University of Nevada, Reno

# On the Solvability of Inverse Problems Arising From the Two-Layer Lorenz '96 System 

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics
by

Beau James Smith

Eric John Olson/Dissertation Advisor

May 2023

## THE GRADUATE SCHOOL

We recommend that the dissertation prepared under our supervision by

BEAU JAMES SMITH
entitled

## On the Solvability of Inverse Problems Arising From the Two-Layer Lorenz '96 System

be accepted in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY
Eric Olson, Ph.D.
Advisor
Bruce Blackadar, Ph.D.
Committee Member
Paul Hurtado, Ph.D.
Committee Member
Deena Schmidt, Ph.D.
Committee Member
Aleksey Telyakovskiy, Ph.D.
Committee Member
George Bebis, Ph.D.
Graduate School Representative
Markus Kemmelmeier, Ph.D., Dean
Graduate School
May, 2023

# University of Nevada, Reno 

# Abstract 

Doctor of Philosophy

by Beau James Smith

The two-layer Lorenz '96 model consists of two linearly coupled systems of ODEs with two distinct time scales. This simple model was designed to reflect the patterns of local instability and growth represented by the interaction of planetary and synoptic dynamics with mesoscale motions and convective clouds. Under the assumption that the large-amplitude variables in the first layer are fully observed, we consider two inverse problems. The first is to estimate the unobserved values of the second layer in the case where the dynamics are known; the second is to solve for both the unobserved small scales and the unknown dynamics governing them. For simplicity, we assume that the dynamics governing the small scales take on a parameterized form with a single unknown parameter. In this case, our goal is to simultaneously estimate that parameter and the unobserved small scales.

Our study begins with a verification that the dynamics in the two-layer Lorenz '96 model are dynamically interesting enough to merit further investigation. We then develop algorithms to solve the two types of inverse problems mentioned above and find theoretical conditions under which those algorithms allow us to estimate the unobserved small scales and, optionally, the unknown parameter. We begin by proving that directly inserting the observations into the model as it is being integrated in time results in synchronization that allows recovery of the unobserved small scales over time. We then make a novel use of derivative information-i.e., the rate at which the observations change over time - to obtain new forms of data assimilation that allow solving the inverse problem faster, under less stringent conditions, and when the parameter governing the small scales is unknown.

Throughout we confirm our theoretical results with numerical experiments and remark that solving the inverse problem numerically turns out to be possible even when the system does not satisfy the hypotheses required by our theory.

To my mother, Cindy Fox, for her love and encouragement.

## Acknowledgements

I wish to thank Dr. Colin Grudzien for serving on my defense committee and spending a year and a half collaborating with me on this project. His contributions, advice, and profound knowledge of statistics has not only influenced the direction of this research but also inspired a whole array of possibilities for future work.

Also, I wish to thank Dr. Judith Fredrickson, professor emeritus at Truckee Meadows Community College, under whom I learned C++ programming in 2010 and 2011. Her computer programming classes prepared me for the coding that my master's thesis, several other math classes, and this dissertations needed. Her accessible teaching style and her skillfully crafted assignments prepared me to write, modify, and understand the code that this work demanded.

Finally, I wish to thank my dissertation advisor, Dr. Eric Olson. I appreciate his thorough explanations of the concepts, his helping me become a better programmer and mathematician, and his commitment to my academic success. Because of him, I am the first student at the University of Nevada, Reno to graduate with a Ph.D. in math; this is equally his honor and my privilege.

## Contents

Abstract ..... i
Acknowledgements ..... iii
List of Tables ..... vi
List of Figures ..... vii
1 Introduction ..... 1
2 Background Theory ..... 8
2.1 The Gauss-Newton Method ..... 8
2.2 Autocorrelation ..... 9
2.3 Akaike and Bayesian Information Criteria ..... 10
3 Preliminaries ..... 14
3.1 Change of Variables ..... 18
3.2 The Decorrelation Timescale ..... 21
4 Stochastic Parameterization ..... 24
4.1 Variance Analysis of the Parameterization ..... 27
4.2 Bayesian Selection of Polynomial Degree ..... 29
5 Rigorous Analytic Bounds ..... 31
5.1 Bounds on the Reference Solution ..... 32
5.2 Supporting Numerics ..... 45
6 Observations of $X$ with Known Dynamics ..... 49
6.1 Coupling on $\boldsymbol{X}$ ..... 50
6.2 Effects of Arbitrary Coupling Matrix $\sigma$ ..... 59
6.3 On the Use of Derivative Information ..... 60
6.4 Coupling on $\boldsymbol{X}$ and Its Higher Derivatives ..... 65
7 Coupling on $X$ with Unknown Dynamics ..... 71
7.1 The Effects of Approximate Dynamics ..... 72
7.2 Solving for $c$ ..... 77
8 Near-Continuous-in-Time Observations ..... 85
9 Conclusion ..... 91
10 Future Work ..... 93
A Numerical Methods ..... 95
B Code Samples ..... 97
B. 1 Definitions ..... 97
B. 2 Random Initial Conditions ..... 99
B. 3 Data Sampling ..... 99
B. 4 Finding $c$ ..... 100

## List of Tables

4.1 Comparison of $p$-values for various values of $c$. Here, $p_{i j}$ represents the ANOVA analysis of degree- $i$ vs. degree- $j$ regression polynomials.28

## List of Figures

1.1 Global circulation of Earth's atmosphere ([11]). ..... 4
3.1 How the oscillators in the two-layer Lorenz '96 system are coupled using the standard coupling matrix $\sigma$ given in Equation (3.1). The directional couplings between $X_{k}$ and $X_{k-2}$ in the large scales and $Y_{j}$ and $Y_{j+2}$ in the small scales are not shown. ..... 15
3.2 A typical point in the phase space of the trajectory of a solution, with $c=20, F=$ $20, m_{1}=20$, and $m_{2}=0.002$. The red dots have been positioned horizontally near the black dots that they interact with through the coupling matrix $\sigma$. ..... 21
3.3 The component $X_{1}(t)$ of $\boldsymbol{X}$ with $t \in[T, T+10], h=1 / 2048, F=20, m_{1}=$ $20, m_{2}=0.002$, and $c=20$. Note this trajectory is representative of $X_{k}$ for other values of $k$. ..... 22
3.4 Autocorrelation of $X_{1}(t)$ vs. time lag for the two-layer Lorenz ' 96 system with $F=20, m_{1}=20, m_{2}=0.002$, and $c=20$. The shaded region indicates the region where the autocorrelation is between -0.05 and 0.05 ; the vertical line in red at a time lag of 7.5 shows the point at which the autocorrelation exceeds the shaded region. ..... 23
3.5 Time lag for $c$ between 4 and 100 , with $F=20, m_{1}=20$, and $m_{2}=0.002$ such that the magnitude of the autocorrelation of $X_{1}(t)$ is less than 0.05 . ..... 23
4.1 Stochastic parameterization using $F=20, m_{1}=20, m_{2}=0.002$, and $c=20$. Here, $\tau$ is the effect of $\boldsymbol{V}$ on the first oscillator influenced by the position of the first oscillator. ..... 26
4.2 Stochastic parameterization similar to Figure 4.1, except with $c=60$. ..... 26
4.3 Standard deviation $s$ vs. $c$ for polynomial regressions of degrees 1-7. ..... 27
$4.4 \mathrm{BIC}_{n}-\min \mathrm{BIC}_{n}$ for $c=4, \ldots, 100$, where $\mathrm{BIC}_{n}$ is the BIC of the regression polynomial of degree $n$ with $n=1, \ldots, 7$. Notice the regression polynomial is of high degree when $c \ll 30$ and of low degree when $c \gg 30$. ..... 29
4.5 Degree of the regression polynomial vs. $c$, where the degree is the first value $n$ such that $\mathrm{BIC}_{n}-\min \mathrm{BIC}_{n} \leq 100$. When $c$ is large, the effect of the small oscillators is not very complicated; thus, a linear or quadratic regression polynomial is sufficient to model the effects of the small scales. ..... 30
5.1 Comparison of $\psi$, its terms, and the bound in Lemma 5.1, with $c=20, F=$ $20, m_{1}=20$, and $m_{2}=0.002$. ..... 34
5.2 Log-log plot of the a priori bounds $\min \left(M_{1}, M_{2}\right)$ and $M_{3}$ compared with the numerical a posteriori bounds $P_{2}$ and $P_{3}$. Here, $M_{1}=\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{m_{2}}}($ with $\delta=1)$, $M_{2}=\frac{\|\boldsymbol{F}\|}{m_{2} \sqrt{c}}, M_{3}=\min \left\{\frac{1}{\sqrt{c m_{2}}}, \sqrt{\frac{m_{1}}{2}}\right\} \frac{\|\boldsymbol{F}\|}{\sqrt{c m_{2}}}, P_{2}(c) \approx \lim \sup _{t \rightarrow \infty}\|\boldsymbol{V}(t)\|$, and $P_{3}(c) \approx \limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty}$, and $c=1.04^{n}$ with $n=0,4, \ldots, 160$.
5.3 Log-log plot of the numerically computed $\max \|\boldsymbol{X}\|$ versus $c$ compared with the theoretical lower bound $\omega$ on that maximum for $F=20, m_{1}=20$, and $m_{2}=0.002$. 48
6.1 Time evolution of the absolute error in $\boldsymbol{v}$ for our standard choice of parameters $F=20, m_{1}=20$, and $m_{2}=0.002$, with $c$ varying between 19 and 21 .
6.2 Least squares fit $1 / T(c) \approx 0.069 c-1.537$, where $T(c)$ is the time it takes for the approximating solution to synchronize to within $10^{-6}$ with the reference solution when $F=20, m_{1}=20, m_{2}=0.002$, and $c$ varying between 20 and 50 .
6.3 Box plots of $\|\boldsymbol{V}-\boldsymbol{v}\|$ for 4 randomly determined $\sigma$ with $F=20, m_{1}=20, m_{2}=$ 0.002 , and $c$ varying between 10 and 25 .
6.4 Box plots of $\|Q \boldsymbol{V}-\boldsymbol{q}\|$ for 4 randomly determined $\sigma$ with $F=20, m_{1}=20, m_{2}=$ 0.002 , and $c$ varying between 10 and 25 .
6.5 A representative graph of the absolute error in $\boldsymbol{q}$ for our standard choice of parameters and $c$ varying between 20 and 22 . Note the convergence occurring when $c=21$ instead of $c=20$ as in Figure 6.1. Repeated runs with different initial conditions produced similar results.
6.6 Absolute error in $\boldsymbol{q}$ for our standard choice of parameters and $h=1 / 4096, T=$ 400 , and $c$ varying between 19,20 , and 21.
6.7 Absolute error in $\boldsymbol{q}$ for our standard choice of parameters and $c$ varying between 20 and 22. Note the convergence occurring when $c=21$ instead of $c=20$ as in Figure 6.1.
6.8 A representative graph of the absolute error in $\boldsymbol{y}$ for 3 Gauss-Newton iterations, $m=5$, our standard choice of parameters, and $c$ varying between 18 and 22 . Note the convergence occurring around $c=19.5$ instead of $c=20$ as in Figure 6.1.
7.1 Statistics of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (7.1c) (left) and in the approximation $P \boldsymbol{V}+\boldsymbol{q}$ of $\boldsymbol{V}$ given by Equation (7.18c) (right) over 3600 independent samples. The relative error is less than the red dots $95 \%$ of the time and less than the aqua dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error.
7.2 Statistics, given a fixed arbitrary coupling matrix $\sigma$, of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (7.1c) (left) and in the approximation $P \boldsymbol{V}+\boldsymbol{q}$ of $\boldsymbol{V}$ given by Equation (7.18c) (right) over 3600 independent samples. The relative error is less than the red dots $95 \%$ of the time and less than the aqua dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error. . . .
7.3 Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=21$, and our standard parameter regime. Sometimes, the solution blows up (left), and other times it simply fails to converge (right).
7.4 Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=23$, and our standard parameter regime.
7.5 Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=24$, and our standard parameter regime.
7.6 Percent of approximating solutions computed using $c_{1}(t)$ compared with those computed using $c_{2}(t)$ that blow up on the time interval [0,20] from an ensemble of 100 reference solutions.
7.7 Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, \tilde{c}=24$, and our standard parameter regime.
7.8 Statistics of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (7.29), using $c_{1}(t)$ (left) and $c_{2}(t)$ (right) as defined in Equations (7.25) and (7.28), respectively, with a warmup of $T=50$, a run time of $T=72000$ and sampled every $\Delta t=20$, and our standard parameter regime. The relative error is less than the blue dots $95 \%$ of the time and less than the purple dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error. Note that the whiskers go away when $c \gtrsim 24.5$.
8.1 Time evolution of $\xi_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=21$, and our standard parameter regime. Sometimes, the solution will likely blow up (left), and other times it will simply fail to converge (right). Note that the gray areas in the left figure represent times when $\xi_{3}<10$; these are the times when the solution is likely to blow up if $\xi_{3}$ were allowed to go below 10 .
8.2 Time evolution of $\xi_{1}(t)$ (left) and $\xi_{2}(t)$ (right) as governed by Equation (8.7) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=24$, and our standard parameter regime.
8.3 Statistics of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (8.7), using $\xi_{1}(t)$ as defined in Equation (8.5) (left) and using $\xi_{3}(t)$ as defined in Equation (8.10) (right) with a warmup of $T=50$, a run time of $T=72000$ and sampled every $\Delta t=20$, and our standard parameter regime. The relative error is less than the magenta dots $95 \%$ of the time and less than the red dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error.
A. 1 Evolution in time of $\|\boldsymbol{X}\|$ for $\tilde{c}=9.0$ (blue) and $\tilde{c}=10.0$ (orange) in Equation (A.3). Equations (A.1)-(A.3) are solved using (left) the adaptive TsitPap8 method included in the DifferentialEquations.jl package in Julia, and (right) the classical RK4 method with a fixed step size. The trajectories in the latter are offset by a height of 1 for readability.

## Chapter 1

## Introduction

As the behavior of a physical system becomes more complicated, solving inverse problems related to that system becomes more difficult and, at the same time, more important. For example, a good weather forecast depends on a realistic model of the atmosphere and an accurate observation of current conditions ([12]), but atmospheric dynamics are enormously intricate due to interaction of air parcels at large and small scales across millions of cubic miles of atmosphere, and a model that could adequately represent the atmosphere at extremely fine scales is still too complex for today's most powerful computers. Moreover, perfect observations and small-scale dynamics are still beyond the grasp of today's most sensitive equipment, and even perfect observations may not be accurately represented in an imperfect model. This situation can be thought of as an inverse problem with insufficient knowledge of the system's dynamics, a tractable mathematical problem with important theoretical and practical ramifications.

To illustrate some mathematical aspects of an inverse problem, consider the equation $A \boldsymbol{x}=\boldsymbol{b}$ from linear algebra. A typical task related to this equation is solving for $\boldsymbol{x}$. However, suppose $A$ were the unknown with $\boldsymbol{x}$ and $\boldsymbol{b}$ known. In this case, the observational data are given by the known values of $\boldsymbol{x}$ and $\boldsymbol{b}$, and approximating $A$ is the inverse problem. There are many possible choices for $A$ such that $A \boldsymbol{x}=\boldsymbol{b}$; therefore, finding the correct value of $A$ is an ill-posed problem, as there are too many free parameters.

Suppose we later learn that the same $A$ also satisfies $A \boldsymbol{y}=\boldsymbol{c}$, where $\boldsymbol{y}$ and $\boldsymbol{c}$ are known. Clearly, not all choices of $A$ such that $A \boldsymbol{x}=\boldsymbol{b}$ satisfy $A \boldsymbol{y}=\boldsymbol{c}$. The question becomes how to reconcile our original estimate of $A$ with a new estimate based on $\boldsymbol{y}$ and $\boldsymbol{c}$. Imagine that over time, additional relationships involving $A$ become known and that the vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{c}$, and so forth are only
partially known or include errors. Using such a time series to arrive at a good estimate for $A$ is analogous to the data assimilation techniques generally used to solve inverse problems, with one notable exception: Practical applications typically involve a nonlinear relationship, often governed by differential equations, between the unknowns and the observations.

More concretely, numerical weather prediction is widely treated as an initial value problem with an uncertain initial condition. In this context, estimating the initial conditions from knowledge of atmospheric processes and a time series history of observations (both direct and indirect) has formed the basis of modern weather forecasting. These kinds of inverse problems have important real-world applications.

Recovering the parameters that govern the unobserved variables can be an important prerequisite to approximating the unobserved variables themselves. Since direct observations of those variables are unavailable, we need to infer their state indirectly based on their effects on things we can observe. This makes a reasonable description of their dynamics difficult or impossible to obtain. In particular, unobserved processes are often represented inadequately; even the best model run by the most powerful supercomputer is a cartoon version of the real-world behavior ([4]), and if such a system is used to model complex natural phenomena, real-time prediction of the behavior of those phenomena can be fraught with errors that grow over time. A simulation with thousands or millions of variables would be needed to adequately represent the system and reduce the size of the errors and their rate of propagation, but real-time filtering is generally limited to only a few dozen variables ([18]). Moreover, the best computations are further complicated by difficulties inherent in data assimilation algorithms; Kalman filters can be applied to nonlinear models but become suboptimal when the models are linearized ([24]), and particle filters can be reasonably applied only to small-scale systems, not to large ones with millions of variables that cannot all be sampled ([18]).

Because physical and discrete parameterizations cannot always be resolved, stochastic parameterizations have been used to estimate those unknowns, recover the parameters, and account for model uncertainty and error. These parameterizations have been shown to improve the accuracy of a model ([26]), represent the uncertainty in a model ([3]), and account for model errors ([17]). Combined with standard techniques such as inflation localization, stochastic parameterization has been shown to improve weather and climate forecasts ([4]). Thus, if one constructs a model that reasonably mimics the dynamics of the atmosphere, one could use stochastic parameterization to estimate the unknown dynamics. In turn, stochastic representation of model uncertainty can make the forecast ensemble more reliable and improve the climatological mean state ([20]).

At this point, it should be noted that characterizing the mean state of the unobserved dynamics and how that affects the observable quantities of interest is different from the inverse problem of approximating the details of the unobserved state. In this research, we use stochastic parameterization as a measure of dynamical complexity. Although these results are interesting on their own, our primary focus is solving the inverse problem of recovering the details of the unobserved variables and the dynamics which govern them.

One simple model related to weather forecasting is a system that [15] created to explore certain properties of mid-latitude atmospheric dynamics:

$$
\begin{equation*}
\frac{d X_{k}}{d t}=X_{k-1}\left(X_{k+1}-X_{k-2}\right)-X_{k}+F, \quad k=1, \ldots, K \tag{1.1}
\end{equation*}
$$

where $X_{-1}=X_{K+1}, X_{0}=X_{K}$, and $X_{K+1}=X_{1}$. Here, $F$ is a time-independent forcing constant. As [16] observed:

For very small values of $F$, all solutions decay to the steady solution $X_{1}=\ldots=$ $X_{K}=F$, while, when $F$ is somewhat larger, most solutions are periodic, but for still larger values of $F$ (dependent on $K$ ) chaos ensues.

The quadratic terms simulate advection. Note that the indices to these terms are not symmetric. From a physical point of view, this may indicate a direction of the convection (e.g., west to east); from an algebraic point of view, the indices have been arranged to obtain cancellations representing conservation of energy (this will be discussed in Lemma 3.1).

The constant and linear terms simulate external forcing and internal dissipation. Intuitively, if $X_{k}$ represents planetary and synoptic dynamics, then $F$ represents large-scale forces such as gravity and heating of the earth from the sun. The subscript $k$ is then the azimuthal angle in the mid-latitude cell (see Figure 1.1) divided as $K$ longitudinal sectors of a latitude circle. Dimensionally, it is not unreasonable to think of $X_{k}$ as representing a velocity; however, [16] notes that $X_{k}$ could be any atmospheric quantity (e.g., pressure, temperature, or humidity). It should again be emphasized that the Lorenz ' 96 model is meant only to represent the kinds of dynamics that would occur in such a latitude circle and is not a realistic model of the actual weather; rather, one can think of this model as an imitation of atmospheric dynamics in the middle latitudes.


Figure 1.1: Global circulation of Earth's atmosphere ([11]).

However, this model does not consider the interaction of large- and small-scale dynamics. To reasonably mimic this interplay, [16] adapted (1.1) as follows:

$$
\begin{align*}
\frac{d X_{k}}{d t} & =-X_{k-1}\left(X_{k-2}-X_{k+1}\right)-X_{k}+F-\frac{h c}{b} \sum_{j=J(k-1)+1}^{k J} Y_{j}, & k=1, \ldots, K  \tag{1.2a}\\
\frac{d Y_{j}}{d t} & =-c b Y_{j+1}\left(Y_{j+2}-Y_{j-1}\right)-c Y_{j}+\frac{h c}{b} X_{\lfloor(j-1) / J\rfloor+1}, & j=1, \ldots, J K \tag{1.2b}
\end{align*}
$$

where the $X_{k}$ are slow, large-amplitude variables that represent planetary and synoptic dynamics; the $Y_{j}$ are fast, small-amplitude variables that represent mesoscale and microscale dynamics ([16]); and $\lfloor s\rfloor$ denotes the greatest integer less than or equal to $s$. The last terms in Equations (1.2a) and (1.2b) couple the system. Here, $h$ is the strength of the coupling and $c$ and $b$ are scaling factors which control the amplitude of $\boldsymbol{Y}$ and the rate at which $\boldsymbol{Y}$ fluctuates. In practice, $b, c, h, F, J$, and $K$ are chosen empirically such that the dynamics of $\boldsymbol{Y}$ are mesoscale compared to the large-scale motion of $\boldsymbol{X}$ : e.g., the amplitudes of $\boldsymbol{Y}$ are one magnitude smaller and the fluctuations of motion a magnitude faster. As the dissipation in the $\boldsymbol{Y}$ equation plays a significant role in our theory and numerical results, one of our first tasks is to rescale the equations so the parameter $c$ only appears in the term $c Y_{j}$.

We again remark that the first terms of the right-hand sides of the equations are quadratic and correspond to energy-preserving advection. Note the indices in the advection term for $\boldsymbol{Y}$ are a
mirror image of the advection for $\boldsymbol{X}$. This is reminiscent of how the small eddies near a large eddy spin in the opposite direction. The linear terms correspond to dissipation of energy and to the coupling already mentioned. The constant forcing term $F$ controls the magnitude of energy injected. Note the simplifying assumption that energy is injected only into the $X_{k}$ variables.

Note that the form of the coupling is energy-conserving. Typically, a two-scale decomposition of a nonlinear PDE causes the scales to be coupled in the nonlinear terms; if these nonlinear terms represent convection, the nonlinear coupling is critical to recovering small-scale motion from observations of the large scales ([19]). In the Lorenz '96 system, the coupling occurs only in the linear terms. This allows us to interpret the system as a model of two physical processes coupled together by an energy-conserving force when the parameters are specified appropriately. Note also that the Lorenz ' 96 system is controlled by several constants that determine the resulting dynamical complexities.

Given an inverse problem of any complexity, one is faced with the question of whether one can recover not only the unobserved variables but also parameters which govern the unobserved dynamics. In particular, we want to know what parameters will cause the fast variables to be exponentially determined by the slow variables. For the two-layer Lorenz ' 96 system, our goal is to identify conditions that allow effective solving of the following two inverse problems: Use observations of $\boldsymbol{X}$ to approximate $\boldsymbol{Y}$; and use observations of $\boldsymbol{X}$ to recover the dissipation in the $\boldsymbol{Y}$ equation while simultaneously approximating $\boldsymbol{Y}$. Note that to treat the second type of inverse problem, Section 3.1 rescales the two-layer Lorenz ' 96 model in order to separate the parameter governing the dissipation from the other terms. This allows us to consider solving for the rate of energy dissipation and the small scales at the same time.

Note that solving for the parameters in the three-equation Lorenz ' 63 model was considered by Carlson et. al. [6]; in that work, the parameter being solved for appeared in the equations governing the observations. In our case, the dissipation parameter we solve for appears only in the equation governing the small scales $\boldsymbol{Y}$, whereas we observe only the large scales $\boldsymbol{X}$. We therefore introduce a new technique based on the derivative information of $\boldsymbol{X}$ that applies to the two-layer Lorenz '96 system and that appears to be generally applicable to other inverse problems.

We can make some interesting observations with a rescaled version of the two-layer Lorenz '96 system, particularly about the choice of the dissipation and coupling parameters. Changing the dissipation parameter changes the rate at which the small scales dissipate the energy from the large scales. The quadratic terms represent conservative advection which conserve the standard
energy. Thus, the coupling parameters represent the rates of energy drawn from the large scale and energy injected into the small scale.

To make these observations, we assume that the fine scales are unobserved. As the real dynamics of these scales are unknown, we expect some model error and imperfections in data assimilation. To get around this problem, various approaches have been employed: discrete-time parameterization ([17]), stochastic parameterization ([3]), covariance localization ([1]), and additive and multiplicative covariance inflation ([10], [2]).

To study the difficulties of such a task in a specific example, we consider Wilks's technique of stochastic parameterization ([26]) applied to the two-layer Lorenz '96 system. He constructed a reduced model for the observed large-scale dynamics and represented the effects of the unobserved variables through fourth-order polynomial regressions. We use this technique to explore parameter regimes that lead to systems for which the inverse problem is interesting. After characterizing the parameter regimes for which the fast variables play an important dynamical role, we then focus on when the inverse problem can be solved.

In Chapter 4, we use stochastic parameterization to determine the complexity of the dynamics $c$ that govern the small-amplitude variables. Specifically, for a range of values of $c$ we calculate the Bayesian information criterion to determine the degree of the regression polynomial that best models the effects of the small-amplitude variables on the large-amplitude ones. From these results, we infer that higher-degree polynomials correspond to greater dynamical complexity and less solvability. In Chapter 5, we find theoretical conditions under which a simple coupling scheme allows the recovery of $\boldsymbol{V}$ over time from observations of $\boldsymbol{X}$, an confirm those conditions using numerical experiments.

In Section 6.1, we prove that directly inserting the observations into the model as it is being integrated in time results in synchronization that allows recovery of the unobserved small scales over time. In Section 6.4, we make a novel use of derivative information-i.e., the rate at which the observations change over time - to obtain new forms of data assimilation that allow solving the inverse problem faster, under less stringent conditions; we do the same in Section 7.2 when the parameter governing the small scales is unknown. We repeat these experiments in Chapter 8 but under the assumption that observations of $\boldsymbol{X}$ are only approximately continuous in time. Our goal is twofold: to identify the conditions that generate effective stochastic parameterizations of this system, and to determine how those conditions are related to the coupled dynamics.

In particular, we want to know what parameters will cause the fast variables to be exponentially determined by coupling on the slow variables. We study this in two contexts: when the dissipation is known, and when the dissipation and the small-amplitude variables are unknown.

## Chapter 2

## Background Theory

In this chapter, we present background information relating to the Gauss-Newton method for minimizing a nonlinear least squares problem, autocorrelation, and the Akaike and Bayesian information criteria. We present these well-known results for completeness and ease of reference.

### 2.1 The Gauss-Newton Method

Given a linear system $A \boldsymbol{x}=\boldsymbol{b}$ with no exact solution, we look for an $\boldsymbol{x}$ that minimizes $J=$ $\|A \boldsymbol{x}-\boldsymbol{b}\|$. Minimizing $J$ is a well known problem in linear algebra, and the minimizer satisfies the normal equations

$$
\begin{equation*}
\boldsymbol{x}=\left(A^{\top} A\right)^{-1} A^{\top} \boldsymbol{b} \quad \text { or, equivalently, } \quad \boldsymbol{x}=R^{-1} Q^{\top} \boldsymbol{b} \tag{2.1}
\end{equation*}
$$

where $A=Q R$ is the reduced QR factorization of $A$. Due to the nonlinear terms $B$ and $\widetilde{B}$ in the two-layer Lorenz '96 system, we will need to solve nonlinear least squares problems. To this end, we describe the Gauss-Newton method for solving nonlinear least squares problems by means of a sequence of linear least squares solutions.

Generally, finding a minimum involves looking for the critical points: in this case, the values of $\boldsymbol{x}$ such that $\nabla J=0$. One could solve the equation by Newton's method; however, this method is often inefficient for systems of many equations, as it requires that the Hessian be calculated. The Gauss-Newton method avoids this problem by requiring only the Jacobian. In particular,
suppose

$$
\begin{equation*}
J(\boldsymbol{x})=\|\boldsymbol{\Phi}(\boldsymbol{x})\|, \quad \text { where } \quad \boldsymbol{\Phi}(\boldsymbol{x}) \in \mathbb{R}^{p} \tag{2.2}
\end{equation*}
$$

Here, we assume $\boldsymbol{\Phi}$ is a differentiable function of $\boldsymbol{x}$ which may not be linear (e.g., a vector-valued polynomial). Let $D \boldsymbol{\Phi}$ be the Jacobian matrix given by $[D \boldsymbol{\Phi}]_{i j}=\partial \Phi_{i}(\boldsymbol{x}) / \partial x_{j}$ and let $\boldsymbol{x}^{(i)}$ be an approximation of the minimizer of $\boldsymbol{x}$. Following Solomon [23], the Gauss-Newton method may be described by first linearizing $\Phi(\boldsymbol{x})$ around $\boldsymbol{x}^{(0)}$ to obtain

$$
\left.\Phi(x) \approx \Phi\left(\boldsymbol{x}^{(0)}\right)+D \Phi\left(\boldsymbol{x}^{(0)}\right)\right)\left(\boldsymbol{x}-\boldsymbol{x}^{(0)}\right) .
$$

Then let $J_{0}$ be the corresponding version of $J$ about $\boldsymbol{x}^{(0)}$ given by

$$
J_{0}(\boldsymbol{x})=\frac{1}{2}\left\|\boldsymbol{\Phi}\left(\boldsymbol{x}^{(0)}\right)+D \boldsymbol{\Phi}\left(\boldsymbol{x}^{(0)}\right)\left(\boldsymbol{x}-\boldsymbol{x}^{(0)}\right)\right\|^{2}
$$

Minimizing $J_{0}$ is a linear least squares problem in which the solutions may be obtained by Equation (2.1), where $A=-D \boldsymbol{\Phi}\left(\boldsymbol{x}^{(0)}\right)$ and $\boldsymbol{b}=\boldsymbol{\Phi}\left(\boldsymbol{x}^{(0)}\right)$.

Let $\boldsymbol{x}^{(0)}$ be an approximation of $\boldsymbol{x}$. The Gauss-Newton method evolves $\boldsymbol{x}^{(0)}$ forward by solving the sequence of linear least squares problems

$$
\begin{equation*}
\boldsymbol{x}^{(n+1)}=\boldsymbol{x}^{(n)}-\left[\left(D \boldsymbol{\Phi}\left(\boldsymbol{x}^{(n)}\right)^{\top} D \boldsymbol{\Phi}\left(\boldsymbol{x}^{(n)}\right)\right]^{-1}\left(D \boldsymbol{\Phi}\left(\boldsymbol{x}^{(n)}\right)\right)^{\top} \boldsymbol{\Phi}\left(\boldsymbol{x}^{(n)}\right), \quad \text { where } \quad n \geq 0\right. \tag{2.3}
\end{equation*}
$$

or, equivalently, using the reduced QR factorization of $D \boldsymbol{\Phi}\left(\boldsymbol{x}^{(n)}\right)$. In general, $\boldsymbol{x}^{(0)}$ must be sufficiently close to the minimizer of the original problem for convergence to occur, and the convergence to that solution may be superlinear (ideally approaching the rate of Newton's method). For a discussion of the convergence of the general Gauss-Newton algorithm, see [22] and references therein. Note that most of the literature discusses regularization techniques to ensure convergence (e.g., Levenberg-Marquardt); in our work, such techniques are not needed.

The Gauss-Newton method is used in Section 6.3 to solve algebraic constraints that appear when the two-layer Lorenz ' 96 system is coupled on $\boldsymbol{X}$ and its first $n$ derivatives.

### 2.2 Autocorrelation

It is well known that nonlinear dynamics can lead to complex time-dependent behavior. For example, the original Lorenz system ([14]) consisting of three coupled ODEs was shown to
possess a Smale horseshoe using rigorous numerics (see [25] and [9]) for the standard choice of parameters. The two-layer Lorenz '96 system studied in this dissertation is significantly more complicated, and similar rigorous numerics appear to be out of reach. However, numerical simulations suggest chaotic behavior typical of nonlinear dynamics.

Over time, dissipative nonlinear systems tend to forget the state of the initial condition and enter into what appears to be a state in phase space that is statistically independent of that initial condition. To measure the rate at which this happens, we shall use the autocorrelation.

In the time-discrete case, consider a time series $x_{n}$ sampled at points $t_{n}$ uniformly in time for $n=1, \ldots, N$. The autocovariance of $x_{n}$ is given by

$$
\begin{equation*}
\gamma_{\ell}=\frac{1}{N-\ell} \sum_{n=1}^{N-\ell}\left(x_{n}-\langle x\rangle_{0}\right)\left(x_{n+\ell}-\langle x\rangle_{\ell}\right) \tag{2.4}
\end{equation*}
$$

where

$$
\begin{equation*}
\langle x\rangle_{m}=\frac{1}{N-\ell} \sum_{n=1}^{N-\ell} x_{n+m} \tag{2.5}
\end{equation*}
$$

Given the way the averages have been defined in Equation (2.5), we note that

$$
\begin{equation*}
\frac{1}{N-\ell} \sum_{n=1}^{N-\ell}\left(x_{n}-\langle x\rangle_{0}\right)=0 \quad \text { and } \quad \frac{1}{N-\ell} \sum_{n=1}^{N-\ell}\left(x_{n+\ell}-\langle x\rangle_{\ell}\right)=0 . \tag{2.6}
\end{equation*}
$$

Therefore, we can write (2.4) as

$$
\begin{equation*}
\gamma_{\ell}=\frac{1}{N-\ell} \sum_{n=1}^{N-\ell}\left(x_{n} x_{n+\ell}-\langle x\rangle_{0}\langle x\rangle_{\ell}\right) . \tag{2.7}
\end{equation*}
$$

With the above definition of $\gamma_{\ell}$, we may write the autocorrelation function as $\Gamma_{\ell}=\gamma_{\ell} / \gamma_{0}$.

In practice, we look for the value $L$ such that $\left|\Gamma_{\ell}\right|<\epsilon$ for some small $\epsilon$ and all $\ell>L$. From this, we infer a timescale over which samples of $x_{n}$ appear independent.

### 2.3 Akaike and Bayesian Information Criteria

The BIC is similar to the Akaike information criterion (AIC). In this section, we give a brief description of both, the ideas behind them, and how to compute them. We then explain why we choose the BIC for our analysis.

When selecting a regression model to fit the data, one seeks a model that minimizes error without overfitting the training data. But such a model is not unique; one can choose from a collection of candidate models that appear to minimize error and adequately fit the data. Inspired by Wilks's use of stochastic parameterization in [26], we use a polynomial regression to parameterize the effects of the small scales on the large scales. This allows us to characterize the dynamical complexity of the interaction between the small and large scales as follows: We use the AIC and BIC to balance the goodness of fit with the degrees of freedom in this parameterization, and infer that the degrees of freedom in a statistically good parameterization represent the dynamical complexity of the small-large scale interactions.

Specifically, let $y$ be the response variable and $x$ be the explanatory variable scalars. Consider a family of functional relationships $y \approx g_{d}(x ; \theta)$. For each $d$, let $\hat{\theta}$ correspond to the best fit. Now search for the value of $d$ that best characterizes $y$ in terms of $x$. To do this, we take a sample of responses $y_{k}$ and assume $y_{k}=f\left(x_{k}\right)+\epsilon_{k}$, where $f$ is the true relationship between $x$ and $y$, and the $\epsilon_{k}$ are samples from independent and identically distributed normal random variables. Denote the joint probability distribution underlying the $\epsilon_{k}$ 's by $\rho$. Similarly, for each $d$, define $\rho_{d}$ as the joint probability distribution underlying $y_{k}-g_{d}\left(x_{k}\right)$. The task of choosing $d$ is now balanced between the number of degrees of freedom represented by $g_{d}$, and how close $g_{d}$ is to $f$.

In 1973, Akaike proposed that the Kullback-Leibler divergence between the probability distributions $\rho$ and $\rho_{d}$ underlying $f$ and $g_{d}$, respectively, could be used to select $d$ (see [8]). Using the $\log$ likelihood function $\log (\mathcal{L}(\hat{\boldsymbol{\theta}}))$ corresponding to $\rho_{d}$ of the best fit, he estimated

$$
E_{\epsilon} E_{\delta}\left[\log \left(\rho_{d}(\boldsymbol{\delta} ; \hat{\boldsymbol{\theta}}(\boldsymbol{\epsilon}))\right)\right]
$$

where $E_{\epsilon}$ and $E_{\delta}$ are calculated with respect to $\rho$. The maximum is overestimated by $\log (\mathcal{L}(\hat{\boldsymbol{\theta}}))$ on the size of $K$, where $K=\operatorname{dim}(\boldsymbol{\theta})+1$ is the number of estimable parameters, with the variance being the additional parameter. Thus,

$$
E_{\epsilon} E_{\delta}\left[\log \left(\rho_{d}(\boldsymbol{\delta} ; \hat{\boldsymbol{\theta}}(\boldsymbol{\epsilon}))\right)\right] \approx \log (\mathcal{L}(\hat{\boldsymbol{\theta}}))-K
$$

From this, Akaike defined an information criterion

$$
\begin{equation*}
\mathrm{AIC}=-2 \log (\mathcal{L}(\hat{\boldsymbol{\theta}} ; \boldsymbol{x}, \boldsymbol{y}))+2 K \tag{2.8}
\end{equation*}
$$

where $\mathcal{L}$ is a density function that measures the probability of $\hat{\boldsymbol{\theta}}$ given the data $\boldsymbol{y}$, and $2 K$ is the penalty term. Note that minimizing the AIC to select the degree $d$ of the polynomial
model involves minimizing the bias-corrected likelihood estimator. This, in turn, minimizes the Kullback-Leibler divergence between $\rho$ and $\rho_{d}$.

The AIC is an attractive selection method due to its efficiency and its selection of a model that minimizes the Kullback-Leibler divergence from the truth $f$, but a significant caveat is its lack of consistency: Even if we assume that the true model exists and is a candidate (i.e., $f(x)=g_{d}(x, \boldsymbol{\theta})$ for some values of $d$ and $\left.\boldsymbol{\theta}\right)$, the AIC will not always choose the values of $d$ and $\boldsymbol{\theta}$ that will match $f$ in the limit of large sample sizes. However, the BIC (defined below) is consistent.

$$
\begin{equation*}
\mathrm{BIC}=-2 \log (\mathcal{L}(\hat{\boldsymbol{\theta}} ; \boldsymbol{x}, \boldsymbol{y}))+K \log (N), \quad \text { where } \quad N=\text { number of observations. } \tag{2.9}
\end{equation*}
$$

We do not discuss the BIC's derivation here, but we note that the larger penalty term $K \log (N)$ means the free parameters are penalized more harshly than in the AIC. As Burnham and Anderson point out [5], this larger penalty term is needed for idealized asymptotic consistency: i.e., if the true model is in the set of candidate models, the BIC will asymptotically select that model with probability 1 as $N \rightarrow \infty$. In our case, none of the polynomial models under consideration reflect this sort of truth. However, even though we do not assume in our work that the true model is a candidate, we exploit the BIC's larger penalty term because we have thousands of data points in our sample and $N$ is large. This offsets the disadvantage that there is nothing in the theory or the derivation of the BIC that addresses the bias-variance tradeoffs considered in the AIC. In particular, the BIC may select a model that has poorer fit, in which case we may select a model using the AIC.

In the applications we consider, we assume a polynomial model for which the errors $\epsilon_{k}$ are normal, independent, and identically distributed. Thus, $y_{k}=g_{d}\left(x_{k}\right)+\epsilon_{k}$ with $g_{d}(x ; \boldsymbol{\theta})=$ $b_{d} x^{d}+\cdots+b_{1} x+b_{0}$, where $\boldsymbol{\theta}=\left(s^{2}, b_{0}, b_{1}, \ldots, b_{d}\right)$. Here, $x_{k}$ represents a time series of observations of the first oscillator $X_{1}$ in the two-layer Lorenz ' 96 system, and $y_{k}$ is the observed influence of the second layer on the first. To avoid confusion, we let $s^{2}$ refer to the variance, as we use $\sigma$ to refer to the standard coupling matrix given by Equation (3.1). Thus, the likelihood function takes the form

$$
\mathcal{L}\left(s^{2}, b_{0}, b_{1}, \ldots, b_{d} ; \boldsymbol{x}, \boldsymbol{y}\right)=\left(2 \pi s^{2}\right)^{-N / 2} \exp \left(-\frac{1}{2 s^{2}} \sum_{k=1}^{N}\left(y_{k}-g_{d}\left(x_{k}\right)\right)^{2}\right)
$$

The maximum likelihood estimators $\hat{b}_{i}$ of the $b_{i}$ are given by the polynomial regression, and $s^{2}$ is obtained from $\epsilon_{k}$ as

$$
\hat{s}^{2}=\frac{1}{N} \sum_{k=1}^{N} \epsilon_{k}^{2} .
$$

We remark that the number of estimable parameters in Equation (2.8) is $K=d+2$.

## Chapter 3

## Preliminaries

Before we proceed with our research, we rewrite the two-layer Lorenz ' 96 system so that it is more notationally convenient and allows for generalizations. We then rescale the $\boldsymbol{Y}$ equation to separate the parameter controlling the dissipation in the small scales from the other parameters. This allows us to consider a number of questions for our research that would not be possible otherwise. Finally, we characterize how the timescales in our rescaled equations depend on the dissipation.

To begin, we represent the coupling terms in Equations (1.2a)-(1.2b) as

$$
\frac{h c}{b} X_{\lfloor(j-1) / J\rfloor+1}=\frac{h c \sqrt{J}}{b}(\sigma \boldsymbol{X})_{j} s \quad \text { and } \quad \frac{h c}{b} \sum_{j=J(k-1)+1}^{k J} Y_{j}=\frac{h c \sqrt{J}}{b}\left(\sigma^{\top} \boldsymbol{Y}\right)_{k}
$$

where

$$
\sigma^{\top}=\frac{1}{\sqrt{J}}\left[\begin{array}{cccccccccccccccc}
1 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0  \tag{3.1}\\
0 & 0 & 0 & \cdots & 0 & 1 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
& & \vdots & & & & & \vdots & & & \ddots & & & \vdots & & \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & 1 & \cdots & 1
\end{array}\right] \in \mathbb{R}^{K \times J K}
$$

Note that $\sigma$ is a matrix with orthonormal columns and therefore $\sigma^{\top} \sigma=I$. To achieve this, we divide by $\sqrt{J}$. Thus, $\sigma: \mathbb{R}^{J K} \rightarrow \mathbb{R}^{K}$ is an isometry: i.e., it preserves lengths and angles. In particular, $\|\sigma\|_{2}=1$. We want a matrix that represents the coupling in the rewritten two-layer Lorenz '96 system. The coupling represented by $\sigma$ may be visualized schematically using the
diagram from [26] (see Figure 3.1). Also note that the coupling is symmetric through rotations but not through reflections.


Figure 3.1: How the oscillators in the two-layer Lorenz '96 system are coupled using the standard coupling matrix $\sigma$ given in Equation (3.1). The directional couplings between $X_{k}$ and $X_{k-2}$ in the large scales and $Y_{j}$ and $Y_{j+2}$ in the small scales are not shown.

For most of this study, we will employ $\sigma$ as given in Equation (3.1). Much of our theoretical analysis relies on the fact that $\sigma^{\top} \sigma=I$; therefore, any suitably sized $\sigma$ with orthonormal columns satisfies the hypotheses of our theorems. This allows us to consider how different couplings between the two layers affect our ability to recover the small scales from observations of the large scales (see Section 6.2).

Following the notation employed by Law et. al. [13] in analogy with standard notation for the convective terms in equations governing fluid dynamics (see, for example, [7]), we set $\boldsymbol{F}=$ $(F, F, \ldots, F) \in \mathbb{R}^{k}$, let $\boldsymbol{x}, \boldsymbol{X} \in \mathbb{R}^{K}$ and $\boldsymbol{y}, \boldsymbol{Y} \in \mathbb{R}^{J K}$. Define $B_{k}: \mathbb{R}^{K} \times \mathbb{R}^{K} \rightarrow \mathbb{R}$ and $\widetilde{B}_{j}: \mathbb{R}^{J K} \times$ $\mathbb{R}^{J K} \rightarrow \mathbb{R}$ as

$$
\begin{align*}
B_{k}(\boldsymbol{X}, \boldsymbol{x}) & =X_{k-1}\left(x_{k-2}-x_{k+1}\right),  \tag{3.2a}\\
\widetilde{B}_{j}(\boldsymbol{Y}, \boldsymbol{y}) & =Y_{j+1}\left(y_{j+2}-y_{j-1}\right), \tag{3.2b}
\end{align*}
$$

where $k=1,2, \ldots, K$ and $j=1,2, \ldots, J K$. Recall that the definition of $X_{k}$ is to be extended to all values of $k$ in (3.2a) by letting $X_{k-K}$ and $X_{k+K}$ equal $X_{k}$. The same extension holds for $x_{k}$, and a similar one for period $J K$ holds for $Y_{j}$ and $y_{j}$ in Equation (3.2b).

We note $B$ and $\widetilde{B}$ have the following properties:
Lemma 3.1. Let $\boldsymbol{W}, \boldsymbol{X} \in \mathbb{R}^{K}$ and $\boldsymbol{Y}, \boldsymbol{Z} \in \mathbb{R}^{J K}$. Then

1. $B(\boldsymbol{X}, \boldsymbol{X}) \cdot \boldsymbol{X}=0$.
2. $\widetilde{B}(\boldsymbol{Y}, \boldsymbol{Y}) \cdot \boldsymbol{Y}=0$.
3. $B$ and $\widetilde{B}$ are bilinear: i.e., for vectors $\boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{X}_{\boldsymbol{i}} \in \mathbb{R}^{K}$ and $\boldsymbol{y}_{\boldsymbol{i}}, \boldsymbol{Y}_{\boldsymbol{i}} \in \mathbb{R}^{J K}$ with $i=1,2$ and constants $a, b, c, d \in \mathbb{R}$,

$$
\begin{align*}
B\left(a \boldsymbol{X}_{\mathbf{1}}+b \boldsymbol{X}_{\mathbf{2}}, c \boldsymbol{x}_{\mathbf{1}}+d \boldsymbol{x}_{\mathbf{2}}\right) & =a c B\left(\boldsymbol{X}_{\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}\right)+a d B\left(\boldsymbol{X}_{\mathbf{1}}, \boldsymbol{x}_{\mathbf{2}}\right) \\
& +b c B\left(\boldsymbol{X}_{\mathbf{2}}, \boldsymbol{x}_{\mathbf{1}}\right)+b d B\left(\boldsymbol{X}_{\mathbf{2}}, \boldsymbol{x}_{\mathbf{2}}\right)  \tag{3.3}\\
\widetilde{B}\left(a \boldsymbol{Y}_{\mathbf{1}}+b \boldsymbol{Y}_{\mathbf{2}}, c \boldsymbol{y}_{\mathbf{1}}+d \boldsymbol{y}_{\mathbf{2}}\right) & =a c \widetilde{B}\left(\boldsymbol{Y}_{\mathbf{1}}, \boldsymbol{y}_{\mathbf{1}}\right)+a d \widetilde{B}\left(\boldsymbol{Y}_{\mathbf{1}}, \boldsymbol{y}_{\mathbf{2}}\right) \\
& +b c \widetilde{B}\left(\boldsymbol{Y}_{\mathbf{2}}, \boldsymbol{y}_{\mathbf{1}}\right)+b d \widetilde{B}\left(\boldsymbol{Y}_{\mathbf{2}}, \boldsymbol{y}_{\mathbf{2}}\right) \tag{3.4}
\end{align*}
$$

4. $[B(\boldsymbol{X}, \boldsymbol{W})+B(\boldsymbol{W}, \boldsymbol{X})] \cdot \boldsymbol{W} \leq 2\|\boldsymbol{X}\|\|\boldsymbol{W}\|^{2}$.
5. $[\widetilde{B}(\boldsymbol{Y}, \boldsymbol{Z})+\widetilde{B}(\boldsymbol{Z}, \boldsymbol{Y})] \cdot \boldsymbol{Y} \leq 2\|\boldsymbol{Y}\|\|\boldsymbol{Z}\|^{2}$.

Proof. Although the results of Lemma 3.1 for the convection-like bilinear term in the Lorenz '96 models are generally well known, we present details here to obtain explicit bounds and illustrate algebraic relations that we will use in later analysis.

To prove part 1 , note that $X_{0}=X_{K}, X_{1}=X_{k+1}$, and so on. Thus, we see

$$
\begin{aligned}
B(\boldsymbol{X}, \boldsymbol{X}) \cdot \boldsymbol{X} & =\sum_{k=1}^{K} X_{k-1}\left(X_{k-2}-X_{k+1}\right) X_{k} \\
& =\sum_{k=1}^{K}\left(X_{k-1} X_{k-2} X_{k}-X_{k-1} X_{k+1} X_{k}\right) \\
& =\sum_{k=1}^{K} X_{k-1} X_{k-2} X_{k}+\sum_{k=1}^{K} X_{k-1} X_{k+1} X_{k} \\
& =\sum_{k=1}^{K} X_{k-2} X_{k-1} X_{k}+\sum_{k=1}^{K} X_{k-2} X_{k-1} X_{k} \\
& =0
\end{aligned}
$$

The proof of part 2 is similar.

Equation (3.3) is established by expanding the expression $B_{k}\left(a \boldsymbol{X}_{\mathbf{1}}+b \boldsymbol{X}_{\mathbf{2}}, c \boldsymbol{x}_{\mathbf{1}}+d \boldsymbol{x}_{\mathbf{2}}\right)$ for an arbitrary $k$. The calculation is easy to confirm:

$$
\begin{gathered}
B_{k}\left(a \boldsymbol{X}_{\mathbf{1}}+b \boldsymbol{X}_{\mathbf{2}}, c \boldsymbol{x}_{\mathbf{1}}+d \boldsymbol{x}_{\mathbf{2}}\right)=\left(a X_{1_{k-1}}+b X_{2_{k-1}}\right)\left(c x_{1_{k-2}}+d x_{2_{k-2}}-c x_{1_{k+1}}-d x_{2_{k+1}}\right) \\
=a X_{1_{k-1}}\left(c x_{1_{k-2}}-c x_{1_{k+1}}\right)+a X_{1_{k-1}}\left(d x_{2_{k-2}}-d x_{2_{k+1}}\right) \\
+b X_{2_{k-1}}\left(c x_{1_{k-2}}-c x_{1_{k+1}}\right)+b X_{2_{k-1}}\left(d x_{2_{k-2}}-d x_{2_{k+1}}\right) \\
=a c B\left(\boldsymbol{X}_{1}, \boldsymbol{x}_{1}\right)+a d B\left(\boldsymbol{X}_{1}, \boldsymbol{x}_{2}\right) \\
+b c B\left(\boldsymbol{X}_{2}, \boldsymbol{x}_{1}\right)+b d B\left(\boldsymbol{X}_{2}, \boldsymbol{x}_{2}\right)
\end{gathered}
$$

A similar calculation establishes Equation (3.4).
The proof of part 4 is similar to that of part 5 , which we prove below. To do this, note

$$
\begin{align*}
& {[\widetilde{B}(\boldsymbol{Y}, \boldsymbol{Z})+\widetilde{B}(\boldsymbol{Z}, \boldsymbol{Y})] \cdot \boldsymbol{Y}} \\
& \quad= \\
& \quad \sum_{j=1}^{J K}\left[-Y_{j+1}\left(Z_{j+2}-Z_{j-1}\right) Y_{j}-Z_{j+1}\left(Y_{j+2}-Y_{j-1}\right) Y_{j}\right]  \tag{3.5}\\
& \quad= \\
& \sum_{j=1}^{J K}\left[-Y_{j+1} Z_{j+2} Y_{j}+Y_{j+1} Z_{j-1} Y_{j}-Z_{j+1} Y_{j+2} Y_{j}+Z_{j+1} Y_{j-1} Y_{j}\right]
\end{align*}
$$

Notice that the first and fourth summands cancel due to telescoping. Thus, the expression becomes

$$
\begin{equation*}
[\widetilde{B}(\boldsymbol{Y}, \boldsymbol{Z})+\widetilde{B}(V, \boldsymbol{Y})] \cdot \boldsymbol{Y}=\sum_{j=1}^{J K}\left[Y_{j+1} Z_{j-1} Y_{j}-Z_{j+1} Y_{j+2} Y_{j}\right] \tag{3.6}
\end{equation*}
$$

By the triangle inequality, followed by Cauchy-Schwarz,

$$
\begin{align*}
{[\widetilde{B}(\boldsymbol{Y}, \boldsymbol{Z})+\widetilde{B}(\boldsymbol{Z}, \boldsymbol{Y})] \cdot \boldsymbol{Y} } & \leq \sum_{j=1}^{J K}\left|Y_{j+1} Z_{j-1} Y_{j}\right|+\left|-Z_{j+1} Y_{j+2} Y_{j}\right| \\
& \leq 2\|\boldsymbol{Z}\|_{\infty}\|\boldsymbol{Y}\|^{2} \\
& \leq 2\|\boldsymbol{Z}\|\|\boldsymbol{Y}\|^{2} \tag{3.7}
\end{align*}
$$

### 3.1 Change of Variables

Our motivation for defining the generalized system below in Equation (3.9) is as follows: First, this form of the system describes the essential dynamical properties of the system, including the energy-preserving, componentwise bilinear form, the linear dissipation and injection of energy, for arbitrary forms of coupling between the layers. Second, the role of the parameter $c$ becomes evident as the magnitude of the dissipation of energy in the fast variables, while $m_{1}$ and $m_{2}$ define the (possibly distinct) draw and amplification, respectively, of energy in its transfer from the slow layer to the fast layer. Note that separating $c$ from the other parameters allows us to think about the equations in a more physical way; this makes our approach novel (to our knowledge) and closer to the way practical problems are formulated.

Definition 1 (Generalized Two-Layer Lorenz '96). Define a pair of arbitrary componentwise bilinear forms $B: \mathbb{R}^{p} \times \mathbb{R}^{p} \rightarrow \mathbb{R}^{p}$ and $\tilde{B}: \mathbb{R}^{q} \times \mathbb{R}^{q} \rightarrow \mathbb{R}^{q}$ that satisfy the orthogonality property:

$$
\begin{array}{rll}
B(\boldsymbol{X}, \boldsymbol{X}) \cdot \boldsymbol{X}=0 & \text { for } & \boldsymbol{X} \in \mathbb{R}^{p} \\
\tilde{B}(\boldsymbol{V}, \boldsymbol{V}) \cdot \boldsymbol{V}=0 & \text { for } & \boldsymbol{V} \in \mathbb{R}^{q} . \tag{3.8b}
\end{array}
$$

Define $\boldsymbol{\sigma}$ as an arbitrary coupling matrix of size $p \times q$. Then the coupled system of nonlinear ordinary differential equations with dynamic state vectors $\boldsymbol{X}, \boldsymbol{V}$, time variable $t$, and the free parameters $c, m_{1}, m_{2}$ is as follows:

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \boldsymbol{\sigma}^{\top} \boldsymbol{V}  \tag{3.9a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \boldsymbol{\sigma} \boldsymbol{X} \tag{3.9b}
\end{align*}
$$

We refer to this as the generalized Lorenz '96 system.

We emphasize in Equation (3.9b) that the parameter $c$ now appears in the term $c \boldsymbol{V}$ controlling the rate of dissipation of energy. The original parameters $b$ and $h$ have been replaced by $m_{1}$ and $m_{2}$. Dimensionally, $m_{1}$ and $m_{2}$ function as masses in the energy equation for the rescaled system. The ability to take the masses constant while varying the dissipation allows us to explore a number of novel theoretical and numerical questions related to $c$. Note also that $\boldsymbol{X}$ remains unchanged from Lorenz's original formulation but $\boldsymbol{Y}$ has been replaced by a rescaled version $\boldsymbol{V}$.

We proceed to derive the generalized Lorenz '96 system as follows. Using the definitions of $B$, $\widetilde{B}$, and $\sigma$ given in Equations (3.1) and (3.2), we rewrite equations (1.2a) and (1.2b) to obtain

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})-\boldsymbol{X}=\boldsymbol{F}-\frac{h c \sqrt{J}}{b} \sigma^{\top} \boldsymbol{Y}  \tag{3.10a}\\
& \frac{d \boldsymbol{Y}}{d t}+c b \widetilde{B}(\boldsymbol{Y}, \boldsymbol{Y})-c \boldsymbol{Y}=\frac{h c \sqrt{J}}{b} \sigma \boldsymbol{X} \tag{3.10b}
\end{align*}
$$

Note Equations (3.10a) and (3.10b) can be written into the form of Definition 1 by rescaling with $\beta>0$ such that $\boldsymbol{Y}=\beta \boldsymbol{V}$ to yield a system of the form in Definition 1.

We do this by first setting $\boldsymbol{Y}=\beta \boldsymbol{V}$. This yields

$$
\begin{align*}
\frac{d \boldsymbol{V}}{d t} & =\frac{1}{\beta}\left[-c b \beta^{2} \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-c \beta \boldsymbol{V}+\frac{h c \sqrt{J}}{b} \sigma \boldsymbol{X}\right]  \tag{3.11}\\
& =-c b \beta \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-c \boldsymbol{V}+\frac{h c \sqrt{J}}{b \beta} \sigma \boldsymbol{X} .
\end{align*}
$$

Equation (3.11) yields the rescaled $\boldsymbol{V}$ equation for arbitrary parameters:

$$
\begin{equation*}
\frac{d \boldsymbol{V}}{d t}=-c b \beta \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-c \boldsymbol{V}+\frac{h c \sqrt{J}}{b \beta} \sigma \boldsymbol{X} . \tag{3.12}
\end{equation*}
$$

To match the terms in (3.12), we enforce $\beta=\frac{1}{c b}$ and obtain

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{h \sqrt{J}}{b^{2}} \sigma^{\top} \boldsymbol{V},  \tag{3.13a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=h c^{2} \sqrt{J} \sigma \boldsymbol{X} . \tag{3.13b}
\end{align*}
$$

To obtain the $\frac{1}{m_{2}}$ in (3.9b), we can change the standard parameter choice to $h=\frac{1}{m_{2} c^{2} \sqrt{J}}$, where $m_{2}$ is free. The system becomes

$$
\begin{align*}
\frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X} & =\boldsymbol{F}-\frac{1}{b^{2} c^{2} m_{2}} \sigma^{\top} \boldsymbol{V}  \tag{3.14a}\\
\frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V} & =\frac{1}{m_{2}} \sigma \boldsymbol{X} . \tag{3.14b}
\end{align*}
$$

Finally, if we take $b^{2}=\frac{m_{1}}{m_{2} c^{2}}$, we obtain equations (3.9a) and (3.9b).
The resulting choice of parameters is thus

$$
\begin{equation*}
\beta=\frac{1}{c b}, \quad h=\frac{1}{m_{2} c^{2} \sqrt{J}}, \quad b=\sqrt{\frac{m_{1}}{m_{2} c^{2}}}, \tag{3.15}
\end{equation*}
$$

and $\boldsymbol{Y}$ is rescaled by a factor of $\sqrt{m_{2} / m_{1}}$ to obtain $\boldsymbol{V}$.

In addition to fewer parameters and a simpler expression, a significant advantage of the rescaled system given in Definition 1 is that $\boldsymbol{F}$ is no longer the only isolated parameter. If we attempt to alter the dynamics of the system by controlling the dynamics of $\boldsymbol{X}$ via $\boldsymbol{F}$, many solutions would become constant (see the text following Equation (1.1) above).

The generalized system allows us to control the rate of dissipation of energy in the second layer via the parameter $c$ without affecting the coupling between the layers. The advantage of this approach will be further supported by Proposition 5.8 , which shows that the dynamics of $\boldsymbol{X}$ remain nontrivial as $c \rightarrow \infty$ while $m_{1}$ and $m_{2}$ are held constant. Thus, we can analyze large values of $c$ without causing the dynamics of $\boldsymbol{X}$ to vanish.

Note also that $m_{1}$ and $m_{2}$ play the role of masses for the nonlinear oscillators represented by $\boldsymbol{X}$ and $\boldsymbol{V}$, respectively. In particular, the energy function of the system now appears as

$$
\begin{equation*}
\psi=m_{1}\|\boldsymbol{X}\|^{2}+m_{2}\|\boldsymbol{V}\|^{2} \tag{3.16}
\end{equation*}
$$

Lorenz [16] focused on the parameters $K=36, J=10, c=10, b=10$, and $h=1.0$, with $F$ varying between $8,10,15$, and 18 . Wilks [26] retained the traditional values of $c, b$, and $h$ and set $K=8, J=32$, and considered $F=18$ and 20 . Referring back to the original parameter choices of $h=1, c=10$, and $b=10$, we follow Wilks [26] by setting $K=8$ and $J=32$. From Equation (3.15), we obtain

$$
m_{1}=m_{2} b^{2} c^{2} \approx 17.6777, \quad \text { where } \quad m_{2}=\frac{1}{h c^{2} \sqrt{J}} \approx 0.001768
$$

which we simplify by rounding $m_{1}$ to 20 and $m_{2}$ to 0.002 . Thus, for the remainder of this study, we vary $c$ and focus on the parameter regime

$$
\begin{equation*}
F=20, \quad m_{1}=20, \quad \text { and } \quad m_{2}=0.002 \tag{3.17}
\end{equation*}
$$

Figure 3.2 illustrates a typical point in the phase space of the trajectory of a solution, with $c=20, F=20, m_{1}=20$, and $m_{2}=0.002$. The black dots represent the values of $\boldsymbol{X}$, and the red dots represent the values of $\boldsymbol{V}$.

We note for this choice of parameters with $\boldsymbol{\sigma}$ defined as in (3.1) that $\boldsymbol{X}$ appears to undergo complex time-dependent motion (see Figure 3.3).


Figure 3.2: A typical point in the phase space of the trajectory of a solution, with $c=$ $20, F=20, m_{1}=20$, and $m_{2}=0.002$. The red dots have been positioned horizontally near the black dots that they interact with through the coupling matrix $\sigma$.

For this figure and all the other calculations which appear in this dissertation, we will use the classic RK4 method with a fixed step size, usually of $h=1 / 2048$. Note that this $h$ was chosen to be small enough so that the statistics of our results do not depend on it and our solutions remain stable and appear sufficiently accurate for all choices of $c$. Our reasons for not using an adaptive method are explained in Appendix A.

### 3.2 The Decorrelation Timescale

To characterize the decorrelation timescale of the motion represented in Figure 3.3, we consider the autocorrelation of the component $x_{n}=X_{1}\left(t_{n}\right)$ for $n=1, \ldots, 288001$, where $t_{n}$ is the sequence of times separated by 0.25 . Figure 3.4 illustrates the autocorrelation as a function of $\Gamma_{\ell}$ when $F=20, m_{1}=20, m_{2}=0.002$, and $c=20$. Note that when the time lag is greater than 7.5, $\left|\Gamma_{\ell}\right| \leq 0.05$. In Figure 3.5, we characterize as a function of $c$ the time lag $T(c)$ such that $\left|\Gamma_{\ell}\right| \leq 0.05$ for all $t_{\ell} \geq T(c)$. Note that $T(c)$ generally increases until $c \approx 23$, reaching a maximum value of 10.75 , and decreases after that. Therefore, to obtain a time lag that is larger than the decorrelation timescale for all values of $c$ under consideration, we sample the trajectories of our deterministic nonlinear dynamical system every 20 units in time. As a result, we have


Figure 3.3: The component $X_{1}(t)$ of $\boldsymbol{X}$ with $t \in[T, T+10], h=1 / 2048, F=20, m_{1}=$ $20, m_{2}=0.002$, and $c=20$. Note this trajectory is representative of $X_{k}$ for other values of $k$.
$N=288001 / 80 \approx 3600$ samples which appear linearly independent and, in particular,

$$
\begin{equation*}
\left|\Gamma_{\ell}\right| \lesssim 0.01818 \quad \text { for } \quad t_{\ell} \geq 20 \quad \text { and } \quad c \in\{4, \ldots, 100\} \tag{3.18}
\end{equation*}
$$

Note that if $x_{n}$ for $n=1, \ldots, 3600$ were truly independent, statistical simulation using 100,000 trials suggests the probability $P\left(\left|\Gamma_{\ell}\right|>0.01818\right) \approx 0.28$. Thus, the upper bound of 0.01818 in (3.18) is reasonably close to 0 given the sample size in our autocorrelation computations.

We emphasize our motivation for defining the generalized system in Equation (3.9) as follows: First, this form of the system describes the essential dynamical properties of the system, including the energy-preserving, componentwise bilinear form, the linear dissipation and injection of energy, for arbitrary forms of coupling between the layers. Second, the role of the parameter $c$ becomes evident as the magnitude of the dissipation of energy in the fast variables, while $m_{1}$ and $m_{2}$ define the (possibly distinct) draw and amplification, respectively, of energy in its transfer from the slow layer to the fast layer. Third, with respect to any such system, we can conclude for which relative scales of dissipation and amplification that the fast layer dynamics will be exponentially determined by the state of the slow layer. We use the method of stochastic parameterization to further characterize the complex time-dependent behavior observed in Figures 3.2 and 3.3.


Figure 3.4: Autocorrelation of $X_{1}(t)$ vs. time lag for the two-layer Lorenz '96 system with $F=20, m_{1}=20, m_{2}=0.002$, and $c=20$. The shaded region indicates the region where the autocorrelation is between -0.05 and 0.05 ; the vertical line in red at a time lag of 7.5 shows the point at which the autocorrelation exceeds the shaded region.


Figure 3.5: Time lag for $c$ between 4 and 100 , with $F=20, m_{1}=20$, and $m_{2}=0.002$ such that the magnitude of the autocorrelation of $X_{1}(t)$ is less than 0.05 .

## Chapter 4

## Stochastic Parameterization

In our analysis and simulation of the two-layer Lorenz '96 system, it will be useful to have a way of characterizing the dynamical complexity of the interaction between the small and large scales. To this end, we consider Wilks's stochastic parameterization of the small scales and use the Bayesian information criterion (BIC) to select the degree of the polynomial model in that parameterization.

Wilks [26] used Equations (1.2) to study stochastic parameterization of the effects of unresolved mesoscale variables. In terms of the notation used here, he parameterized the effects of $\tau=$ $\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}$ via the model

$$
\begin{equation*}
\frac{d \boldsymbol{X}}{d t}=B(\boldsymbol{X}, \boldsymbol{X})-\boldsymbol{X}+\boldsymbol{F}-G(\boldsymbol{X}) \tag{4.1}
\end{equation*}
$$

and called $\tau$ the tendencies. Thus, $\tau$ is modeled by $G(\boldsymbol{X})$. Since $G(\boldsymbol{X})$ treats the effects of $\boldsymbol{V}$ on $\boldsymbol{X}$ and ignores the state of $\boldsymbol{V}$, provided the detailed motion of $\boldsymbol{V}$ is significant, then the modeling of $\tau$ by $G(\boldsymbol{X})$ must include an error. Wilks considered the polynomial model

$$
\begin{equation*}
G(\boldsymbol{X})=\left(g_{d}\left(\boldsymbol{X}_{1}\right), \ldots, g_{d}\left(\boldsymbol{X}_{K}\right)\right)+\boldsymbol{e}(t), \quad \text { where } \quad g_{d}(x)=b_{d} x^{d}+\cdots+b_{1} x+b_{0} . \tag{4.2}
\end{equation*}
$$

with $d=4$ and $\boldsymbol{e}$ a stochastic term that represents the modeling error. Note that the same polynomial is used for each of the vector components of $G$ because the rotational symmetry in Figure 3.1 implies all the $\boldsymbol{X}$ oscillators are equivalent. Other choices of $\sigma$ may not have this symmetry.

Of interest in our research is the degree of the polynomial $g_{d}$. In particular, we shall use the ANOVA test and Bayesian information criterion (BIC) to characterize the degree of the polynomial that best models the effects of $\boldsymbol{V}$ on $\boldsymbol{X}$. If this polynomial is of high degree, we infer that the dynamics which govern that interaction are complicated, whereas if the degree is smaller (e.g., less than 4), we presume the dynamics to be simpler.

In passing, we know there are other aspects of stochastic parameterization that could be used to characterize how important the interaction is between the small and large scales: e.g., one could imagine measuring the difference between the motion of $\boldsymbol{X}$ as influenced by the actual tendencies compared to the motion of $\boldsymbol{X}$ when the tendencies are modeled by $G(\boldsymbol{X})$ using a Hamming-like information-theoretic notion of distance. The technical difficulties of creating a suitable notion of distance are beyond the scope of this research.

One could also compare the length $C$ of the curve $\left(x, g_{d}(x)\right)$ for $x \in[-5,15]$ to the distance $L$ between the endpoints $\left(-5, g_{d}(-5)\right)$ and $\left(15, g_{d}(15)\right)$. The ratio $C / L$ is called the tortuosity. The larger the tortuosity, the more the polynomial oscillates, and the more complicated the relationship between $\boldsymbol{V}$ and $\boldsymbol{X}$. Even just the variance around the fit contains information about how important the motion of $\boldsymbol{V}$ is to the dynamics of $\boldsymbol{X}$. While we shall look at this variance (actually standard deviation) in detail, we omit a discussion of the tortuosity in order to proceed directly to the BIC.

Our goal, then, is to vary $c$ and report the degree of the corresponding polynomial $g_{d}$ selected using the BIC. First, we repeat Wilks's experiment for our choice of parameters given in Equation (3.17) with varying values of $c$. In our simulations, we use a time step size of $h=1 / 2048$, and we warm up the solution for each value of $c$ for time $T=50$ or, equivalently, 102400 time steps. In light of the autocorrelation results discussed in Section 3.2, we sample the trajectory at intervals of $\Delta t=20$ to obtain 3600 samples with the independence suggested by the bound given in (3.18).

These parameterizations demonstrate a strong dependence of the unresolved variables on the resolved variables, but they also suggest a dependence on the dissipation. When $c=20$ (weak dissipation), the tendencies deviate more, and when $c=60$ (strong dissipation), the tendencies deviate less and the fit looks more like a straight line (see Figures 4.1 and 4.2). These parameterizations suggest a direct correlation between dissipation and dependence: As the dissipation increases, the unresolved variables depend more strongly on the resolved variables.

These graphs suggest that for modest values of $c$ between 4 and 20, a linear regression is not sufficient to capture the $\boldsymbol{X}$ dependency in the unresolved tendencies. However, when $c$ is large


Figure 4.1: Stochastic parameterization using $F=20, m_{1}=20, m_{2}=0.002$, and $c=20$. Here, $\tau$ is the effect of $\boldsymbol{V}$ on the first oscillator influenced by the position of the first oscillator.


Figure 4.2: Stochastic parameterization similar to Figure 4.1, except with $c=60$.
(e.g., $c=100$ ), a linear regression may be enough. This motivates us to explore how $c$ affects the degree of the best polynomial model.

### 4.1 Variance Analysis of the Parameterization

Define the standard deviation $s=s(c)$ from the regression polynomial of degree $d$ as

$$
\begin{equation*}
s(c)=\sqrt{\frac{1}{N-p} \sum_{n=1}^{N}\left(G_{1}\left(t_{n}\right)-g_{d}\left(\boldsymbol{X}_{1}\left(t_{n}\right)\right)^{2}\right.}=\sqrt{\frac{1}{N-p} \sum_{n=1}^{N}\left(e_{1}\left(t_{n}\right)\right)^{2}} \tag{4.3}
\end{equation*}
$$

where

$$
\begin{equation*}
G(t)=\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}(t) \tag{4.4}
\end{equation*}
$$

Here, $p=d+1$ to account for the $y$-intercept, and $t_{1}$ represents time sufficiently far in the future such that the solution $\boldsymbol{X}, \boldsymbol{V}$ reflects the long-time behavior of the parameters, particularly the value of $c$; moreover, the spacing between the times $t_{n+1}-t_{n}=0.25$ is large enough that the values of the unresolved tendencies appear independent. Note that $s(c)$ is computed based only on the tendencies of the first oscillator, $\boldsymbol{X}_{1}$. Given the symmetry in $\sigma$ (see Figure 3.1), $\boldsymbol{X}_{k}$ for any fixed $k$ should have identical statistics. At the same time, due to the coupling between adjacent oscillators, $\boldsymbol{X}_{k}$ and $\boldsymbol{X}_{k+1}$ will be correlated. To avoid this correlation, we use only one oscillator when computing $s(c)$.


Figure 4.3: Standard deviation $s$ vs. $c$ for polynomial regressions of degrees 1-7.

By calculating the tendencies $G\left(t_{n}\right)$ for $c=1,2, \ldots, 100$ and obtaining the standard deviations of the tendencies from the regression polynomials of degrees 1-7 (see Figure 4.3), we see that the differences in the standard deviations are negligible for $c>30$. Therefore, increasing the degree of the polynomial when $c$ is large does not significantly decrease the deviation of the tendencies
about the best fit. Before focusing further on the degree of the polynomial, we observe that for all polynomials regardless of degree, standard deviation of the tendencies about the fit decreases as $c$ increases. Therefore, not only may a polynomial of lower degree be needed when $c$ is large, but that polynomial fits the tendencies with less error. This is consistent with our intuition that as $c$ increases, the dynamical effects of $\boldsymbol{V}$ are more predictable and, therefore, solving the inverse problem to find $\boldsymbol{V}$ becomes less important.

Although Figure 4.3 suggests that higher-order regressions are unnecessary for large $c$, it turns out that the standard deviation as defined in Equation (4.3) may not fully reflect the significant contributions of higher-order terms to the regression. To further explore these contributions, we consider the following hypothesis test.

Recall that $g_{d}$ is the $d^{\text {th }}$-degree regression polynomial whose coefficients $b_{0}, b_{1}, \ldots, b_{d}$ minimize the deviation $s(c)$. We hypothesize the following:

$$
\begin{aligned}
H_{0}: & \text { Only } b_{0}, b_{1} \neq 0 \\
H_{1, d}: & b_{i} \neq 0 \text { for all } i=2, \ldots, d,
\end{aligned}
$$

where $d$ is the degree of the regression polynomial.

To test $H_{0}$, we perform an ANOVA analysis on regression models of degrees 1-7. Table 4.1 shows that the $p$-values are all below the $5 \%$ significance criterion for $c \leq 50$, indicating that the higher-order terms of the regression polynomial are nonzero; thus, a linear regression polynomial is insufficient to model the data. However, when $c=100$, all the $p$-values are above the $5 \%$ significance criterion, meaning that we fail to find a linear polynomial insufficient.

|  | $c=10$ | $c=20$ | $c=50$ | $c=100$ |
| :---: | :---: | :---: | :---: | :---: |
| $p_{12}$ | $2.2 \times 10^{-16}$ | $2.2 \times 10^{-16}$ | $8.335 \times 10^{-6}$ | 0.1573 |
| $p_{13}$ | $2.2 \times 10^{-16}$ | $2.2 \times 10^{-16}$ | $2.582 \times 10^{-8}$ | 0.2386 |
| $p_{14}$ | $2.2 \times 10^{-16}$ | $2.2 \times 10^{-16}$ | $5.350 \times 10^{-8}$ | 0.3835 |
| $p_{15}$ | $2.2 \times 10^{-16}$ | $2.2 \times 10^{-16}$ | $1.427 \times 10^{-7}$ | 0.3743 |
| $p_{16}$ | $2.2 \times 10^{-16}$ | $2.2 \times 10^{-16}$ | $6.052 \times 10^{-8}$ | 0.4157 |
| $p_{17}$ | $2.2 \times 10^{-16}$ | $2.2 \times 10^{-16}$ | $1.265 \times 10^{-8}$ | 0.4730 |

Table 4.1: Comparison of $p$-values for various values of $c$. Here, $p_{i j}$ represents the ANOVA analysis of degree- $i$ vs. degree- $j$ regression polynomials.

This analysis was performed by calculating a set of tendencies using the RK4 method with our standard parameters, using the tendencies to create linear and polynomial regression models of degrees $2-7$, and comparing the linear model against each of the polynomial models. In particular, we calculated the $p$-values using the standard ANOVA $F$-test.

### 4.2 Bayesian Selection of Polynomial Degree

The ANOVA analysis rejects the null hypothesis and suggests that a nonlinear regression polynomial is necessary when $c$ is small to sufficiently model the effects of $\boldsymbol{V}$ on $\boldsymbol{X}$, but the test does not specify the degree of the regression polynomial. We use the BIC to determine this. In particular, for each $c$, we calculate the BIC for each regression polynomial of degrees $1-7$, and we define the degree of the regression polynomial as the degree such that the BIC is minimized. Unfortunately, a log-linear plot of BIC v. $c$ is similar to Figure 4.3, and our ANOVA analysis suggests a linear polynomial is insufficient to model the data; thus, we cannot use the BIC alone to determine the degree of the regression polynomial. To get around this problem, we subtract the minimum BIC for each BIC value. The results are seen in Figures 4.4 and 4.5. For $c<20$, the degree of the regression polynomial is clearly higher order (i.e., 6 or 7 ). The polynomial becomes simpler between $c=20$ and $c=40$, and is linear or quadratic when $c>40$. Clearly, large values of $c$ are needed to justify the use of a linear regression polynomial, confirming the results of our ANOVA analysis.


Figure 4.4: $\mathrm{BIC}_{n}-\min \mathrm{BIC}_{n}$ for $c=4, \ldots, 100$, where $\mathrm{BIC}_{n}$ is the BIC of the regression polynomial of degree $n$ with $n=1, \ldots, 7$. Notice the regression polynomial is of high degree when $c \ll 30$ and of low degree when $c \gg 30$.

As illustrated in Table 4.1 and Figures 4.3-4.5, the effects of the small scales can be effectively modeled by a linear or quadratic regression polynomial when $c$ is large. Intuitively, this reflects the effect that when $c$ is large, the linear term in Equation (3.9b) dominates the nonlinear term, and the small-scale oscillators in the second layer essentially stop oscillating. While this is an interesting observation about stochastic parameterization and how the mean state of the $\boldsymbol{V}$


Figure 4.5: Degree of the regression polynomial vs. $c$, where the degree is the first value $n$ such that $\mathrm{BIC}_{n}-\min \mathrm{BIC}_{n} \leq 100$. When $c$ is large, the effect of the small oscillators is not very complicated; thus, a linear or quadratic regression polynomial is sufficient to model the effects of the small scales.
variables affects the $\boldsymbol{X}$ variables, another aspect of the above statistical study is to characterize the values of $c$ for which the small-scale oscillators play an important dynamical role.

The ANOVA $F$-test, AIC, and BIC tests show that values of $c$ around 20 are in the parameter regime where the motion of the small scales is important when the other parameters are $F=$ $20, m_{1}=20$, and $m_{2}=0.002$. For the remainder of this dissertation, we focus on recovering the small scales for the parameter regime where $c \approx 20$. These are the parameters for which the data assimilation problem is interesting.

## Chapter 5

## Rigorous Analytic Bounds

We have made a preliminary exploration of the dynamics of the two-layer Lorenz '96 system, and we have identified parameters for which the data assimilation problem is interesting. With respect to any complex dynamical system, it is useful to know which relative scales of dissipation and amplification will cause the fast layer dynamics to be exponentially determined by the state of the slow layer. Thus, we proceed to our main goal of recovering the values of $\boldsymbol{V}$ from observations of $\boldsymbol{X}$. In this section, our focus is on developing new mathematical analysis that yields rigorous bounds under which a simple coupling scheme allows the recovery of $\boldsymbol{V}$ over time, and we further illustrate our theory using numerics. Although our techniques are based on well-known energy methods, their application here uniquely depends on our rescaling of the $\boldsymbol{V}$ equation and on the structure of the coupling between the two layers.

In particular, we wish to identify the choices of $c, m_{1}$, and $m_{2}$ for which the dynamics of the system Equation (3.9) in the faster variables will be exponentially determined by the slower variables. using the system

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{5.1a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{5.1b}\\
& \frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{5.1c}
\end{align*}
$$

The above system defines an ODE governing the evolution in time of $\|\boldsymbol{V}-\boldsymbol{v}\|$, where

$$
\begin{equation*}
\boldsymbol{V}=\left(V_{1}, \ldots, V_{J K}\right), \quad \boldsymbol{v}=\left(v_{1}, \ldots, v_{J K}\right) \tag{5.2}
\end{equation*}
$$

and $\|\cdot\|$ is the Euclidean norm unless otherwise noted. We say $\boldsymbol{V}$ and $\boldsymbol{v}$ are synchronized if $\lim _{t \rightarrow \infty}\|\boldsymbol{V}-\boldsymbol{v}\|=0$. The synchronization is governed by Equations (5.1b) and (5.1c).

Beforehand, we need to show a bound on the free-running solution denoted by $\boldsymbol{X}$ and $\boldsymbol{V}$. This is provided in the following section and illustrated numerically in Section 5.2. A rigorous proof that $\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ when $c$ is large enough is provided in Section 6.1.

### 5.1 Bounds on the Reference Solution

Before setting bounds on $\boldsymbol{X}$ and $\boldsymbol{V}$ separately, we define the function

$$
\begin{equation*}
\psi=m_{1}\|\boldsymbol{X}\|^{2}+m_{2}\|\boldsymbol{V}\|^{2} \tag{5.3}
\end{equation*}
$$

to simultaneously represent the behavior of $\boldsymbol{X}$ and $\boldsymbol{V}$ and the effects of the transfer of energy between the large and small scales. With this, we first prove the following lemma for large $c$ :

Lemma 5.1. Let $\psi$ be as above. Suppose $c>1-\frac{1}{2 \delta}$ with $\delta>\frac{1}{2}$. Then

$$
\begin{equation*}
\psi \leq \psi_{0} e^{-\alpha t}+\frac{\nu}{\alpha}\left(1-e^{-\alpha t}\right) \tag{5.4}
\end{equation*}
$$

for

$$
\begin{equation*}
\alpha=2-\frac{1}{\delta} \quad \text { and } \quad \nu=m_{1} \delta\|\boldsymbol{F}\|^{2} . \tag{5.5}
\end{equation*}
$$

Proof. We first dot Equation (3.9a) on the right by $\boldsymbol{X}$ and Equation (3.9b) on the right by $\boldsymbol{V}$ to obtain

$$
\begin{align*}
\frac{d\|\boldsymbol{X}\|^{2}}{d t} & =-2\|\boldsymbol{X}\|^{2}+2 \boldsymbol{X} \cdot \boldsymbol{F}-\frac{2}{m_{1}} \boldsymbol{X} \cdot \sigma^{\top} \boldsymbol{V}  \tag{5.6a}\\
\frac{d}{d t}\|\boldsymbol{V}\|^{2} & =-2 c\|\boldsymbol{V}\|^{2}+\frac{2}{m_{2}} \boldsymbol{V} \cdot \sigma \boldsymbol{X} \tag{5.6b}
\end{align*}
$$

Because $\frac{d \psi}{d t}=2 m_{1}\|\boldsymbol{X}\|+2 m_{2}\|\boldsymbol{V}\|$ and $-\boldsymbol{X} \cdot \sigma^{\top} \boldsymbol{V}+\boldsymbol{V} \cdot \sigma \boldsymbol{X}=0$, we can combine Equations (5.6a) and (5.6b) into a single equation:

$$
\begin{equation*}
\frac{1}{2} \frac{d \psi}{d t}=-m_{1}\|\boldsymbol{X}\|^{2}+m_{1} \boldsymbol{X} \cdot \boldsymbol{F}-m_{2} c\|\boldsymbol{V}\|^{2} \tag{5.7}
\end{equation*}
$$

By the Cauchy-Schwarz inequality, $|\boldsymbol{X} \cdot \boldsymbol{F}| \leq\|\boldsymbol{X}\|\|\boldsymbol{F}\|$. Take $\delta>0$ and apply Young's inequality to get $\|\boldsymbol{X}\|\|\boldsymbol{F}\| \leq \frac{1}{2 \delta}\|\boldsymbol{X}\|^{2}+\frac{\delta}{2}\|\boldsymbol{F}\|^{2}$. Thus, (5.7) becomes

$$
\begin{equation*}
\frac{1}{2} \frac{d \psi}{d t} \leq-m_{1}\|\boldsymbol{X}\|^{2}\left(1-\frac{1}{2 \delta}\right)+\frac{m_{1} \delta}{2}\|\boldsymbol{F}\|^{2}-m_{2} c\|\boldsymbol{V}\|^{2} \tag{5.8}
\end{equation*}
$$

Because $c>1-\frac{1}{2 \delta}$,

$$
\begin{equation*}
\frac{1}{2} \frac{d \psi}{d t} \leq-\left(1-\frac{1}{2 \delta}\right) \psi+\frac{m_{1} \delta}{2}\|\boldsymbol{F}\|^{2} \tag{5.9}
\end{equation*}
$$

Thus, $\dot{\psi} \leq-\alpha \psi+\nu$, where $\alpha$ and $\nu$ are as in (5.5).
Next, we show $\frac{d}{d t} \psi e^{\alpha t} \leq \nu e^{\alpha t}$. We multiply

$$
\begin{equation*}
\frac{d}{d t} \psi e^{\alpha t}=\dot{\psi} e^{\alpha t}+\psi \alpha e^{\alpha t} \leq(-\alpha \psi+\nu) e^{\alpha t}+\psi \alpha e^{\alpha t}=\nu e^{\alpha t} \tag{5.10}
\end{equation*}
$$

through by the integrating factor $e^{\alpha t}$ and integrate by parts from 0 to $t$ to get

$$
\begin{equation*}
\psi e^{\alpha t}-\psi_{0} \leq \frac{\nu}{\alpha}\left(e^{\alpha t}-1\right) \tag{5.11}
\end{equation*}
$$

Dividing by the integrating factor and rearranging the inequality yields our final result.

Notice in Equation (5.7) that as $c \rightarrow \infty$, we see $\frac{d \psi}{d t} \rightarrow-\infty$. Thus, as long as $\|\boldsymbol{V}\|$ is nonzero, we have $\|\psi\| \rightarrow 0$. Also, because $\alpha>0$, the differential inequality (5.11) will be used later to prove $\psi$ is bounded.

For small $c$, we have the following bound on $\psi$ :
Lemma 5.2. Let $\psi$ be as before. Suppose $0<c<1-\frac{1}{2 \delta}$ with $\delta>\frac{1}{2}$. Then

$$
\begin{equation*}
\psi \leq \psi_{0} e^{-\alpha t}+\frac{\nu}{\alpha}\left(1-e^{-\alpha t}\right) \tag{5.12}
\end{equation*}
$$

for

$$
\begin{equation*}
\alpha=2 c \quad \text { and } \quad \nu=m_{1} \delta\|\boldsymbol{F}\|^{2} \tag{5.13}
\end{equation*}
$$

Proof. The proof is the same as that of Lemma 5.1 up to and including Equation (5.8). Then $0>-c \geq \frac{1}{2 \delta}-1$ implies

$$
\begin{align*}
\frac{1}{2} \frac{d \psi}{d t} & \leq-m_{1}\|\boldsymbol{X}\|^{2} c+\frac{m_{1} \delta}{2}\|\boldsymbol{F}\|^{2}-m_{2} c\|\boldsymbol{V}\|^{2} \\
& =-c \psi+\frac{m_{1} \delta}{2}\|\boldsymbol{F}\|^{2} \tag{5.14}
\end{align*}
$$

The rest of the proof proceeds as in Lemma 5.1.

To compare the bound in Lemma 5.1 with numerical simulations, note first that

$$
\begin{equation*}
\limsup _{t \rightarrow \infty} \psi(t)=\frac{\nu}{\alpha}=\frac{m_{1} \delta\|\boldsymbol{F}\|^{2}}{2-\frac{1}{\delta}} \tag{5.15}
\end{equation*}
$$

Minimizing the right-hand side with respect to $\delta$ under the condition that $\delta>1 / 2$ yields the bound $m_{1}\|\boldsymbol{F}\|^{2}$ when $\delta=1$. Figure 5.1 shows that for $c=20, F=20, m_{1}=20$, and $m_{2}=0.002$, the numerical value for $\psi$ is about one decimal order of magnitude smaller than this bound. Since we are primarily interested in recovering the small scales, we also include the computed values for $m_{2}\|\boldsymbol{V}\|^{2}$ and $m_{2}\|\boldsymbol{V}\|_{\infty}^{2}$ for reference to illustrate how much of the total energy is in the small scales. Since the estimate

$$
\begin{equation*}
\|\boldsymbol{V}\|^{2}=\sum_{j=1}^{J K}\left|V_{j}\right|^{2} \leq \sum_{j=1}^{J K}\|\boldsymbol{V}\|_{\infty}^{2}=J K\|\boldsymbol{V}\|_{\infty}^{2} \tag{5.16}
\end{equation*}
$$

will be considered later in the analysis, we also show the time evolution of $m_{2} J K\|\boldsymbol{V}\|_{\infty}^{2}$ in Figure 5.1. Note that the magnitude of this term is similar to that of the total energy.


Figure 5.1: Comparison of $\psi$, its terms, and the bound in Lemma 5.1, with $c=20, F=$ $20, m_{1}=20$, and $m_{2}=0.002$.

These bounds on $\psi$ prove that $\boldsymbol{X}$ and $\boldsymbol{V}$ remain finite for fixed values of $\alpha$ and $\nu$, allowing us to derive specific bounds on $\boldsymbol{V}$. The first is a bound for large $c$ :

Proposition 5.3. Under the assumption $c>1-\frac{1}{2 \delta}$ with $\delta>\frac{1}{2}$ of Lemma 5.1,

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\| \leq M_{1}, \quad \text { where } \quad M_{1}=\|\boldsymbol{F}\| \sqrt{\frac{m_{1} \delta}{\left(2-\frac{1}{\delta}\right) m_{2}}} . \tag{5.17}
\end{equation*}
$$

Moreover, under the assumption $0<c<1-\frac{1}{2 \delta}$ with $\delta>\frac{1}{2}$ of Lemma 5.2,

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\| \leq M_{1}^{\prime}, \quad \text { where } \quad M_{1}^{\prime}=\|\boldsymbol{F}\| \sqrt{\frac{m_{1} \delta}{2 m_{2} c}} \tag{5.18}
\end{equation*}
$$

Proof. By definition of $\psi$ and the result of Lemma 5.1,

$$
\begin{equation*}
m_{2}\|\boldsymbol{V}\|^{2} \leq \psi \leq\left(\psi_{0}-\frac{\nu}{\alpha}\right) e^{-\alpha t}+\frac{\nu}{\alpha} \tag{5.19}
\end{equation*}
$$

Taking the lim sup through the inequality yields

$$
\begin{equation*}
\limsup _{t \rightarrow \infty} m_{2}\|\boldsymbol{V}\|^{2} \leq \limsup _{t \rightarrow \infty} \psi \leq \limsup _{t \rightarrow \infty}\left(\psi_{0}-\frac{\nu}{\alpha}\right) e^{-\alpha t}+\frac{\nu}{\alpha} \tag{5.20}
\end{equation*}
$$

Then because $\limsup _{t \rightarrow \infty} e^{-\alpha t}=0$ and $\frac{\nu}{\alpha}$ is a constant, the inequality becomes

$$
\begin{equation*}
\limsup _{t \rightarrow \infty} m_{2}\|\boldsymbol{V}\|^{2} \leq \frac{\nu}{\alpha} \tag{5.21}
\end{equation*}
$$

Dividing both sides by $m_{2}$ yields

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\|^{2} \leq \frac{\nu}{m_{2} \alpha}=\frac{m_{1} \delta\|\boldsymbol{F}\|^{2}}{\left(2-\frac{1}{\delta}\right) m_{2}} \tag{5.22}
\end{equation*}
$$

Taking the square root of both sides yields the first result of our proposition.

Next, by definition of $\psi$ and the result of Lemma 5.2, we write (5.21) as

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\|^{2} \leq \frac{\nu}{m_{2} \alpha}=\frac{m_{1} \delta\|\boldsymbol{F}\|^{2}}{2 m_{2} c} \tag{5.23}
\end{equation*}
$$

Taking the square root of both sides yields the second result.

We remark that the case under which (5.18) holds requires $c<1$. For the range of parameters defined in (3.15) and further studied in Section 4, the interesting values of $c$ are much larger. In particular, for much of our computations, we focus on $c \approx 20$, which means the assumptions of (5.17) hold.

The following proposition establishes a bound on $\boldsymbol{V}$.

Proposition 5.4. When $c>1$,

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\| \leq M_{2}, \quad \text { where } \quad M_{2}=\frac{\|\sigma\|_{2}\|\boldsymbol{F}\|}{m_{2} \sqrt{c}} \tag{5.24}
\end{equation*}
$$

Proof. By (5.17),

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\left(m_{1}\|\boldsymbol{X}\|^{2}+m_{2}\|\boldsymbol{V}\|^{2}\right) \leq \frac{m_{1} \delta\|\boldsymbol{F}\|^{2}}{2-\frac{1}{\delta}} . \tag{5.25}
\end{equation*}
$$

Taking $\delta=1$ yields

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\left(m_{1}\|\boldsymbol{X}\|^{2}+m_{2}\|\boldsymbol{V}\|^{2}\right) \leq m_{1}\|\boldsymbol{F}\|^{2} \tag{5.26}
\end{equation*}
$$

This implies $\|\boldsymbol{X}\| \leq(1+\epsilon)\|\boldsymbol{F}\|$ for sufficiently large $t \geq T$.

Next, recall Equation (5.6b). By the Cauchy-Schwarz inequality, then Young's inequality and the fact that $c>1$,

$$
\begin{align*}
\frac{d}{d t}\|\boldsymbol{v}\|^{2} & =-2 c\|\boldsymbol{V}\|^{2}+\frac{2}{m_{2}} \boldsymbol{V} \cdot \sigma \boldsymbol{X} \\
& \leq-2 c\|\boldsymbol{V}\|^{2}+c\|\boldsymbol{V}\|^{2}+\frac{1}{m_{2}^{2}}\|\sigma\|_{2}^{2}\|\boldsymbol{X}\|^{2} \\
& \leq-c\|\boldsymbol{V}\|^{2}+\frac{(1+\epsilon)^{2}\|\sigma\|_{2}^{2}\|\boldsymbol{F}\|^{2}}{m_{2}^{2}} \tag{5.27}
\end{align*}
$$

for $t \geq T$. Multiplying through by the integrating factor $e^{c t}$ yields

$$
\begin{equation*}
\frac{d}{d t}\|\boldsymbol{v}\|^{2} e^{c t} \leq e^{c t} \frac{(1+\epsilon)^{2}\|\sigma\|_{2}^{2}\|\boldsymbol{F}\|^{2}}{m_{2}^{2}}, \quad t \geq T \tag{5.28}
\end{equation*}
$$

and integrating over the interval $[T, t]$ yields

$$
\begin{equation*}
\|\boldsymbol{V}\|^{2} e^{c t}-\|\boldsymbol{V}(T)\|^{2} e^{c T} \leq \frac{1}{c}\left(e^{c t}-e^{c T}\right) \frac{(1+\epsilon)^{2}\|\sigma\|_{2}^{2}\|\boldsymbol{F}\|^{2}}{m_{2}^{2}} \tag{5.29}
\end{equation*}
$$

Isolating $\|\boldsymbol{V}\|^{2}$ yields

$$
\begin{equation*}
\|\boldsymbol{V}\|^{2} \leq\|\boldsymbol{V}(T)\|^{2} e^{-c(t-T)}+\frac{1}{c}\left(1-e^{-c(t-T)}\right) \frac{(1+\epsilon)^{2}\|\sigma\|_{2}^{2}\|\boldsymbol{F}\|^{2}}{m_{2}^{2}} \tag{5.30}
\end{equation*}
$$

Taking $t \rightarrow \infty$ yields

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\|^{2} \leq \frac{(1+\epsilon)^{2}\|\sigma\|_{2}^{2}\|\boldsymbol{F}\|^{2}}{c m_{2}^{2}} \quad \text { for all } \quad \epsilon>0 \tag{5.31}
\end{equation*}
$$

Letting $\epsilon \rightarrow 0$ and taking the square root of both sides yields the result.

The next proposition establishes a bound on the time average of $\|\boldsymbol{V}\|_{\infty}$.

## Proposition 5.5.

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty} \leq M_{3}, \quad \text { where } \quad M_{3}=\min \left\{\frac{1}{\sqrt{c m_{2}}}, \sqrt{\frac{m_{1}}{2}}\right\} \frac{\|\boldsymbol{F}\|}{\sqrt{c m_{2}}} \tag{5.32}
\end{equation*}
$$

Proof. Recall Equation (3.9b), from the rescaled system:

$$
\begin{equation*}
\frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{5.33}
\end{equation*}
$$

Solve for $c \boldsymbol{V}$, and take the dot product of both sides with respect to $\boldsymbol{V}$. This yields

$$
\begin{equation*}
c \boldsymbol{V} \cdot \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \cdot \boldsymbol{V}-\frac{d \boldsymbol{V}}{d t} \cdot \boldsymbol{V}-\widetilde{B}(\boldsymbol{V}, \boldsymbol{V}) \cdot \boldsymbol{V} \tag{5.34}
\end{equation*}
$$

Note $\widetilde{B}(\boldsymbol{V}, \boldsymbol{V}) \cdot \boldsymbol{V}=0$ and $\boldsymbol{V} \cdot \boldsymbol{V}=\|\boldsymbol{V}\|^{2}$. Thus, the equation further simplifies:

$$
\begin{equation*}
c\|\boldsymbol{V}\|^{2}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \cdot \boldsymbol{V}-\frac{1}{2} \frac{d}{d t}\|\boldsymbol{V}\|^{2} . \tag{5.35}
\end{equation*}
$$

Next, we will integrate both sides with respect to $t$ and apply the Cauchy-Schwarz inequality, then Young's inequality. This yields

$$
\begin{align*}
c \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t & \leq \frac{1}{m_{2}}\left[\int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t\right]^{1 / 2}-\frac{1}{2}\|\boldsymbol{V}(s)\|^{2}+\frac{1}{2}\left\|\boldsymbol{V}_{0}\right\|^{2} \\
& \leq \frac{1}{m_{2}}\left[\frac{1}{2 c m_{2}} \int_{0}^{s}\|\sigma \boldsymbol{X}\| d t+\frac{c m_{2}}{2} \int_{0}^{s}\|\boldsymbol{V}\| d t\right]-\frac{1}{2}\|\boldsymbol{V}\|^{2}+\frac{1}{2}\left\|\boldsymbol{V}_{0}\right\|^{2}  \tag{5.36}\\
& =\frac{1}{2 c m_{2}^{2}} \int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t+\frac{c}{2} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t-\frac{1}{2}\|\boldsymbol{V}\|^{2}+\frac{1}{2}\left\|\boldsymbol{V}_{0}\right\|^{2} .
\end{align*}
$$

Notice we can simplify this inequality by subtracting $\frac{c}{2} \int_{0}^{s}\|\boldsymbol{V}\|^{2}$ from both sides. This yields

$$
\begin{equation*}
\frac{c}{2} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t \leq \frac{1}{2 c m_{2}^{2}} \int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t-\frac{1}{2}\|\boldsymbol{V}\|^{2}+\frac{1}{2}\left\|\boldsymbol{V}_{0}\right\|^{2} \tag{5.37}
\end{equation*}
$$

Next, we divide both sides by $c s$ and multiply both sides by 2 . This yields

$$
\begin{equation*}
\frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t \leq \frac{1}{c^{2} m_{2}^{2} s} \int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t-\frac{1}{s}\|\boldsymbol{V}\|^{2}+\frac{1}{s}\left\|\boldsymbol{V}_{0}\right\|^{2} \tag{5.38}
\end{equation*}
$$

Now let $s \rightarrow \infty$. Then $\frac{1}{s}\|\boldsymbol{V}(s)\|^{2}$ and $\frac{1}{s}\left\|\boldsymbol{V}_{0}\right\|^{2}$ vanish, and the inequality becomes

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t \leq \frac{1}{c^{2} m_{2}^{2}} \limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t \tag{5.39}
\end{equation*}
$$

Let us isolate and determine bounds for $\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t$ and $\lim \sup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2}$. We will start with Equations (5.1a) and (5.1b):

$$
\begin{align*}
\frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X} & =-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}+\boldsymbol{F}  \tag{5.40}\\
\frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V} & =\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{5.41}
\end{align*}
$$

Dotting (5.1a) with $\boldsymbol{X}$ and (5.1b) with $\boldsymbol{V}$ yields

$$
\begin{align*}
\frac{1}{2} \frac{d\|\boldsymbol{X}\|^{2}}{d t}+\|\boldsymbol{X}\|^{2} & =-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V} \cdot \boldsymbol{X}+\boldsymbol{F} \cdot \boldsymbol{X}  \tag{5.42}\\
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2}+c\|\boldsymbol{V}\|^{2} & =\frac{1}{m_{2}} \sigma \boldsymbol{X} \cdot \boldsymbol{V} \tag{5.43}
\end{align*}
$$

Next, we multiply (5.42) by $m_{1}$ and (5.43) by $m_{2}$, and we add (5.42) and (5.43). Then because $\sigma^{\top} \boldsymbol{V} \cdot \boldsymbol{X}=\sigma \boldsymbol{X} \cdot \boldsymbol{V}$, we get

$$
\begin{equation*}
m_{1}\|\boldsymbol{X}\|^{2}+c m_{2}\|\boldsymbol{V}\|^{2}=m_{1} \boldsymbol{F} \cdot \boldsymbol{X}-\frac{m_{1}}{2} \frac{d\|\boldsymbol{X}\|^{2}}{d t}-\frac{m_{2}}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2} \tag{5.44}
\end{equation*}
$$

Next, we integrate from 0 to $s$, multiplying by $\frac{1}{s}$, and then apply the Cauchy-Schwarz and Young inequalities to get

$$
\begin{align*}
& \frac{1}{s} \int_{0}^{s} m_{1}\|\boldsymbol{X}\|^{2}+c m_{2}\|\boldsymbol{V}\|^{2} d t  \tag{5.45}\\
& \leq \frac{1}{s} \int_{0}^{s} m_{1}\|\boldsymbol{F}\|\|\boldsymbol{X}\| d t-\frac{m_{1}}{2 s}\left(\|\boldsymbol{X}(s)\|-\left\|\boldsymbol{X}_{0}\right\|\right)-\frac{m_{2}}{2 s}\left(\|\boldsymbol{V}(s)\|-\left\|\boldsymbol{V}_{0}\right\|\right) \\
& \quad \leq \frac{1}{s}\left[\frac{m_{1}}{2} \int_{0}^{s}\|\boldsymbol{F}\|^{2} d t+\frac{m_{1}}{2} \int_{0}^{s}\|\boldsymbol{X}\|^{2} d t\right]-\frac{m_{1}}{2 s}\left(\|\boldsymbol{X}(s)\|-\left\|\boldsymbol{X}_{0}\right\|\right)-\frac{m_{2}}{2 s}\left(\|\boldsymbol{V}(s)\|-\left\|\boldsymbol{V}_{0}\right\|\right)
\end{align*}
$$

Let $s \rightarrow \infty$. Then

$$
\begin{equation*}
\limsup _{s \rightarrow \infty}\left[\frac{m_{1}}{s} \int_{0}^{s}\|\boldsymbol{X}\|^{2}+\frac{c m_{2}}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t\right] \leq \frac{m_{1}}{2}\|\boldsymbol{F}\|^{2}+\limsup _{s \rightarrow \infty} \frac{m_{1}}{2 s} \int_{0}^{s}\|\boldsymbol{X}\|^{2} d t . \tag{5.46}
\end{equation*}
$$

Subtracting $\lim \sup _{s \rightarrow \infty} \frac{m_{1}}{2 s} \int_{0}^{s}\|\boldsymbol{X}\|^{2} d t$ from both sides yields

$$
\begin{equation*}
\limsup _{s \rightarrow \infty}\left[\frac{m_{1}}{2 s} \int_{0}^{s}\|\boldsymbol{X}\|^{2}+\frac{c m_{2}}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t\right] \leq \frac{m_{1}}{2}\|\boldsymbol{F}\|^{2} \tag{5.47}
\end{equation*}
$$

Because both these integrals are nonnegative, we have

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{X}\|^{2} d t \leq\|\boldsymbol{F}\|^{2} \tag{5.48}
\end{equation*}
$$

and, as a bonus,

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t s \leq \frac{m_{1}}{2 c m_{2}}\|\boldsymbol{F}\|^{2} . \tag{5.49}
\end{equation*}
$$

By Equation (5.48) and the fact that $\|\sigma \boldsymbol{X}\| \leq\|\sigma\|_{2}\|\boldsymbol{X}\|$, we get

$$
\begin{equation*}
\frac{1}{c^{2} m_{2}^{2}} \limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\sigma \boldsymbol{X}\|^{2} d t \leq \frac{\|\sigma\|_{2}^{2}}{c^{2} m_{2}^{2}}\|\boldsymbol{F}\|^{2} \tag{5.50}
\end{equation*}
$$

Next, by Jensen's inequality,

$$
\begin{equation*}
\left(\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\| d t\right)^{2} \leq \limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t \tag{5.51}
\end{equation*}
$$

and Equation (5.49) becomes

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\| d t \leq\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{2 c m_{2}}} . \tag{5.52}
\end{equation*}
$$

Recall that our goal was to get a bound on $\|\boldsymbol{V}\|_{\infty}$. To do this, note that $\|\boldsymbol{V}\|_{\infty} \leq\|\boldsymbol{V}\|$. Therefore,

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty} \leq\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{2 c m_{2}}} \tag{5.53}
\end{equation*}
$$

Then by (5.39) and (5.48), we have

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|^{2} d t \leq \frac{\|\sigma\|_{2}^{2}\|\boldsymbol{F}\|^{2}}{c^{2} m_{2}^{2}} \tag{5.54}
\end{equation*}
$$

By Jensen's inequality, we have

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty} d t \leq \limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\| d t \leq \frac{\|\sigma\|_{2}\|\boldsymbol{F}\|}{c m_{2}} \tag{5.55}
\end{equation*}
$$

Combining (5.53) with (5.55), and noting $\|\sigma\|_{2}=1$ yields

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty} \leq M_{3}, \quad \text { where } \quad M_{3}=\min \left\{\frac{1}{\sqrt{c m_{2}}}, \sqrt{\frac{m_{1}}{2}}\right\} \frac{\|\boldsymbol{F}\|}{\sqrt{c m_{2}}} \tag{5.56}
\end{equation*}
$$

This result shows that the size of $\boldsymbol{V}$ is affected by decreasing $F$ (injecting less energy into the large scales) or increasing $c$ (dissipating more energy out of the small scales). If $F$ is small, this affects the dynamics of the $X_{k}$ by decreasing their energy to the point where they could transition from a chaotic to a non-chaotic regime. If $c$ is large, the chaotic dynamics of the $X_{k}$ can be shown to persist (see Proposition 5.8), and the motion of the small scales follows the coupling from $X_{k}$ more closely due to the increased dissipation.

The following proposition shows that as $c \rightarrow \infty, \boldsymbol{X}$ remains nontrivial and $\boldsymbol{V}$ vanishes. Specifically, we show that the limiting solution exists and is actually a solution to the one-layer system. Such a a result is intuitively plausible because of the way we rescaled the original two-layer system. Since it is not always true that the limiting solution satisfies the limit equations, having a mathematical proof in our case is not only significant but illustrates the value of the rescaling in Chapter 3.1. Moreover, this result provides some theoretical support to the statistical observations in Section 4, which show that the effects of the small scales on the large scales may be parameterized by a first-degree polynomial when $c$ is large.

Given initial conditions $\boldsymbol{X}(0) \in \mathbb{R}^{K}$ and $\boldsymbol{V}(0) \in \mathbb{R}^{J K}$, let $\boldsymbol{X}_{\boldsymbol{c}}$ and $\boldsymbol{V}_{\boldsymbol{c}}$ be the corresponding solutions to the two-layer Lorenz '96 system given by Equations (3.9). Further, we denote as $\boldsymbol{X}_{\infty}$ the solution to the one-layer Lorenz '96 system given by Equation (1.1) with the same $\boldsymbol{X}(0)$ as the initial condition. First, we prove

Proposition 5.6. $\boldsymbol{X}_{\infty}(t)=\lim _{c \rightarrow \infty} \boldsymbol{X}_{\boldsymbol{c}}(t)$ for any fixed $t>0$.

Proof. Recall that $\boldsymbol{X}_{\infty}$ is the solution of

$$
\begin{equation*}
\frac{d \boldsymbol{X}_{\infty}}{d t}+B\left(\boldsymbol{X}_{\infty}, \boldsymbol{X}_{\infty}\right)+\boldsymbol{X}_{\infty}=\boldsymbol{F} \quad \text { with } \quad \boldsymbol{X}_{\infty}(0)=\boldsymbol{X}_{\boldsymbol{c}}(0) \tag{5.57}
\end{equation*}
$$

Let $\boldsymbol{W}=\boldsymbol{X}_{\infty}-\boldsymbol{X}_{\boldsymbol{c}}$. Then $\boldsymbol{W}$ satisfies

$$
\begin{equation*}
\frac{d \boldsymbol{W}}{d t}+B\left(\boldsymbol{X}_{\infty}, \boldsymbol{X}_{\infty}\right)-\widetilde{B}\left(\boldsymbol{X}_{\boldsymbol{c}}, \boldsymbol{X}_{\boldsymbol{c}}\right)+\boldsymbol{W}=\frac{1}{m_{2}} \sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}} \quad \text { and } \quad \boldsymbol{W}_{\mathbf{0}}=\mathbf{0} \tag{5.58}
\end{equation*}
$$

Adding and subtracting $B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{X}_{\boldsymbol{c}}\right)$ from $B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{X}_{\boldsymbol{\infty}}\right)-\widetilde{B}\left(\boldsymbol{X}_{\boldsymbol{c}}, \boldsymbol{X}_{\boldsymbol{c}}\right)$ yields $B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{W}\right)+$ $B\left(\boldsymbol{W}, \boldsymbol{X}_{\boldsymbol{c}}\right) ;$ to this, we subtract and add $B\left(\boldsymbol{W}, \boldsymbol{X}_{\boldsymbol{c}}\right)$ to obtain $B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{W}\right)-B(\boldsymbol{W}, \boldsymbol{W})+$ $B\left(\boldsymbol{W}, \boldsymbol{X}_{\boldsymbol{\infty}}\right)$. Equation (5.57) thus becomes

$$
\begin{equation*}
\frac{d \boldsymbol{W}}{d t}+\boldsymbol{W}=B(\boldsymbol{W}, \boldsymbol{W})-B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{W}\right)-B\left(\boldsymbol{W}, \boldsymbol{X}_{\boldsymbol{\infty}}\right)+\frac{1}{m_{2}} \sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}} \tag{5.59}
\end{equation*}
$$

Dotting it through by $\boldsymbol{W}$ yields

$$
\begin{equation*}
\frac{1}{2} \frac{d\|\boldsymbol{W}\|^{2}}{d t}+\|\boldsymbol{W}\|^{2}=B(\boldsymbol{W}, \boldsymbol{W}) \cdot \boldsymbol{W}-B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{W}\right) \cdot \boldsymbol{W}-B\left(\boldsymbol{W}, \boldsymbol{X}_{\boldsymbol{\infty}}\right) \cdot \boldsymbol{W}+\frac{1}{2 m_{2}} \sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}} \cdot \boldsymbol{W} \tag{5.60}
\end{equation*}
$$

By Lemma 3.1, $B(\boldsymbol{W}, \boldsymbol{W}) \cdot \boldsymbol{W}=0$; thus,

$$
\begin{equation*}
\frac{1}{2} \frac{d\|\boldsymbol{W}\|^{2}}{d t}+\|\boldsymbol{W}\|^{2}=-B\left(\boldsymbol{X}_{\boldsymbol{\infty}}, \boldsymbol{W}\right) \cdot \boldsymbol{W}-B\left(\boldsymbol{W}, \boldsymbol{X}_{\boldsymbol{\infty}}\right) \cdot \boldsymbol{W}+\frac{1}{2} \sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}} \cdot \boldsymbol{W} \tag{5.61}
\end{equation*}
$$

Next, by the triangle, Cauchy-Schwarz, and Young inequalities,

$$
\begin{equation*}
\frac{1}{2} \frac{d\|\boldsymbol{W}\|^{2}}{d t}+\|\boldsymbol{W}\|^{2} \leq 2\left\|\boldsymbol{X}_{\boldsymbol{\infty}}\right\|\|\boldsymbol{W}\|^{2}+\frac{1}{2 m_{2}^{2}}\left\|\sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}}\right\|^{2}+\frac{1}{2}\|\boldsymbol{W}\|^{2} \tag{5.62}
\end{equation*}
$$

To bound $\left\|\boldsymbol{X}_{\infty}\right\|$, we now employ estimates similar to those in Lemma 5.1 as follows. We take the inner product of Equation (5.57) with $\boldsymbol{X}_{\boldsymbol{\infty}}$ to obtain

$$
\begin{equation*}
\frac{1}{2} \frac{d\left\|\boldsymbol{X}_{\infty}\right\|^{2}}{d t}+\left\|\boldsymbol{X}_{\infty}\right\|^{2}=\boldsymbol{F} \cdot \boldsymbol{X}_{\infty} \leq \frac{1}{2}\|\boldsymbol{F}\|^{2}+\frac{1}{2}\left\|\boldsymbol{X}_{\infty}\right\|^{2} \tag{5.63}
\end{equation*}
$$

Multiplying through by the integrating factor $e^{t}$ and integrating from 0 to $t$ yields

$$
\begin{align*}
\left\|\boldsymbol{X}_{\boldsymbol{\infty}}\right\|^{2} & \leq\left\|\boldsymbol{X}_{\infty}(0)\right\|^{2} e^{-t}+\left(1-e^{-t}\right)\|\boldsymbol{F}\|^{2} \\
& \leq\|\boldsymbol{X}(0)\|^{2}+\|\boldsymbol{F}\|^{2} \tag{5.64}
\end{align*}
$$

Multiplying (5.62) through by 2 , isolating $\frac{d \boldsymbol{W}}{d t}$, and factoring out $\|\boldsymbol{W}\|^{2}$ on the right yields

$$
\begin{align*}
\frac{d\|\boldsymbol{W}\|^{2}}{d t} & \leq\left(4\left\|\boldsymbol{X}_{\infty}\right\|-1\right)\|\boldsymbol{W}\|^{2}+\frac{1}{2 m_{2}^{2}}\left\|\sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}}\right\|^{2} \\
& \leq\left(4 \sqrt{\|\boldsymbol{X}(0)\|^{2}+\|\boldsymbol{F}\|^{2}}-1\right)\|\boldsymbol{W}\|^{2}+\frac{1}{2 m_{2}^{2}}\left\|\sigma^{\top} \boldsymbol{V}_{\boldsymbol{c}}\right\|^{2} \tag{5.65}
\end{align*}
$$

By Proposition 5.4 for $t \geq T$ sufficiently large, $\left\|\boldsymbol{V}_{\boldsymbol{c}}(t)\right\|^{2} \leq 2\|\boldsymbol{F}\|^{2} /\left(m_{2}^{2} c\right)$. Since $\left\|\boldsymbol{V}_{\boldsymbol{c}}(t)\right\|$ is continuous on $[0, T]$, then by the Maximum Value Theorem $\left\|\boldsymbol{V}_{\boldsymbol{c}}(t)\right\|$ attains its maximum $\boldsymbol{V}_{\max }$ on $[0, T]$. For easier analysis, we denote

$$
\begin{equation*}
\kappa=4 \sqrt{\|\boldsymbol{X}(0)\|^{2}+\|\boldsymbol{F}\|^{2}}-1 \quad \text { and } \quad \beta_{c}=\frac{1}{2 m_{2}^{2}} \max \left\{\frac{2\|\boldsymbol{F}\|^{2}}{m_{2}^{2} c}, \boldsymbol{V}_{\max }\right\} \tag{5.66}
\end{equation*}
$$

Then for all $t \geq 0$,

$$
\begin{equation*}
\frac{d\|\boldsymbol{W}\|^{2}}{d t} \leq \kappa\|\boldsymbol{W}\|^{2}+\beta_{c} \tag{5.67}
\end{equation*}
$$

which we multiply through by $e^{-\kappa t}$ to obtain

$$
\begin{equation*}
\frac{d\|\boldsymbol{W}\|^{2}}{d t} e^{-\kappa t} \leq e^{-\kappa t} \beta_{c} \tag{5.68}
\end{equation*}
$$

Integrating from 0 to $t$ yields

$$
\begin{equation*}
\|\boldsymbol{W}\|^{2} e^{-\kappa t}-\|\boldsymbol{W}(0)\|^{2} \leq \frac{1}{\kappa}\left(1-e^{-\kappa t}\right) \beta_{c} \tag{5.69}
\end{equation*}
$$

Because $\boldsymbol{W}(0)=0$, the second term vanishes. Then multiplying by $e^{\kappa t}$ yields

$$
\begin{equation*}
\|\boldsymbol{W}\|^{2} \leq \frac{1}{\kappa}\left(e^{\kappa t}-1\right) \beta_{c} \tag{5.70}
\end{equation*}
$$

Finally, by Proposition 5.4 and because $t$ is fixed, $\beta_{c} \rightarrow 0$ as $c \rightarrow \infty$. Thus, $\lim _{c \rightarrow \infty}\|\boldsymbol{W}(t)\|=0$, and so $\lim _{c \rightarrow \infty}\left\|\boldsymbol{X}_{\infty}(t)-\boldsymbol{X}_{\boldsymbol{c}}(t)\right\|=0$, for any fixed $t$. This completes the proof.

The existence of $\boldsymbol{X}_{\infty}$ as defined in Proposition 5.6 allows us to obtain a lower bound on it.
Proposition 5.7. Let $\boldsymbol{X}_{\infty}$ be a solution to Equation (1.1). Then

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\left\|\boldsymbol{X}_{\infty}\right\| \geq \omega, \quad \text { where } \quad \omega=\frac{-1+\sqrt{1+8 F_{k}}}{4} \tag{5.71}
\end{equation*}
$$

Proof. For contradiction, suppose that $\lim \sup _{t \rightarrow \infty}\left\|\boldsymbol{X}_{\infty}\right\|<\omega$. Therefore, there is some $T$ such that $\left\|\boldsymbol{X}_{\infty}(s)\right\|<\omega$ for all $s \geq T$. For notational convenience, let $\boldsymbol{X}=\boldsymbol{X}_{\boldsymbol{\infty}}$. First, we multiply Equation (1.1) through by the integrating factor $e^{t}$ and rearrange to yield

$$
\begin{equation*}
\frac{d \boldsymbol{X}}{d t} e^{t}=e^{t}(\boldsymbol{F}-B(\boldsymbol{X}, \boldsymbol{X})) \tag{5.72}
\end{equation*}
$$

Integrating on $[T, t]$ yields

$$
\begin{equation*}
\boldsymbol{X}(t) e^{t}-\boldsymbol{X}(T) e^{T}=\int_{T}^{t} e^{s}(\boldsymbol{F}-B(\boldsymbol{X}(s), \boldsymbol{X}(s))) d s \tag{5.73}
\end{equation*}
$$

By part 4 of Lemma 3.1, we have $\|B(\boldsymbol{X}(s), \boldsymbol{X}(s))\| \leq 2 \omega^{2}$. Thus, for all $k=1, \ldots, K$,

$$
\begin{equation*}
F_{k}-2 \omega^{2} \leq F_{k}-B_{k}(\boldsymbol{X}(s), \boldsymbol{X}(s)) \leq F_{k}+2 \omega^{2} \tag{5.74}
\end{equation*}
$$

which we multiply through by $e^{s}$ and integrate to yield

$$
\begin{equation*}
\int_{T}^{t} e^{s}\left(F_{k}-2 \omega^{2}\right) d s \leq \int_{T}^{t} e^{s}\left(F_{k}-B_{k}(\boldsymbol{X}(s), \boldsymbol{X}(s))\right) d s \leq \int_{T}^{t} e^{s}\left(F_{k}+2 \omega^{2}\right) d s \tag{5.75}
\end{equation*}
$$

By applying (5.73) and calculating the outer integrals, we rewrite this as

$$
\begin{equation*}
\left(e^{t}-e^{T}\right)\left(F_{k}-2 \omega^{2}\right) \leq X_{k}(t) e^{t}-X_{k}(T) e^{T} \leq\left(e^{t}-e^{T}\right)\left(F_{k}+2 \omega^{2}\right) \tag{5.76}
\end{equation*}
$$

We isolate $X_{k}(t)$ to yield

$$
\begin{equation*}
X_{k}(T) e^{-(t-T)}+\left(1-e^{-(t-T)}\right)\left(F_{k}-2 \omega^{2}\right) \leq X_{k}(t) \tag{5.77}
\end{equation*}
$$

Now take $t \rightarrow \infty$. Then

$$
\begin{equation*}
0<F_{k}-2 \omega^{2} \leq \limsup _{t \rightarrow \infty} X_{k}(t) \tag{5.78}
\end{equation*}
$$

Next, because $\lim \sup _{t \rightarrow \infty}\|\boldsymbol{X}(s)\|<\omega$,

$$
\begin{equation*}
F_{k}-2 \omega^{2} \leq \limsup _{t \rightarrow \infty} X_{k}(t) \leq \limsup _{t \rightarrow \infty}\|\boldsymbol{X}(t)\|<\omega \tag{5.79}
\end{equation*}
$$

Finally, since $\omega$ satisfies $2 \omega^{2}+\omega-F_{k}=0$, then substituting for $F_{k}$ on the left side of (5.79) yields $\omega<\omega$, a contradiction. Thus, $\lim \sup _{t \rightarrow \infty}\|\boldsymbol{X}(t)\| \geq \omega$.

We now prove a generalization of Proposition 5.7 that applies to $\boldsymbol{X}_{\boldsymbol{c}}$.

Proposition 5.8. For all solutions $\boldsymbol{X}_{\boldsymbol{c}}(t)$ and $\boldsymbol{V}_{\boldsymbol{c}}(t)$ of Equations (3.9),

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\left\|\boldsymbol{X}_{\boldsymbol{c}}\right\| \nrightarrow 0 \quad \text { and } \quad \limsup _{t \rightarrow \infty}\left\|\boldsymbol{V}_{\boldsymbol{c}}\right\| \rightarrow 0 \quad \text { as } \quad c \rightarrow \infty \tag{5.80}
\end{equation*}
$$

Proof. Note $\lim \sup _{t \rightarrow \infty}\left\|\boldsymbol{V}_{\boldsymbol{c}}\right\|=0$ as $c \rightarrow \infty$ as an immediate consequence of Proposition 5.4. For notational convenience, let $\boldsymbol{X}=\boldsymbol{X}_{\boldsymbol{c}}$. For the analysis of $\boldsymbol{X}$, suppose for contradiction there is a function $\gamma:[0, \infty) \rightarrow[0, \infty)$ such that $\lim _{c \rightarrow \infty} \gamma(c)=0$ and $\lim _{\sup _{t \rightarrow \infty}}\|\boldsymbol{X}\|=\gamma(c)$. Notice that for sufficiently large $t$ depending on $c$, we have $\|\boldsymbol{X}\| \leq 2 \gamma(c)$.

Now we multiply Equation (3.9a) through by $e^{t}$ and rearrange to yield

$$
\begin{equation*}
\frac{d \boldsymbol{X}}{d t} e^{t}=e^{t}\left(\boldsymbol{F}-B(\boldsymbol{X}, \boldsymbol{X})-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}\right) \tag{5.81}
\end{equation*}
$$

Integrating on $[T, t]$ yields

$$
\begin{equation*}
\boldsymbol{X}(t) e^{t}-\boldsymbol{X}(T) e^{T}=\int_{T}^{t} e^{s}\left(\boldsymbol{F}-B(\boldsymbol{X}(s), \boldsymbol{X}(s))-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}\right) d s \tag{5.82}
\end{equation*}
$$

Let $\omega$ be given as in Proposition 5.7. Also recall $M_{2}=\frac{\|\boldsymbol{F}\|}{m_{2} \sqrt{c}}$. Choose $c$ so large that

$$
\begin{equation*}
2 \gamma(c)<\omega \leq F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}} \tag{5.83}
\end{equation*}
$$

Note that such a $c$ exists because the definition of $\omega$ implies that $\omega<F_{k}$. Choose $T$ large enough so that $\|\boldsymbol{X}(s)\|<2 \gamma(c)$ for all $s \geq T$. By Proposition 5.4 and part 4 of Lemma 3.1, we have $\left\|B(\boldsymbol{X}(s), \boldsymbol{X}(s))+\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}\right\| \leq 8 \gamma(c)^{2}+\frac{M_{2}}{m_{1}}$. Thus, for all $k=1, \ldots, K$,

$$
\begin{equation*}
F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}} \leq F_{k}-B_{k}(\boldsymbol{X}(s), \boldsymbol{X}(s)) \leq F_{k}+8 \gamma(c)^{2}+\frac{M_{2}}{m_{1}} \tag{5.84}
\end{equation*}
$$

which we multiply through by $e^{s}$ and integrate to yield

$$
\begin{equation*}
\int_{T}^{t} e^{s}\left(F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}}\right) d s \leq \int_{T}^{t} e^{s}\left(F_{k}-B_{k}(\boldsymbol{X}(s), \boldsymbol{X}(s))\right) d s \leq \int_{T}^{t} e^{s}\left(F_{k}+8 \gamma(c)^{2}+\frac{M_{2}}{m_{1}}\right) d s \tag{5.85}
\end{equation*}
$$

By applying (5.82) and calculating the outer integrals, we rewrite this as

$$
\begin{equation*}
\left(e^{t}-e^{T}\right)\left(F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}}\right) \leq X_{k}(t) e^{t}-X_{k}(T) e^{T} \leq\left(e^{t}-e^{T}\right)\left(F_{k}+8 \gamma(c)^{2}+\frac{M_{2}}{m_{1}}\right) \tag{5.86}
\end{equation*}
$$

We isolate $X_{k}(t)$ to yield

$$
\begin{equation*}
X_{k}(T) e^{-(t-T)}+\left(1-e^{-(t-T)}\right)\left(F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}}\right) \leq X_{k}(t) \tag{5.87}
\end{equation*}
$$

Now take $t \rightarrow \infty$. Then

$$
\begin{equation*}
\omega \leq F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}} \leq \limsup _{t \rightarrow \infty} X_{k}(t) \tag{5.88}
\end{equation*}
$$

Next, because $\|\boldsymbol{X}(s)\|<2 \gamma(c)<\omega$,

$$
\begin{equation*}
\omega \leq F_{k}-8 \gamma(c)^{2}-\frac{M_{2}}{m_{1}} \leq \limsup _{t \rightarrow \infty}\|\boldsymbol{X}\|<\omega \tag{5.89}
\end{equation*}
$$

a contradiction. Therefore, $\lim \sup _{t \rightarrow \infty}\|\boldsymbol{X}\| \nrightarrow 0$ as $c \rightarrow \infty$.

Note that a more careful inspection of the proof of Proposition 5.8 reveals that

$$
\begin{equation*}
\liminf _{c \rightarrow \infty} \limsup _{t \rightarrow \infty}\left\|\boldsymbol{X}_{\boldsymbol{c}}\right\| \geq \omega \tag{5.90}
\end{equation*}
$$

where $\omega$ is as defined in Proposition 5.7. For the development of the present research, this is a minor point which needs no further elaboration.

### 5.2 Supporting Numerics

To illustrate and check the results of Propositions 5.4-5.8, we perform some numerical simulations. As these propositions all involve time limits, we choose large $T=8000$ for these simulations. Note this is 400 time intervals of size $\Delta t=20$ that correspond to the decorrelation time found in (3.18). Even though there is no weather forecasting model involved at this point, we solve the two-layer Lorenz '96 model in Equations (3.9) using the classic RK4 method with $h=1 / 2048$. This is to ensure that these computations follow the same discrete dynamics we will later employ in Sections 4 and 5, when we couple into an approximating solution.

First, we discuss the sharpness of the bound in Proposition 5.4:

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\| \leq M_{2}, \quad \text { where } \quad M_{2}=\frac{\|\sigma\|_{2}\|\boldsymbol{F}\|}{m_{2} \sqrt{c}} . \tag{5.91}
\end{equation*}
$$

To analyze the sharpness of this bound, we note $\|\sigma\|_{2}=1$ and consider the set $\left\{c=1.04^{n}: n=0,4, \ldots, 160\right\}$. Note that

$$
\begin{equation*}
\sup \left\{\limsup _{t \rightarrow \infty}\|\boldsymbol{V}\|: \text { all } \boldsymbol{X}(0), \boldsymbol{V}(0)\right\} \leq \sup \{\|\boldsymbol{V}\|: \boldsymbol{X}, \boldsymbol{V} \in \mathcal{A}\} \tag{5.92}
\end{equation*}
$$

where $\mathcal{A}$ is the global attractor for the two-layer Lorenz ' 96 system. To approximate the lim sup appearing in (5.91), which holds for any initial condition $\boldsymbol{X}(0), \boldsymbol{V}(0)$, we take the maximum of $\|\boldsymbol{V}\|$ over the interval $[T, 2 T]$ starting with one arbitrary initial condition. Thus, we replace $M_{2}$ by the a posteriori numerical bound

$$
\begin{equation*}
P_{2}(c)=\max \{\|\boldsymbol{V}(T+k h)\|: k=1,2, \ldots, T / h\} \approx \limsup _{t \rightarrow \infty}\|\boldsymbol{V}(t)\| \tag{5.93}
\end{equation*}
$$

Figure 5.2 plots these bounds as a function of $c \in[1,100]$ and compares them to the corresponding theoretical bound on $P_{1}$ given by $M_{1}$ in (5.17) and the theoretical bound on $P_{2}$ given by $M_{3}$ in (5.32).

Although our computation involves only one initial condition, the trajectory is followed for such a long time that we visit approximately 400 independent points on the global attractor that represent the lim sup in (5.91) starting from many different initial conditions. This computational shortcut of taking the maximum over one trajectory instead of the lim sup over many initial
conditions is typically justified by appealing to some sort of ergodicity; i.e., time averages are equivalent to spatial averages. In this case, we have no such theory; however, the chaotic motion observed in the $\boldsymbol{X}$ variable, and consequently $\boldsymbol{V}$, along with the decorrelation interval $\Delta t=20$, lead us to believe that the samples appearing in (5.93) represent the desired bounds on the attractor.

Figure 5.2 illustrates that $\max \|\boldsymbol{V}\|<M_{2}$, and that $M_{2}(c) \approx 10\|\boldsymbol{V}(c)\|$ for all $c$. This suggests that the a priori bound $M_{2}$ is generous and the actual simulations lead to much smaller values of $\|\boldsymbol{V}\|$. This factor-of-10 difference will become important when interpreting Theorem 3 and related results on recovering the state of $\boldsymbol{V}$ from the observations of $\boldsymbol{X}$.

We also discuss the sharpness of the bound in Proposition 5.5:

$$
\begin{equation*}
\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty} \leq M_{3}, \quad \text { where } \quad M_{3}=\min \left\{\frac{1}{\sqrt{c m_{2}}}, \sqrt{\frac{m_{1}}{2}}\right\} \frac{\|\boldsymbol{F}\|}{\sqrt{c m_{2}}} \tag{5.94}
\end{equation*}
$$

We choose $c$ as before and consider the numerical bound

$$
\begin{equation*}
P_{3}(c)=\frac{h}{T} \sum_{k=1}^{T / h}\|\boldsymbol{V}(T+k h)\|_{\infty} \approx \limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty} \tag{5.95}
\end{equation*}
$$

Figure 5.2 illustrates that max $\frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty}<M_{3}$, and that $M_{3}(c) \approx \frac{10}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty}$ for all $c$. This, too, suggests that $M_{3}$ is a soft bound. Moreover, there is a bend in $M_{3}$ around $c=56$; more interestingly, the values of $\frac{1}{s} \int_{0}^{s}\|V\|_{\infty}$ bend earlier, near $c=15$.

The reader should note that additional experiments were performed for larger $T$ and yielded similar results to those in Figure 5.2, and that other experiments using different parameter regimes were not considered in this study. Thus, the reader should bear in mind that different parameter regimes may yield sharper or softer bounds than seen here.

Finally, we discuss the claim in Proposition 5.8 that $\|\boldsymbol{X}\| \nrightarrow 0$ as $c \rightarrow \infty$. We let $c$ and $T$ be as before, and we warm up Equations (3.9) for time $T$ and evolve forward using the fourth-order Runge-Kutta method. Figure 5.3 shows that $\|\boldsymbol{X}\|$ tends downward for small $c$ but tends upward thereafter. Not only does $\boldsymbol{X}$ remain nontrivial as $c \rightarrow \infty$, but the theoretical lower bound $\omega$ is significantly less than the numerically realized energy levels of $\boldsymbol{X}$. We remark that 100 different simulations were performed using random initial conditions and produced similar results to those in Figure 5.3.


Figure 5.2: Log-log plot of the a priori bounds $\min \left(M_{1}, M_{2}\right)$ and $M_{3}$ compared with the numerical a posteriori bounds $P_{2}$ and $P_{3}$. Here, $M_{1}=\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{m_{2}}}$ (with $\delta=1$ ), $M_{2}=\frac{\|\boldsymbol{F}\|}{m_{2} \sqrt{c}}, M_{3}=\min \left\{\frac{1}{\sqrt{c m_{2}}}, \sqrt{\frac{m_{1}}{2}}\right\} \frac{\|\boldsymbol{F}\|}{\sqrt{c m_{2}}}, P_{2}(c) \approx \lim _{\sup _{t \rightarrow \infty}\|\boldsymbol{V}(t)\|}$, and $P_{3}(c) \approx$ $\limsup _{s \rightarrow \infty} \frac{1}{s} \int_{0}^{s}\|\boldsymbol{V}\|_{\infty}$, and $c=1.04^{n}$ with $n=0,4, \ldots, 160$.


Figure 5.3: Log-log plot of the numerically computed max $\|\boldsymbol{X}\|$ versus $c$ compared with the theoretical lower bound $\omega$ on that maximum for $F=20, m_{1}=20$, and $m_{2}=0.002$.

## Chapter 6

## Observations of $X$ with Known

## Dynamics

In this section, we focus on the first of the two inverse problems which form the focus of this research. As previously stated, we seek to determine the dynamical conditions under which the small variables $\boldsymbol{V}$ can be determined from the large variables $\boldsymbol{X}$. But because we assume $\boldsymbol{V}$ is unobserved, we must estimate it by obtaining other information from the system. To do that, we compare it to a set of variables governed by Equation (3.9b) and are assumed to be observable. This set of variables $\boldsymbol{v}$ we call the reference solution. In this chapter, we identify conditions under which the reference solution $\boldsymbol{v}$ and unobserved solution $\boldsymbol{V}$ become asymptotically identical in time. As before, we govern $\boldsymbol{X}, \boldsymbol{V}$, and $\boldsymbol{v}$ by the respective equations

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.1a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{6.1b}\\
& \frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{6.1c}
\end{align*}
$$

and we look for parameter values that cause synchronization.

In this chapter, we assume $c$ is known, and we fix values of $c$ and evolve the system forward in time to determine whether convergence occurs. Section 6.1 provides theoretical bounds on the three-equation system defined by Equations (6.1a)-(6.1c) and conditions for the convergence of $\|\boldsymbol{V}-\boldsymbol{v}\|$, and it concludes with numerical results to confirm those bounds. Recursion relations
for coupling on $\boldsymbol{X}$ and its derivatives are provided in Section 6.3 for the purpose of finding a lower value of $c$ that establishes convergence; using these, we provide numerical simulations, and we compare and contrast the values of $c$ that cause synchronization to occur.

### 6.1 Coupling on $X$

Theorem 6.1. $\|\boldsymbol{v}\|$ is uniformly bounded in time regardless of whether $\boldsymbol{v}$ is synchronized by $\boldsymbol{V}$. Specifically, we have

$$
\begin{equation*}
\|\boldsymbol{v}\|^{2} \leq\left\|\boldsymbol{v}_{0}\right\|^{2}+\frac{L_{1}}{c^{2}} \quad \text { for all } \quad t, \quad \text { and } \quad \limsup _{t \rightarrow \infty}\|\boldsymbol{v}\|^{2} \leq \frac{L_{2}}{c} \tag{6.2}
\end{equation*}
$$

where $L_{1}$ and $L_{2}$ are given in (6.6) and (6.8), respectively.

Proof. Let $\psi$ be as before. By Lemma 5.1, $\psi \leq \psi_{0} e^{-\alpha t}+\frac{\nu}{\alpha}\left(1-e^{-\alpha t}\right)$. If $c>1-\frac{1}{2 \delta}$, take $\alpha, \nu$ to be as in Lemma 5.1; otherwise, if $0<c<1-\frac{1}{2 \delta}$, we take $\alpha, \nu$ to be as in Lemma 5.2. For definiteness, we assume $c>1-\frac{1}{2 \delta}$. Note that the proof in the other case is similar for different $L_{1}$ and $L_{2}$.

Next, we take (6.1c) and dot it with $\boldsymbol{v}$ to get

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2}+c\|\boldsymbol{v}\|^{2}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \cdot \boldsymbol{v} \tag{6.3}
\end{equation*}
$$

Applying the Cauchy-Schwarz inequality, then Young's inequality, yields

$$
\begin{align*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2}+c\|\boldsymbol{v}\|^{2} & \leq \frac{1}{m_{2}}\|\sigma \boldsymbol{X}\|\|\boldsymbol{v}\| \\
& \leq \frac{1}{2 c m_{2}}\|\sigma \boldsymbol{X}\|^{2}+\frac{c}{2}\|\boldsymbol{v}\|^{2} \tag{6.4}
\end{align*}
$$

Rearranging (6.4) yields

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2}+\frac{c}{2}\|\boldsymbol{v}\|^{2} \leq \frac{1}{2 m_{2} c}\|\sigma \boldsymbol{X}\|^{2} \leq \frac{1}{2 m_{2} c} \frac{\psi}{m_{1}} \leq \frac{L_{1}}{2 c}, \tag{6.5}
\end{equation*}
$$

where

$$
\begin{align*}
L_{1} & =\frac{1}{m_{1}} \sup _{t \geq 0}\left\{\psi_{0} e^{-\alpha t}+\frac{\nu}{\alpha}\left(1-e^{-\alpha t}\right)\right\} \frac{1}{m_{2}} \\
& \leq \frac{1}{m_{1} m_{2}}\left(\psi_{0}+\frac{\nu}{\alpha}\right) \\
& =\left(\frac{1}{m_{2}}\|\boldsymbol{X}(0)\|^{2}+\frac{1}{m_{1}}\|\boldsymbol{V}(0)\|^{2}+\frac{\delta\|\boldsymbol{F}\|^{2}}{\left(2-\frac{1}{\delta}\right) m_{2}}\right) . \tag{6.6}
\end{align*}
$$

Next, we multiply (6.5) through by the integrating factor $e^{c t}$. Integrating yields $\|\boldsymbol{v}\|^{2} e^{c t}-$ $\left\|\boldsymbol{v}_{\mathbf{0}}\right\|^{2} \leq \frac{L_{1}}{c}\left(e^{c t}-1\right)$. Dividing this through by $e^{c t}$ and rearranging yields

$$
\begin{equation*}
\|\boldsymbol{v}\|^{2} \leq\left\|\boldsymbol{v}_{\mathbf{0}}\right\|^{2} e^{-c t}+\frac{L_{1}}{c^{2}}\left(1-e^{-c t}\right) \tag{6.7}
\end{equation*}
$$

Notice the left side is bounded by $\left\|\boldsymbol{v}_{\boldsymbol{0}}\right\|^{2}+L_{1} / c^{2}$.

To obtain an asymptotic bound as $t \rightarrow \infty$, let $\epsilon>0$ and $T$ be so large that for $t \geq T$,

$$
\begin{equation*}
\frac{d}{d t}\|\boldsymbol{v}\|^{2}+c\|\boldsymbol{v}\|^{2} \leq(1+\epsilon) L_{2}, \quad \text { where } \quad L_{2}=\frac{\delta\|\boldsymbol{F}\|^{2}}{\left(2-\frac{1}{\delta}\right) m_{2}} \tag{6.8}
\end{equation*}
$$

Now integrating from $T$ to $t$ yields

$$
\begin{equation*}
\|\boldsymbol{v}\|^{2} \leq\|\boldsymbol{v}(T)\|^{2} e^{-c(t-T)}+\frac{(1+\epsilon) L_{2}}{c}\left(1-e^{-c(t-T)}\right) \tag{6.9}
\end{equation*}
$$

Therefore,

$$
\begin{equation*}
\limsup _{t \rightarrow \infty}\|\boldsymbol{v}\|^{2} \leq \frac{(1+\epsilon) L_{2}}{c} \tag{6.10}
\end{equation*}
$$

As $\epsilon$ is arbitrary, we obtain $L_{2} / c$ for the upper bound in (6.10). Thus, $\|\boldsymbol{v}\|^{2}$ is uniformly bounded in time even when $\boldsymbol{v}$ is not synchronized by $\boldsymbol{V}$.

Proposition 6.2. As before, define $\boldsymbol{\Delta}=\boldsymbol{V}-\boldsymbol{v}$. Then for all $\epsilon>0$, there is a $T$ such that

$$
\begin{equation*}
\frac{1}{2} \frac{d\|\boldsymbol{\Delta}\|^{2}}{d t} \leq\left(2 M_{\epsilon}-c\right)\|\boldsymbol{\Delta}\|^{2} \quad \text { for } \quad t \geq T \tag{6.11}
\end{equation*}
$$

where $M_{\epsilon}=M_{1}(1+\epsilon)$.

Proof. Recall Equations (6.1b) and (6.1c):

$$
\begin{aligned}
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \\
& \frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X}
\end{aligned}
$$

Next, we subtract $d \boldsymbol{v} / d t$ from $d \boldsymbol{V} / d t$. The resulting equation is

$$
\begin{equation*}
\frac{d \boldsymbol{\Delta}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})=-c \boldsymbol{\Delta} \tag{6.13}
\end{equation*}
$$

Next, we take the dot product of each side with the quantity $\boldsymbol{\Delta}$. The equation becomes

$$
\begin{equation*}
\frac{1}{2} \frac{d\|\boldsymbol{\Delta}\|^{2}}{d t}=-c\|\boldsymbol{\Delta}\|^{2}-[\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})] \cdot \boldsymbol{\Delta} \tag{6.14}
\end{equation*}
$$

We now add and subtract $\widetilde{B}(\boldsymbol{v}, \boldsymbol{V})$ inside the brackets. The expression becomes

$$
\begin{equation*}
[\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+\widetilde{B}(\boldsymbol{v}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})] \cdot \boldsymbol{\Delta}=[\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{\Delta})] \cdot \boldsymbol{\Delta} \tag{6.15}
\end{equation*}
$$

Adding and subtracting $\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})$ yields

$$
\begin{equation*}
[\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{\Delta})+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})] \cdot \boldsymbol{\Delta} \tag{6.16}
\end{equation*}
$$

By Lemma 3.1, $\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{\Delta})$ disappears. By Equation (3.6), we thus have

$$
\begin{equation*}
[\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V})+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})] \cdot \boldsymbol{\Delta}=\sum_{j=1}^{J K}\left[\Delta_{j+1} V_{j-1} \Delta_{j}-V_{j+1} \Delta_{j+2} \Delta_{j}\right] \tag{6.17}
\end{equation*}
$$

By Proposition 5.3 and the fact that $\left|V_{j}\right| \leq\|\boldsymbol{V}\|$, there exists $T$ such that

$$
\begin{equation*}
\left|V_{j}(t)\right| \leq M_{1}(1+\epsilon)=M_{\epsilon} \quad \text { for all } \quad t \geq T \tag{6.18}
\end{equation*}
$$

Then by the Cauchy-Schwarz inequality, we have

$$
\begin{align*}
{[\widetilde{B}(\boldsymbol{\Delta},} & \boldsymbol{V})+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})] \cdot \boldsymbol{\Delta} \\
\quad & \sum_{j=1}^{J K}\left[\left|\Delta_{j+1}\right|\left|V_{j-1}\right|\left|\Delta_{j}\right|+\left|V_{j+1}\right|\left|\Delta_{j+2}\right|\left|\Delta_{j}\right|\right] \\
\leq & \sum_{j=1}^{J K}\left[\left|\Delta_{j+1}\right|\left|\Delta_{j}\right|+\left|\Delta_{j+2}\right|\left|\Delta_{j}\right|\right] M_{\epsilon} \\
\leq & 2\|\boldsymbol{\Delta}\|^{2} M_{\epsilon} \tag{6.19}
\end{align*}
$$

Thus, Equation (6.14) becomes

$$
\begin{align*}
\frac{1}{2} \frac{d\|\boldsymbol{\Delta}\|^{2}}{d t} & \leq-c\|\boldsymbol{\Delta}\|^{2}+2\|\boldsymbol{\Delta}\|^{2} M_{\epsilon} \\
& =\left(2 M_{\epsilon}-c\right)\|\boldsymbol{\Delta}\|^{2} \tag{6.20}
\end{align*}
$$

for $t \geq T$. This completes the proof.

This proposition implies
Theorem 6.3. If $c>2 M_{1}$, then $\|\boldsymbol{\Delta}\| \rightarrow 0$ as $t \rightarrow \infty$.

We can use this fact to determine the parameter regimes in which convergence is theoretically guaranteed to occur. First, suppose $c$ and $\delta$ are as in Lemma 5.1. Then because $M_{\epsilon}>M_{1}$ for all $\epsilon$, we have after minimizing over $\delta$

$$
\begin{equation*}
c>\min _{\delta>\frac{1}{2}}\left\{2 \delta\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{(2 \delta-1) m_{2}}}\right\}=2\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{m_{2}}} \quad \text { and } \quad c>\frac{1}{2} . \tag{6.21}
\end{equation*}
$$

Note that the minimum occurs when $\delta=1$. This is the estimate that corresponds to our standard parameters $F=20, m_{1}=20$, and $m_{2}=0.002$.

Similar work reveals that if $c$ and $\delta$ are as in Lemma 5.2 , we get convergence if

$$
\begin{equation*}
1-\frac{1}{2 \delta}>c>2 M=2\|\boldsymbol{F}\| \sqrt{\frac{m_{1} \delta}{2 c m_{2}}} \tag{6.22}
\end{equation*}
$$

or, equivalently,

$$
\begin{equation*}
1-\frac{1}{2 \delta}>c>\sqrt[3]{\frac{2\|\boldsymbol{F}\|^{2} m_{1} \delta}{m_{2}}} \tag{6.23}
\end{equation*}
$$

In this case, $c$ must be small, which after optimizing with respect to $\delta$ yields the condition that

$$
\begin{equation*}
\frac{\|\boldsymbol{F}\|^{2} m_{1}}{m_{2}}<\frac{27}{256}, \quad \text { where } \quad \delta=2 \tag{6.24}
\end{equation*}
$$

However, because the range of values for $c, F, m_{1}$, and $m_{2}$ falls outside the dynamically interesting range identified in Section 4, we forgo further analysis of the values of $c$ which satisfy (6.23).

The proof of Proposition 6.2 leading to Theorem 6.3 also implies something about the rate of convergence. We state this in

Corollary 6.4. Suppose $c$ satisfies (6.21). Then for all $\epsilon>0$, there is a constant $C$ such that

$$
\begin{equation*}
\|\boldsymbol{\Delta}\|^{2} \leq C \exp \left(\left[4(1+\epsilon)\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{m_{2}}}-2 c\right] t\right) \tag{6.25}
\end{equation*}
$$

Proof. For any $\epsilon>0$, there is a $T$ so large that Proposition 6.2 holds as in Equation (6.18) and the asymptotic bounds given in Proposition 5.3 and Theorem 6.1 imply

$$
\begin{equation*}
\|\boldsymbol{V}(T)\| \leq(1+\epsilon) M_{1} \quad \text { and } \quad\|\boldsymbol{v}(T)\| \leq(1+\epsilon) L_{2} \tag{6.26}
\end{equation*}
$$

Integrate the differential inequality in (6.11) from $T$ to $t$ after multiplying $e^{\left(4 M_{\epsilon}-2 c\right) t}$. Thus,

$$
\begin{equation*}
\|\boldsymbol{\Delta}(t)\|^{2} \exp \left(\left[2 c-4(1+\epsilon)\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{m_{2}}}\right] t\right)-\|\boldsymbol{\Delta}(T)\|^{2} \leq 0 \tag{6.27}
\end{equation*}
$$

By (6.26), it follows that

$$
\begin{equation*}
\|\boldsymbol{\Delta}(T)\| \leq\|\boldsymbol{V}(T)\|+\|\boldsymbol{v}(T)\| \leq(1+\epsilon)\left(M_{1}+L_{2}\right) \tag{6.28}
\end{equation*}
$$

Rearranging terms and setting $C=\|\boldsymbol{V}(T)\|+\|\boldsymbol{v}(T)\| \leq(1+\epsilon)\left(M_{1}+L_{2}\right)$ completes the proof.

Theoretically, provided that $c>2 M_{1}$, as $c$ increases, the rate of convergence between $\boldsymbol{V}$ and $\boldsymbol{v}$ increases exponentially. Recall that our standard parameterization is $F=20, c=20, m_{1}=20$, and $m_{2}=0.002$, where $c$ is the magnitude of the dissipation of energy in the fast variables. When $c$ is large, the dissipation is strong; when $c$ is small, the dissipation is weak. To numerically illustrate the effects of changing $c$, we warm up the reference solution $\boldsymbol{X}, \boldsymbol{V}$ starting from a random initial condition for $T=50$ units of time so that it reflects the long-time behavior governed by $c$. At this point, we start coupling $\boldsymbol{v}$ as given by Equation (6.1c). For definiteness, we denote $t_{0}=0$ to be the time at which we start coupling, and we initialize $\boldsymbol{v}=\mathbf{0}$. As Figure
6.1 shows, a unit difference in $c$ can make recovery impossible, make it possible and slow, or make it possible and rapid.

Note also that synchronization of $\boldsymbol{v}$ with $\boldsymbol{V}$ occurs for values of $c$ much smaller than the theoretical bound $2 M_{1}$ given in Theorem 6.3. Using the a priori estimate for $M_{1}$ as in Corollary 6.4, our parameter regime given in Equation (3.17) implies that

$$
\begin{equation*}
c=21 \ll 2\|\boldsymbol{F}\| \sqrt{\frac{m_{1}}{m_{2}}}=40 \cdot \sqrt{\frac{20 K}{0.002}} \approx 11313.71 \tag{6.29}
\end{equation*}
$$

Similarly, the a posteriori bounds in Figure 5.2 suggest that $\lim \sup _{t \rightarrow \infty}\|\boldsymbol{V}\| \lesssim 478$. And again,

$$
\begin{equation*}
c=21<2 M_{1} \approx 2 \cdot 478=956 . \tag{6.30}
\end{equation*}
$$



Figure 6.1: Time evolution of the absolute error in $\boldsymbol{v}$ for our standard choice of parameters $F=20, m_{1}=20$, and $m_{2}=0.002$, with $c$ varying between 19 and 21 .

Figure 6.1 as well as Corollary 6.4 suggest that the recovery time decreases as $c$ increases. To confirm this, we define

$$
T(c)=\left\{\begin{array}{l}
\tau-T:\|\boldsymbol{V}-\boldsymbol{v}\|<10^{-6} \text { for } t \in[\tau, 3 T] \text { and } \tau \in[T, 2 T]  \tag{6.31}\\
\infty \text { if } \tau>2 T \text { or if }\|\boldsymbol{V}-\boldsymbol{v}\| \text { never goes below } 10^{-6}
\end{array}\right.
$$

where $c$ is the dissipation, $T$ is the length of the warmup period, and $\boldsymbol{X}(T)$ and $\boldsymbol{V}(T)$ are points on the attractor with $\boldsymbol{V}(T)=0$.

We now fit the exponential rate of convergence suggested by Corollary 6.4 to the numerical data for $T(c)$. First assume

$$
10^{-6}=\|\boldsymbol{V}-\boldsymbol{v}\| \approx C \exp ((M-2 c) t) \quad \text { when } \quad t=T(c)
$$

Consequently, solving for $T(c)$ yields

$$
T(c) \approx \frac{\log \left(10^{-6} / C\right)}{M-2 c}
$$

or

$$
\frac{1}{T(c)} \approx m c+b, \quad \text { where } \quad m=-\frac{2}{\log \left(10^{-6} / C\right)} \quad \text { and } \quad b=\frac{M}{\log \left(10^{-6} / C\right)} .
$$

Through a least squares fit involving 10 independent samples of $T(c)$ for values of $c$ ranging from 20 up to 50 , we find

$$
m \approx 0.0691 \quad \text { and } \quad b \approx 1.5375
$$

Solving back for $C$ and $M$ yields

$$
C \approx 1.928 \quad \text { and } \quad M \approx 22.25
$$

Note that 22.25 is slightly larger than the minimum value of $c$ observed for which synchronization occurs at all. Figure 6.2 compares this least squares fit to the data and shows that for $c \gg 22.25$ the driven solution $\boldsymbol{v}$ rapidly synchronizes with the free-running solution $\boldsymbol{V}$ over time. On average, one can expect rapid recovery of $\boldsymbol{V}$ for large values of $c$.

Also, recall that $m_{1}$ is the magnitude of the energy drawn from the slow layer to the fast layer, and $m_{2}$ is the magnitude of the energy amplified in its transfer from the slow layer to the fast layer. If $m_{1}>m_{2}$, more energy is being transferred to the fast layer; if $m_{1}<m_{2}$, more energy is being amplified in the slow layer; if $m_{1}=m_{2}$, the energy being transferred and the energy being amplified are equal. Thus, Equation and (6.21) and Proposition 6.2 tell us that $\|\boldsymbol{\Delta}\| \rightarrow 0$ when dissipation is greater than the ratio of energy drawn to energy transferred.

We now turn to an improvement of Theorem 6.1.


Figure 6.2: Least squares fit $1 / T(c) \approx 0.069 c-1.537$, where $T(c)$ is the time it takes for the approximating solution to synchronize to within $10^{-6}$ with the reference solution when $F=20, m_{1}=20, m_{2}=0.002$, and $c$ varying between 20 and 50.

Theorem 6.5. Let $\boldsymbol{V}$ and $\boldsymbol{v}$ be as in Equations (6.1b) and (6.1c), respectively. Define $\boldsymbol{\Delta}=$ $\boldsymbol{V}-\boldsymbol{v}$. Then $\|\boldsymbol{\Delta}\| \rightarrow 0$ as $t \rightarrow \infty$, provided

$$
\begin{equation*}
\limsup _{t \rightarrow \infty} \frac{1}{t} \int_{0}^{t}\|\boldsymbol{V}\|_{\infty} d s<\frac{c}{2} \tag{6.32}
\end{equation*}
$$

Proof. Let $\boldsymbol{\Delta}$ be defined as above. Then by the work in Proposition 6.2, Equation (6.14) results, along with the inequality

$$
\begin{align*}
\frac{1}{2} \frac{d\|\boldsymbol{\Delta}\|^{2}}{d t} & =\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta}) \cdot \boldsymbol{\Delta}+\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V}) \cdot \boldsymbol{\Delta}-c\|\boldsymbol{\Delta}\|^{2} \\
& \leq 2 \cdot \max \left\{\left|V_{i}\right|: i=1, \ldots, q\right\}\|\boldsymbol{\Delta}\|^{2}-c\|\boldsymbol{\Delta}\|^{2} \\
& =2\|\boldsymbol{V}\|_{\infty}\|\boldsymbol{\Delta}\|^{2}-c\|\boldsymbol{\Delta}\|^{2} \tag{6.33}
\end{align*}
$$

From this, we get

$$
\begin{equation*}
\left(\frac{d}{d t}+2 c-4\|\boldsymbol{V}\|_{\infty}\right)\|\boldsymbol{\Delta}\|^{2} \leq 0 \tag{6.34}
\end{equation*}
$$

Next, let $\tau$ be the integrating factor $\exp \left(\int_{0}^{t}\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right) d s\right)$. Then $\tau^{\prime}(t)=\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right) \tau(t)$ and, by (6.34),

$$
\begin{align*}
\frac{d}{d t}\left[\tau(t)\|\boldsymbol{\Delta}\|^{2}\right] & =\tau^{\prime}(t)\|\boldsymbol{\Delta}\|^{2}+\frac{d\|\boldsymbol{\Delta}\|^{2}}{d t} \tau(t) \\
& =\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right) \tau(t)\|\boldsymbol{\Delta}\|^{2}+\frac{d\|\boldsymbol{\Delta}\|^{2}}{d t} \tau(t) \\
& =\tau(t)\left(\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right)\|\boldsymbol{\Delta}\|^{2}+\frac{d\|\boldsymbol{\Delta}\|^{2}}{d t}\right) \\
& \leq 0 \tag{6.35}
\end{align*}
$$

We integrate this from 0 to $t$ to yield the inequality $\tau(t)\|\boldsymbol{\Delta}\|^{2}-\tau(0)\left\|\boldsymbol{\Delta}_{0}\right\|^{2} \leq 0$. Then $\tau(t)\left\|\boldsymbol{\Delta}_{0}\right\|^{2} \leq \tau(0)\left\|\boldsymbol{\Delta}_{0}\right\|^{2}$. Because $\tau(0)=1$, we can rearrange this inequality as

$$
\begin{equation*}
\|\boldsymbol{\Delta}\|^{2} \leq \exp \left(-\int_{0}^{t}\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right) d s\right)\left\|\boldsymbol{\Delta}_{0}\right\|^{2} \tag{6.36}
\end{equation*}
$$

Finally, we multiply the integral by $t / t$ and rearrange it as follows:

$$
\begin{equation*}
\int_{0}^{t}\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right) d s=\frac{t}{t} \int_{0}^{t}\left(2 c-4\|\boldsymbol{V}\|_{\infty}\right) d s=2 t\left(c-\frac{2}{t} \int_{0}^{t}\|\boldsymbol{V}\|_{\infty} d s\right) \tag{6.37}
\end{equation*}
$$

Then when we apply (6.32) to (6.36) and let $t \rightarrow \infty$, the result follows.

To compare the conditions which appear in Theorems 6.3 and 6.5 to the threshold value of $c$ found in Figure 6.1, after which $\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ numerically, we define the a posteriori bounds

$$
\begin{equation*}
P_{2}(c)=\max \left\{\|\boldsymbol{V}(t)\|_{2}: t \in[T, 2 T]\right\}, \quad P_{3}(c)=\frac{1}{T} \int_{T}^{2 T}\|\boldsymbol{V}\|_{\infty} d t \tag{6.38}
\end{equation*}
$$

Figure 5.2 plots these bounds as a function of $c \in[1,100]$ and compares them to the corresponding theoretical bound on $P_{2}$ given by $M_{1}$ in (5.17) and the theoretical bound on $P_{3}$ given by $M_{3}$ in (5.94). The line $c / 2$ crosses the data represented by $P_{3}(c)$ at $c \approx 60$, and the hypotheses of Theorem 6.5 empirically hold. Therefore, when $c \gtrsim 60$, Theorem 6.5 implies that $\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ as $t \rightarrow \infty$. On the other hand, we have a theoretical guarantee when $c / 2>M_{3}$ or equivalently when $c>200$ ish that $\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ as $t \rightarrow \infty$.

### 6.2 Effects of Arbitrary Coupling Matrix $\sigma$

This section is an exploration of how the details of the coupling between the two layers affects the ability to recover $\boldsymbol{V}$ from the observations of $\boldsymbol{X}$. From a physical point of view, this is changing the dynamics of the problem; from a mathematical point of view, it is made possible through our identification of the coupling matrix $\sigma$ with orthonormal columns. In particular, by changing $\sigma$, we get to explore whether Lorenz's original choice for $\sigma$ makes it easier or more difficult to recover $\boldsymbol{V}$ : i.e., which conditions on $c$ lead to the synchronization of $\boldsymbol{v}$ with $\boldsymbol{V}$ over time.


Figure 6.3: Box plots of $\|\boldsymbol{V}-\boldsymbol{v}\|$ for 4 randomly determined $\sigma$ with $F=20, m_{1}=20, m_{2}=$ 0.002 , and $c$ varying between 10 and 25 .

We defined $\sigma$ as in Equation (3.1) to represent the coupling in the original two-layer Lorenz '96 system. Here, we let $\sigma$ be an arbitrary coupling matrix with orthonormal columns, and we numerically determine the critical value of $c$. We again emphasize that the theory in Chapter 5 applies to any such matrix.

To create $\sigma$, we initialize a matrix of uniformly distributed random values in $[-1 / 2,1 / 2]$ using the Mersenne twister (the default pseudorandom number generator in Julia), perform a QR decomposition, and define $\sigma$ to be the matrix consisting of the first eight columns of $Q$-i.e., $\sigma$ is obtained from the reduced $Q R$ decomposition of a random matrix.


Figure 6.4: Box plots of $\|Q \boldsymbol{V}-\boldsymbol{q}\|$ for 4 randomly determined $\sigma$ with $F=20, m_{1}=20, m_{2}=$ 0.002 , and $c$ varying between 10 and 25 .

Figures 6.3 and 6.4 show a substantial improvement in the critical value of $c$ when $\sigma$ is a random coupling matrix. When the coupling is on $\boldsymbol{X}$ alone as given by Equation (6.1c), synchronization occurs around $c=17$, considerably less than when we couple using our traditional $\sigma$. Interestingly, when the coupling is on $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$ as given by Equation (6.42c), synchronization occurs around $c=18$-still less than when we couple using our traditional $\sigma$ but more than when we use Equation (6.1c). We performed this simulation for 10 different random $\sigma$, and we obtained similar results each time. While it is conceivable that a coupling matrix similar to (3.1) might be obtained randomly, this did not happen in practice. These graphs illustrate that a random coupling matrix will likely lead to a smaller value of $c$ needed to achieve synchronization. Since changing the physics of the problem is impossible in real applications, we resume using our regular $\sigma$ in the next section.

### 6.3 On the Use of Derivative Information

In Section 4, we showed that the value of $c$ influences how strongly the unresolved variables depend on the resolved ones. In particular, when $c$ is large, $\boldsymbol{V}$ depends more strongly on $\boldsymbol{X}$, and thus a better estimate of $\boldsymbol{V}$ can be obtained, when $c$ is large. Although we know the
theoretical conditions under which convergence occurs, our numerical experiments have not yet demonstrated convergence for the smallest possible values of $c$.

One possible way to improve the estimates of the dynamics and the recovery of $c$ is to assume $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$ are known. From a mathematical point of view, if exact observations of $\boldsymbol{X}$ are known continuously in time, then we have already observed $\dot{\boldsymbol{X}}$ and all its derivatives. In practice, there are many obstacles to inferring $\dot{\boldsymbol{X}}$ from the observations of $\boldsymbol{X}$. First, those observations may include noise with the effect that they would not be differentiable even if observed continuously in time. Another impediment to knowing $\dot{\boldsymbol{X}}$ is that real observational data is never continuous in time but only near-continuous in time at best. In this case, one can approximate the derivative information by means of finite differences. We defer this line of study until Section 8, where it becomes a crucial tool when solving for $c$ if both $c$ and $\boldsymbol{V}$ are unknown.

To make use of the exact derivative information, note the observation of $\boldsymbol{X}$ over time implies $\dot{\boldsymbol{X}}$ is known. The first differential equation then implies that we also know $\sigma^{\top} \boldsymbol{V}$. In particular, let $P=\sigma \sigma^{\top}$ and $Q=I-P$. We remark that $P$ is an orthogonal projection onto a subspace of the phase space of $\boldsymbol{V}$, and $Q$ is the orthogonal complement of $P$. We define a new variable, $\boldsymbol{q}$, and replace Equation (6.1c) with $d \boldsymbol{q} / d t$. Thus, let $\boldsymbol{q}=Q \boldsymbol{v}$. By orthonormality of $\sigma$, we have $Q \sigma \boldsymbol{X}=0$. Thus, if we multiply (6.1c) leftwise by $Q$, we obtain

$$
\begin{equation*}
Q \dot{\boldsymbol{v}}+Q \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c Q \boldsymbol{v}=\mathbf{0} \tag{6.39}
\end{equation*}
$$

and by definition of $\boldsymbol{q}$,

$$
\begin{equation*}
\frac{d \boldsymbol{q}}{d t}+Q \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c \boldsymbol{q}=\mathbf{0} \tag{6.40}
\end{equation*}
$$

We now use $\boldsymbol{v}=P \boldsymbol{V}+\boldsymbol{q}$ for the recovery of $\boldsymbol{V}$. Thus, we obtain

$$
\begin{equation*}
\frac{d \boldsymbol{q}}{d t}+Q \widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \boldsymbol{V}+\boldsymbol{q})+c \boldsymbol{q}=0 \tag{6.41}
\end{equation*}
$$

and we obtain the three-equation system with coupling on $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$ :

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.42a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{6.42b}\\
& \frac{d \boldsymbol{q}}{d t}+Q \widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \boldsymbol{V}+\boldsymbol{q})+c \boldsymbol{q}=0 . \tag{6.42c}
\end{align*}
$$

A similar argument to the proof in Theorem 6.3 (also see Proposition 6.2) shows that $\|\boldsymbol{q}-Q \boldsymbol{V}\| \rightarrow$ 0 as $t \rightarrow \infty$ provided $c$ is large enough. In fact, the same condition on $c$ given in Theorem 6.3 is enough to theoretically guarantee convergence. Before proceeding, we check this through simulations.


Figure 6.5: A representative graph of the absolute error in $\boldsymbol{q}$ for our standard choice of parameters and $c$ varying between 20 and 22 . Note the convergence occurring when $c=21$ instead of $c=20$ as in Figure 6.1. Repeated runs with different initial conditions produced similar results.

Interestingly, we do not see an improvement in the numerical value of $c$ for which synchronization occurs. As shown by the numerical computations in Figure 6.5, synchronization of $\boldsymbol{q}$ with $Q \boldsymbol{V}$ appears to require larger values of $c$. The difficulty seems to be that we have ignored how $P \boldsymbol{V}$ relates to the equation governing $\boldsymbol{X}$. Equation (6.42c) is also worse in a statistical sense than Equation (6.1c) when the value of $c$ is unknown. This will be discussed further in Section 7.

The fact that Equation (6.42c) performs worse than the simpler coupling considered before is somewhat surprising, as we are using more information about the observations: in particular, we have replaced $\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})$ with $\widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \boldsymbol{V}+\boldsymbol{q})$, which should be more accurate because $P \boldsymbol{V}$, given by the derivative information, is correct. This shall be remedied in Section 6.4, where we project Equation (6.42a) by $P$ to obtain algebraic constraints governing the evolution of $P \boldsymbol{V}+\boldsymbol{q}$. Before proceeding to the nonlinear optimization techniques discussed in Section 6.4, we next explore a different approach to make use of the algebraic constraints: namely, we employ the differential algebraic equation (DAE) solver given in [21].

Suppose $\boldsymbol{X}, \dot{\boldsymbol{X}}$, and $\ddot{\boldsymbol{X}}$ are known. Then $\sigma^{\top} \boldsymbol{V}$ and $\sigma^{\top} \dot{\boldsymbol{V}}$ are known, and $\boldsymbol{q}=Q \boldsymbol{V}$ is still unknown. Note $\dot{\boldsymbol{q}}=Q \dot{\boldsymbol{V}}$ and dotting Equations (6.42a)-(6.42b) yields

$$
\begin{align*}
\ddot{\boldsymbol{X}}+B(\dot{\boldsymbol{X}}, \boldsymbol{X})+B(\boldsymbol{X}, \dot{\boldsymbol{X}})+\dot{\boldsymbol{X}} & =-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.43a}\\
\ddot{\boldsymbol{V}}+\widetilde{B}(\dot{\boldsymbol{V}}, \boldsymbol{V})+\widetilde{B}(\boldsymbol{V}, \dot{\boldsymbol{V}})+c \dot{\boldsymbol{V}} & =\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{6.43b}
\end{align*}
$$

Dotting Equation (6.43b) and recalling $P+Q=I$ yields

$$
\begin{equation*}
Q \ddot{\boldsymbol{V}}+Q \widetilde{B}(P \dot{\boldsymbol{V}}+\dot{\boldsymbol{q}}, P \boldsymbol{V}+\boldsymbol{q})+\widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \dot{\boldsymbol{V}}+\dot{\boldsymbol{q}})+c Q \dot{\boldsymbol{V}}=\mathbf{0} \tag{6.44}
\end{equation*}
$$

Thus, we yield the four-equation system

$$
\begin{align*}
& \dot{\boldsymbol{X}}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.45a}\\
& \dot{\boldsymbol{V}}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{6.45b}\\
& \dot{\boldsymbol{\eta}}+Q \ddot{\boldsymbol{V}}+Q \widetilde{B}(P \dot{\boldsymbol{V}}+\dot{\boldsymbol{q}}, P \boldsymbol{V}+\boldsymbol{q}),+\widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \dot{\boldsymbol{V}}+\dot{\boldsymbol{q}})+c Q \dot{\boldsymbol{V}}=\mathbf{0}  \tag{6.45c}\\
& \dot{\boldsymbol{q}}=\boldsymbol{\eta} \tag{6.45d}
\end{align*}
$$

Figure 6.6 shows that the approximation $q$ given by Equations (6.45c)-(6.45d) fails to synchronize with $Q \boldsymbol{V}$ over time for the same values of $c$ used in Figure 6.5. It appears that the coupling enhances the propagation of errors over time; in particular, no synchronization was observed, even for $c \gg 20$.


Figure 6.6: Absolute error in $\boldsymbol{q}$ for our standard choice of parameters and $h=1 / 4096, T=$ 400 , and $c$ varying between 19,20 , and 21.

When Equation (6.44) is introduced into the four-equation system as an algebraic constraint, we get the system

$$
\begin{align*}
& \dot{\boldsymbol{X}}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.46a}\\
& \dot{\boldsymbol{V}}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{6.46b}\\
& \dot{\boldsymbol{\eta}}+Q \ddot{\boldsymbol{V}}+Q \widetilde{B}(P \dot{\boldsymbol{V}}+\dot{\boldsymbol{q}}, P \boldsymbol{V}+\dot{\boldsymbol{q}}),+\widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \dot{\boldsymbol{V}}+\dot{\boldsymbol{q}})+c Q \dot{\boldsymbol{V}}=\mathbf{0}  \tag{6.46c}\\
& \dot{\boldsymbol{q}}=\boldsymbol{\eta}  \tag{6.46d}\\
& \boldsymbol{\eta}+Q \widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \boldsymbol{V}+\boldsymbol{q})+c \boldsymbol{q}=\mathbf{0} \tag{6.46e}
\end{align*}
$$

which is evolved forward in time by a DAE solver. Figure 6.7 shows that convergence occurs numerically when $c \gtrsim 21$. As this is no better than the results in Figure 6.5 or the results in Figure 6.1 without using derivative information, we look for a better way of using that information.


Figure 6.7: Absolute error in $\boldsymbol{q}$ for our standard choice of parameters and $c$ varying between 20 and 22. Note the convergence occurring when $c=21$ instead of $c=20$ as in Figure 6.1.

Coupling on higher derivatives of $\boldsymbol{X}$ and incorporating additional constraints produces mixed results when using the DAE solver. We see failure to achieve convergence when coupling on $\boldsymbol{X}$ and its first two derivatives, and convergence at undesirably higher levels of $c$ when another constraint is added. Whether or not additional constraints will lower the critical value of $c$ is unclear on the basis of these numerical results, but it is an inquiry that warrants further investigation, which we discuss in the following section.

### 6.4 Coupling on $\boldsymbol{X}$ and Its Higher Derivatives

Recall the two-layer Lorenz '96 system:

$$
\begin{align*}
\frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X} & =\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.47a}\\
\frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V} & =\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{6.47b}
\end{align*}
$$

Here, $\boldsymbol{F}$ is a constant vector and $\sigma \in \mathbb{R}^{J K \times K}$ is a matrix with orthonormal columns, which in the classical case are given by interpolation and averaging. Given continuous observations of $\boldsymbol{X}$ in time, the goal is to recover $\boldsymbol{V}$. Note the observation of $\boldsymbol{X}$ over time implies $d^{m} \boldsymbol{X} / d t^{m}$ is known for all orders of derivatives. The first differential equation then implies that we also know $d^{m} \sigma^{\top} \boldsymbol{V} / d t^{m}$.

We first compute the derivatives of $\boldsymbol{X}$ and $\boldsymbol{V}$ to carry out any further operations involving those quantities. We do this recursively by setting

$$
\beta_{m}=\frac{d^{m} B(\boldsymbol{X}, \boldsymbol{X})}{d t^{m}} \quad \text { and } \quad \widetilde{\beta}_{m}=\frac{d^{m} \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})}{d t^{m}}
$$

and then noting that

$$
\begin{align*}
\frac{d \boldsymbol{X}}{d t} & =-\beta_{0}-\boldsymbol{X}+\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{6.48}\\
\frac{d \boldsymbol{V}}{d t} & =-\widetilde{\beta}_{0}-c \boldsymbol{V}+\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{6.49}
\end{align*}
$$

and for $m \geq 1$ that

$$
\begin{align*}
& \frac{d^{m+1} \boldsymbol{X}}{d t^{m+1}}=-\beta_{m}-\frac{d^{m} \boldsymbol{X}}{d t^{m}}-\frac{1}{m_{1}} \sigma^{\top} \frac{d^{m} \boldsymbol{V}}{d t^{m}}  \tag{6.50}\\
& \frac{d^{m+1} \boldsymbol{V}}{d t^{m+1}}=-\widetilde{\beta}_{m}-c \frac{d^{m} \boldsymbol{V}}{d t^{m}}+\frac{1}{m_{2}} \sigma \frac{d^{m} \boldsymbol{X}}{d t^{m}} \tag{6.51}
\end{align*}
$$

To finish, we use the binomial theorem for derivatives to explicitly compute $\beta_{m}$ and $\widetilde{\beta}_{m}$ for all values of $m$. In particular,

$$
\beta_{m}=\sum_{j=0}^{m}\binom{m}{j} B\left(\frac{d^{m-j} \boldsymbol{X}}{d t^{m-j}}, \frac{d^{j} \boldsymbol{X}}{d t^{j}}\right) \quad \text { and } \quad \widetilde{\beta}_{m}=\sum_{j=0}^{m}\binom{m}{j} \widetilde{B}\left(\frac{d^{m-j} \boldsymbol{V}}{d t^{m-j}}, \frac{d^{j} \boldsymbol{V}}{d t^{j}}\right)
$$

Consequently, given all derivatives of $\boldsymbol{X}$ and $\boldsymbol{V}$ up to and including order $m$, we recursively obtain derivatives of order $m+1$ for $\boldsymbol{X}$ and $\boldsymbol{V}$. We now use this derivative information to help
recover the complete dynamics of $\boldsymbol{V}$ from continuous observations of $\boldsymbol{X}$.
Upon defining $P=\sigma \sigma^{\top}$ and $Q=I-P$, we obtain

$$
\begin{align*}
P \dot{\boldsymbol{V}}+P \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c P \boldsymbol{V} & =\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{6.52}\\
Q \dot{\boldsymbol{V}}+Q \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c Q \boldsymbol{V} & =\mathbf{0} \tag{6.53}
\end{align*}
$$

Observations of $\boldsymbol{X}, \sigma^{\top} \boldsymbol{V}$, and $\sigma^{\top} \dot{\boldsymbol{V}}$ turn the first equation into an algebraic constraint on $\boldsymbol{V}$. The second describes how the unobserved quantity $Q \boldsymbol{V}$ evolves in time.

We therefore set $\boldsymbol{v}=P \boldsymbol{V}+M \boldsymbol{y}$ where the matrix $M$ represents an orthogonal basis for $\operatorname{col}(Q)$. Noting that $Q=M M^{\top}, M^{\top} M=I$, and $\boldsymbol{y}=M^{\top} \boldsymbol{V}$, we obtain

$$
\begin{align*}
\sigma^{\top} \dot{\boldsymbol{V}}+\sigma^{\top} \widetilde{B}(P \boldsymbol{V}+M \boldsymbol{y}, P \boldsymbol{V}+M \boldsymbol{y})+c \sigma^{\top} \boldsymbol{V} & =\frac{1}{m_{2}} \boldsymbol{X}  \tag{6.54}\\
\dot{y}+M^{\top} \widetilde{B}(P \boldsymbol{V}+M \boldsymbol{y}, P \boldsymbol{V}+M \boldsymbol{y})+c \boldsymbol{y} & =\mathbf{0} \tag{6.55}
\end{align*}
$$

The first equation is an quadratic constraint on the unknown $y$, and the second governs the evolution of that unknown.

The algebraic constraint on $\boldsymbol{y}$ involves $K$ equations with $J K-K$ unknowns. In the standard setting where $J=32$ and $K=8$, this constraint corresponds to 8 equations with 248 unknowns. Additional constraints are obtained by differentiation.

Again using the binomial theorem for derivatives, we obtain

$$
\rho_{m}=\frac{d^{m} \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})}{d t^{m}}=\sum_{j=0}^{m}\binom{m}{j} \widetilde{B}\left(\frac{d^{m-j} \boldsymbol{v}}{d t^{m-j}}, \frac{d^{j} \boldsymbol{v}}{d t^{j}}\right)
$$

Now since

$$
\begin{align*}
\dot{\boldsymbol{v}} & =P \dot{\boldsymbol{V}}+M \dot{\boldsymbol{y}}=P \dot{\boldsymbol{V}}-M\left(M^{\top} \widetilde{B}(P \boldsymbol{V}+M \boldsymbol{y}, P \boldsymbol{V}+M \boldsymbol{y})+c \boldsymbol{y}\right) \\
& =P \dot{\boldsymbol{V}}-Q(\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c \boldsymbol{v})=P \dot{\boldsymbol{V}}-Q\left(\rho_{0}+c \boldsymbol{v}\right) \tag{6.56}
\end{align*}
$$

we find for $m \geq 1$ that

$$
\begin{equation*}
\frac{d^{m+1} \boldsymbol{v}}{d t^{m+1}}=P \frac{d^{m+1} \boldsymbol{V}}{d t^{m+1}}-Q\left(\rho_{m}+c \frac{d^{m} \boldsymbol{v}}{d t^{m}}\right) \tag{6.57}
\end{equation*}
$$

With the time derivatives of $\boldsymbol{X}, \boldsymbol{V}$, and $\boldsymbol{v}$ in hand, we obtain the constraints

$$
\varphi_{2}(\boldsymbol{y})=\sigma^{\top}\left(\dot{\boldsymbol{V}}+\rho_{0}+c \boldsymbol{V}\right)-\frac{1}{m_{2}} \boldsymbol{X}
$$

and for $n \geq 1$ that

$$
\varphi_{n+2}(\boldsymbol{y})=\sigma^{\top}\left(\frac{d^{n+1} \boldsymbol{V}}{d t^{n+1}}+\rho_{n}+c \frac{d^{n} \boldsymbol{V}}{d t^{n}}\right)-\frac{1}{m_{2}} \frac{d^{n} \boldsymbol{X}}{d t^{n}}
$$

By construction we know that $\varphi_{m}\left(M^{\top} \boldsymbol{V}\right)=0$ for $m \geq 2$. Note that the subscript $m$ corresponds to the polynomial degree of the constraint.

We further include a Tikhonov regularizing term given by $\varphi_{1}(\boldsymbol{y})=\boldsymbol{y}-\boldsymbol{y}_{\boldsymbol{p}}$ of degree one where $\boldsymbol{y}_{\boldsymbol{p}}$ is the predicted value of $\boldsymbol{y}$ that will be corrected by the other constraints. If the predicted $\boldsymbol{y}_{\boldsymbol{p}}=M^{\top} \boldsymbol{V}$ then also $\varphi_{1}\left(M^{\top} \boldsymbol{V}\right)=0$. This motivates the following data-assimilation algorithm to recover the complete value of $\boldsymbol{V}$.

Given the prediction $\boldsymbol{y}_{\boldsymbol{p}}\left(t_{n}\right)$, minimize

$$
J(\boldsymbol{y})=\frac{1}{2} \sum_{m=1}^{d} \gamma_{m}\left\|\varphi_{m}(\boldsymbol{y})\right\|^{2}=\frac{1}{2}\|\Phi(\boldsymbol{y})\|^{2} \quad \text { where } \quad \Phi(y)=\left[\begin{array}{c}
\gamma_{1} \varphi_{1}(\boldsymbol{y}) \\
\vdots \\
\gamma_{d} \varphi_{d}(\boldsymbol{y})
\end{array}\right]
$$

to obtain the corrected value $\boldsymbol{y}_{\boldsymbol{c}}\left(t_{n}\right)$. Here the weights $\gamma_{d}$ are tunable with

$$
\sum_{m=1}^{d} \gamma_{m}=1 \quad \text { and } \quad \gamma_{1} \ll 1
$$

to discount the regularizing term in favor of other constraints.
After finding the minimizer $y_{c}\left(t_{n}\right)$, we solve the differential equation

$$
\dot{y}+M^{\top} \widetilde{B}(P \boldsymbol{V}+M \boldsymbol{y}, P \boldsymbol{V}+M \boldsymbol{y})+c \boldsymbol{y}=0
$$

on the interval $\left[t_{n}, t_{n+1}\right]$ with initial condition $\boldsymbol{y}\left(t_{n}\right)=\boldsymbol{y}_{\boldsymbol{c}}\left(t_{n}\right)$ to obtain the next prediction $\boldsymbol{y}_{\boldsymbol{p}}\left(t_{n+1}\right)=\boldsymbol{y}\left(t_{n+1}\right)$. This process is repeated with the new prediction $\boldsymbol{y}_{\boldsymbol{p}}\left(t_{n+1}\right)$.

Minimizing $J$ is a nonlinear least-squares problem. Under the assumption that $\boldsymbol{y}_{\boldsymbol{p}}$ is close to the desired minimum, we suppose the Gauss-Newton method will converge. Set $\boldsymbol{y}_{\mathbf{0}}=\boldsymbol{y}_{\boldsymbol{p}}$ and for


Figure 6.8: A representative graph of the absolute error in $\boldsymbol{y}$ for 3 Gauss-Newton iterations, $m=5$, our standard choice of parameters, and $c$ varying between 18 and 22 . Note the convergence occurring around $c=19.5$ instead of $c=20$ as in Figure 6.1.
$\ell=0,1, \ldots$ linearize $\Phi(\boldsymbol{y})$ around $\boldsymbol{y}_{\boldsymbol{\ell}}$ to obtain

$$
\Phi(\boldsymbol{y}) \approx \Phi\left(\boldsymbol{y}_{\ell}\right)+D \Phi\left(\boldsymbol{y}_{\ell}\right)\left(\boldsymbol{y}-\boldsymbol{y}_{\ell}\right)
$$

where $D$ is differentiation with respect to $\boldsymbol{y}$ and $D \Phi\left(y_{\ell}\right) \in \mathbb{R}^{(J K-K+d K) \times(J K-K)}$. Since $D \varphi_{1}(\boldsymbol{y})=$ $I$ the columns of $D \Phi\left(\boldsymbol{y}_{\ell}\right)$ are always linearly independent.

Let $J_{\ell}$ be the corresponding version of $J$ about $\boldsymbol{y}_{\ell}$ given by

$$
J_{\ell}(\boldsymbol{y})=\frac{1}{2}\left\|\Phi\left(\boldsymbol{y}_{\ell}\right)+D \Phi\left(\boldsymbol{y}_{\ell}\right)\left(\boldsymbol{y}-\boldsymbol{y}_{\ell}\right)\right\|^{2}
$$

Minimizing $J_{\ell}$ is a linear least squares problem. Upon defining $y_{\ell+1}$ to be the minimizer of $J_{\ell}$, iteration hopefully leads to sequence which converges to the minimizer of $J(\boldsymbol{y})$.

To avoid rounding errors write $\boldsymbol{x}=\boldsymbol{y}_{\boldsymbol{\ell}}-\boldsymbol{y}$ and solve the overconstrained problem

$$
D \Phi\left(\boldsymbol{y}_{\ell}\right) \boldsymbol{x}=\Phi\left(\boldsymbol{y}_{\ell}\right) \quad \text { in Julia as } \quad \boldsymbol{x}=D \Phi\left(\boldsymbol{y}_{\ell}\right) \backslash \Phi\left(\boldsymbol{y}_{\ell}\right) .
$$

It then follows that $\boldsymbol{y}_{\ell+\boldsymbol{1}}=\boldsymbol{y}_{\boldsymbol{\ell}}-\boldsymbol{x}$.

The only thing left to carry out this procedure is the computation of

$$
D \Phi(\boldsymbol{y})=\left[\begin{array}{c}
\gamma_{1} D \varphi_{1}(\boldsymbol{y}) \\
\vdots \\
\gamma_{d} D \varphi_{d}(\boldsymbol{y})
\end{array}\right]
$$

which we now describe in detail. We obtain

$$
\begin{align*}
& D \boldsymbol{v}=D(P \boldsymbol{V}+M \boldsymbol{y})=M  \tag{6.58}\\
& D \dot{\boldsymbol{v}}=D\left(P \dot{\boldsymbol{V}}-Q\left(\rho_{0}+c \boldsymbol{v}\right)\right)=-Q\left(D \rho_{0}+c D \boldsymbol{v}\right) \tag{6.59}
\end{align*}
$$

and for $n \geq 1$ that

$$
\begin{equation*}
D \frac{d^{m+1} \boldsymbol{v}}{d t^{m+1}}=-Q\left(D \rho_{m}+c D \frac{d^{m} \boldsymbol{v}}{d t^{m}}\right) \tag{6.60}
\end{equation*}
$$

where

$$
D \rho_{m}=\sum_{j=0}^{m}\binom{m}{j}\left\{\widetilde{B}\left(D \frac{d^{m-j} \boldsymbol{v}}{d t^{m-j}}, \frac{d^{j} \boldsymbol{v}}{d t^{j}}\right)+\widetilde{B}\left(\frac{d^{m-j} \boldsymbol{v}}{d t^{m-j}}, D \frac{d^{j} \boldsymbol{v}}{d t^{j}}\right)\right\}
$$

Note that when $A \in \mathbb{R}^{p \times q}$ and $\boldsymbol{v} \in \mathbb{R}^{p}$, the notation above means

$$
\begin{align*}
& \widetilde{B}(A, \boldsymbol{v})=\left[\widetilde{B}\left(\boldsymbol{a}_{\mathbf{1}}, \boldsymbol{v}\right)|\cdots| \widetilde{B}\left(\boldsymbol{a}_{\boldsymbol{q}}, \boldsymbol{v}\right)\right]  \tag{6.61}\\
& \widetilde{B}(\boldsymbol{u}, A)=\left[\widetilde{B}\left(\boldsymbol{u}, \boldsymbol{a}_{\boldsymbol{1}}\right)|\cdots| \widetilde{B}\left(\boldsymbol{u}, \boldsymbol{a}_{\boldsymbol{q}}\right)\right] \tag{6.62}
\end{align*}
$$

where

$$
A=\left[a_{1}|\cdots| a_{q}\right] .
$$

Having recursively obtained the above derivatives, one may now compute

$$
\begin{align*}
& D \varphi_{1}(\boldsymbol{y})=D\left(\boldsymbol{y}-\boldsymbol{y}_{\boldsymbol{p}}\right)=I,  \tag{6.63}\\
& D \varphi_{2}(\boldsymbol{y})=D\left(\sigma^{\top}\left(\dot{\boldsymbol{V}}+\rho_{0}+c \boldsymbol{V}\right)-\frac{1}{m_{2}} \boldsymbol{X}\right)=\sigma^{\top} D \rho_{0}, \tag{6.64}
\end{align*}
$$

and for $m \geq 1$,

$$
\begin{equation*}
D \varphi_{m+2}(\boldsymbol{y})=\sigma^{\top} D \rho_{m} \tag{6.65}
\end{equation*}
$$

To test the efficacy of these constraints, we calculate $\|M \boldsymbol{V}-\boldsymbol{y}\|$ using $m=5$ constraints. Figure 6.8 shows the trajectories of $\|M \boldsymbol{V}-\boldsymbol{y}\|$ for $c$ between 18 and 22 . Convergence occurs around $c=19.5$, noticeably earlier than when we couple on $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$. Further experiments might be needed to determine whether larger numbers of constraints or Newton iterations will lower $c$.

## Chapter 7

## Coupling on $X$ with Unknown Dynamics

So far, we established conditions under which $\lim _{t \rightarrow \infty}\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ and developed recursion relations for coupling on $\boldsymbol{X}$ and its derivatives. These results were created under the assumption that the dynamical equations which govern the motion of $\boldsymbol{X}$ and $\boldsymbol{V}$ are known, and that only the values of $\boldsymbol{V}$ are unknown. In practical problems, it can happen that these dynamics are unknown or only known to an approximation. We now turn to the second inverse problem in which both $c$ and $\boldsymbol{V}$ are unknown.

Typically, the equations governing the observed quantities are better known than the equations governing the unobserved quantities. Therefore, we assume that the equations governing the motion of $\boldsymbol{X}$ are known exactly but the equations governing the motion of $\boldsymbol{V}$ are known up to some approximation. Even if the form of the equation governing the motion of $\boldsymbol{V}$ is known, some parameters may not be known - in this case, $c, m_{2}, \sigma$, and the dimension of the system $J K$. We consider here the case in which only the parameter $c$ is unknown. This represents the simplest case when the dynamics governing the motion of the unobserved state $\boldsymbol{V}$ are known only approximately. The goal of this chapter is to use observations of $\boldsymbol{X}$ to infer not only the state of $\boldsymbol{V}$ but the value of $c$ which governs the motion of $\boldsymbol{V}$.

### 7.1 The Effects of Approximate Dynamics

We begin by considering coupling on $\boldsymbol{X}$ as in Section 4.1, except with the exact value of $c$ denoted by $\tilde{c}$ in the equation governing $\boldsymbol{V}$. Thus,

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{7.1a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+\tilde{c} \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{7.1b}\\
& \frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{7.1c}
\end{align*}
$$

The following proposition establishes a bound on $\boldsymbol{V}-\boldsymbol{v}$ when $c$ is unknown.

Proposition 7.1. Define $\boldsymbol{\Delta}=\boldsymbol{V}-\boldsymbol{v}$. Then

$$
\begin{equation*}
\|\boldsymbol{\Delta}\|^{2} \leq\left\|\boldsymbol{\Delta}_{\mathbf{0}}\right\|^{2} e^{-(c-4 M) t}+\frac{|c-\tilde{c}|^{2}}{c(c-4 M)} J K M^{2}\left[1-e^{-(c-4 M) t}\right] \tag{7.2}
\end{equation*}
$$

where $M$ is a uniform bound on $\|\boldsymbol{V}\|$.

Proof. Let $\boldsymbol{\Delta}$ be as above. Then

$$
\begin{equation*}
\frac{d \boldsymbol{\Delta}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+\tilde{c} \boldsymbol{V}-c \boldsymbol{v}=0 \tag{7.3}
\end{equation*}
$$

Adding and subtracting $\tilde{c} \boldsymbol{V}$ yields

$$
\begin{equation*}
\frac{d \boldsymbol{\Delta}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+(c-\tilde{c}) \boldsymbol{V}+c \boldsymbol{\Delta}=0 \tag{7.4}
\end{equation*}
$$

Adding and subtracting $\widetilde{B}(\boldsymbol{V}, \boldsymbol{v})$ yields

$$
\begin{equation*}
\frac{d \boldsymbol{\Delta}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})-\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{v})+(c-\tilde{c}) \boldsymbol{V}+c \boldsymbol{\Delta}=0 \tag{7.5}
\end{equation*}
$$

Adding and subtracting $\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V})$ yields

$$
\begin{equation*}
\frac{d \boldsymbol{\Delta}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})+\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{\Delta})+(c-\tilde{c}) \boldsymbol{V}+c \boldsymbol{\Delta}=0 \tag{7.6}
\end{equation*}
$$

Next, we dot both sides with $\boldsymbol{\Delta}$. By Lemma $3.1, \widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{\Delta}) \cdot \boldsymbol{\Delta}$ disappears, and

$$
\begin{equation*}
\frac{1}{2} \frac{d\|\boldsymbol{\Delta}\|^{2}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta}) \cdot \boldsymbol{\Delta}+\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V}) \cdot \boldsymbol{\Delta}+(c-\tilde{c}) \boldsymbol{V} \cdot \boldsymbol{\Delta}+c\|\boldsymbol{\Delta}\|^{2}=0 \tag{7.7}
\end{equation*}
$$

Now we estimate $\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V}) \cdot \boldsymbol{\Delta}$ and $\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta}) \cdot \boldsymbol{\Delta}$. By part 5 of Lemma 3.1,

$$
\begin{equation*}
|[\widetilde{B}(\boldsymbol{\Delta}, \boldsymbol{V})+\widetilde{B}(\boldsymbol{V}, \boldsymbol{\Delta})] \cdot \boldsymbol{\Delta}| \leq 2\|\boldsymbol{V}\|\|\boldsymbol{\Delta}\|^{2} \tag{7.8}
\end{equation*}
$$

Thus, (7.7) becomes

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{\Delta}\|^{2}+(c-\tilde{c}) \boldsymbol{V} \cdot \boldsymbol{\Delta}+c\|\boldsymbol{\Delta}\|^{2} \leq 2\|\boldsymbol{V}\|\|\boldsymbol{\Delta}\|^{2} \tag{7.9}
\end{equation*}
$$

and we rearrange it and apply the Cauchy-Schwarz inequality to obtain

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{\Delta}\|^{2}+(c-2\|\boldsymbol{V}\|)\|\boldsymbol{\Delta}\|^{2} \leq(c-\tilde{c}) \boldsymbol{V} \cdot \boldsymbol{\Delta} \leq|c-\tilde{c}|\|\boldsymbol{V}\|\|\boldsymbol{\Delta}\| . \tag{7.10}
\end{equation*}
$$

To this we apply Young's inequality with $\epsilon=c$, and we get

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{\Delta}\|^{2}+(c-2\|\boldsymbol{V}\|)\|\boldsymbol{\Delta}\|^{2} \leq \frac{|c-\tilde{c}|^{2}}{2 c}\|\boldsymbol{V}\|^{2}+\frac{c}{2}\|\boldsymbol{\Delta}\|^{2} . \tag{7.11}
\end{equation*}
$$

Rearranging and multiplying through by 2 yields

$$
\begin{equation*}
\frac{d\|\boldsymbol{\Delta}\|^{2}}{d t}+(c-4\|\boldsymbol{V}\|)\|\boldsymbol{\Delta}\|^{2} \leq \frac{|c-\tilde{c}|^{2}}{c}\|\boldsymbol{V}\|^{2} \tag{7.12}
\end{equation*}
$$

Define $\theta(t)=\frac{|c-\tilde{c}|^{2}}{c}\|\boldsymbol{V}\|^{2}$ and note $\|\boldsymbol{V}\| \leq M$. Then

$$
\begin{equation*}
\frac{d\|\boldsymbol{\Delta}\|^{2}}{d t}+(c-4 M)\|\boldsymbol{\Delta}\|^{2} \leq \theta(t) . \tag{7.13}
\end{equation*}
$$

Now let $e^{(c-4 M) t}$ be an integrating factor. Then

$$
\begin{equation*}
\frac{d}{d t} e^{(c-4 M) t}\|\boldsymbol{\Delta}\|^{2} \leq e^{(c-4 M) t} \theta(t) \tag{7.14}
\end{equation*}
$$

and integrating this result yields

$$
\begin{equation*}
\|\boldsymbol{\Delta}\|^{2} \leq\left\|\boldsymbol{\Delta}_{\mathbf{0}}\right\|^{2} e^{-(c-4 M) t}+e^{-(c-4 M) t} \int_{0}^{t} s e^{(\tilde{c}-4 M) t} \theta(s) d s \tag{7.15}
\end{equation*}
$$

Next, note $\|\boldsymbol{V}\|^{2}=\sum_{j=1}^{J K}\left|V_{i}\right|^{2} \leq \sum_{j=1}^{J K} M^{2}=J K M^{2}$. Thus,

$$
\begin{equation*}
\theta(s)=\frac{|c-\tilde{c}|^{2}}{c}\|\boldsymbol{V}\|^{2} \leq \frac{|c-\tilde{c}|^{2}}{c} J K M^{2} \tag{7.16}
\end{equation*}
$$

and so

$$
\begin{equation*}
e^{-(c-4 M) t} \frac{1}{c-4 M}\left(e^{(c-4 M) t}-1\right) \frac{|c-\tilde{c}|^{2}}{c} J K M^{2}=\frac{|c-\tilde{c}|^{2}}{c(c-4 M)} J K M^{2}\left(1-e^{-(c-4 M) t}\right) \tag{7.17}
\end{equation*}
$$

The result follows.

The results of Proposition 7.1 are meaningful when $c>4 M$. As noted in (6.29) and (6.30), whether $M$ represents an a priori or an a posteriori bound on $\boldsymbol{V}$, much smaller values of $c$ work numerically. We now illustrate the result of Proposition 7.1 when there is a mismatch between our approximation of $c$ and the exact value $\tilde{c}$.

We take $\tilde{c}=21$ and vary $c$ in a neighborhood about $\tilde{c}$. Following the procedure used in Section 4 to obtain 3600 independent samples, we use a time step size of $h=1 / 2048$ and warm up the reference solution $\boldsymbol{V}$ for each value of $c$ for time $T=50$. Then we initialize $\boldsymbol{v}=P \boldsymbol{V}$ and warm up the approximating solution until $\boldsymbol{v}$ no longer reflects the initial condition $\boldsymbol{v}_{\mathbf{0}}$ and begins to resemble its long-time behavior.

Again, the autocorrelation results discussed in Section 3.2 lead us to sample the trajectory at intervals of $\Delta t=20$. The graph on the left in Figure 7.1 illustrates the resulting statistics of the relative error $\|\boldsymbol{V}-\boldsymbol{v}\| /\|\boldsymbol{V}\|$ for $c$ between 20.5 and 21.5.

It is notable that when the approximate value of $c$ differs by as little as 0.05 , the whiskers extend to $100 \%$ relative error, which means once in a while, the approximating solution $\boldsymbol{v}$ loses synchronicity with $\boldsymbol{V}$ with the result that our recovery of the mesoscale and microscale state represented by $\boldsymbol{v}$ is occasionally quite wrong. On the other hand, the red dots show that $95 \%$ of the time, the relative errors in our approximation are much smaller. The aqua dots tell a similar story for $90 \%$ of the time. Having an occasional outlier in the error means that any subsequent predictions based on that particular estimated state $\boldsymbol{v}$ of $\boldsymbol{V}$ will be worse than expected. In this section, we seek not only to estimate the unknown value of $c$ but to reduce the size of these outliers.

The coupling given in Equation (6.42c) appears less stable with respect to errors in our estimate of $c$ : i.e., the sizes of the outliers are still large. Recall the equations governing the evolution of
$\boldsymbol{q}$ are given by

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{7.18a}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+\tilde{c} \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{7.18b}\\
& \frac{d \boldsymbol{q}}{d t}+Q \widetilde{B}(P \boldsymbol{V}+\boldsymbol{q}, P \boldsymbol{V}+\boldsymbol{q})+c \boldsymbol{q}=\mathbf{0} \tag{7.18c}
\end{align*}
$$

where the value of $c$ in the last equation represents an approximate value. We follow the same warmup procedure as used in Figure 7.1.

In Figure 7.1, the graph on the right illustrates the resulting statistics of the relative error $\|Q \boldsymbol{V}-\boldsymbol{q}\| /\|\boldsymbol{V}\|$ for $c$ between 20.5 and 21.5. Note that the maximum size of the outliers still exceeds $100 \%$ and the red and aqua dots are slightly higher than those in the graph on the left.


Figure 7.1: Statistics of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (7.1c) (left) and in the approximation $P \boldsymbol{V}+\boldsymbol{q}$ of $\boldsymbol{V}$ given by Equation (7.18c) (right) over 3600 independent samples. The relative error is less than the red dots $95 \%$ of the time and less than the aqua dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error.

To further illustrate the effects of not knowing the exact dynamics governing $\boldsymbol{V}$-i.e., not knowing the correct value of $c$-we repeat the simulations for Figure 7.1 and for Figure 7.2 with the


Figure 7.2: Statistics, given a fixed arbitrary coupling matrix $\sigma$, of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (7.1c) (left) and in the approximation $P \boldsymbol{V}+\boldsymbol{q}$ of $\boldsymbol{V}$ given by Equation (7.18c) (right) over 3600 independent samples. The relative error is less than the red dots $95 \%$ of the time and less than the aqua dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error.
random coupling matrices $\sigma$ described in Section 6.2. As before, we create $\sigma$ by initializing a matrix of random values, perform a QR decomposition, and take $\sigma$ to be the matrix consisting of the first eight columns of $Q$. Although the critical value of $c$ at which synchronization occurs is generally smaller for an arbitrary coupling matrix, we again set $\tilde{c}=21$ and allow our approximation of $c$ to vary between 20.5 and 21.5.

Figure 7.2 shows a substantial improvement to the relative error when given a typical arbitrary coupling matrix $\sigma$. We performed this simulation for 10 different random $\sigma$, and we obtained similar results each time. Unlike the results in Figure 7.1, the relative errors on the right appear smaller than those on the left.

While it is interesting that these random $\sigma$ show substantial improvement to the relative error in the approximating solution, it is worth remembering what $\sigma$ in Equation (3.1) represents physically: the coupling between the slow, large-amplitude variables and the fast, small-amplitude variables. This coupling is given by the physical laws of nature. Though it would be convenient
to change $\sigma$, mathematically, we generally cannot do so, as we cannot change the laws of nature to make our inverse problem easier to solve.

### 7.2 Solving for $c$

In this section and the rest of this research, we return to the standard coupling matrix given in Equation (3.1). In this case, even a close approximation of $c$ leads to occasional errors approaching $100 \%$ as shown in Figure 7.1. To remedy this defect, we introduce an equation for dynamically tuning $c$ as the observations of $\boldsymbol{X}$ are assimilated.

Assume $\boldsymbol{X}$ is known and $\boldsymbol{V}$ and $\tilde{c}$ are not. Because we know $\boldsymbol{X}$, we also know $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$. We remark that the use of derivative information to solve for $c$ appears related to the fact that $c$ does not appear directly in the equation for our observations $\boldsymbol{X}$. In particular, other methods in the literature typically assume that the unknown parameter appears in the equation governing the observations. Thus, to our knowledge, the algorithm developed in this section is original.

To start, we know

$$
\begin{equation*}
\sigma^{\top} \boldsymbol{V}=-m_{1}(\dot{\boldsymbol{X}}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}-\boldsymbol{F}) \tag{7.19}
\end{equation*}
$$

and

$$
\begin{equation*}
\sigma^{\top} \dot{\boldsymbol{V}}=-m_{1}(\ddot{\boldsymbol{X}}+B(\boldsymbol{X}, \dot{\boldsymbol{X}})+B(\dot{\boldsymbol{X}}, \boldsymbol{X})+\dot{\boldsymbol{X}}) . \tag{7.20}
\end{equation*}
$$

To solve for the unknown $\tilde{c}$, rearrange (7.18a) as

$$
\begin{equation*}
\tilde{c} \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}-\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\dot{\boldsymbol{V}} \tag{7.21}
\end{equation*}
$$

and we multiply through by $\sigma^{\top}$ to obtain

$$
\begin{equation*}
\tilde{c} \sigma^{\top} \boldsymbol{V}=\frac{1}{m_{2}} \boldsymbol{X}-\sigma^{\top} \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\sigma^{\top} \dot{\boldsymbol{V}} \tag{7.22}
\end{equation*}
$$

If we dot Equation (7.21) on the right with $\sigma^{\top} \boldsymbol{V}$, we get

$$
\begin{equation*}
\tilde{c}\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}=\frac{1}{m_{2}} \boldsymbol{X} \cdot \sigma^{\top} \boldsymbol{V}-\sigma^{\top} \widetilde{B}(\boldsymbol{V}, \boldsymbol{V}) \cdot \sigma^{\top} \boldsymbol{V}-\sigma^{\top} \dot{\boldsymbol{V}} \cdot \sigma^{\top} \boldsymbol{V} \tag{7.23}
\end{equation*}
$$

Dividing through by $\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}$ yields

$$
\begin{equation*}
\tilde{c}=\frac{\left[\frac{1}{m_{2}} \boldsymbol{X}-\sigma^{\top} \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\sigma^{\top} \dot{\boldsymbol{V}}\right] \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}} \tag{7.24}
\end{equation*}
$$

Because $\boldsymbol{V}$ is unknown and has to be inferred, Equation (7.24) does not allow us to solve exactly for $\tilde{c}$. As already mentioned, $\sigma^{\top} \boldsymbol{V}$ and $\sigma^{\top} \dot{\boldsymbol{V}}$ are determined through $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$. However, $\sigma^{\top} \widetilde{B}(\boldsymbol{V}, \boldsymbol{V})$ on the right-hand side of Equation (7.24) is unknown. Under the assumption that $\boldsymbol{v} \approx \boldsymbol{V}$, we replace this term with $\sigma^{\top} \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})$ to approximate $\tilde{c}$ using

$$
\begin{equation*}
c_{1}(t)=\frac{\left[\frac{1}{m_{2}} \boldsymbol{X}-\sigma^{\top} \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})-\sigma^{\top} \dot{\boldsymbol{V}}\right] \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}} \tag{7.25}
\end{equation*}
$$

We remark that the denominator in the expression for $c_{1}(t)$ must be nonzero. However, from an analytic point of view, it is unclear whether $\boldsymbol{V}$ passes through $\mathbf{0}$ as the system evolves forward in time. In general, this question is related to whether there are points on the attractor with $\boldsymbol{V}=\mathbf{0}$ and, specifically, whether there are points with $\sigma^{\top} \boldsymbol{V}=\mathbf{0}$.

Now substitute this approximation into (7.1c) to obtain

$$
\begin{equation*}
\frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+c_{1}(t) \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{7.26}
\end{equation*}
$$

Figure 7.3 shows that convergence is achieved for $\tilde{c}=22$. However, for $\tilde{c}=21$, not only does $c_{1}(t)$ fail to recover the unknown value of $\tilde{c}$, but it eventually goes negative and $\boldsymbol{v}$ blows up.

Before proceeding, it is instructive to examine the expression for $c_{1}(t)$ given in Equation (7.25) to understand what terms cancel in order to gain $\tilde{c}$, and to highlight the fact that the approximation deteriorates as the denominator decreases. To do this, we substitute Equation (7.1b) to obtain

$$
\begin{align*}
c_{1}(t) & =\frac{\left[\frac{1}{m_{2}} \boldsymbol{X}-\sigma^{\top} \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+\sigma^{\top}\left(\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+\tilde{c} \boldsymbol{V}-\frac{1}{m_{2}} \sigma \boldsymbol{X}\right)\right] \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}}  \tag{7.27}\\
& =\frac{\left[\sigma^{\top}(\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v}))+\tilde{c} \sigma^{\top} \boldsymbol{V}\right] \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}} \\
& =\frac{\sigma^{\top}(\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})) \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}}+\tilde{c} .
\end{align*}
$$

Although an analytic treatment of this term appears beyond reach, we can numerically observe the denominator and note that when it gets smaller, our approximation of $\tilde{c}$ may become less accurate. Also note that the better the cancellation between the two bilinear terms $\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})$ and $\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})$, the better our approximation of $\tilde{c}$ will be. Ideally, $\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ hand in hand with $c_{1}(t) \rightarrow \tilde{c}$. We emphasize, however, that unexpected smallness of $\left\|\sigma^{\top} \boldsymbol{V}\right\|$ might spoil this even when $\tilde{c}$ is large.


Figure 7.3: Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=21$, and our standard parameter regime. Sometimes, the solution blows up (left), and other times it simply fails to converge (right).

Figure 7.3 illustrates the evolution of $c_{1}(t)$ when $\tilde{c}=21$. In the graph on the left, the solution blows up before $t=20$. Out of 100 random trials, this happened $89 \%$ of the time. Numerically, what seems to happen is that our approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ degrades to the point where $c_{1}(t)<$ 0 . Subsequently, energy is pumped into Equation (7.26): i.e., the damping is converted into exponential growth, and the approximation $\boldsymbol{v}$ becomes even worse. At this point, we reach the stability limits of our numerical scheme, and the floating point arithmetic overflows to the point that we get NaN. We remark that no blowup occurred in the graph on the right, although it may at a later time if a longer simulation is run.

Although $\tilde{c}=21$ resulted in convergence when $c$ was known, as shown in Figures 6.1 and 6.5 having to estimate $c$ and $\boldsymbol{V}$ does not result in good estimates of convergence when we use Equation (7.26). This is not entirely unexpected. When $c$ is larger, there is greater dissipation, and it becomes easier to solve for $c$. This is illustrated in Figures 7.4 and 7.5.


Figure 7.4: Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=23$, and our standard parameter regime.

Even though the calculations in these figures indicate a good skill in recovering both $\tilde{c}$ and $\boldsymbol{V}$, we still have the difficulty that the denominator in $c_{1}(t)$ is not well controlled. In particular, the approximating solution sometimes blows up when the experiment is performed with the same parameters but the reference solution is warmed up from a different random initial condition. Before considering changes to regularize the denominator, we first explore the idea of improving the cancellation in the numerator.

For example, more cancellation may occur if, instead of $\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})$, we used $\widetilde{B}(P \boldsymbol{V}+Q \boldsymbol{v}, P \boldsymbol{V}+Q \boldsymbol{v})$ for the approximation to obtain

$$
\begin{equation*}
c_{2}(t)=\frac{\sigma^{\top}(\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})-\widetilde{B}(P \boldsymbol{V}+Q \boldsymbol{v}, P \boldsymbol{V}+Q \boldsymbol{v})) \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}}+\tilde{c} . \tag{7.28}
\end{equation*}
$$



Figure 7.5: Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=24$, and our standard parameter regime.

Approximating Solutions That Blow Up


Figure 7.6: Percent of approximating solutions computed using $c_{1}(t)$ compared with those computed using $c_{2}(t)$ that blow up on the time interval [0,20] from an ensemble of 100 reference solutions.

We again remark that $P \boldsymbol{V}$ is known from $\dot{\boldsymbol{X}}$ and $Q \boldsymbol{v}$ is obtained through the evolution of Equation (7.26) with $c_{1}$ replaced by $c_{2}$.

To compare the approximations of $\tilde{c}$ obtained from $c_{1}$ and $c_{2}$ for each value of $\tilde{c}$ in the range from 21 to 26 , we consider an ensemble of 100 reference solutions. For each $\tilde{c}$, each solution in its ensemble is obtained starting with a random initial condition, which is then warmed up for time $T=50$ so that it reflects the long-term dynamics from the choice of $\tilde{c}$ and other parameters. At this point, we start computing an approximating solution $\boldsymbol{v}$ using Equation (7.26) and another one using the same equation except with $c_{2}$.

Figure 7.6 characterizes how the tendency of the approximating solution to blow up depends on the value of $\tilde{c}$ and whether $c_{1}$ or $c_{2}$ is used to approximate $\tilde{c}$. Note that as $\tilde{c}$ increases, it is less likely that the solution $\boldsymbol{v}$ will blow up. Note also that there is not that much difference between $c_{1}$ and $c_{2}$ in terms of what percentage of approximations blow up. We remark that it is difficult to compare the average error in the two different ways of approximating $\boldsymbol{v}$ given the tendency of the approximations to blow up.

One possible remedy is to introduce a lower bound on our approximation of $\tilde{c}$. For example, if it were known that $\tilde{c} \geq 10$, we could compute $\boldsymbol{v}$ as

$$
\begin{equation*}
\frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+\max \left(10, c_{i}(t)\right) \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{7.29}
\end{equation*}
$$

The advantage of Equation (7.29) is that it ensures there is a dissipation of at least 10 in the $\boldsymbol{v}$ equation, regardless of how bad our approximation $c_{i}(t)$ is. The real problem being avoided here is when $c_{i}(t)<0$, at which point energy is being injected into the system; this problem is resolved by imposing a cutoff. This allows us to prove

Theorem 7.2. Let $\boldsymbol{v}$ be governed by Equation (7.29). Then $\|\boldsymbol{v}\|$ remains uniformly bounded in time regardless of whether $\boldsymbol{v}$ is synchronized by $\boldsymbol{V}$. Specifically, we have

$$
\begin{equation*}
\|\boldsymbol{v}\|^{2} \leq\left\|\boldsymbol{v}_{0}\right\|^{2}+\frac{L_{1}}{c^{2}} \quad \text { for all } \quad t, \quad \text { and } \quad \quad \underset{t \rightarrow \infty}{\limsup }\|\boldsymbol{v}\|^{2} \leq \frac{L_{2}}{c} \tag{7.30}
\end{equation*}
$$

where $L_{1}$ and $L_{2}$ are given in (6.6) and (6.8), respectively.

Proof. The proof is similar to that of Theorem 6.1. Since we have chosen 10 to be the minimum cutoff for our estimates of $\tilde{c}$, it is reasonable to assume $\tilde{c} \geq 10$. In this case, Lemma 5.1 provides bounds on $\psi=m_{1}\|\boldsymbol{X}\|^{2}+m_{2}\|\boldsymbol{V}\|^{2}$. In particular,

$$
\psi \leq \psi_{0} e^{-\alpha t}+\frac{\nu}{\alpha}\left(1-e^{-t}\right), \quad \text { where } \quad \alpha=2-\frac{1}{\delta} \quad \text { and } \quad \nu=m_{1} \delta\|\boldsymbol{F}\|^{2} .
$$

Here, $\delta>\frac{1}{2}$.

Next, we take (7.29) and dot it with $\boldsymbol{v}$ to get

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2}+\max \left(10, c_{i}(t)\right)\|\boldsymbol{v}\|^{2}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \cdot \boldsymbol{v} . \tag{7.31}
\end{equation*}
$$

Consequently,

$$
\begin{equation*}
\frac{1}{2} \frac{d}{d t}\|\boldsymbol{v}\|^{2}+10\|\boldsymbol{v}\|^{2} \leq \frac{1}{m_{2}} \sigma \boldsymbol{X} \cdot \boldsymbol{v} . \tag{7.32}
\end{equation*}
$$

Taking $c=10$ allows us to finish the proof as in Theorem 6.1.

We remark that the above theorem guarantees that there is no blowup anymore. This is further illustrated numerically by Figure 7.7. On the left, the approximation $c_{1}(t)<0$ around $t=3.0$, which leads to blowup of our approximating solution $\boldsymbol{v}$. On the right, the cutoff of $c=10$ takes effect around the same time and prevents the blowup. The grayed-out part of the trajectory illustrates the values of $c_{1}(t)$ that were replaced by $c=10$.


Figure 7.7: Time evolution of $c_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, \tilde{c}=24$, and our standard parameter regime.

Now that we have removed the blowup, we can do a comparison of the trajectories calculated using $c_{1}(t)$ and $c_{2}(t)$. Figure 7.8 shows that in this particular experiment, convergence occurs at $\tilde{c} \approx 24.5$ when either $c_{1}(t)$ or $c_{2}(t)$ is employed; note, however, that in the case of $c_{2}$, the lengths
of the whiskers do not visually decrease to 0 until $\tilde{c} \approx 25$. At the same time, when $\tilde{c}=23.5$, the inner quartile range has already collapsed onto 0 , which suggests that the approximation of $\boldsymbol{V}$ by $\boldsymbol{v}$ is good for much of the time with an occasional loss of synchronization represented by the long whiskers. Again, no outliers have been removed from the box plots. Note that the tops of most whiskers exceed 1.0; this indicates that at some point in time, $\boldsymbol{v}$ lost synchronicity with $\boldsymbol{V}$ and went wrong. This likely occurred because $\max \left(c_{i}(t), 10\right)=10$ numerous times in the calculation. This is important because in an application, losing synchronization means the prediction is poor.


Figure 7.8: Statistics of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (7.29), using $c_{1}(t)$ (left) and $c_{2}(t)$ (right) as defined in Equations (7.25) and (7.28), respectively, with a warmup of $T=50$, a run time of $T=72000$ and sampled every $\Delta t=20$, and our standard parameter regime. The relative error is less than the blue dots $95 \%$ of the time and less than the purple dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error. Note that the whiskers go away when $c \gtrsim 24.5$.

## Chapter 8

## Near-Continuous-in-Time

## Observations

Mathematically knowing $\boldsymbol{X}$ continuously in time determines $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$, but practical observations are only approximately continuous in time and may contain errors. Assuming observations $\boldsymbol{X}_{n}=\boldsymbol{X}\left(t_{n}\right)$ are made at times $t_{n}=t_{0}+\delta n$, we note that $\delta$ must be much smaller than the decorrelation timescale for the observations of $\boldsymbol{X}$ to be considered approximately continuous in time. In this case, we wish to determine whether $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$ may be approximated by finite differences. For example, our approximations to second order can be written as

$$
\begin{equation*}
\dot{\boldsymbol{X}}\left(t_{n}\right) \approx \mathcal{D}_{1} \boldsymbol{X}_{n}=\frac{\boldsymbol{X}_{n+1}-\boldsymbol{X}_{n-1}}{2 \delta} \tag{8.1}
\end{equation*}
$$

and

$$
\begin{equation*}
\ddot{\boldsymbol{X}}\left(t_{n}\right) \approx \mathcal{D}_{2} \boldsymbol{X}_{n}=\frac{\boldsymbol{X}_{n+1}-2 \boldsymbol{X}_{n}+\boldsymbol{X}_{n-1}}{\delta^{2}} \tag{8.2}
\end{equation*}
$$

In the numerics, we take $\delta=h$ and note that changing $h$ and consequently $\delta$ will affect the results. Of course, replacing $h$ by $h / 2$ and setting $\delta=2 h$ would also yield similar results.

From these, we obtain

$$
\begin{equation*}
\sigma^{\top} \boldsymbol{V}\left(t_{n}\right) \approx \mathcal{S}_{1}=-m_{1}\left(\mathcal{D}_{1} \boldsymbol{X}_{n}+B\left(\boldsymbol{X}_{n}, \boldsymbol{X}_{n}\right)+\boldsymbol{X}_{n}-\boldsymbol{F}\right) \tag{8.3}
\end{equation*}
$$

and

$$
\begin{equation*}
\sigma^{\top} \dot{\boldsymbol{V}}\left(t_{n}\right) \approx \mathcal{S}_{2}=-m_{1}\left(\mathcal{D}_{2} \boldsymbol{X}_{n}+B\left(\boldsymbol{X}_{n}, \mathcal{D}_{1} \boldsymbol{X}_{n}\right)+B\left(\mathcal{D}_{1} \boldsymbol{X}_{n}, \boldsymbol{X}_{n}\right)+\mathcal{D}_{1} \boldsymbol{X}_{n}\right) \tag{8.4}
\end{equation*}
$$

and we therefore obtain a piecewise constant approximation for $c_{1}(t)$ given by

$$
\begin{equation*}
c_{1}(t) \approx \xi_{1}\left(t_{n}\right)=\frac{\left[\frac{1}{m_{2}} \boldsymbol{X}_{n}-\sigma^{\top} \widetilde{B}\left(\boldsymbol{v}\left(t_{n}\right), \boldsymbol{v}\left(t_{n}\right)\right)-\mathcal{S}_{2}\right] \cdot \mathcal{S}_{1}}{\left\|\mathcal{S}_{1}\right\|^{2}} \quad \text { for } \quad t \in\left[t_{n}, t_{n+1}\right) \tag{8.5}
\end{equation*}
$$

and for $c_{2}(t)$,

$$
\begin{equation*}
c_{2}(t) \approx \xi_{2}\left(t_{n}\right)=\frac{\left[\frac{1}{m_{2}} \boldsymbol{X}_{n}-\sigma^{\top} \widetilde{B}\left(\sigma \mathcal{S}_{1}+Q \boldsymbol{v}\left(t_{n}\right), \sigma \mathcal{S}_{1}+Q \boldsymbol{v}\left(t_{n}\right)\right)-\mathcal{S}_{2}\right] \cdot \mathcal{S}_{1}}{\left\|\mathcal{S}_{1}\right\|^{2}} \tag{8.6}
\end{equation*}
$$

We note that the approximations for $\dot{\boldsymbol{X}}$ and $\ddot{\boldsymbol{X}}$ involve observations of $\boldsymbol{X}$ in the future. This is not a problem, as we can evolve the approximating solution lagged by one unit of time compared to the free-running solution. In particular, we evolve

$$
\begin{equation*}
\frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+\max \left(\xi_{i}\left(t_{n}\right), 10\right) \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X}_{n} \quad \text { for } \quad t \in\left[t_{n}, t_{n+1}\right) \tag{8.7}
\end{equation*}
$$

noting that $\xi_{1}\left(t_{n}\right)$ involves observations of $\boldsymbol{X}\left(t_{n+1}\right)$.
We note that obtaining an accurate approximation for $\dot{\boldsymbol{X}}$ from observational measurements is a fairly difficult task when there is noise in the measurements. Even more difficult would be an accurate approximation of $\ddot{\boldsymbol{X}}$. We therefore consider a modification of Equation (7.25) to obtain an approximation for the unknown value $\tilde{c}$ that does not involve $\ddot{\boldsymbol{X}}$.

$$
\begin{equation*}
c_{3}(t)=\frac{\left[\frac{1}{m_{2}} \boldsymbol{X}-\sigma^{\top} \widetilde{B}(\boldsymbol{v}, \boldsymbol{v})-\sigma^{\top} \dot{\boldsymbol{v}}\right] \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}} \tag{8.8}
\end{equation*}
$$

Here, we have replaced $\sigma^{\top} \dot{\boldsymbol{V}}$ with $\sigma^{\top} \dot{\boldsymbol{v}}$, which does not involve $\ddot{\boldsymbol{X}}$ and we know exactly through our computation of the approximating solution $\boldsymbol{v}$.

Upon substituting, where $\dot{\boldsymbol{v}}$ is given by Equation (7.29), note that the $\boldsymbol{X}$ and $\widetilde{B}$ terms cancel, and we obtain

$$
\begin{equation*}
c_{3}(t)=\max \left(c_{3}(t), 10\right) \frac{\sigma^{\top} \boldsymbol{v} \cdot \sigma^{\top} \boldsymbol{V}}{\left\|\sigma^{\top} \boldsymbol{V}\right\|^{2}}=\max \left(c_{3}(t), 10\right) \frac{\left\|\sigma^{\top} \boldsymbol{v}\right\|}{\left\|\sigma^{\top} \boldsymbol{V}\right\|} \cos \theta \tag{8.9}
\end{equation*}
$$

where $\theta$ is the angle between $\sigma^{\top} \boldsymbol{V}$ and $\sigma^{\top} \boldsymbol{v}$.
Intuitively, if $\left\|\sigma^{\top} \boldsymbol{V}\right\|>\left\|\sigma^{\top} \boldsymbol{v}\right\|$, this has the tendency to make the approximation $c_{3}(t)$ larger, which subsequently increases the dissipation in the equation governing $\boldsymbol{v}$. This would then decrease $\|\boldsymbol{v}\|$. Similarly, $\left\|\sigma^{\top} \boldsymbol{V}\right\|<\left\|\sigma^{\top} \boldsymbol{v}\right\|$ decreases the dissipation and causes $\|\boldsymbol{v}\|$ to increase.

All this is further affected by the angle $\theta$ between the approximating solution and the reference solution.

Unfortunately, the definition given in (8.8) is circular. In order to compute $\dot{\boldsymbol{v}}(t)$, we need $c_{3}(t)$; but in order to compute $c_{3}(t)$, we need $\boldsymbol{\boldsymbol { v }}(t)$. Since $\tilde{c}$ is a constant, we suppose the approximations are slowly varying; therefore, $c_{3}(t-\delta) \approx c_{3}(t)$.

We now obtain the following piecewise constant approximation of $\tilde{c}$ similar to $\xi_{1}$ given by

$$
\begin{equation*}
c_{3}(t) \approx \xi_{3}\left(t_{n}\right)=\max \left(\xi_{3}\left(t_{n-1}\right), 10\right) \frac{\sigma^{\top} \boldsymbol{v}\left(t_{n}\right) \cdot \mathcal{S}_{1}}{\left\|\mathcal{S}_{1}\right\|^{2}} \quad \text { for } \quad t \in\left[t_{n}, t_{n+1}\right) \tag{8.10}
\end{equation*}
$$

and $\xi_{3}\left(t_{0}\right)=10$. Note that $\xi_{3}$ is now defined recursively and no longer circular.


Figure 8.1: Time evolution of $\xi_{1}(t)$ as governed by Equation (7.26) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=21$, and our standard parameter regime. Sometimes, the solution will likely blow up (left), and other times it will simply fail to converge (right). Note that the gray areas in the left figure represent times when $\xi_{3}<10$; these are the times when the solution is likely to blow up if $\xi_{3}$ were allowed to go below 10 .

Figure 8.1 illustrates the evolution of $\max \left(10, \xi_{1}(t)\right)$ when $\tilde{c}=21$. In the graph on the left, the trajectory repeatedly goes below 10 , meaning that there are multiple instances when the solution could blow up if not for the lower limit of 10 on our approximation of $\tilde{c}$. Note that Theorem 7.2 still applies with slight modifications in this case. Out of 100 random trials, the majority of
the trajectories repeatedly went below 10. Similar results occurred when 100 random trials were performed with both $\xi_{2}$ and $\xi_{3}$.

As in Chapter 7, we now compare the trajectories calculated using $\xi_{1}, \xi_{2}$, and $\xi_{3}$, and we look for trajectories that converge for $\tilde{c}=23$ and $\tilde{c}=24$. However, as shown in Figure 8.2, it is difficult to determine whether convergence occurs solely on the basis of trajectories of $\max \left(10, \xi_{1}\right)$ and the relative errors. In Chapter 7, the relative errors reach numerical zero, but here, we do not observe this. Instead, the relative error lingers around $10^{-5}$; this occurs in the cases of apparent convergence and obvious failure to converge.

The reader should note that these figures are representative of the results in 100 random trials; i.e., for $\xi_{i}$ with $i$ from 1 to 3 , and for large values of $\tilde{c}$, we did not observe convergence to numerical zero. This is due to the inherent errors in approximating the solution using finite differences at discrete points in time.


Figure 8.2: Time evolution of $\xi_{1}(t)$ (left) and $\xi_{2}(t)$ (right) as governed by Equation (8.7) with a warmup of $T=50$, a run time of $T=20, h=1 / 2048, \tilde{c}=24$, and our standard parameter regime.

To better understand the regimes in which numerical convergence is likely to occur when using $\xi_{1}$, we perform statistics on the errors for values of $\tilde{c}$ varying from 21 to 35 over very long timescales. As Figure 8.3 illustrates, synchronization is less likely to occur for $\tilde{c}<27$ and can
be expected, if not assumed, to occur for $\tilde{c}>27$. The reader should note that results similar to $\xi_{1}$ are obtained with $\xi_{2}$. However, $\xi_{3}$ demonstrates the improved ability to recover both the unknown parameter $\tilde{c}$ along with the small scales represented by $\boldsymbol{V}$. In particular, the error levels when $\xi_{3}$ is used and $\tilde{c} \approx 24$ are about the same as when $\xi_{1}$ is used and $\tilde{c} \approx 25$.

Plausibly, the improvement with $\xi_{3}$ is due to not using the finite difference approximation $\mathcal{D}_{2}$ for $\ddot{\boldsymbol{X}}$ and subsequently $\mathcal{S}_{2}$. Even though $\sigma^{\top} \dot{\boldsymbol{v}}$ may be a poor approximation of $\sigma^{\top} \dot{\boldsymbol{V}}$, substituting the exact value of $\dot{\boldsymbol{v}}$ results in cancellations in the calculation of $\xi_{3}(t)$ and avoids the additional errors in $\mathcal{S}_{2}$. Although the improvement between $\tilde{c}=24$ and $\tilde{c}=25$ may seem modest, the advantages of $\xi_{3}$ are likely to become more pronounced when practical observations include errors that make the approximation of the second derivative even worse. Adding noise to the discrete time to demonstrate even more clearly the superiority of $\xi_{3}$ is a topic that we save for future investigation.


Figure 8.3: Statistics of the relative error in the approximation $\boldsymbol{v}$ of $\boldsymbol{V}$ in Equation (8.7), using $\xi_{1}(t)$ as defined in Equation (8.5) (left) and using $\xi_{3}(t)$ as defined in Equation (8.10) (right) with a warmup of $T=50$, a run time of $T=72000$ and sampled every $\Delta t=20$, and our standard parameter regime. The relative error is less than the magenta dots $95 \%$ of the time and less than the red dots $90 \%$ of the time. The tops of the whiskers represent the maximum relative error.

## Chapter 9

## Conclusion

In this research, we have considered two inverse problems: estimating $\boldsymbol{V}$ from observations of $\boldsymbol{X}$ when $c, F, m_{1}$, and $m_{2}$ are known; and simultaneously estimating $\boldsymbol{V}$ and an unknown $c$ from observations of $\boldsymbol{X}$. Throughout, we take $F=20, m_{1}=20$, and $m_{2}=0.002$, and we vary $c$. These questions were made possible by first rescaling the equations to isolate the weights in the coupling from the dissipation.

Part of the study involved choosing suitable values for $c, F, m_{1}$, and $m_{2}$ such that the resulting dynamics of the two-layer Lorenz '96 system are interesting and reflect the physics of synoptic and mesoscale motion in realistic atmospheric problems. One of the features of such problems is the tendency of the mesoscales to affect the motion of the large scales. These tendencies were parameterized by Wilks using a fourth-degree polynomial in order to determine the model error when substituting the one-layer Lorenz model in place of the two-layer; instead of model error, our research focuses on solving inverse problems.

For our solution techniques and results to be interesting, they need to be effective when the underlying dynamics governing $\boldsymbol{X}$ and $\boldsymbol{V}$ are complicated. This is the case when the effects of $\boldsymbol{V}$ on $\boldsymbol{X}$ are significant and unpredictable as seen by the tendencies and the error in the polynomial fit of the tendencies. Therefore, rather than fixing the degree of the polynomial parameterization, we instead use the BIC to select the degree. When the selected degree is at least 4, we infer that the mesoscales are playing an important role in the dynamics; in particular, values of $c \lesssim 30$ lead to an interesting inverse problem.

In passing, we note that the BIC depends on independent observations. To this end, we considered the autocorrelation of the large slow scales represented by $\boldsymbol{X}$. The autocorrelation was less than 0.01818 for $t_{\ell} \geq 20$, which is essentially the same as if the observations were truly independent.

We discovered that when $c$ is known, the inverse problem of recovering $\boldsymbol{V}$ from observations of $\boldsymbol{X}$ is solvable for $c \gtrsim 19.5$ using a synchronization approach that involves driving an approximating solution $\boldsymbol{v}$ with the observations. The inverse problem when both $c$ and $\boldsymbol{V}$ are unknown is solvable for $c \gtrsim 24$. In particular, we obtain an approximation for $c$ from derivative information of the observations, and subsequently we can recover these two quantities simultaneously when $c \gtrsim 24$.

We note that there are other data assimilation techniques that could be used to help recover both $c$ and $\boldsymbol{V}$ : e.g., 4D-variational methods and particle filters based on Bayesian statistics. Even so, some of the basic building blocks by which these methods work rely on the synchronization effects demonstrated in the present research; thus, we hope that the simple techniques of solving the inverse problem considered in this dissertation form an important foundation upon which more complicated techniques may be based.

Our final results consider the case when $c$ and $\boldsymbol{V}$ are unknown and $\boldsymbol{X}$ is observed at discrete moments in time. We have shown that the same method of solving for $c$ when the observations are continuous in time works when the derivatives of $\boldsymbol{X}$ are replaced with finite differences. Further simplifications allowed us to work with only the finite difference approximation of the first derivative. In particular, we obtain partial synchronization of $\boldsymbol{v}$ with $\boldsymbol{V}$ and an estimate of $c$ to an accuracy within $1 \%$ that could be useful in real-world applications. To that end, there are many ways this research could be extended, which we discuss in the next section.

## Chapter 10

## Future Work

Throughout our work, we observed $\boldsymbol{X}$ and tried to determine $\boldsymbol{V}$, and we found a parameter regime in which synchronization occurred. But supposing $\boldsymbol{X}$ does not determine $\boldsymbol{V}$ via the coupling, could we still determine $c$ ? Perhaps one could do this by looking at the coupled
 efficiently find $c$.

Another extension involves the structure of $\boldsymbol{X}$. Suppose $\boldsymbol{X}$ has a coherent structure but $c$ is such that $\|\boldsymbol{V}-\boldsymbol{v}\| \nrightarrow 0$. Could we introduce something into the assimilation step (e.g., a Kalman filter or some sort of smoothing) to make $\|\boldsymbol{V}-\boldsymbol{v}\| \rightarrow 0$ ?

Yet another extension involves using a variational method to solve for $c$. The advantage might be that one could identify the exact value of $c$ when $c$ is not as large as in the sequential method explored in Chapter 7.

In Section 6.2, it was shown that choosing $\sigma$ to be a random matrix with orthonormal columns made the inverse problem significantly easier to solve. This leads one to ask how much the original coupling matrix would have to be modified to obtain the same results as the random matrix. Perhaps only a partially randomized matrix with a few oscillators coupled in different ways would be enough to obtain an inverse problem that is easier to solve.

Convection in the two-layer Lorenz '96 system is represented by two terms $B$ and $\widetilde{B}$. These terms are notable for being mirror images of each other. Physically, this mirror image makes sense in the sense of two interlocking gears spinning in opposite directions. One interesting question for further research is what happens if the same convection term is used for both layers. Do any of
the results reported in this dissertation depend on the mirror-image relationship between $B$ and $\widetilde{B}$ ?

Another interesting line of further research is one where the unobserved variables represented by $\boldsymbol{V}$ are modeled using approximate dynamics with reduced dimensionality. This is a generalization of solving for $c$, in which the actual dynamics are unknown and other parameters have to be tuned. Intuitively, if we want to view the fine scale as an arbitrary dimensional refinement of the scales that exist interpolated between the large scale, we may ask that the weights scale proportionately to the resolution of the fine scale. For example, the least-squares approach could be generalized to find parameters in the model of the hidden dynamics which may or may not be physical.

In Section 6.4, we used Tikhonov regularization to solve the sequence of least squares problems that occur in the Gauss-Newton algorithm when solving the algebraic constraints. It may be possible to enhance the efficiency of this technique and extend it to PDE problems by using the singular value decomposition instead of Tikhonov regularization.

In the case where $\boldsymbol{X}$ is observed continuously (and derivative information about $\boldsymbol{X}$ and $\sigma^{\top}$ is therefore known), there are two possibilities for future work. One could reduce the number of small oscillators per large oscillator: i.e., choose a smaller value of $J$ so that each algebraic constraint resolves a larger proportion of the unresolved scales per degree of the polynomial. Alternatively, one could perform the same calculations for our usual value of $J$, but using extended-precision arithmetic; this way, the rounding errors that appear when evaluating high-degree polynomial constraints and their derivatives are lessened.

Note also in the context of a partial differential equation that governs motion with a wide range of length scales, the algebraic constraints which appear through the nonlinear term naturally involve the most significant of the unobserved scales first. Again, this may lead to better performance of the constraints in real applications, and numerical experiments will be needed to test this hypothesis.

## Appendix A

## Numerical Methods

When numerically solving a coupled system of equations, one must ask if an adaptive method is feasible. Even if a method is easy to use and of a higher order than a classical method, the use of different time steps at different points in the computation can lead to rounding errors which cause mathematically identical computations to diverge. To illustrate the problem, we consider the system

$$
\begin{align*}
& \frac{d \boldsymbol{X}}{d t}+B(\boldsymbol{X}, \boldsymbol{X})+\boldsymbol{X}=\boldsymbol{F}-\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}  \tag{A.1}\\
& \frac{d \boldsymbol{V}}{d t}+\widetilde{B}(\boldsymbol{V}, \boldsymbol{V})+c \boldsymbol{V}=\frac{1}{m_{2}} \sigma \boldsymbol{X}  \tag{A.2}\\
& \frac{d \boldsymbol{v}}{d t}+\widetilde{B}(\boldsymbol{v}, \boldsymbol{v})+\tilde{c} \boldsymbol{v}=\frac{1}{m_{2}} \sigma \boldsymbol{X} \tag{A.3}
\end{align*}
$$

where $m_{1}=20, m_{2}=0.002, F=20$, and $c=20$.
Note $\boldsymbol{X}$ and $\boldsymbol{V}$ are coupled in both directions and $\boldsymbol{v}$ is coupled with $\boldsymbol{X}$ but not vice versa. We perform two simulations with the same initial conditions and parameter regimes except $\tilde{c}=9$ in the first and $\tilde{c}=10$ in the second. We expect these computations to be mathematically identical due to the one-way coupling of $\boldsymbol{X}$ and $\boldsymbol{v}$, but when the computations are performed with the TsitPap8 method (see [21]), the trajectories diverge. Similar results are obtained using the TsitPap5 method. However, if we perform the same two simulations using the classic RK4 method with a fixed step size, the calculations of $\boldsymbol{X}$ and $\boldsymbol{V}$ are identical (see Figure A.1).

The difference in these two sets of computations is caused by an inherent difficulty in using an adaptive method to compute Equations (A.1)-(A.3). Different changes in the step size based


Figure A.1: Evolution in time of $\|\boldsymbol{X}\|$ for $\tilde{c}=9.0$ (blue) and $\tilde{c}=10.0$ (orange) in Equation (A.3). Equations (A.1)-(A.3) are solved using (left) the adaptive TsitPap8 method included in the DifferentialEquations.jl package in Julia, and (right) the classical RK4 method with a fixed step size. The trajectories in the latter are offset by a height of 1 for readability.
on the computation of $\boldsymbol{v}$ create different rounding errors, which accumulate over time and cause the reference solution $\boldsymbol{X}$ to differ. The trajectories illustrated in Figure A. 1 were obtained using absolute and relative tolerances of $10^{-8}$. Similar results occurred with other tolerances.

Since it is not physical if the prediction affects the physical process that one has been observing, from a mathematical point of view the simulation run on the weather forecasting computer should not affect the real weather. Using an RK4 method, shown for the fixed step size of $h=0.0001$ in Figure A.1, avoids this problem, as the two simulations have the same discrete dynamics for $\boldsymbol{X}$ and $\boldsymbol{V}$ at each step. For this reason, the rest of the computations which appear in this paper were performed using RK4 with a fixed step size.

## Appendix B

## Code Samples

In this appendix, we provide a few excerpts of programs written in Julia that were used for the simulations which appear in this dissertation.

## B. 1 Definitions

This program contains the definitions of commonly used constants and functions. These were employed throughout much of our research and have been combined into a single file. Note that sigma on line 54 was converted into a sparse matrix; this saved significant computational time.

```
1 using Random, Plots, DelimitedFiles, SparseArrays, LinearAlgebra
2 using Statistics, Polynomials
kplus1 = [[2:K;]; 1]
{ } _ { 5 } \text { kminus1 = [K; [1:K-1;]]}
6 kminus2 = [K-1; K; [1:K-2;]]
jplus1 = [[2:JK;]; 1]
jplus2 = [[3:JK;]; 1; 2]
10 jminus1 = [JK; [1:JK-1;];]
11
12 function B(X, Y)
13 return X[kminus1].*(Y[kminus2] - Y[kplus1])
```

```
end
function Bt(X, Y)
    return X[jplus1].*(Y[jplus2] - Y[jminus1])
end
function dXdt(X,V)
    return B(X,X) - X .+ F - 1.0/m1*sigma'V
    end
function dVdt(X,V,c)
        return Bt(V,V) - c*V + 1.0/m2*sigma*X
    end
    function rk4twolayer(X, V, h, ct)
        k1x = dXdt(X, V)
        k1V = dVdt(X, V, ct)
        k2x = dXdt(X + h/2*k1x, V + h/2*k1V)
        k2V = dVdt(X + h/2*k1x, V + h/2*k1V, ct)
        k3x = dXdt(X + h/2*k2x, V + h/2*k2V)
        k3V = dVdt(X + h/2*k2x, V + h/2*k2V, ct)
        k4x = dXdt(X + h*k3x, V + h*k3V)
        k4V = dVdt(X + h*k3x, V + h*k3V, ct)
        X .= X + h/6*(k1x + 2*k2x + 2*k3x + k4x)
        V .= V + h/6*(k1V + 2*k2V + 2*k3V + k4V)
end
    const J = 32
    const K = 8
    const JK = J * K
global sigma = zeros(JK, K)
for k = 1:K
        for j = J*(k-1)+1:J*k
        sigma[j, k] = 1.0/sqrt(J)
```

```
5 2 ~ e n d
53 end
54 sigma = sparse(sigma)
55
56 const m1 = 20.0
57 const m2 = 0.002
58 const F = 20.0
```


## B. 2 Random Initial Conditions

Many of our experiments considered an ensemble of long-time randomly chosen reference solutions. These were created by choosing a random initial condition given by the function below, which was then warmed up for $T=50$ units of time. Note that $\boldsymbol{X}$ and $\boldsymbol{V}$ are uniformly distributed on the interval $[-1 / 2,1 / 2]$; in practice, however, this is a minor point, as the warmup time coupled with the nonlinear chaotic dynamics causes the initial distribution to be forgotten.

```
1 function makeIC(seed)
2 Random.seed!(seed)
3 X = rand(Float64, K) .- 0.5
4 V = rand(Float64, JK) .- 0.5
    return X, V
end
```


## B. 3 Data Sampling

This program calculates the tendencies as discussed in Section 4. Note that line 20 writes out samples of $\frac{1}{m_{1}} \sigma^{\top} \boldsymbol{V}$ every 40960 time steps, which translates to $\Delta t=20$.

```
function drdt(V)
    return 1.0/m1*sigma'V
end
function getRegData(c)
    T = 2048*2*360000
    h = 1/2048
8 XO, VO = makeIC(1)
```

```
    dr = zeros(K)
    Xdr = zeros(2, length(XO))
    X = copy(X0)
    V = copy(V0)
    dr = copy(dr0)
    for n = 1:11*T
        rk4twolayer(X, V, h, c)
        dr = copy(drdt(V))
        if(n == T)
                Xdr = [X[1] dr[1]]
            end
            if(n % 40960 == 0 && n > T)
            Xdr = [Xdr; [X[1] dr[1]]]
        end
    end
    writedlm("output_f_20_c_"*string(c)*".txt", Xdr)
end
gc = parse(Int64, ARGS[1])
println("Running with c = ", gc)
getRegData(gc)
```


## B. 4 Finding $c$

This program calculates $c_{1}(t)$ from Section 7.2 and plots the relative errors in $\|\boldsymbol{V}-\boldsymbol{v}\| /\|\boldsymbol{V}\|$ and $\left|\tilde{c}-c_{1}\right| / \tilde{c}$. The definitions of commonly used constants and functions are given in B.1.

```
function dvdt(X, V, v)
    dV = dVdt(X, V, ct)
    c = ((X/m2 - sigma'*(Bt(v, v) + dV))'*(sigma'V))[1]/norm(sigma'V)^2
    return dVdt(X, v, c)
end
function rk4unknownC(X, V, v, h)
    k1x = dXdt(X, V)
    k1V = dVdt(X, V, ct)
    k1v = dvdt(X, V, v)
```

```
11
k2x = dXdt (X + h/2*k1x, V + h/2*k1V)
    k2V = dVdt(X + h/2*k1x, V + h/2*k1V, ct)
    k2v = dvdt(X + h/2*k1x, V + h/2*k1V, v + h/2*k1v)
    k3x = dXdt(X + h/2*k2x, V + h/2*k2V)
    k3V = dVdt(X + h/2*k2x, V + h/2*k2V, ct)
    k3v = dvdt( X + h/2*k2x, V + h/2*k2V, v + h/2*k2v)
    k4x = dXdt(X + h*k3x, V + h*k3V)
    k4V = dVdt(X + h*k3x, V +h*k3V, ct)
    k4v = dvdt(X + h*k3x, V + h*k3V, v + h*k3v)
    X . = X + h/6*(k1x + 2*k2x + 2*k3x + k4x)
    V .= V + h/6*(k1V + 2*k2V + 2*k3V + k4V)
    v .= v + h/6*(k1v + 2*k2v + 2*k3v + k4v)
end
function findC(X, V, h, steps, imod)
    cvals = zeros(steps\divimod+1)
    vNorms = zeros(steps\divimod+1)
    for _ in 1:2048*50
        rk4twolayer(X, V, h, ct)
    end
    v = P*V
    for i in 1:steps+1
        c = estimateC(X, V, v)
        if i % imod == 0
                    cvals[i\divimod] = abs(c - ct)/ct
                vNorms[i\divimod] = norm(V - v)/norm(V)
            end
            rk4unknownC(X, V, v, h)
        end
        return cvals, vNorms
end
function makePlots(ct)
    data = readdlm("data-ct$ct.txt")
```

```
    plot(data[:,1], data[:,2], xlabel = "Time", ylabel = "Relative Error",
            xlims = (0, 100), ylims = (1e-20, 1), yscale=:log10, label = "||v - v|| / ||V|
    display(plot!(data[:,1], data[:,3], xlabel = "Time", ylabel = "Relative Error",
            title = "Relative Errors when ct = $ct", legend=:bottomleft,
            xlims = (0, 100), ylims = (1e-20, 1), yscale=:log10, label = "|ct - c1| / ct")
    savefig("plots-$ct.pdf")
end
function runfindC()
    X, V = makeIC()
    steps = 2048*100
    T = 1:steps
    h = 1/2048
    imod = 1
    cvals, vNorms = findC(X, V, h, steps, imod)
    writedlm("data-ct$ct.txt", [(0:imod:steps)*h cvals vNorms])
end
cn = 21.5:0.5:26
for c in cn
    global ct = c
    println("c = ", ct)
    runfindC()
    makePlots(ct)
end
```


## Bibliography

[1] Anderson, J. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D: Nonlinear Phenomena, pages 99-111.
[2] Anderson, J. and Anderson, S. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Montly Weather Review, pages 2741-2758.
[3] Arnold, H., Moroz, I., and Palmer, T. (2013). Stochastic parametrizations and model uncertainty in the Lorenz '96 system. Philosophical Transactions of the Royal Society A, pages $1-22$.
[4] Berner, J. et al. (2017). Stochastic parameterization: Toward a new view of weather and climate models. Bulletin of the American Meteorological Society, pages 565-588.
[5] Burnham, K. and Anderson, D. (2002). Model Selection and Multimodel Interference: A Practical Information-Theoretic Approach. Springer, 2 edition.
[6] Carlson, E., Hudson, J., Larios, A., Martinez, V., Ng, E., and WhiteHead, J. (2021). Dynamically learning the parameters of a chaotic system using partial observations. AIMS, 42(8):3809-3839.
[7] Constantin, P. and Foias, C. (1988). Navier-Stokes equations. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL.
[8] deLeeuw, J. (1992). Breakthroughs in Statistics Volume I: Foundations and Basic Theory, pages 599-609. Springer.
[9] Galias, Z. and Zgliczyński, P. (1998). Computer assisted proof of chaos in the Lorenz equations. Physica D: Nonlinear Phenomena, 115:165-188.
[10] Hamill, T. and Whitaker, J. (2005). Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. Monthly Weather Review, pages 3132-3147.
[11] Kaidor (2013). Earth global circulation. Wikimedia Commons, CC BY-SA 3.0.
[12] Kalnay, E. (2003). Atmospheric modeling, data assimilation and predictability. Cambridge University Press.
[13] Law, K., Sanz-Alonso, D., Shukla, A., and Stuart, A. (2015). Filter accuracy for the Lorenz 96 model: Fixed versus adaptive observation operators. Physica D, pages 1-13.
[14] Lorenz, E. (1963). Deterministic nonperiodic flow. Journal of the Atmospheric Sciences, 20:130-141.
[15] Lorenz, E. (1995). Predictability: a problem partly solved. In Seminar on Predictability, 4-8 September 1995, volume 1, pages 1-18, Shinfield Park, Reading. ECMWF, ECMWF.
[16] Lorenz, E. (1996). Predictability: a problem partly solved. Predictability of Weather and Climate, pages 40-58.
[17] Lu, F., Tu, X., and Chorin, A. (2017). Accounting for model error from unresolved scales in ensemble Kalman filters by stochastic parameterization. Monthly Weather Review, pages 3709-3723.
[18] Majda, A. and Harlim, J. (2012). Filtering Complex Turbulent Systems. Cambridge University Press.
[19] Olson, E. and Titi, E. (2008). Determining modes and Grashof number in 2D turbulence-a numerical case study. Theoretical and Computational Fluid Dynamics, 22(5):327-339.
[20] Palmer, T. et al. (2005). Representing model uncertainty in weather and climate prediction. Annual Review of Earth and Planetary Sciences, pages 163-193.
[21] Rackauckas, C. and Nie, Q. (2017). DifferentialEquations.jl-a performant and feature-rich ecosystem for solving differential equations in Julia. Journal of Open Research Software, 5(1).
[22] Schaback, R. (1985). Convergence analysis of the general Gauss-Newton algorithm. Numerische Mathematik, 46:281-309.
[23] Solomon, J. (2015). Numerical algorithms: methods for computer vision, machine learning, and graphics. CRC Press, 1 edition.
[24] Speyer, J. and Chung, W. (2008). Stochastic Processes, Estimation, and Control. SIAM.
[25] Tucker, W. (2002). A rigorous ODE solver and Smale's 14th problem. Foundations of Computational Mathematics, 2:53-117.
[26] Wilks, D. (2005). Effects of stochastic parametrizations in the Lorenz '96 system. Quarterly Journal of the Royal Meteorological Society, pages 389-407.

