ϵ, we have

$$
\begin{array}{rcl}
\frac{1}{\epsilon^2} : \qquad L_2 u_0 & = & 0, \qquad\qquad\qquad\qquad (B.6) \\[2mm]
\frac{1}{\epsilon} : \quad L_1 u_0 + L_2 u_1 & = & 0, \qquad\qquad\qquad\qquad (B.7) \\[2mm]
1 : \qquad - \partial_t u_0 & = & L_0 u_0 + L_1 u_1 + L_2 u_2, \quad\; (B.8)
\end{array}
$$

where L_2 is the operator associated with the OU process for $\vec{y}$. Hence using (B.6), we get that $u_0 = u_0(\vec{x}, t)$, since the elliptic operator L_2 is a function of $\vec{y}$ only. Therefore, u_0 is a constant w.r.t. to the $\vec{y}$ variables. Now define operator $\mathbf{P}$ as

$$\mathbf{P} = \int d\vec{y} \, \rho(\vec{y}|\vec{x}), \tag{B.9}$$

where $\mathbf{P}$ is averaging w.r.t to $\vec{y}$. Using (A.4), we know that $\mathbf{P} L_2 = 0$. Applying $\mathbf{P}$ to (B.7) and using the fact that $\mathbf{P} L_2 = 0$, we have that $\mathbf{P} L_1 = 0$. This is the compatibility condition. In our case, all systems under consideration have this compatibility condition. If $\mathbf{P} L_1 \neq 0$, this is exactly the case of averaging and is not considered in

our applications. Considering (B.7) again, we get that

$$u_1 = -L_2^{-1} L_1 u_0. \tag{B.10}$$

Applying $\mathbf{P}$ on (B.8) and using (B.10), we get

$$\mathbf{P}(-\partial_t u_0) = \mathbf{P}(L_0 u_0) + \mathbf{P}(L_1 u_1) + \mathbf{P}(L_2 u_2), \tag{B.11}$$

$$-\partial_t u_0 = L_0 u_0 + \mathbf{P}(-L_1 L_2^{-1} L_1 u_0) + 0. \tag{B.12}$$

To explain L_2^{-1}, consider

$$\partial_t g(\vec{y}, t) = L_2 g(\vec{y}, t) + f(\vec{y}),$$

$$\tag{B.13}$$

$$g(\vec{y}, 0) = 0.$$

Define $\bar{g}(\vec{y}) = \lim_{t \to \infty} g(\vec{y}, t)$. If $\partial_t g = 0$, then $\bar{g} = -L_2^{-1} f$ from (B.13). Solving for g in (B.13) we get

$$g(\vec{y}, t) = \int_0^t ds \; e^{L_2(t-s)} f(\vec{y_t}). \tag{B.14}$$

Given a backward equation $\partial_t h = L_2 h$, we know that the solution is $h = e^{L_2 t}$ for time t. Also from (A.2), we know that $h(\vec{X_t}|\vec{X_0}) = \mathbf{E}_\mathbf{w}(\vec{X_t}|\vec{X_0})$. Hence from (B.14) and setting $t - s = \tau$ we have

$$e^{L_2 \tau} f(y) = \mathbf{E}_\mathbf{w}[f(y_\tau)|y_0 = y]. \tag{B.15}$$

Therefore

$$g(\vec{y}, t) \;=\; \int_0^t d\tau \; \mathbf{E_w}[f(y_\tau)|y_0 = y], \tag{B.16}$$

$$\rightarrow \int_0^\infty d\tau \; \mathbf{E_w}[f(y_\tau)|y_0 = y], \tag{B.17}$$

$$= \; -L_2^{-1} f(\vec{y_t}), \tag{B.18}$$

where CF is the correlation function. Notice that the last implication is using $\bar{g} = -L_2^{-1} f$. For illustration purposes, consider the case where $L_1 = \partial_x f(y)$, then

$$\mathbf{P}(-L_1 L_2^{-1} L_1 u_0) \;=\; \int_0^\infty d\tau \; f(y_0)\mathbf{E_w} f(y_\tau), \tag{B.19}$$

$$= \int_0^\infty d\tau \; \mathbf{E_w} f(y_0) f(y_\tau), \tag{B.20}$$

$$= \int_0^\infty d\tau \; \langle f(y_0) f(y_\tau) \rangle, \tag{B.21}$$

$$= \int_0^\infty d\tau \; CF_{f(y)}, \tag{B.22}$$

where the last implication is true iff $f(y)$ is linear.

B.2   Triad Case

To illustrate $\mathbf{P}(-L_1 L_2^{-1} L_1 u_0)$ for a simple equation, consider a triad model

$$
\begin{aligned}
dx &= B_1 y_1 y_2 dt, \\
dy_1 &= B_2 x y_2 dt - \gamma_1 y_1 dt + \sigma_1 dW_1, \\
dy_2 &= B_3 x y_1 dt - \gamma_2 y_2 dt + \sigma_2 dW_2.
\end{aligned}
\tag{B.23}
$$

Then

$$\mathbf{P}(-L_1 L_2^{-1} L_1 u_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dy_1 \, dy_2 \, \pi(y_1, y_2) L_1 L_2^{-1} L_1 u_0, \qquad \text{(B.24)}$$

where $\pi(y_1, y_2)$ is the joint normal distribution of y_1 and y_2. In the case of the triad model

$$L_1 = B_1 y_1 y_2 \partial_x + B_2 x y_2 \partial_{y_2} + B_3 x y_1 \partial_{y_2}. \qquad \text{(B.25)}$$

Now using (B.25) in (B.24) and using the fact $B_2 x y_2 \dfrac{\partial u_0}{\partial y_1} = B_3 x y_1 \dfrac{\partial u_0}{\partial y_2} = 0$, we have three integrals. We consider the first integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dy_1 \, dy_2 \, \pi(y_1, y_2) \left(B_1 y_1 y_2 \frac{\partial}{\partial x} \right) B_1 y_1 y_2 \frac{\partial u_0}{\partial x},$$

$$= B_1^2 \frac{\partial^2 u_0}{\partial x} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dy_1 \, dy_2 \, \pi(y_1, y_2) y_1 y_2 L_2^{-1} y_1 y_2 \quad \text{(since } u_0 = u_0(x, t)\text{)},$$

$$= -B_1^2 \frac{\partial^2 u_0}{\partial x} \int_0^{\infty} dt \, \langle y_1(0) y_1(t) \rangle \, \langle y_2(0) y_2(t) \rangle \quad \text{(from (B.19) to (B.21))},$$

$$= -B_1^2 \frac{\partial^2 u_0}{\partial x} \int_0^{\infty} dt \, \frac{\sigma_1^2}{2\gamma_1} e^{-\gamma_1 t} \frac{\sigma_2^2}{2\gamma_2} e^{-\gamma_2 t},$$

$$= -\frac{1}{\gamma_1 + \gamma_2} B_1^2 \frac{(\sigma_1 \sigma_2)^2}{4 \gamma_1 \gamma_2} \frac{\partial^2 u_0}{\partial x}.$$

Similarly, we can calculate the other two integrals to get

$$-\frac{\partial u_0}{\partial t} = \left(x B_1 B_2 \frac{\sigma_2^2}{2\gamma_2} \frac{1}{\gamma_1 + \gamma_2} + x B_1 B_3 \frac{\sigma_1^2}{2\gamma_1} \frac{1}{\gamma_1 + \gamma_2} \right) \frac{\partial u_0}{\partial x}$$

$$+ \frac{1}{2} \left(B_1^2 \frac{(\sigma_1 \sigma_2)^2}{2\gamma_1 \gamma_2} \frac{1}{\gamma_1 + \gamma_2} \right) \frac{\partial^2 u_0}{\partial x^2}. \qquad \text{(B.26)}$$

Therefore, the reduced equation for the x variable becomes

$$dx = \left(B_1 B_2 \frac{\sigma_2^2}{2\gamma_2} \frac{1}{\gamma_1 + \gamma_2} + B_1 B_3 \frac{\sigma_1^2}{2\gamma_1} \frac{1}{\gamma_1 + \gamma_2} \right) x \; dt + B_1 \frac{\sigma_1 \sigma_2}{\sqrt{2\gamma_1 \gamma_2}} \frac{1}{\sqrt{\gamma_1 + \gamma_2}} dW. \text{(B.27)}$$