UNIVERSITY OF CALIFORNIA, SAN DIEGO

Synchronization and Statistical Methods for the Data Assimilation of HVc Neuron Models

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Physics

by

Mark Kostuk

Committee in charge:

Professor Henry D. I. Abarbanel, Chair
Professor Timothy Gentner
Professor Herbert Levine
Professor Gabriel Silva
Professor David Tytler

2012

The dissertation of Mark Kostuk is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2012

para mon buen pleito.

For the seeker of knowlege, the abyss of the unknown
expands exponentially faster than the island of understanding.

–MK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Of course, *none* of this would have been possible without the loving support of my friends, family and co-workers –thank you for advising, for being patient, and for encouraging me to continue this endeavor. The many discussions of relevant physical problems and inspirational paradoxes that we have shared have helped me to become the scientist that I am today.

A special thanks to the Margoliash Lab, and to C. D. Meliza in particular, for their hospitality, insight and for the wealth of quality experimental data that they have provided.

CURRICULUM VITAE

EDUCATION

2012	Doctor of Philosophy in Physics	University of California, San Diego
2001	Bachelor of Science in Physics and Chemistry	Union College, Schenectady, NY

PUBLICATIONS

Kostuk, M., Toth, B. A., Meliza, C. D., Margoliash, D., and Abarbanel, Henry D. I. **Dynamical estimation of neuron and network properties II: path integral Monte Carlo methods** *Biological Cybernetics* Vol 106, Issue 3, p155-167. (2012)

Toth, B. A., Kostuk, M., Meliza, C. D., Margoliash, D., and Abarbanel, Henry D. I. **Dynamical estimation of neuron and network properties I: variational methods** *Biological Cybernetics* Vol 105, p217-237. (2011)

Abarbanel, H. D. I., Bryant, P., Gill, P. E., Kostuk, M., Rofe, J., Singer, Z., Toth, B., and Wong, E. **Dynamical Parameter and State Estimation in Neuron Models.** *The Dynamic Brain: An Exploration of Neuronal Variability and its Functional Significance* Glanzman, D. and Mingzhou Ding, eds.; Oxford University Press. (2011)

Abarbanel, H. D. I., Kostuk, M. and Whartenby, W. **Data Assimilation with Regularized Nonlinear Instabilities** *Quarterly Journal of the Royal Meteorological Society*, Vol 136 (648) p769-783. (2010)

Abarbanel, H. D. I., Creveling, D. R., Farsian, R., and Kostuk, M. **Dynamical State and Parameter Estimation** *SIAM Journal of Applied Dynamical Systems*, Vol 8, Issue 4, p1341-1381. (2009)

Abarbanel, H. D. I., Kostuk, M., Margoliash, D., and Meliza, C. D. **Data-Assimilated Neuron Models of a Population of Neurons in the High Vocal Center of the Zebra Finch** *in progress (2012)*

FELLOWSHIPS AND AWARDS

2008-2011	San Diego Fellowship
2006-2007	GAANN Fellowship
1997-2001	William H. Wright Chemistry Scholarship
2011	American Institute of Chemists Award
1999	Joseph Henry Turnbull Physics Prize

ABSTRACT OF THE DISSERTATION

Synchronization and Statistical Methods for the Data Assimilation of HVc Neuron Models

by

Mark Kostuk

Doctor of Philosophy in Physics

University of California, San Diego, 2012

Professor Henry D. I. Abarbanel, Chair

Within the context of data assimilation, we describe the use of chaotic synchronization to overcome instabilities in the search space of the associated optimization problem and use numerical examples to demonstrate how the elimination of positive (conditional) Lyapunov exponents allows one to determine both the number and specific state variable choice of dynamical dimensions that must be measured in order to ensure feasibility of synchronization-based optimization techniques.

We present a novel objective function based upon the chaotic synchronization-error metric that utilizes a strongly-coupled fiducial trajectory as a full dimensional surrogate of the measured data. This synchronization fiducial serves as the metric origin in the space of model dynamical trajectories, thus making the search over model parameters more informative and eliminating the need for collocation of the dynamical variables.

The relationship between this (and other) proposed objective functions and the configuration functions of classical statistical physics are explored, including an additive-noise approximation to errors in the model expressed as a path integral over the joint probability distribution of the dynamics.

From these considerations we create two statistical (derivative-free) numerical optimization algorithms on parallel processors that employ Monte Carlo techniques to evaluate the dis-

tribution of unknown state variables and model parameters; this is done on chaotic and electrophysiological twin-experiments, where we demonstrate how these methods are used to assist in the design of neuron models and stimulus (current injection) protocols. Finally, we report data-assimilation results, including model error estimates, of a model optimized to a current-clamp recording of a neuron from the High Vocal center (HVc) of the zebra finch birdsong neural pathway.

Chapter 1

Data Assimilation and Optimization

Data assimilation is a quintessential problem of inductive science. Scientists express physical phenomena in terms of mathematical models as a means to understanding relevant constituents and to provide future predictions. The ability to create highly specific instances of these models requires the incorporation of measured data from the system in question.

If the the physics of the experimental system is well known then it may be possible to determine the functional form of a model to any desired accuracy. However, even in a situation of such unlikely good fortune as this, in order to make predictions with this model one requires full knowledge of the system state, possibly beyond those variables that can be directly measured. Or perhaps the model contains many unknown parameters due to a large number of interacting particles or subsystems. The obtainment of these unknown variables and any unknown parameters from measured experimental data is the problem of *data assimilation*.

It is a natural consequence of reductionalism that demands are placed upon these models to explain an ever increasing range of experimental results; indeed, first principles can become cumbersome as the complexity of the system increases. As this occurs, the models inevitably contain features that straddle the boundary between established physical principles and effects that are specific to individual experimental instances.

Managing parameter and state optimization in situations where such complexity cannot be avoided is what we attempt to illuminate within this manuscript; specifically by utilizing concepts familiar to nonlinear dynamics such as synchronization and time-delay embedding. Some of these complexity issues include systems that contain chaotic parameter regimes, the

presence of measurement noise, as well as models whose functional form remains an incomplete or incorrect description of nature.

Many approaches to data assimilation, including filters (Judd and Stemler (2009)) and estimators (Cessna et al. (2007)) have been developed for the high-dimensional problem of weather forcasting, in which choatic dynamics are ubiquitous. Among other fields such as geophysical fluid dynamics, state and parameter estimation has been applied to biological systems (Bhalla and Bower (1993); Vanier and Bower (1999); Tabak et al. (2000)). Our attention is focused on neuron systems, where models of single neuron and network properties quickly extend into large-variable descriptions requiring careful treatment (Huys et al. (2006); Beuter et al. (2003)).

1.1 Dynamical Systems

Our investigation is limited to classical (i.e. non-quantum) mechanics, where a dynamical system is represented by a set of equations that describe how the state of the system varies in time. The variables of these equations are *state variables*, that vary in time according to the dynamics; together with any constant parameters they fully specify the system at a given point in time.

With time as a parameterization, the vector of state variables traces out a trajectory through *state space*, which is the space of all possible states of the system. All of the systems considered are assumed to be *bounded*, meaning that physical restrictions on the state variables prevent them from going to infinity as time progresses, even in the limit as time goes to infinity.

The nature of time during the course of our system's evolution is assumed to obey Galilean relativity; that is, there is no absolute reference point in time so that shifting the origin of time does not change the system. This is equivalent to asserting that the system is *autonomous*, meaning that time cannot appear explicitly in the dynamics; only time intervals may have a physical realization.

Futhermore, this discussion is primarily intended for *deterministic* systems; it is assumed that there is no random component to the dynamics. Some of the optimization methods examined herein include stochastic algorithms, and there is considerable discussion in chapter 5 about including a random component to the dynamics as a means to investigate the effects of model error, however these methods are developed with underlying nonlinear deterministic systems in mind,

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{p}) \quad . \tag{1.1}$$

With the discrete version of the dynamics operation over a timestep dt given by the map

$$\mathbf{x}(t+dt) = \mathbf{F} \circ (\mathbf{x}(t), \mathbf{p}) \quad . \tag{1.2}$$

Barring a chaotic caveat, the deterministic nature of the dynamics means that it is permissible to discuss the unique trajectories of $\mathbf{f}$ by their initial conditions and parameters. Using the shorthand notation

$$\begin{aligned} \mathbf{x}_T &= \mathbf{F}^{N_T} \cdot (\mathbf{x}(0), \mathbf{p}) \\ &= \{\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(N_T)\} \quad , \end{aligned} \tag{1.3}$$

where $\mathbf{F}^{N_T}$ represents N_T repeated applications of $\mathbf{F}$, the dynamical map corresponding to the model differential equations $\mathbf{f}$.

1.2 Variational Optimization

It is possible to formulate the data assimilation problem as a special realization of *optimal control* theory, and more generally as a problem for *nonlinear optimization*, where one searches for a solution set that minimizes a cost or *objective* function. Optimal control theory has a rich history (see for ex. Kirk (1970); Bryson and Yu-Chi (1969); Gill et al. (1981); Powell (1982); Barclay et al. (1998)), and from it originates much of the terminology that is used througout.

Algorithmic search routines support a natural division along which methods are commonly separated into two steps: the first is the selection of a potential point for evaluation, and the second is a comparison of two points to determine which is the 'better' one. The methods to accomplish the former task are quite varied and heuristic-based, distinguishing one's overall approach to optimization, while the second challenge is readily met by means of an objective function.

The objective function is a measure which defines what it means for one point to be a more desirable solution than another. The discussion of optimization is limited to minimization problems, where -all else being equal- a point with a lower objective function is considered closer to the solution than one with a large objective function. Clearly, this is both an essential piece of optimization and a very subjective declaration; the best that an optimization algorithm can do is solve the problem that is posed to it, whether or not it is the problem that one intends to solve is relegated to the skill of the inquisitor.

Presented next are some fundamentals of the variational calculus approach to optimization. The discussion remains general, although the emphasis is on how these methods are useful for solving data assimilation problems.

1.2.1 Unconstrained Optimization

Our search variable, ω, is a W-dimensional vector comprising all of the unknown state variables and parameters of our data assimilation problem. For example, one specfic search space that we are interested in includes the model initial conditions and parameters (see chapter 7), in which case $\omega = \{\mathbf{x}(0), \mathbf{p}\}$ and $W = N_x + N_p$, where N_x is the dimension of the dynamical model and N_p is the number of unkown parameters in the model. More commonly in the optimization of dynamical systems, the state variables are *collocated*, where the search is performed over the state variables at each time point $\mathbf{x}(t)$ individually, resulting in $\omega = \{\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(N_T - 1), \mathbf{p}\}$ and $W = N_x \cdot N_T + N_p$. For this section however, the discussion remains general and any specific relationship between the physical model dynamics and the optimization should be considered unspecified. One restriction that may be assumed because of the association with physical dynamical systems is that our *search space*, let us call it Ω, is a continuous[1] and bounded subspace of $\mathbb{R}^W$.

The objective function $\phi(\omega) : \omega \to \mathbb{R}_+$ is used to evaluate the relative fitness of various points $\omega \in \Omega$ in the search space. The simplest formulation of optimization that is of concern is

$$\text{minimize} \quad \phi(\omega) \qquad \omega \in \Omega, \quad \Omega \subseteq \mathbb{R}^W, \tag{1.4}$$

which represents an *unconstrained* problem, as it involves solely the objective function.

The solution to this problem is formally written as ω^*, although keep in mind that this is at best a local solution, for the assertion

$$\omega^* = \omega' : \forall_{\omega' \in \Omega} \quad \phi(\omega') \leq \phi(\omega) \quad ,$$

is a statement of global optimization, and is as impractical as it is desirable.

The variational approach takes a different tack and instead looks to fulfill derivative requirements. The first and second *optimality conditions* are easily understood, for it is obvious even to the causal observer that for a point ω^* to be a local minimum of ϕ the following must be satisfied:

$$i.) \quad \left.\frac{d\phi}{d\omega}\right|_{\omega^*} = 0 \tag{1.5}$$

$$ii.) \quad \left.\frac{d^2\phi}{d\omega^2}\right|_{\omega^*} > 0 \quad . \tag{1.6}$$

These follow from elementary calculus considerations, and are ultimately responsible for the locality restriction on the solution ω^* due to the infinitessimal limit in the definition of the derivative.

[1] The numerical search is done on a finite-state search space, but we generally allow continuity requirements by choosing a suitably small dt_{int}.

1.2.2 Constrained Optimization

Of course, one would like to restrict the region of Ω that must be searched over as much as possible by imposing known relations amongst the search variables through the use of *constraint equations*. The constraints are usually divided into equality constraints $c_{\mathcal{E}}(\omega) = 0$, and inequality constraints $c_{\mathcal{I}}(\omega) > 0$, so that they may be dealt with more efficiently separately. Generally, optimization problems with inequality constraints present additional difficulties to those with equality constraints alone.

With the exception of simple bounds, which maybe linearly scaled away (see section 2.1.2), we are unlikely to encounter many inequality constraints in data assimilation, and for purposes of exposition restrict the discussion to optimization problems with equality constraints alone.

The constrained optimization problem with equality constraints is expressed as

$$\begin{aligned} \text{minimize} \quad & \phi(\omega) && \omega \in \Omega, \quad \Omega \subseteq \mathbb{R}^W \\ \text{subject to} \quad & c_j(\omega) = 0 && j \in 1, \ldots, \mathcal{E} \quad . \end{aligned} \tag{1.7}$$

For a point ω^* to be a solution to this problem, it must satisfy both the constraints and minimize the objective function.

If ω satisfies a constraint, $c_j(\omega) = 0$, then that constraint is *active*, and if ω satisfies all $\mathcal{E}$ of the constraints then ω is said to be a *feasible* point; the set of all feasible points being the feasible region of Ω. In high dimensions and especially with nonlinear constraints, the feasible region need not be simply connected; disjoint 'islands' of feasibility may exist throughout the search space.

Remaining in the feasible region between successive iterations of a search algorithm is not guaranteed if the constraint equations are nonlinear, because the curvature of these equations may depend upon high order derivatives that are not accurately reflected in the perturbation step. Algorithms vary on their approach to finding and staying within the feasible region, yet clearly there is a benefit to using constraints to limit the size of the search space, especially if there are linear constraints as any linearly independent dimensions of the active set of constraints reduces the dimesion of the problem.

1.2.3 Solution Uniqueness

It is commonplace to inquire whether a *linear* system of equations is well-determined. That is, if the constraint equations restrict the solution space to a single point ω^*, then a unique solution exists and the problem is well defined. Too few restrictions and there may exist multiple

degenerate solutions where $\phi(\omega_1) = \phi(\omega_2)$, yet $\omega_1 \neq \omega_2$; too many restrictions and there is a risk that no solution exists in the feasible region.

Independence of the unknown parameters can be algorithmically tested for a linear system, and an assessment of whether the above situation can be satisfied is possible. However, the linear-independence of *nonlinear* variables can only be assessed in a local, point-wise fashion; indeed, the lack of a globally equivalent criterion is arguably what makes nonlinear systems difficult to work with. In general, then, we note that it is not possible to guarantee the uniqueness of a solution to a nonlinear optimization problem.

While it is obvious that redundant parameters should not be present in a model, the possible interdependence of apparently independent parameters of a nonlinear system may involve complicated inverse (and often transcendental) expressions of the state variables making elimination counterproductive from the standpoint of numerical optimization and data assimilation. This represents a realization of the vague notion of complexity mentioned in the introduction.

1.2.4 Lagrange Multipliers

The method of Lagrange multipliers (see for example Arfken and Weber (2005)) is often used in variational optimization to satisfy both feasibility of the constraints (to first order) and optimality of the objective function. This approach recognizes that we seek a solution that is in the feasible region, but where any feasible move must be perpedicular to the gradient of the objective function. For if it were not orthogonal, a move could be made in this feasible direction that would reduce the objective function.

The Lagrange function[2] relates the constraints to the objective function via their gradient information, and is defined as

$$L(\omega, \lambda) = \phi(\omega) - \sum_{j=1}^{\mathcal{E}} \lambda_j c_j(\omega) \quad , \tag{1.8}$$

where λ is the $\mathcal{E}$ dimensional vector of Lagrange multipliers.

The first order optimality condition (1.5) is satisfied if $\frac{\partial \phi}{\partial \omega}$ can be written as a linear combination of the constraint derivatives, $\frac{\partial c}{\partial \omega}$, locally at the solution ω^* according to

$$\left.\frac{\partial \phi}{\partial \omega_i}\right|_{\omega^*} = \sum_{j=1}^{\mathcal{E}} \lambda_j^* \left.\frac{\partial c_j}{\partial \omega_i}\right|_{\omega^*} \quad , \tag{1.9}$$

[2]The term 'Lagrangian' is only used in regards to optimization (where it is applied to the objective function) although it is mathematically equivalent to its usage in determining unknown forces in dynamics situations.

where λ^* is the vector of Lagrange multipliers at the solution. To satisfy the second order optimality condition, consider $\mathcal{H}$ the Hessian of (1.8), written as

$$\mathcal{H} = \frac{\partial^2 \phi}{\partial \omega^2} - \sum_{j=1}^{\mathcal{E}} \lambda_j \frac{\partial^2 c_j}{\partial \omega^2} \quad . \tag{1.10}$$

This contains the second derivative information of the constrained optimization problem, and as per (1.6) must be positive definite.

1.2.5 Uncontrained Penalty Functions

It is possible to formulate the constrained optimization problem (1.7) as an unconstrained problem using Lagrange multipliers by absorbing the constraints into the objective function and creating what is called a *penalty* function (see for example Gill et al. (1981)). Problem (1.7) is then written as

$$\text{minimize} \quad \psi(\omega) = \phi(\omega) + \rho \mathbf{c}^\dagger(\omega) \cdot \mathbf{c}(\omega) \qquad \omega \in \Omega, \quad \Omega \subseteq \mathbb{R}^W \tag{1.11}$$

where $\phi(\omega)$ is the original objective function, $\mathbf{c}$ is the vector of constraint functions and ρ is known as a penalty parameter (where $x^\dagger$ represents transpose). The inner product is taken over the constraints to ensure that any violation is non-negative.

The first order optimality condition (1.5) on this penalty function is

$$\frac{\partial \psi}{\partial \omega_i} = \frac{\partial \phi}{\partial \omega_i} + 2\rho \sum_{j=1}^{E} c_j \frac{\partial c_j}{\partial \omega_i} \quad ,$$

which is driven to equal equation (1.9) in the limit that $\rho \to \infty$, in which case the penalty parameter is associated with the lagrange multiplier at the solution according to

$$2\rho c_j \to -\lambda_j^* \quad .$$

1.2.6 Optimization Software

In addition to the synchronization-inspired statistical optimization scheme developed within this work, the data assimilation experiments reported herein rely upon external algorithms to satisfy the associated optimization problem.

Two useful software implementations with which we have had success are SNOPT (Gill et al. (2005)) and IPOPT (Wächter (2002)). SNOPT is developed and maintained by P. E. Gill here at UCSD, and IPOPT is available as open source software under the Eclipse Public License (see http://www.coin-or.org/Ipopt/).

Both of these algorithms use a variational approach to minimizing the collocated dynamical trajectory subject to the equations of motion as equality (or very nearly equality) constraints between the state vectors at different time points. This results in a matrix of constraint derivatives that is sparse, and these methods take advantage of this sparcity to manage problems with a significant number of variables, i.e. ω is high dimensional.

Chapter 2

Complications

In this chapter we discuss some of the general complications that affect data assimilation regardless of the optimization approach that is used to solve the associated minimization problem. These complications include things such as numerical issues, high dimensionality of the search space and chaotic instabilities in the objective function surface.

2.1 Numerical & Symmetry Considerations

Computations involving values that range over many orders of magnitude are prone to numerical instabilities. Due to finite limitations on precision, floating point arithmetic is not commutative. As a result, when creating numerical algorithms one should add and subtract terms of a similar magnitude, and keep intermediate multiplicitive terms close to unity if possible (Press et al. (2007)).

Errors also arise naturally during progression of the dynamics due to the presence of positive Lyapunov exponents, the focus of section 2.3 and chapter 4. As the potential for poor variable scaling after many computations exists due to either of these complications alone, dealing with both of them in a high dimensional setting can be a very real impediment to effective algorithm design.

2.1.1 Integration

As numerical integration schemes cannot satisfy invariance under all possible symmertries simultaneously, it is imperative to choose numerical methods that reflect the symmetries of the dynamics, and therefore preserve the relevant conserved quantities of the system. As dissipation

in the dynamics prevents the nonlinear systems that we investigate from conserving the first-integral of the motion, we are not concerned with simplectic intergrators. See Hairer et al. (2002) and the references therein.

Time-reversal symmetry is important for the collection of accurate statistics by the Monte Carlo techniques that are employed in chapters 6 and 7; the other concern being speed, this calculation uses an implicit (second order) trapezoidal rule. For the collocated optimization problems presented to SNOPT, and IPOPT, an implicit Simpson integration scheme is used for its time-symmetry, relative speed and accuracy (fourth order).

For the multiple-Markov chain Monte Carlo approach of chapter 7, an explicit fourth order Runge-Kutta scheme is used, again for its accuracy and speed. However here an explicit form is used because the emphasis is on the speed of repeatedly solving the initial value problem and not on preserving time-symmetry. Although a time-symmetric Runge-Kutta scheme of equivalent accuracy is known to exist (Omelan et al. (2003)), it was not used in the results presented here. Also, similar considerations apply for the calculation of conditional Lyapunov exponents in chapter 4 where a fourth order explicit Runge-Kutta scheme is used.

2.1.2 Scaling

All of the optimization variables are constrained by upper and lower bounds, $l \leq \omega \leq u$, assumed to be provided by the user, however, during the optimization process it is best to scale them all to $-1 \leq \omega' \leq 1$. Using a linear transformation $\omega' = a\omega + b$ and fixing the bounds of ω' to ± 1, the new variable becomes

$$\omega' = \frac{2}{u-l}\omega - \frac{u+l}{u-l} \quad . \tag{2.1}$$

In a geometrical sense, one would like to avoid searching for minimizing solutions in long, narrow valleys. This implies that there is a direction (along the valley) that allows for a large search step without changing the objective function significantly, and that there is another direction (up the valley walls) where only small perturbations are permitted. The situation is further complicated in high dimensions as the valley itself is likely to curve in yet other directions.

One would like to scale the search space to turn any such valleys into radially-symmetric basins, with uniform step sizes in various directions leading to approximately the same amount of change in the objective function. As the first derivatives are zero at the solution, the shape of the local basin is given by second derivative information; in variational techniques, the search variables ω are scaled by the square root of the Hessian (1.10) to make this region more radially symmetric, $\omega' \propto \mathcal{H}(\omega)^{1/2}$, see Gill et al. (1981).

To overcome these difficulties for a stochastic search algorithm, where derivatives are not used, the variance of the distribution in different directions may be used to normalize the perturbation step size, as is done in section 7.4.2.

2.2 High Dimensional Spaces

The number of unknown parameters and state variables that must be searched over during the optimization process can quickly become quite large. There are complications to working in high dimensional spaces that follow solely from geometrical considerations and that strain our intuition on such matters as precision, sampling and statistics.

2.2.1 Entropy of a Gridsearch

A simple argument shows how the size of a space that must be searched over for the data assimilation problem can make the task of finding a minimum challenging. Consider trying to estimate a function over a bounded one-dimensional space; with n uniformly separated samples one obtains approximations to the function at a resolution of about $1/n$ over the bounded region.

The difficulty arises when the space is d dimensional, because in order to achieve the equivalent resolution in all variables requires n^d samples. If one selects a single sample at random, it is done so with probability $P = n^{-d}$.

2.2.2 Non-intuitive Statistics

An intuitive notion of being able to sample a probability distribution over a space is that an adequate approximation may be obtained by clustering the sampling about the mean. For instance, the fractional area of a unit circle inscribed within a square of side-length 2 is fairly significant (≈ 0.785), however the fractional hyper-volume of the analogous problem in a d dimensional space goes to *zero* as d goes to infinity [1]. Meaning that one's ability to accurately sample a high dimensional space does not necessarily follow from low-dimensional intuition (Kolmogorov and Rozanov (1960); Friedman (1989)).

[1] The d-volume of a unit hyper-sphere is given by $V_d = \frac{2\pi^{d/2}}{d\Gamma(d/2)}$, while the normalization $\Gamma(d/2) \to \infty$ as $d \to \infty$, see for example Arfken and Weber (2005).

2.2.3 The 'No Free Lunch' Theorem

One should excercise restraint when attempting to compare the relative success of different optimization schemes, for Wolpert and Macready (Wolpert and Macready (1997)) have shown that on average, if an optimization method perfoms well (as in taking less steps to find a solution than randomly guessing) on a certain class of problem, it will perform poorly (worse than randomly guessing) on the compliment of problems.

The arguments are quite general, stemming from information theory and geometry; they require that the search space is finite –certainly true of any computational method on a finite-state machine. They also require that the space of all possible objective functions that map the search space to $\mathbb{R}_+$ be finite, which is again technically satisfied for the same reason.

Perhaps all is not lost, for we are limiting ourselves to the (purposefully vaguely defined) class of optimization problems that can be described by the dynamical approaches to data assimilation herein. Dangling tautologies aside, we push on ahead, but also caution the reader that the success or failure of an optimization routine on one problem does not guarantee a similar result on an another problem, and so comparisons between the relative performance of algorithms should be made carefully and taken lightly.

2.3 Nonlinearity & Chaos

The richness of the physical world cannot be adequately described with only linear dynamical equations. If a function $f \circ x$ is linear in its argument x it allows us to write $f \circ (a+b) = f \circ a + f \circ b$. In multiple dimensions, the set of equations f can be considered as a matrix operator upon the vector x and the edifice of linear algebra may be used to solve the system of equations.

For nonlinear systems the action of f is not uniform over the space of x; it is not seperable in the way shown here so that applications of linear algebra may only be used as a local approximation. Nonlinearity in a system of equations is necessary for chaotic behavior, however it is not sufficient. The system must also have enough dimensions (at least three) to allow for aperiodic behavior in a bounded region (Lorenz (1963); Afraimovich et al. (1977); Strogatz (1994)).

2.3.1 Sensitivity to Initial Conditions

Allowing for aperiodic behavior in a bounded region of state space requires that $\forall_t \mathbf{x}(t) \neq \mathbf{x}(0)$ as the number of repeated applications of the transformation $\mathbf{x}(t+1) = F \circ \mathbf{x}(t) = F^{t+1} \circ \mathbf{x}(0)$

increases to infinity. It is not surprising that the precision demanded of $\mathbf{x}$ to make the distinction $\mathbf{x} + \epsilon \neq \mathbf{x}$ increases as well (leading to the sampling complications mentioned in 2.2.1). Since the dynamics themselves are deterministic this precision requirement can be related back to restrictions on the inital conditions.

2.3.2 Predicting Chaotic Trajectories

It is the future predictions of our model that provide the ultimate test of the success of the data assimilation process, and these predictions beyond the current measurements are what we are truly after. The previous section shows us that regardless of the success of any optimization procedure it is not possible to predict a chaotic system forward in time indefinitely without infinite precision in the state variables.

No matter how much we try to limit the error introduced by f the chaos in the dynamics will exponentiate the size of the error until it cannot be regarded as negligable with respect to the values of the state variables. However, it is possible to estimate the length of a prediction window, during which time the error in the predicted state is less than some threshold.

Given some error in our current time estimation of the dynamical system $\Delta\mathbf{x}(t)$, perhaps at the end of the data assimilation window when $T = t_0 + N_T \cdot dt_{obs}$, it is possible to estimate the error of the state at a later time $\Delta\mathbf{x}(T + \tau)$ as

$$\Delta\mathbf{x}(T + \tau) = \Delta\mathbf{x}(T) \exp(\lambda_1 \cdot \tau) \quad . \tag{2.2}$$

For data assimilation this is best regarded as providing an upper limit to the time that a reasonable prediction can be made,

$$\tau \lesssim \frac{1}{\lambda_1} \ln \frac{\Delta\mathbf{x}(\infty)}{\Delta\mathbf{x}(T)} \quad . \tag{2.3}$$

One may approximate $\Delta\mathbf{x}_\infty$ as the radius of the attractor, or as the standard deviation of the prediction from a final-time distribution over the measured state; this is done for the Lorenz '63 experiment results in section 7.5.1.

2.3.3 Lyapunov Exponents

Aside from the usage just described, familiarity with the construction of the Lyapunov exponents is also useful because it is the basis for the calculation of the *conditional* Lyapunov exponents, which may be used to determine the number of measurements required for synchronization, and therefore for synchronization-based approaches to data assimilation to succeed (see chapter 4).

Converting the dynamical system (1.1) into a discrete time map involves integrating it for a relatively short interval dt, so that we have

$$x_i(t+dt) = F_i \circ (\mathbf{x}(t), \mathbf{p}) \quad . \tag{2.4}$$

The Jacobian of this map

$$DF_{ij}(x) = \frac{\partial x_i(t+dt)}{\partial x_j(t)}\Big|_x \quad , \tag{2.5}$$

is a matrix which governs the linear behavior of displacements about the point $\mathbf{x}$ at which it is evaluated. If one expands the behavior of the system $\mathbf{F}$ about a point in phase space, the Jacobian DF is the matrix coefficient of the linear term in the expansion. As such, it can be viewed as a transformation that will stretch and rotate any perturbation vector to which it is applied.

We can represent the repeated action of this transformation along the trajectory of the dynamical system $\mathbf{x}(t)$ from $t = 0$ to $N_T - 1$ as

$$\mathrm{DF}^{N_T} = \mathrm{DF}\Big|_{\mathbf{x}(N_t-1)} \cdot \mathrm{DF}\Big|_{\mathbf{x}(N-2)} \cdots \mathrm{DF}\Big|_{\mathbf{x}(1)} \cdot \mathrm{DF}\Big|_{\mathbf{x}(0)} \quad . \tag{2.6}$$

Taking the long time limit $N_T \to \infty$ and multiplying by the transpose to ensure that the eigenvalues are real, Oseledec proved (Oseledec (1968); Johnson et al. (1987)) that the eigenvalues of

$$\mathrm{OSL}(\mathbf{x}_T) = \lim_{N_T \to \infty} \left[(\mathrm{DF}^{N_T})^{\dagger} \cdot \mathrm{DF}^{N_T} \right]^{1/2N_T} \quad , \tag{2.7}$$

exist, are constant to within a basin of attraction, and are unchanged by a smooth coordinate transformation. If we write the eigenvalues of this OSL matrix as $\lambda = \{\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_D\}$, they are the Lyapunov exponents of the system.

As it originates from a nonlinear system, the Jacobian matrix is likely to have both positive and negative eigenvalues, and the multiplication of a large number of ill conditioned matricies quickly leads to numerical scaling problems. Thus, it is important to employ QR (or QL) decomposition to each individual Jacobian of equation (2.7) to perform the numerical calculation. The decomposition separates each subsequent Jacobian into a rotation matrix Q which propogates the direction of the eigenvector along the trajectory, and into a triangular scaling matrix R. The Lyapunov exponents can then be summed from the R decompositions along the way and taken from the log of this matrix (see for example Abarbanel (1996)).

2.3.4 Instabilities in the Objective Function

A connection between the Lyapunov exponents of the model and the instabilities in the objective function surface can easily be seen (Abarbanel et al. (2009)) with a least-squares

objective function such as,

$$\phi(\omega) = \frac{1}{N_T} \sum_{t=0}^{N_T-1} \sum_{i=1}^{M} (x_i(t) - y_i(t))^2 \quad ,$$

where the search domain is over the initial conditions and parameters, $\omega = \{\mathbf{x}(0), \mathbf{p}\}$, and the dependence upon the parameters enters through the dynamics.

Choosing one dimension of the model as an example, as per the optimality condition (1.5) the estimation of $x_i(0)$ formally requires the evaluation of

$$\begin{aligned} \frac{\partial \phi}{\partial x_j(0)} &= 0 \\ \frac{\partial \phi}{\partial x_j(0)} &= \frac{2}{N_T - 1} \sum_{t=0}^{N_T-1} \sum_{i=1}^{M} (x_i(t) - y_i(t)) \frac{\partial x_i(t)}{\partial x_j(0)} \quad . \end{aligned} \tag{2.8}$$

The time evolution of the dynamics (equation (2.4)) allows us to express the chain-rule term in (2.8) as

$$\frac{\partial x_i(t)}{\partial x_j(0)} = \left(\frac{\partial F_i \circ \mathbf{x}(t-1)}{\partial x_k(t-1)} \right) \left(\frac{\partial x_k(t-1)}{\partial x_j(0)} \right) \quad ,$$

where $j, k \in (1, N_x)$, and where the Jacobian of the dynamical map $DF(\mathbf{x}(t))$ is contained within this epression.

This operation must be applied repeatedly so that when $t = N_T - 1$, equation (2.8) contains the term

$$\frac{\partial x_j(N_T - 1)}{\partial x_j(0)} = \left(\frac{\partial F_{jT-1} \circ \mathbf{x}(N_T - 2)}{\partial x_{jT-2}(N_T - 2)} \right) \cdots \left(\frac{\partial F_{j2} \circ \mathbf{x}(1)}{\partial x_{j1}(1)} \right) \left(\frac{\partial F_{j1} \circ \mathbf{x}(0)}{\partial x_j(0)} \right) \quad .$$

The connection is now readily apparent between this expression and the Oseledec matrix (2.7), the eigenvalues of which are the Lyapunov exponents. Consequently, the instabilities characteristic to a chaotic system manifest in the attempt to obtain derivative information over the entire trajectory.

2.3.5 A Note on Identical Systems & Twin Experiments

For the sake of more involved discussion it is useful to categorize different problem situations by the relationship between the involved data and model systems. We will consider two systems to be *identical* when they have the same functional form $\mathbf{f} = \mathbf{g}$ and the same value for any constant parameters $\mathbf{p} = \mathbf{q}$; they may have different initial conditions and still be considered identical in this sense. If two systems have different parameter sets but otherwise share functional forms we will refer to them as *functionally equivalent* systems, $\mathbf{f} = \mathbf{g}$, $\mathbf{p} \neq \mathbf{q}$.

True experimental verification of a mathematical model comes from its predictive power, however to demonstrate that our data assimilation algorithms can succeed on similar problems one often constructs *twin experiments* where the 'data' is generated by a functionally equivalent system, the unmeasured dimensions of which are also known to the experimenter to use as validation.

Chapter 3

Relevant Techniques from Nonlinear Dynamics

In this chapter we discuss two phenomena from nonlinear dynamics, namely chaotic synchronization and phase-space embedding, and how they may be used in data assimilation to overcome the problems discussed in the previous chapter. Specifically, we rely upon synchronization to regulate chaotic instabilities in the dynamical trajectories, and comment on the use of proxy-space transformations to provide more dynamical information in light of insufficient measurements for synchronization.

3.1 Chaotic Synchronziation

One of the more interesting discoveries to come out of the study of chaotic dynamical systems is the phenomenon of *synchronization* (Afraimovich et al. (1983); Fujisaka and Yamada (1983); Pecora and Carroll (1990)). This effect demonstrates that one dynamical system, with equations given by $\mathbf{g}$ and referred to as the *receiver*, may be induced to follow a trajectory given by another system $\mathbf{f}$, denoted the *transmitter*. Under suitable conditions, this remarkable effect of trajectory matching occurs in functionally equivalent systems regardless of their initial conditions, parameters, or the presence of chaotic instabilities.

Although other variants exist, the synchronization relevant to data assimilation is unidirectional: in which an experiment is 'transmitting' measurements to a 'receiving' mathematical model through an ad hoc coupling term. Given the present context, the terms *data* system and *model* system are used rather than transmitter and receiver, although the idea remains the same.

The measured state variables of the data system are passed into the model dynamics by means of a coupling term. The strength of this coupling is mitigated by a *control* parameter γ, and the effect of the coupling term is to force the model to follow the data.

3.1.1 Synchronization of Equivalent Systems

In the description of synchronization that follows, the model and data systems are functionally equivalent. This avoids unnecessary complications at the outset, however the notational distinction between the model $\mathbf{f}$ and data $\mathbf{g}$ equations is preserved for later consistency. Furthermore, for illustration purposes the data system is assumed to be completely known to the researcher, akin to a twin experiment.

The data system is given by a set of dynamical equations

$$\frac{d\mathbf{y}}{dt} = \mathbf{g}(\mathbf{y}, \mathbf{q})$$

where the dependence upon the parameter set $\mathbf{q}$ is made explicit and again, the discussion will be restricted to physically bounded deterministic systems. The first N_M dimensions of $\mathbf{y}$ are those that can be measured in an experiment, and are a subset of the entire N_y-dimensional system.

The first N_M equations of the receiving model may be ordered so that they correspond to their analog in the data system. To accomplish synchronization with the data, these N_M equations of the model are also modified to contain a coupling term, one for each dimension of the measured data, as

$$\begin{aligned}
\dot{x}_1 &= f_1(\mathbf{x}, \mathbf{p}) + \gamma_1 \cdot (y_1 - x_1) \\
&\vdots \\
\dot{x}_M &= f_M(\mathbf{x}, \mathbf{p}) + \gamma_M \cdot (y_M - x_M) \quad , \\
\dot{x}_{m+1} &= f_{m+1}(\mathbf{x}, \mathbf{p}) \\
&\vdots \\
\dot{x}_{N_x} &= f_{N_x}(\mathbf{x}, \mathbf{p})
\end{aligned} \tag{3.1}$$

where $\mathbf{f}(\mathbf{x}, \mathbf{p})$ are the uncoupled model equations, and the overdot denotes differentiation with respect to time, $\dot{\mathbf{x}} = \dfrac{d\mathbf{x}}{dt}$.

With the form of the coupling term shown in equation (3.1), the control parameter γ multiplies the separation between the data and model trajectories in a manor resembling dampening. Although other coupling schemes are certainly possible, unless otherwise noted this form will be used throughout.

The control parameter is positive semi-definite $\gamma \geq 0$ and one can see from $\dot{\mathbf{x}} \sim -\gamma \cdot \mathbf{x}$ that it drives $\mathbf{x} \to \mathbf{y}$, at least in those N_M dimensions for which measurements are available. What makes this phenomenon remarkable is that the other dimensions of $\mathbf{x}(t)$ follow their analogous *un-measured* $\mathbf{y}(t)$ counterparts.

For notational convenience γ may be considered a N_x dimensional vector with the $i > N_M$ components zero, so that the set of equations (3.1) may be written more compactly as

$$\frac{d\mathbf{x}}{dt} = \mathfrak{f}(\mathbf{x}, \mathbf{p}, \mathbf{y}, \gamma) \quad , \tag{3.2}$$

using $\mathfrak{f}$ to represent the coupled model equations and $\mathbf{y}$ is the vector of available data.

3.1.2 The Synchronization Error Metric

To address questions regarding the quality of synchronization requires a measure between the two systems. One way to determine the closeness of two dynamical trajectories is to take the average separation of their locations in phase space. Let us define the synchronization error $\mathrm{SE}(\cdot, \cdot)$ between two timeseries as

$$\mathrm{SE}^2(\mathbf{x}_T, \mathbf{y}_T) \equiv \lim_{N_T \to \infty} \frac{1}{N_T + 1} \sum_{t=0}^{N_T} \sum_{i=1}^{N_x} \frac{1}{r_i^2} (x_i(t) - y_i(t))^2 \quad . \tag{3.3}$$

In principle this can use any metric[1] between the vectors at time t, but for simplicity is chosen to be a scaled Euclidean metric. Scaling over the dynamical range $r_i = u_i - l_i$ for the ith state variable weighs the contribution from each dimension in a uniform manor and makes the SE unitless; one may choose $r = 1$ if the model is already adimensionalized.

The infinite time limit $N_T \to \infty$ is part of this defnintion because the synchronization error is a global property of the combined model-data system. This limit eliminates the contribution to SE from short-time transients. The dependence upon the coupling γ is implicit in the dynamics $\mathfrak{f}$ that are used to generate $\mathbf{x}_T$; again, $\mathbf{y}_T$ is the data timeseries. Note that while not all dimensions of $\mathbf{y}$ must be coupled to $\mathbf{x}$, all dimensions must be known to calculate the synchronization error.

If the models are functionally equivalent and contain the same parameters then $\mathrm{SE} \to 0$ as $N_T \to \infty$ *regardless* of initial conditions, if the parameters are different then $\mathrm{SE} \to c$, $c > 0$. This is crucial for the optimization methods that we develop based upon synchronization, and is illustrated in figure 3.1.

[1]The square-root is explicitly taken so that the SE retains the proper qualifications for a metric, noting that the least squares measure fails the triangle inequality.

3.1.3 Synchronization of Non-Equivalent Systems

Throughout this work synchronization will be used to couple data into a model system for the purpose of data assimilation. From the vantage point of doing this for *experimental* data, where one is ignorant of the underlying system and furthermore, where the data system is almost certainly different from the proposed model, the question of whether experimental data will synchronize with the proposed model is really a question of under what conditions incorrect models will synchronize.

However relevant this question may be, it is unfortunately difficult to answer for a few reasons. If the two systems have a different number of dimensions, it is difficult to define a measure between them, except perhaps with the L_∞ norm, which is computationally expensive to implement.

More severly, this is very problem specific. Given our proposed model, we may ask how it synchronizes with different systems, but the question remains, what second system should be chosen? It is for these reasons that we pursue the question of additive noise as a form of model error in chapter 5 and 6, and attempt to avoid this problem altogether with the use of the self-synchronization error in chapter 7.

3.2 Synchronization in Data Assimilation

The idea of using synchronization techniques as a means to overcome chaotic instabilities that are present during the estimation of state variables and model parameters has been in development for quite some time (So et al. (1994); Voss et al. (2004); Sakaguchi (2002); Sorrentino and Ott (2009); Yu et al. (2007)). However, the problem persists that it is not possible to use the synchronization error as given in (3.3) directly as an objective function due to the requirement of knowing the unmeasured data dimesions.

3.2.1 Balanced Synchronization

A delicate balance must be reached between using a coupling that is strong enough to prevent chaotic instabilities in the solution dynamics from manifesting and one that is so powerful that it aligns all trajectories with the data, making the search space singular. For the variational methods that are implemented within this work to find a solution to the data assimilation problem with chaotic, or otherwise nonlinear models, a modified form of the model dynamics (3.2) are used to couple the data to the model (thus allowing for synchronization at large values of the

coupling), along with an objective function that penalizes, or *balances* the use of the coupling term, relative to the mismatch between the model and the data (Abarbanel et al. (2008, 2009)).

An example of such an objective function is the least squares difference between the data $\mathbf{y}$ and the model $\mathbf{x}$, with an additional quadratic penalty term for the coupling γ, such as

$$\begin{aligned} \text{minimize} \quad & \phi = \sum_{t=0}^{N_T-1} \sum_{i=1}^{M} \Big(x_i(t) - y_i(t) \Big)^2 + \gamma^2 \quad . \\ \text{subject to} \quad & \mathbf{x}(t+1) - \mathfrak{F} \circ (\mathbf{x}(t), \mathbf{p}, \mathbf{y}(t), \gamma) \quad . \end{aligned} \tag{3.4}$$

This may be considered a type of penalty function (1.11) (with $\rho = 1$ here), where $\mathfrak{F}$ is used to denote the dynamical map of the coupled equations $\mathfrak{f}$ in (3.1).

The use of the coupling must be regulated because it is an *ad hoc* addition to the model dynamics $\mathbf{f} \to \mathfrak{f}$ which, it is presumed, are already an adequate description of the physical system that created the data; one therefore wishes to use as small a coupling as possible.

For collocation methods, where each $\mathbf{x}(t)$ is taken to be an independent variable, the coupling can effectively be driven to zero between neighboring time points $\mathbf{x}(t)$ and $\mathbf{x}(t+1)$. However the coupling is still required during the optimization procedure to make the search space surrounding the solution $\mathbf{x}(t) \approx \mathbf{y}(t)$ more informative to the derivative-dependent linear algebra operations performed by variational approaches. An added complication is that although formally $\gamma \to 0$, there remains the practical difficulty of how one chooses the relative scaling (i.e. ρ) between the terms.

3.2.2 Properties of the Synchronization Error Metric

The coupled trajectory can be written with explicit dependence on the initial conditions and parameters of the dynamics as the discrete time map $\mathfrak{F}$ as

$$\begin{aligned} \mathbf{x}_T &= \mathbf{x}(0), \mathfrak{F} \circ (\mathbf{x}(0), \mathbf{p}, \mathbf{y}(0), \gamma), \mathfrak{F}^2 \circ (\mathbf{x}(1), \mathbf{p}, \mathbf{y}(1), \gamma), \ldots, \mathfrak{F}^{N_T} \circ (\mathbf{x}(N_T - 1), \mathbf{p}, \mathbf{y}(N_T - 1), \gamma) \\ &= sync(\mathbf{x}(0), \mathbf{p}, \mathbf{y}_T, \gamma) \end{aligned} \tag{3.5}$$

With this understanding, the SE may be expressed in terms of the model initial conditions, coupling and data

$$\mathrm{SE}(\mathbf{x}_T, \mathbf{y}_T) = \mathrm{SE}(\mathbf{x}(0), \mathbf{p}, \mathbf{y}_T, \gamma) \quad , \tag{3.6}$$

a notation that makes the quantities of interest readily available.

Figure 3.1 shows the synchronization error between two trajectories of the Lorenz '63 system with different initial conditions; this particular model is chosen for later reference to the

twin experiment of chapter 7, however presently we are concerned with the general behavior of SE. From this figure one should note two important properties of the synchronization error metric as $\gamma \to \infty$. Firstly, with $\gamma \cdot dt$ greater than 0.08, the SE monotonically decreases for all of the finite N_T evaluations; this indicates that stronger coupling leads to a model trajectory that is closer to the data trajectory (recall that *all* dimensions of the data enter into (3.3) and here are assumed known).

The second feature of the SE we point out is that using more time points $N_T \to \infty$ in the calculation results in consistently reduced values by eliminating the contribution of finite-time transients to the SE.

3.3 Synchronization as a Critical Phenomenon

The image of the synchronization error in figure 3.1 is clearly reminiscent of a critical phase transition (e.g. ferromagetism at the Curie temperature), where there is an emergence of a uniform state from individual constituents (Landau and Lifshitz (1980); Pathria (1996)). Related as synchronization is to the onset of chaos in dynamical systems, it is possible that this is another expression of the Feigenbaum criticality (Feigenbaum (1979); Coppersmith (1999)), or perhaps anther type of bifurcation (Afraimovich and Shilnikov (1974); Afraimovich et al. (1983)).

Regardless, critical phase transitions with the synchronization error as the order parameter have been observed in more traditionally structured systems that interact via coupling as in equation (3.1) and otherwise exhibit spatio-temporal chaos (see for example, Ahlers and Pikovsky (2002); Szendro et al. (2009)).

In the context of chaotic systems, the knowledge of the behavior of the SE near the critical coupling γ_c is perhaps of less interest than knowing the value of γ_c itself. For γ_c represents the smallest value of the coupling between two systems that allows for synchronization of their trajectories in the long time limit. This means that the coupled system is altered the least (from the uncoupled dynamics) that is allowed by chaos. The need to determine γ_c for the model in question as a prerequisite for the success of the subsequent optimization problem is explored in Abarbanel, Creveling, Farsian and Kostuk (Abarbanel et al. (2009)), although presented purely with regards to the success of data assimilation, and not as it relates to critical phenomena.

We return to this analogy again in section 5.5 as a possible means to constructing a path integral representation for data assimilation. Although this connection is not rigorously exhausted, it is provided as motivation for the design of the optimization algorithm in chapter 7, which uses the emergence of this criticality as a marker to terminate the search procedure.

3.4 Phase Space Reconstruction

There is another phenomenon from nonlinear dynamics that is useful for data assimilation known as phase-space reconstruction. Meaning that it is possible to reconstruct higher dimensional information from a measured timeseries by expressing it in a time-delay or *proxy-space* coordinate system. Takens and Ruelle used a period-double approximation to turbulent chaos to place an upper bound of $2d + 1$ on the number of time-delay dimensions required to guarantee an embedding of the original d-dimensional system (Ruelle and Takens (1971); Takens (1981)); The embedding theorem of Takens is the extension of this to nonlinear systems in general (Takens (1981)).

Packard *et. al* conjecture that a time-delay coordinate system is sufficient to capture some geometrically invariant properties of the dynamical system, such as the topology and dimension of the attractor (Packard et al. (1980)). They also demonstrate that the dimension of an unknown system may be determined from the smallest time-delay reconstruction (equation 3.7) that collapses the conditional probability distribution $P(x(t)|s_k(t), \tau)$ to a delta function (in the noise-free limit).

An appropriate delay size τ can be chosen by minimizing the mutual information between the original trajectory and the one projected onto time-delay coordinates (Fraser and Swinney (1986)). It is also possible to choose non-uniform spacings between the time-delayed dimensions to more-efficiently capture dynamics with multiple timescales (Hirata et al. (2006)). One can use this reconstruction to determine estimates of the Lyapunov exponents of a measured signal, provided that sufficient data is available (thus approximating $N_T \to \infty$)(Abarbanel et al. (1992)). Additionally, this method can also be used to distinguish between high dimensional deterministic systems and types of noise (Kennel and Isabelle (1992)).

3.4.1 Time-delay Coordinates

With the emphasis throughout this work on using synchronization to assist in the data assimilation problem, one must consider the problem that the coupled data from the measured dimension may be insufficient to achieve synchronization. The number of required measured dimensions needed to allow for synchronization may be determined for a model as described in chapter 4, however can anything be done if one is not able to measure all of these variables?

Assume for the moment that one can only measure one dimension of the dynamics $x_1(t)$; According to the embedding theorem, there is sufficient information contained in this timeseries to reconstruct the dynamics in the time-delayed space of the system.

Consider the d-dimensional vector formed from the time-delay values of x_1,

$$\mathbf{s}(t) = (x_1(t), x_1(t-\tau), \ldots, x_1(t-(d-1)\tau))^\dagger \quad . \tag{3.7}$$

Since the delayed dynamics may be written as

$$\mathbf{s}(t+1) = \mathbf{G} \circ \mathbf{s}(t) \quad , \tag{3.8}$$

and due to the nature of how this space is constructed, these dynamics form a rule to take the set of delay coordinates into the next measured state $x_1(t+1)$ as the first dimension of $\mathbf{s}(t+1)$.

A method is proposed in Abarbanel et al. (2009) to use this phase space embedding to achieve an enhanced version of synchronization that relies upon applying the coupling to the transformed, proxy-space expression (3.8) of the dynamics. The practical difficulties encountered with performing this transformation (more specifically, the reverse transformation $\mathbf{s} \to \mathbf{x}$) on high dimensional systems, and when multi-step integration routines are in place, make this approach challenging and is not pursued further in this work.

3.4.2 Proxy Space Inspired Objective Functions

In addition to improving the synchronization between two systems with insufficient measurements, the information contained in the time-delay coordinates of the measured variables can be applied more directly to the minimization problem of data assimilation.

For the variational approach as described in chapter 1, the use of time-delay information manifests as additional constraints of the form

$$x_i(t) - F_i^n \circ \mathbf{x}(t - n\tau) = 0 \qquad n > 1 \quad , \tag{3.9}$$

where $\mathbf{F}^n$ implies n repetitions of the dynamical map.

Although this may appear redundant for a collocated problem that already contains the $n = 1$ case as a constraint, we point out that (3.9) only connects the variables at two time points, namely t and $t - n\tau$, and *skips* over the intermediary times, enforcing longer-term correlations between the data and model values. It should not be surprising that if n is chosen too large, chaotic instabilities will occur in the evaluation of these constraints in exactly the same way as described in section 2.3.4.

On the other hand, if n is chosen very small ($n \approx 1$) then the benefit to using time-delay constraints becomes very slight compared to simple model error constraints. Additionally, every n creates off-diagonal entries in the Lagrangian of the objective function, so that the incorporation of many different n-constraints may have serious consequences for the sparcity of the resultant

Lagrangian. As the two variational solvers that we utilize rely upon sparse matrix techniques for increasing the speed of the linear algebra calculations, this is a relevant concern to using time-delays.

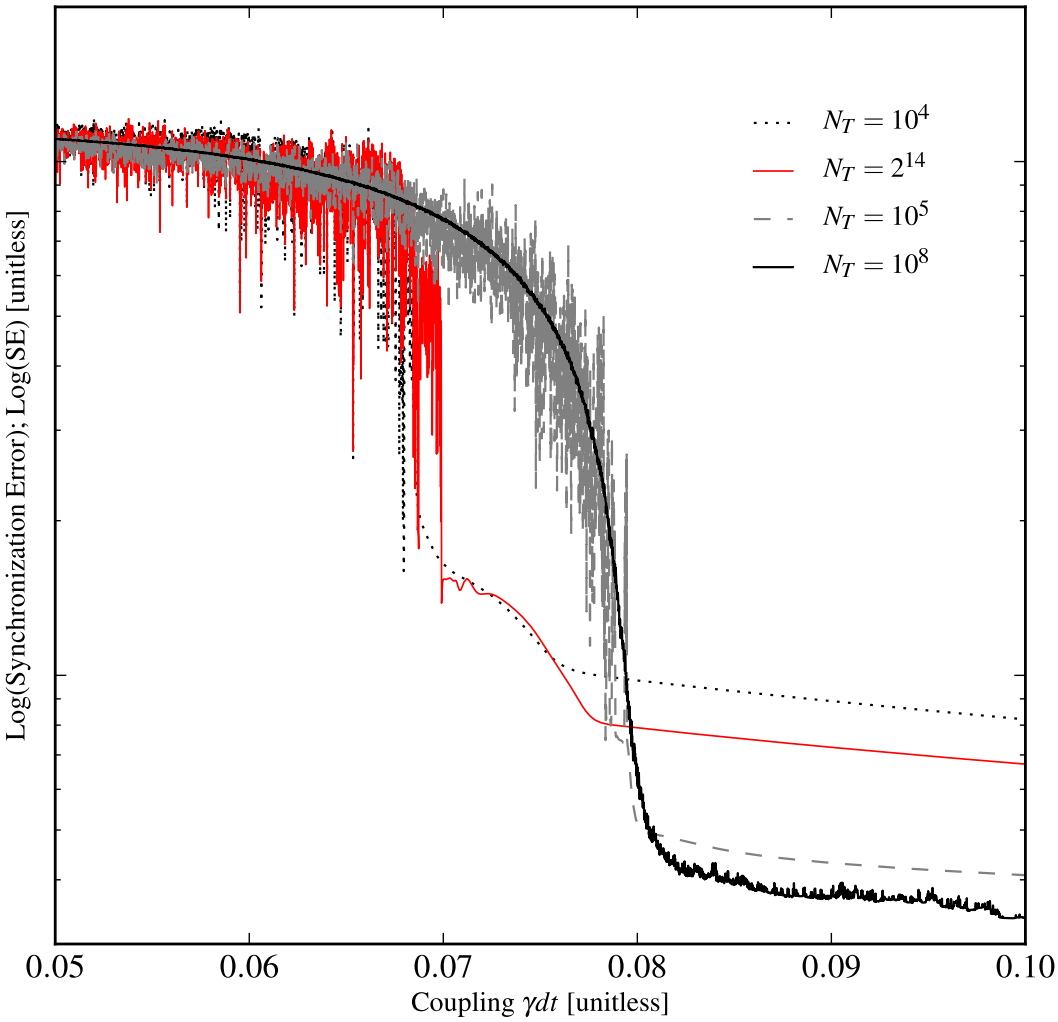

Figure 3.1: The synchronization error (SE), equation (3.3), of the Lorenz '63 system is shown as a function of coupling strength γ for timeseries of various lengths. The two trajectories have differing initial conditions, yet identical parameters ($\mathbf{f} = \mathbf{g}, \mathbf{p} = \mathbf{q}$), which are integrated using an explicit RK4 scheme with timestep of $dt_{int} = 0.01$; the total number of timepoints N_T over which the SE is calculated is given in the legend. The parameters that are used are listed in table 7.1, and are the same as those used for the 'data' in the twin experiment of section 7.5. The emergence of a bifurcation at $\gamma_c dt_{int} \approx 0.08$ is suggestive of a critical phenomenon as $N_T \to \infty$.

Chapter 4

Determining the Necessary Measurements for Synchronization

Synchronization may be used in a number of ways to assist the optimization step of the data assimilation problem. Therefore ascertaining whether a dynamical system is capable of synchronizing to the given data trajectory becomes paramount for the problem of data assimilation.

Not knowing the full data system makes a rigorous determination of synchronization success impossible for a true experimental situation. However, the information gained from examining the *self*-synchronization case, when $\mathbf{g} = \mathbf{f}$, is quite useful nonetheless and can be regarded as necessary (though perhaps not sufficient) to describing the more general case.

With these considerations, we will answer the following question in a model specific way: *given a model with coupling,* $\mathfrak{f}(\mathbf{x}_T, \mathbf{p}, \mathbf{y}_T, \gamma)$*, what measurements must be supplied to it, and at what coupling strength* γ_c*, so that it will synchronize with data* $\mathbf{y}_T$ *generated from* $\mathbf{g}(\mathbf{y}_T, \mathbf{q})$*, when* $\mathbf{g} = \mathbf{f}$. An equivalent way to think about this is to ask what dimensions of γ must be non-zero in order for self-synchronization to occur?

The investigations show that the search-space region around the synchronization manifold (where $\mathbf{x}_T \approx \mathbf{y}_T$) is smoothed when the largest *conditional* Lyapunov exponent is negative (Rulkov et al. (1995)).

4.1 The Synchronization Manifold

The relevant space of the combined data-model system in which the synchronized trajectory evolves is $\mathbb{R}^{N_x} \otimes \mathbb{R}^{N_y}$, and the *synchronization manifold* is the subset of this space where the model is equivalent to the data; i.e. where $\mathbf{x}_T = \mathbf{y}_T$. The distance of the model trajectory from this manifold is measured by the synchronization error, equation (3.3).

One should not think of this manifold as a simple projection of the model dynamics into the higher dimensional combined space because the synchronization coupling term alters the dynamics of the model system (see equation (3.1)) so that the actual trajectory of the synchronized model is a mixture of the data and model attractors (Abarbanel et al. (1995)). It is true that in the identical case, in which $\mathbf{f} = \mathbf{g}$ and $\mathbf{p} = \mathbf{q}$, the model is innately capable of reproducing the data trajectory; the difference between it and the data is their starting location, and this is equivalent to a time-translation between their attractors. However, despite the existence of such a translation, it remains to be determined if the data that is passed to the model is capable of enforcing it.

We note that since this combined surface is distinct from the native (un-coupled) model trajectory due to blending with the data, the Lyapunov exponents of the model are no longer the appropriate measure for determining the stability of the synchronized system.

4.1.1 Stability of the Synchronization Manifold

Ensuring the stability of the space surrounding the synchronization manifold is important because it allows the alignment of the model to the data, and also because otherwise chaotic instabilities can mask possible solutions to the data assimilation problem.

For mismatched data-model systems, or even functionally equivalent systems with different parameter sets, the underlying chaotic dynamics makes it unlikely that the synchronization manifold is stable. Meaning that a small perturbation from this region will push a dynamical trajectory farther away, especially perturbations in directions that are not measured. From the viewpoint of a model attempting to follow a data trajectory, this is precisely what the unmeasured dimensions of the model dynamics do.

However, as mentioned above this manifold is dependent upon the specific data that is presented to the model, therefore when $\mathbf{f} \neq \mathbf{g}$ its stability can only truly be tested on a case by case basis. It is for this reason that the restriction $\mathbf{f} = \mathbf{g}$ is used to address the question of stability, thereby making it a question of the relevant measurements necessary for self-synchronization of a single model choice $\mathbf{f}$.

The synchronization-error surface suffers from the same precarious situation as the least-squares objective function as described in section 2.3.4. This should not be surprising given the similarity between the form of the synchronization error (3.3) and a least squares objective function. An image of this irregular surface is given in figure 4.2.

The ability of the measured data dimensions, given a strong enough ($\gamma >> \gamma_c$) coupling, to smooth the synchronization manifold means that they are necessary for synchronization-based data assimilation techniques to be successful; otherwise the surrounding surface is riddled with local minima. Combining the implications of figure 4.2 with those of figure 3.1 one may infer that when the coupling is at the minimum value required for stablization, i.e. $\gamma = \gamma_c$, then the only point in the search space that will not have a large synchronization error will be the 'solution' $\{\mathbf{x}(0), \mathbf{p}\} = \{\mathbf{y}(0), \mathbf{q}\}$, and the surrounding space will be un-informative. Increasing the coupling from γ_c increases the size of the stablized region about the solution.

4.1.2 Conditional Lyapunov Exponents

Recognizing from the self-synchronization case that the instabilities in the synchronization manifold come from the presence of chaos in the dynamics, it follows that if our measured data is coupled into the model in such a way that the Lyapunov exponents of the resultant, coupled data-model system, are reduced to negative values, then the region near the this manifold will attract nearby trajectories and become stable to small pertubations. As these experiments are dependent upon the presented data, they are referred to as *conditional* exponents.

The calculation of conditional Lyapunov exponents (CLEs) proceeds similarly to that of standard Lyapunov exponents for a dynamical system. The main difference occurs at the beginning, with the inclusion of the data timeseries $\mathbf{y}_T$ that is driving the model system $\mathbf{f}$,

$$\frac{d\mathbf{x}}{dt} = \mathbf{f} \circ (\mathbf{x}, \mathbf{p}) + \gamma(\mathbf{y} - \mathbf{x}) \quad . \tag{4.1}$$

Here, recall that only N_M dimensions of γ are non-zero; the non-zero elements of γ corresponding to the measured dimensions of $\mathbf{y}_T$.

The stability of the synchronization manifold is then the condition that the real part of the largest CLE, λ_1 is sufficiently negative to prevent the divergence of nearby trajectories. In practice, simply being negative at all is usually adequate, however there may be situations where $\lambda_1 < 0$ is insufficient (Rulkov et al. (1995))

To investigate the infinite coupling limit, where the tracking of the data by an auxiliary system has the best chance of succeeding, the system takes on the value in the measurement $x_M = y_M$ (as per eqn. (3.1)), and these degrees of freedom are eliminated. Since the dynamics

of the system $\mathbf{F}$ no longer depend on those variables that have been replaced,

$$\mathbf{F} \circ (x_1, \ldots, x_{N_x}) \rightarrow \mathbf{F} \circ (y_1, \ldots, y_M, x_{M+1}, \ldots, x_{N_x}), \tag{4.2}$$

the rows and columns of the Jacobian $\partial f_i / \partial x_j$, with $i, j \in 1, \ldots, M$, are zero. The remaining $N_M + 1$ to N_x uncoupled dimensions define the reduced, or sub-Jacobian, matrix of the system (Pecora and Carroll (1990)). It is now possible to continue the CLE calculation analogously to the Lyapunov exponent calculation in section 2.3.3 where the decomposition is done on this reduced dimensional matrix.

4.2 Numerical Examples

The following numerical results are similar to those in Abarbanel, Kostuk and Whartenby (Abarbanel et al. (2010)), in which the connection between positive conditional Lyapunov exponents and the stability of the synchronization manifold is explored. Additionally in Abarbanel et al. (2010), the optimization software SNOPT (Gill et al. (2005)) is used to solve data assimilation twin experiments on the Lorenz '96 Model by collocating the timeseries of dynamical variables and using a multi-dimensional chaotic synchronization term to regulate search instabilities.

4.2.1 The Lorenz '96 Model

Exploring the situations discussed above using some numerical examples, our system of interest is the Lorenz '96 system. This system is a toy model that was presented by Lorenz at a conference in 1996 (Lorenz (1996)). This model is a ring (one spatial dimension with periodic boundary conditions) composed of K similar equations:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + \nu \qquad i \in 1, \ldots, K, \tag{4.3}$$

where $x_0 = x_K$ and $x_{-1} = x_{K-1}$, making the system a K-dimensional set of differential equations. The model has a single forcing parameter ν that drives it to chaos when ν is larger than a critical value ν_{crit}; this critical value depends upon K, however we use $\nu = 8.17$ which is greater than ν_{crit} for all of the work that follows.

The physical interpretation of this model could loosely be described as 'weather on a circle'. It describes the motion of a single quantity x that has some natural rotation around a bounded one-dimensional space, with driving force ν. The nonlinearity is provided by a term resembling a spatial discretization of advection.

The number of equations, K, represents the spatial discretization of the system around the ring. We exploit the dimensional-flexibility of this system to demonstrate the effects of increasing complexity on the stability of the synchronization manifold.

Adding a synchronization coupling vector γ to this model, and a driving data signal given by $\mathbf{y}_T$ we have,

$$\mathfrak{f}_i = \frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + \nu + \gamma_i(y_i - x_i) \qquad i \in 1, \ldots, K, \tag{4.4}$$

Specifically choosing the $K = 5$ system and a coupling vector $\gamma = \{\gamma_1, 0, 0, 0, 0\}$ results in the following response system

$$\begin{aligned}
\dot{x}_1 &= x_5(x_2 - x_4) - x_1 + \nu + \gamma_1(y_1 - x_1) \\
\dot{x}_2 &= x_1(x_3 - x_5) - x_2 + \nu \\
\dot{x}_3 &= x_2(x_4 - x_1) - x_3 + \nu \\
\dot{x}_4 &= x_3(x_5 - x_2) - x_4 + \nu \\
\dot{x}_5 &= x_4(x_1 - x_3) - x_5 + \nu \qquad .
\end{aligned} \tag{4.5}$$

The data signal $y_1(t)$ is taken from an identical Lorenz '96 system with the same forcing parameter ν, but with different initial conditions than the response system; i.e., $y(0) \neq x(0)$.

Insufficient Measurements

As it happens, the coupling scheme presented to the model in equation (4.5) is insufficient to stablize the synchronization manifold. This can be seen in the top panel of figure 4.1, where we see how the spectrum of conditional Lyapunov exponents of the singly coupled Lorenz '96 $K = 5$ model is affected by the finite value of the coupling parameter γ. The largest conditional Lyapunov exponent, shown with diamonds (and a red line), never becomes negative denoting a failure of synchronization for all coupling strengths. Finite values of γ, expressed as unitless values of γdt, are used and the exponents of the full Jacobian are calculated. Beyond $\gamma dt \geq 1$ this approach becomes numerically unstable; however one is able to verify that the CLEs of the four remaining –that is, uncoupled– state variables indeed approach those calculated using the sub-Jacobian method (representing infinite coupling strength) which is used to calculate the CLEs in table 4.1.

In table 4.1 and all of the tables that follow, the models are integrated for 10^5 time points, after which the Lyapunov and conditional Lyapunov exponents are calculated along subsequent trajectories of 10^9 points. All integration is done using an explicit fourth-order Runge-Kutta integration scheme with a timestep of $dt_{int} = 0.01$ per point. The Jacobian is numerically

calculated based upon an averaged offset of both $\pm 10^{-8}$ in all directions. The eigenvalues of the Oseledec matrix that is formed from the reduced Jacobians, as described in the previous section, are determined using a standard QR decomposition algorithm as found in Numerical Recipies (Press et al. (2007)).

Sufficient Measurements

Taking symmetery into consideration, there are two unique ways to assimilate two measurements into the Lorenz '96 $K = 5$ model; coupling through x_1 and x_2 or through x_1 and x_3. We chose the latter, where $\gamma = \{\gamma, 0, \gamma, 0, 0\}$, for the bottom panel of figure 4.1. In this coupling scenario the largest conditional Lyapunov exponent becomes negative; indicating that the proposed criterion for stability of the synchronization manifold is now satisfied.

Table 4.1: The Lyapunov exponents of the autonomous and conditional Lyapunov exponents of the Lorenz '96 $K = 5$ model (with $\nu = 8.17$) for various coupling schemes. Both of the dual coupling schemes result in stabilized synchronization manifolds, while the single coupling scheme does not.

autonomous (LE)	single coupling (CLE) y_1	dual coupling (CLE) y_1, y_2	dual coupling (CLE) y_1, y_3
0.527	0.491	-0.117	-0.106
10^{-7}	-0.298	-1.201	-1.089
-0.432	-0.832	-1.871	-2.183
-1.303	-3.550	-	-
-3.793	-	-	-

4.2.2 Projection of the Synchronization Manifold Surface

We now examine the effect that a negative λ_1 has on the structure of the surface near the synchronization manifold. Consider a higher dimensional Lorenz '96 model, with $K = 9$. A projection of the synchronization error for this model and a single driving variable coupled through x_1 is shown in the upper left panel of figure 4.2 and in table 4.2 we have listed both the Lyapunov exponents of the autonomous $K = 9$ model, as well as the CLEs that result from this single coupling. The response system $\mathbf{x}$ is started completely on the synchronization manifold with the exception of the x_4 direction, whose deviation from this manifold is given on the abscissa. That is, $x_i(0) = y_i(0)$ for all $i \neq 4$.

Notice that there does exist a sharp minimum at $x_4(0) = y_4(0)$ indicating the presence of the synchronization manifold. However, with a single coupling there remains a positive exponent, indicating the lingering instabilities surrounding the synchronization manifold which results in the nearby surface being self-similar in nature. In the upper right panel of figure 4.2 we present a hundred-fold enlargement of the region around $x_4(0) = y_4(0)$. Once again, a minimum indicating the existence of the synchronization manifold is present, however it is not any easier to find within a smaller bounded region.

Table 4.2: The Lyapunov exponents and singly-coupled conditional Lyapunov exponents of the Lorenz '96 $K = 9$ model (with $\nu = 8.17$). The single coupling term reduces the positive Lyapunov exponents, however it provides insufficient stablization for synchronization to be successful. (see the upper two panels of figure 4.2.)

autonomous (LE)	single coupling y_1
1.268	1.012
0.585	0.447
10^{-6}	-0.092
-0.208	-0.628
-0.759	-1.125
-1.245	-1.681
-1.815	-2.459
-2.573	-3.632
-4.252	-

Judging by the conditional Lyapunov exponents in Table 4.3, *none* of the simple coupling schemes that incorporate two observations into the Lorenz '96 $K = 9$ model will result in synchronization because they all contain at least one positive exponent.

It is not until three observations are used that regularization can occur for this system The CLEs that result from some of these triple coupling schemes are listed in table 4.4. We have left out some of the schemes that are redundant based upon the symmetry of this model.

We also note that some coupling schemes are more successful than others at minimizing the conditional Lyapunov exponents. For example the y_1, y_2 and y_3 coupling arrangement does not lead to a completely negative spectrum of CLEs, whereas the y_1, y_2 and y_7 arrangement does. This is to be expected because the manor in which the measurements are passed thoughout the dynamics is not uniform.

If we examine a driven model that does have a completely negative spectrum of conditional Lyapunov exponents, such as the y_1, y_2 and y_7 coupling scheme, the synchronization error

Table 4.3: The conditional Lyapunov exponents of the Lorenz '96 $K = 9$ model with two coupled observations. Each column represents a different coupling scheme with two pieces of information $y_1(t)$ and $y_i(t)$, for $i = 2, 3, 4, 5$. Those not listed are similar to one that is listed via symmetry. All of these dual-coupling schemes are incapable of synchronizing the $K = 9$ model.

y_1, y_2	y_1, y_3	y_1, y_4	y_1, y_5
0.162	0.789	0.638	0.799
-0.208	0.189	0.058	0.265
-0.581	-0.347	-0.458	-0.380
-0.909	-0.880	-0.953	-0.930
-1.281	-1.425	-1.449	-1.511
-1.794	-2.225	-2.087	-2.342
-2.543	-3.414	-3.061	-3.213
-	-	-	-
-	-	-	-

Table 4.4: The conditional Lyapunov exponents of the Lorenz '96 $K = 9$ model with three coupled observations. Each column represents a different coupling scheme. The choice of variables to observe and couple is important; some couplings, e.g. y_1, y_2, y_7, are capable of synchronizing, while others are not.

y_1, y_2, y_3	y_1, y_2, y_7	y_1, y_2, y_8	y_1, y_3, y_5	y_1, y_3, y_7	y_1, y_4, y_7
0.106	-0.070	-0.024	0.661	0.501	0.025
-0.308	-0.553	-0.471	0.013	-0.076	-0.421
-0.736	-0.837	-0.846	-0.618	-0.732	-0.986
-1.130	-1.207	-1.137	-1.226	-1.265	-1.257
-1.653	-1.517	-1.523	-2.083	-1.991	-1.625
-2.436	-2.128	-2.311	-3.216	-2.906	-2.202
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

that results from the phase space neighboring the synchronization manifold is now smooth and capable of being searched for a minimum. This regularization effect is shown in the bottom panel of figure 4.2.

4.2.3 Parameter Dependence of the Conditional Lyapunov Exponents

It is also possible to examine the effect that the forcing parameter has on the spectrum of conditional Lyapunov exponents. As we increase the forcing parameter ν we expect that the model exhibits more chaotic behavior, which should manifest itself in the CLE spectrum as an

increasing number of positive exponents. The CLE spectrum for the singly-coupled Lorenz '96 $K = 5$ model, as in equation (4.5), is presented in figure 4.3. Here the spectrum is calculated using the sub-Jacobian method and so one conditional Lyapunov exponent is identically zero (not shown), corresponding to the coupled dimension, y_1. The remainder of the spectrum is completely negative for $\nu \lesssim 7$, indicating that this single measured state variable is sufficient to allow for synchronization if the parameter of the model is below this range. Above this forcing value there exists a positive exponent, signifying a failure of synchronization with the coupling of this single measurement; this is the regime presented above where $\nu = 8.17$. If the model has an even greater forcing, $\nu \gtrsim 18$, there exist multiple positive CLEs that could require even more measurements for regularization. One can clearly see degeneracies in the eigenvalues in the region with $\nu \lesssim 5$, and there is a prominent bifurcation at $\nu \approx 2.5$.

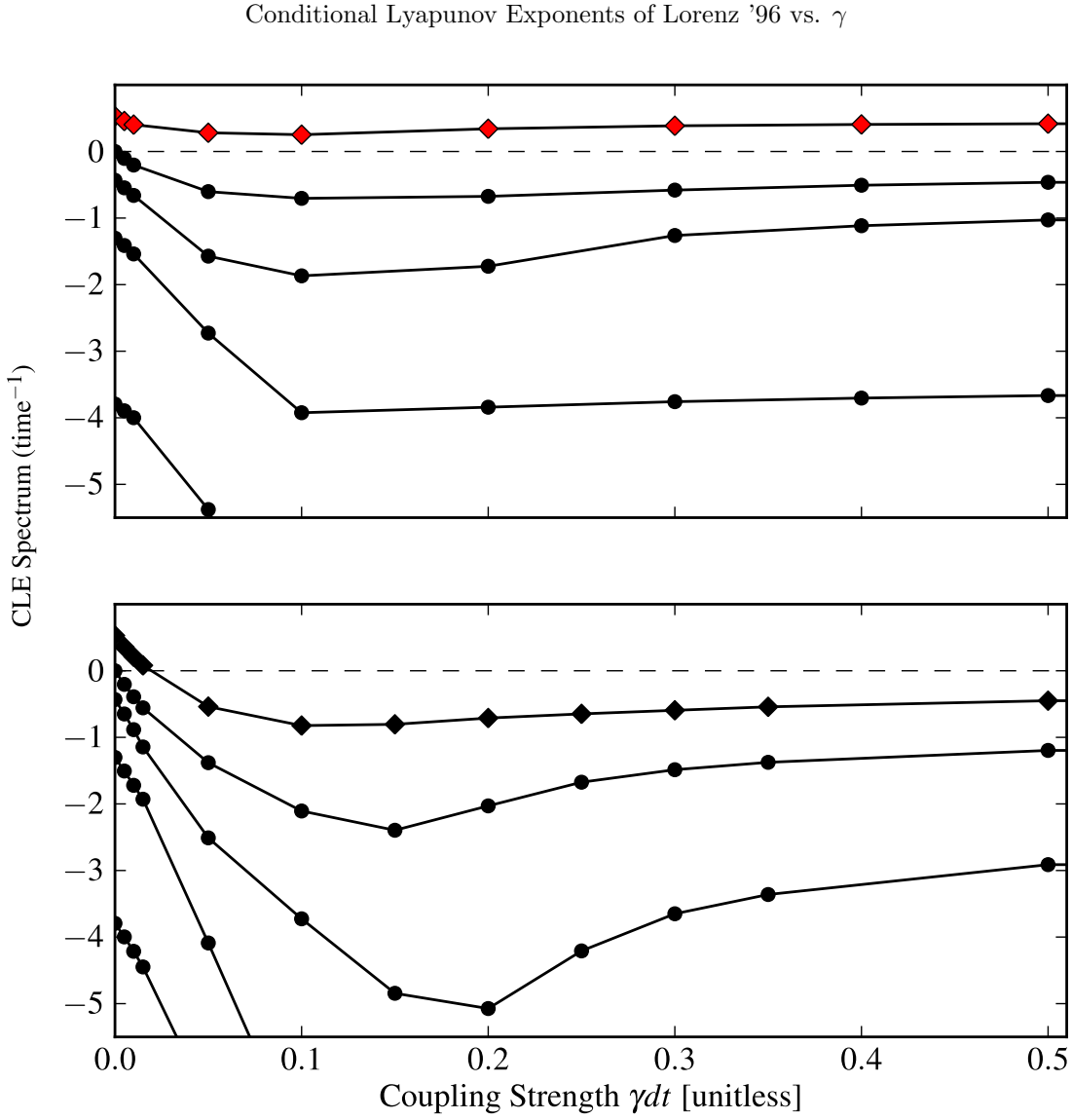

Figure 4.1: Each panel shows the effect of varying the strength of the coupling parameter γ on the conditional Lyapunov exponent (CLE) spectrum of the Lorenz '96 $K = 5$ model with (upper panel:) a single coupled measurement, as in equation (4.5), which is insufficient to make this system's single positive CLE (shown with red diamonds) negative for any value of the control parameter. And with (lower panel:) two coupled measurements, $y_1(t)$ and $y_3(t)$, which is sufficient to reduce the largest CLE below zero. Note in each how the exponent associated with the coupled state-variable(s) goes (off the graph) as $-\gamma$ for large γ. (The lines are present to guide the eyes.)

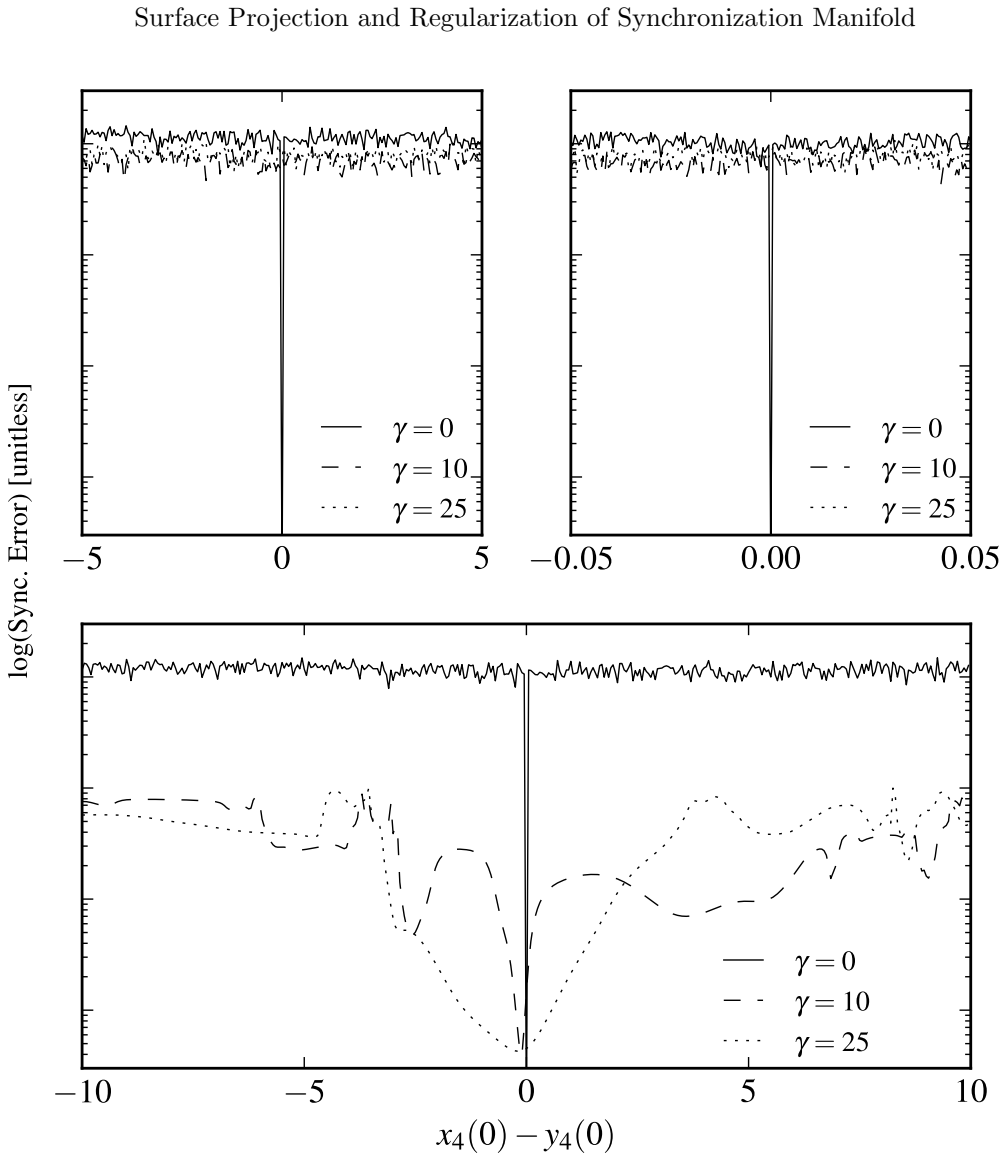

Figure 4.2: The synchronization error of the Lorenz '96 $K = 9$ model that results from three values of the coupling parameter $\gamma = 0$, 10, and 25 are plotted in each panel; The upper two panels use $\{\gamma, 0, 0, 0, 0, 0, 0, 0, 0\}$, while the lower panel uses coupling $\{\gamma, \gamma, 0, 0, 0, 0, \gamma, 0, 0\}$ (through x_1, x_2, x_7). The surface is projected along deviations to the uncoupled dimension x_4. A sharp minimum exists at the synchronization manifold itself, but the surrounding space is un-informative without the aid of the coupling. Note the different scale in the upper right panel.

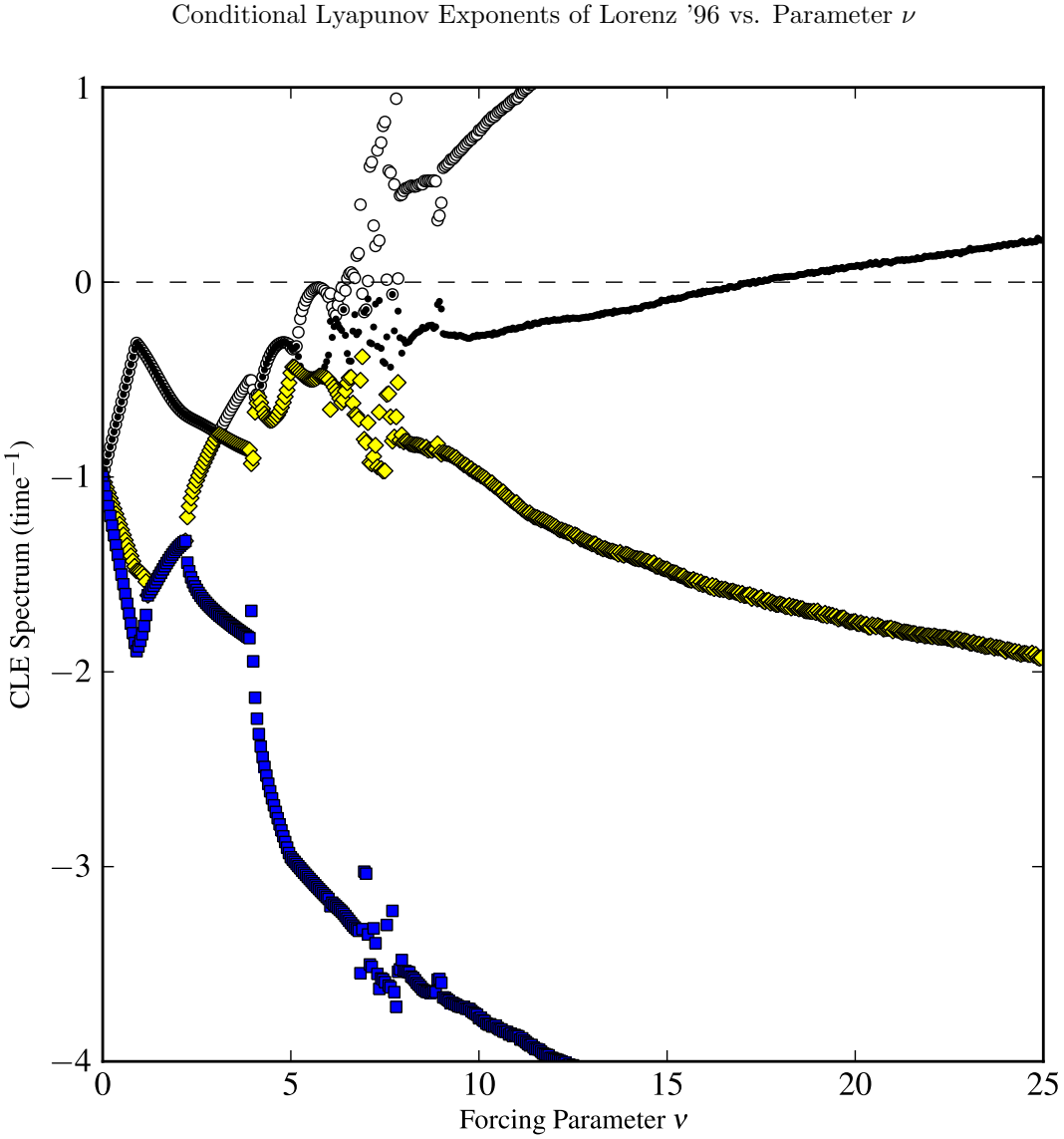

Figure 4.3: The spectrum of conditional Lyapunov exponents (CLEs) of the Lorenz '96 $K = 5$ model with a single coupling (see equation (4.5)) for various values of the forcing parameter ν. The emergence of additional positive CLEs (at $\nu \gtrsim 7.18$) as the forcing is increased indicates the failure of this single coupling to regularize the synchronization manifold in this parameter regime. Note only the four CLEs from the uncoupled eigenvectors are plotted.

Chapter 5

The Objective Function as a Path Integral

In this chapter we will explore an interpretation of the objective function as a (discretized) path integral. Since the mathematics of physical phenomena ranging from fluid phase transitions to particle collision crossections are representable as a path integral (Feynman and Hibbs (1965); Domb and Green (1976); Zinn-Justin (2002)) exploring this connection allows for deeper insight into the data assimilation problem and for the broader application of the techniques developed herein.

The problem of minimizing the energy of a physical system to uncover its ground state(s) is a centuries old quest, and the interpretation of an objective function as such an energy function has lead to useful statistical approaches to solving the more general optimization problem. In this regard, the Boltzmann distribution serves as the keystone by connecting a wide range of systems through a common mathematical framework, and providing a meaningful context to the statistical optimization algorithms developed for parallel computation described in chapters 6 and 7.

The Bayesian approach that we follow is motivated by the observation that for any real-world application of data assimilation, there is inevitably noise in the measurement and the model is at best only an approximation. As there is no measure on the space of models, one must account for discrepancies between the model prediction and the measured data. One way to do this is to assume that there is noise in the model dynamics, the physical origin of the noise being degrees of freedom in the system that are not accounted for by the model. We address additive noise only.

5.1 The Boltzmann Distribution

In the micro-canonical ensemble of statistical mechanics, the Boltzmann distribution is the probility distribution $\pi_B(\omega)$ on the phase space Ω that maximizes the entropy of an ensemble of states with a given average energy (Pathria (1996)).

With no other restrictions[1], the *a priori* assignment of equal probabilities to all states $\omega \in \Omega$ defines the entropy of a distribution $\pi(\omega)$ as

$$H = -\sum_{\omega \in \Omega} \pi(\omega) \ln \pi(\omega) \quad ,$$

where the summation is performed over all (assumed finite) states ω in the space Ω (Boltzmann (1868); Pauli (2000)). If the logarithm is base two then H is the entropy of Ω measured in bits (Cover and Thomas (2006)).

The maximization of H gives all states a probability $\pi_B(\omega) = N(\Omega)^{-1}$ where $N(\Omega)$ is the cardinality of the states $\{\omega\}$ in Ω. If Ω is the search space that our objective function $\phi(\omega)$ is defined on, then the additional requirement on $\pi_B(\omega)$ that the average objective function be $\langle \phi(\omega) \rangle = \sum_{\omega \in \Omega} \phi(\omega) \pi_B(\omega)$ leads itself to a constrained optimization problem that may be solved via Lagrange multipliers (see for example, Otten and van Ginneken (1989)).

The resultant Lagrange multiplier β appears in the Boltzmann distribution as

$$\pi_B(\omega) = \frac{\exp(-\beta \phi(\omega))}{\sum \omega' \in \Omega \exp(-\beta \phi(\omega'))} \quad , \tag{5.1}$$

and obtains the physical interpretation as (scaled) inverse temperature if ϕ is taken to be an energy.

Once again, it is relevant for our discussion because this analogy between the objective function of a minimization problem and an energy of a classical statistical system is very useful for the understanding of statistical approaches to optimization.

5.2 Bayesian Formulation

For problems with many unknown variables and parameters, or when it is likely that they share complicated dependencies, one naturally looks to the statistics about the solution as a characterization of error. It is possible to use the definition of conditional probability to define such a characterization, which, following the ideas of Lorenc (1986), Eyink et al. (2004),

[1] Of course, since $\pi_B(\omega)$ is a probability distribution we require $0 < \pi_B(\omega) < 1$ and the normalization $\sum_{\omega \in \Omega} \pi_B(\omega) = 1$.

and Apte et al. (2007), ultimately leads to a path integral measure for the problem of data assimilation (Restrepo (2008); Abarbanel (2009)).

5.2.1 Conditional Probability

The definition of conditional probability asserts that the probability of both events A *and* B occuring is given by the probability of event B occuring independently, multiplied by the conditional probability of event A occuring *given that* B occurs. It is written as

$$P(A|B) \cdot P(B) = P(A \cap B) \quad , \tag{5.2}$$

where $P(\cdot)$ denotes probability, $A|B$ represents the conditional, and the union of events is denoted by $\cap$. Bayes' rule follows immediately from the commutivity of the intersection of events; since the joint probabilities are equal, $P(A \cap B) = P(B \cap A)$ it follows that

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \quad ,$$

which is known as Bayes' rule.

It is possible to phrase the problem of data assimilation by asking, what is the probability that the model will be in state $\mathbf{x}(t)$, given that a measurement $\mathbf{y}(t)$ has been made of the N_M measureable dimensions? Expressed as a conditional probability following equation (5.2), this may be written as

$$P(\mathbf{x}(t)|\mathbf{y}(t)) = \frac{P(\mathbf{x}(t) \cap \mathbf{y}(t))}{P(\mathbf{y}(t))} \quad . \tag{5.3}$$

Since it is beneficial to use an entire timeseries of measurements when addressing the above question, consider the series of measurements from 1 to t denoted $\mathbf{y}_t = \{\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(t)\}$, to formulate

$$P(\mathbf{x}(t)|\mathbf{y}_t) = \frac{P(\mathbf{x}(t) \cap \mathbf{y}_t)}{P(\mathbf{y}_t)} \quad , \tag{5.4}$$

which represents the probability that the system is in state $\mathbf{x}$ at time t conditioned upon all of the available measurements $\mathbf{y}_t$.

Recognizing that the set of measurements is seperable,

$$\mathbf{y}_{t-1} \cap \mathbf{y}(t) = \{\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(t)\} \quad ,$$

allows for the explicit re-introduction of the measurement at $\mathbf{y}(t)$ using equation (5.2) again to write

$$P(\mathbf{x}(t) \cap \mathbf{y}_t) = P(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1}) \cdot P(\mathbf{y}_{t-1})$$
$$P(\mathbf{y}_t) = P(\mathbf{y}(t)|\mathbf{y}_{t-1}) \cdot P(\mathbf{y}_{t-1}) \quad ,$$

and substituting this into equation (5.4) results in

$$P(\mathbf{x}(t)|\mathbf{y}_t) = \frac{P(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_t)} \quad .$$

5.2.2 Conditional Mutual Information

An application of unity allows us to express $P(\mathbf{x}(t)|\mathbf{y}_t)$ in terms of the conditional mutual information as

$$P(\mathbf{x}(t)|\mathbf{y}_t) = \Big[\frac{P(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_{t-1}) \cdot P(\mathbf{x}(t)|\mathbf{y}_{t-1})}\Big] P(\mathbf{x}(t)|\mathbf{y}_{t-1}) \quad . \tag{5.5}$$

Where the term in brackets is the exponential of the conditional mutual information (CMI) between the measurements and the model dynamics (Fano (1961)),

$$\mathrm{CMI}(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1}) = \log\Big[\frac{P(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_{t-1}) \cdot P(\mathbf{x}(t)|\mathbf{y}_{t-1})}\Big] \quad . \tag{5.6}$$

Note that this is slighlty different from some definitions of conditional mutual information, such as in Cover and Thomas (2006), in which a summation is performed over all possible states.

5.3 The Marginal Distribution

The normalization of probability over an intermediate distribution, along with the definition of conditional probability, leads to a useful identity known as the Chapman-Kolmogorov equation (see for example Van Kampen (2007)). Expressed in a form that is relevant to the discussion at hand, it is

$$P(\mathbf{x}(t)|\mathbf{y}_{t-1}) = \int \mathrm{d}^{N_x}\mathbf{x}(t-1) \;\; P(\mathbf{x}(t)|\mathbf{x}(t-1)) \cdot P(\mathbf{x}(t-1)|\mathbf{y}_{t-1}) \quad , \tag{5.7}$$

and requires that the transition probabilities in question are independent; i.e. that the dynamics

$$\mathbf{x}(t+1) = \mathbf{F} \circ \mathbf{x}(t)$$

satisfies the Markov condition. In the case of deterministic dynamics, the transition probability for the state to advance in time is given by a delta function on the discretized form of the differential equations,

$$P(\mathbf{x}(t+1)|\mathbf{x}(t)) = \delta^{N_x}(\mathbf{x}(t+1)|F \circ \mathbf{x}(t)) \quad . \tag{5.8}$$

The delta function is relaxed in the following section to account for errors in the model, however for now, note that the knowledge of the dynamics and the Chapman-Kolmogorov equation (5.7),

allows one to take the influence of the measurements on the model at a prior time, $P(\mathbf{x}(t-1)|\mathbf{y}_{t-1})$, and calculate the conditional probability distribution of the model at the current time, $P(\mathbf{x}(t)|\mathbf{y}_{t-1})$. This approach is used by some on-line filters (Judd and Stemler (2009)) to advance an estimate of the model state.

Alternatively, one may combine (5.7) and (5.5) to obtain a recursion relation

$$P(\mathbf{x}(t)|\mathbf{y}_t) = \Big[\frac{P(\mathbf{x}(t)\cap\mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_{t-1})\cdot P(\mathbf{x}(t)|\mathbf{y}_{t-1})}\Big]\int \mathrm{d}^{N_x}\mathbf{x}(t-1)\;\; P(\mathbf{x}(t)|\mathbf{x}(t-1))\cdot P(\mathbf{x}(t-1)|\mathbf{y}_{t-1}) \quad , \tag{5.9}$$

relating the probability that the model is in state $\mathbf{x}(t)$ conditioned upon the measurements $\{\mathbf{y}(1),\ldots,\mathbf{y}(t)\}$, to the same quantity at a previous time, $P(\mathbf{x}(t-1)|\mathbf{y}_{t-1})$.

Recursively applying this equation gives another exponentiated conditional mutual information term as well as an additional integral over a second intermediary time point, which results in a relationship between $P(\mathbf{x}(t)|\mathbf{y}_t)$ and $P(\mathbf{x}(t-2)|\mathbf{y}_{t-2})$. Explicitly, this is

$$\begin{aligned} P(\mathbf{x}(t)|\mathbf{y}_t) = &\Big[\frac{P(\mathbf{x}(t)\cap\mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_{t-1})\cdot P(\mathbf{x}(t)|\mathbf{y}_{t-1})}\Big] \\ &\times\int \mathrm{d}^{N_x}\mathbf{x}(t-1)\mathrm{d}^{N_x}\mathbf{x}(t-2)\;\; \Big[\frac{P(\mathbf{x}(t-1)\cap\mathbf{y}(t-1)|\mathbf{y}_{t-2})}{P(\mathbf{y}(t-1)|\mathbf{y}_{t-2})\cdot P(\mathbf{x}(t-1)|\mathbf{y}_{t-2})}\Big] \\ &\times P(\mathbf{x}(t)|\mathbf{x}(t-1))\cdot P(\mathbf{x}(t-1)|\mathbf{x}(t-2))\cdot P(\mathbf{x}(t-2)|\mathbf{y}_{t-2}) \quad , \end{aligned}$$

where two instances of the discrete dynamics can be seen, advancing $\mathbf{x}(t-2)\to\mathbf{x}(t-1)\to\mathbf{x}(t)$. The e^{CMI} term at $t-1$ cannot escape integration because it involves quantities at an intermediary timestep, unlike the analogous term at t, which is the final time in the series.

Continually applying (5.9) from time $t=N_T$ to time 0 results in

$$P(\mathbf{x}(N_T)|\mathbf{y}_T) = \int\prod_{t=1}^{N_T}\mathrm{d}^{N_x}\mathbf{x}(t-1)\;\; \Big[\frac{P(\mathbf{x}(t)\cap\mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_{t-1})\cdot P(\mathbf{x}(t)|\mathbf{y}_{t-1})}\Big]P(\mathbf{x}(t)|\mathbf{x}(t-1))\cdot P(\mathbf{x}(0)) \quad . \tag{5.10}$$

where the $\mathrm{CMI}(\mathbf{x}(t)\cap\mathbf{y}(t)|\mathbf{y}_{t-1})$ term has been placed inside the integration for compactness. Also note that, since no measurement is made at $\mathbf{y}(0)$, the conditional is dropped from the final term in the sequence to become $P(\mathbf{x}(0)|\mathbf{y}(0))\to P(\mathbf{x}(0))$. This equation (5.10) is the marginal distribution of $\mathbf{x}(t)$ conditioned upon the entire timeseries of measurements, $\mathbf{y}_T = \{\mathbf{y}(0),\mathbf{y}(1),\ldots,\mathbf{y}(N_T)\}$.

If instead, the intermediate distributions are not integrated over, one obtains an expression for the probability distribution of the model state vector during the entire data window conditioned upon the entire timeseries of measurements,

$$P(\mathbf{x}_T|\mathbf{y}_T) = \prod_{t=1}^{N_T}\Big[\frac{P(\mathbf{x}(t)\cap\mathbf{y}(t)|\mathbf{y}_{t-1})}{P(\mathbf{y}(t)|\mathbf{y}_{t-1})\cdot P(\mathbf{x}(t)|\mathbf{y}_{t-1})}\Big]P(\mathbf{x}(t)|\mathbf{x}(t-1))\cdot P(\mathbf{x}(0)) \quad ,$$

where $\mathbf{x}_T = \{\mathbf{x}(0), \ldots, \mathbf{x}(N_T)\}$. Incorporating the conditional mutual information in a explicit way, this equation may equivalently be written as

$$P(\mathbf{x}_t|\mathbf{y}_t) = \prod_{t'=1}^{t} e^{CMI(\mathbf{x}(t')\cap\mathbf{y}(t')|\mathbf{y}_{t'-1})} e^{\log P(\mathbf{x}(t')|\mathbf{x}_{t'-1})} e^{\log P(\mathbf{x}(0))} \quad . \tag{5.11}$$

In the following chapter, this distribution is estimated using a Metropolis-Hastings Monte Carlo algorithm for the case of additive Gaussian noise in both the measurements and model dynamics.

5.4 Approximating the Effective Action

Borrowing from the ideas statistical physics presented in section 5.1 as a means to express the probability distribution of equation (5.11) in a manageable way, and from the work of Restrepo (2008) and Abarbanel (2009), consider an expression for this probability distribution such as

$$P(\mathbf{x}_T|\mathbf{y}_T) \propto e^{-A_0(\mathbf{x}_T,\mathbf{y}_T)} \quad , \tag{5.12}$$

where the normalization factor is not of concern because of the Metropolis Hastings method of generating samples in accordance with the desired distribution; additive constants to A_0 may also be omitted for the same reason.

By comparison with (5.11), we may write

$$A_0(\mathbf{x}_T, \mathbf{y}_T) = -\sum_{t=0}^{N_T} \mathrm{CMI}(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1}) - \sum_{t=0}^{N_T} \log P(\mathbf{x}(t)|\mathbf{x}_{t-1}) \quad , \tag{5.13}$$

where it is assumed that $P(\mathbf{x}(0))$ is a uniform distribution over the state space, and therefore may be dropped as a constant. In the dynamics literature, a function such as A_0 is referred to as the *effective action* of the trajectory $\mathbf{x}_T$ (Hochberg et al. (1999); Abarbanel (2009)).

The first term in the action measures the total conditional mutual information shared between the model state and the measurement, conditioned upon all of the previous measurements until that time point. While the second term measures the conditional probability over the entire path that the dynamics are propagating the model state from one time point to the next.

5.4.1 Additive Noise as Measurement Error

The first term in (5.13) may be approximated to a usable computational form if we assume that any discrepancy between the model state and the measurement is due to noise in the measurement; any error in the model is relegated to the second term. Furthermore, taking

this noise to be a multivariate Gaussian distribution, the conditional probability at a single time point t is

$$P(\mathbf{x}(t)|\mathbf{y}(t)) = (2\pi)^{-M/2}|\mathbb{Q}|^{-1/2}\exp\Big(-\frac{1}{2}\sum_{m=1}^{M}(x_m(t)-y_m(t))^{\dagger}\mathbb{Q}^{-1}(x_m(t)-y_m(t))\Big) \quad ,$$

where $\mathbb{Q}$ is $N_M \times N_M$ dimensional covariance matrix of the noise, and $x^{\dagger}$ represents transpose. The summation is over the number of measurements, N_M, and again the normalization factor will eventually be of little consequence.

Additionally, it is assumed that the measurement noise is uncorrelated; a commonly adopted premise in the laboratory. This way, the conditional probability $P(A|B)$ reduces to $P(A)$ and, by first applying the identity $P(A \cap B|C) = P(A|C) \cdot P(B|A \cap C)$, the conditional mutual information reduces to

$$\mathrm{CMI}(\mathbf{x}(t) \cap \mathbf{y}(t)|\mathbf{y}_{t-1}) \approx \log P(\mathbf{x}(t)|\mathbf{y}(t)) - \log P(\mathbf{y}(t)) \quad . \tag{5.14}$$

Since $P(\mathbf{y}(t))$ does not depend upon the search variables it may be regarded as a constant for the purposes of optimization, and therefore also dropped.

The noise covariance matrix $\mathbb{Q}$ is assumed to be diagonal so that the vector σ_m^2 may be used to denote the variance in the N_M measurements. Depending upon the nature of the measurement noise this may not be an appropriate assumption, however only a single measurement[2] is used for the neuron models that we assimilate and so remains valid.

5.4.2 Additive Noise as Model Error

Relaxing the distribution that connects the dynamical state between time t and $t+1$ in equation (5.8) from a delta function to one with finite variance may be used to incorporate additive noise into the time-evolution of the dynamics. The differential equations (1.1) that describe the dynamics of our model are extended to include noise

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{p}) + \eta(t) \quad , \tag{5.15}$$

where the noise $\eta(t)$ is assumed to be Gaussian-distributed with mean zero, however this restriction may be relaxed (Restrepo (2008)). The noise is also uncorrolated in time, i.e.

$$\langle \eta(t)\eta(t') \rangle = \delta^{N_x}(t,t') \quad ,$$

preserving the assumption of Markovian dynamics.

[2]making σ_m^2 a scalar so that the subscript simply denotes 'measurement noise'

The discretization of the dyamics in time is accomplished with the trapezoidal rule, which is implicit (i.e. time-symmetric) and accurate to second order in the integration timestep dt_{int} (Hairer et al. (2002)). Discretizing equation (5.15) using this rule gives

$$x_i(t+1) = x_i(t) + \frac{dt}{2}\Big(f_i(x_i(t)) + f_i(x_i(t+1))\Big) + \frac{dt}{2}\Big(\eta_i(t) + \eta_i(t+1)\Big) \quad ,$$

Separating out the noise-free dynamical map, and abreviating it as TRP allows us to write

$$\begin{aligned} \mathrm{TRP} &= x_i(t+1) - x_i(t) - \frac{dt}{2}\Big(f_i(x_i(t)) + f_i(x_i(t+1))\Big) \\ \mathrm{TRP} &= \frac{dt}{2}\Big(\eta_i(t) + \eta_i(t+1)\Big) \quad , \end{aligned} \tag{5.16}$$

which appears in the approximation below.

As the deterministic dynamics must approach a delta function transition probability in the noise-free limit as given by equation (5.8), we may again approximate the transition probability as the limit of a multivariate Gaussian distribution over the dynamical equations. This is by no means the only way to achieve a delta function in the limit, however it mirrors the discussion concerning the conditional mutual information term, and again allows for the error in the model to be interpreted as noise, this time in the dynamics. The single time transition probability is then given by

$$P(\mathbf{x}(t+1)|\mathbf{x}(t)) = \\ (2\pi)^{-N_x/2}|\mathbb{Q}_f|^{-1/2}\exp\Big(-\frac{1}{2}\sum_{i=1}^{N_x}\mathrm{TRP}^{\dagger}(x_i(t),x_i(t+1))\mathbb{Q}_f^{-1}\,\mathrm{TRP}(x_i(t),x_i(t+1))\Big) \tag{5.17}$$

where $\mathrm{TRP}(x_i(t), x_i(t+1))$ is given by equation (5.16). The different dimensions of model error are assumed to be independent, resulting in a covariance matrix of the dynamical noise $\mathbb{Q}_f$ that is diagonal, and given by the vector of model error variance σ_f^2. Typically, the different dimensions of σ_f^2 are taken to be uniform for convenience.

5.4.3 Computational Form of the Action

Combining the approximations from equations (5.14) and (5.17), allows us to write the action equation (5.13) as the discretized sum,

$$A_0(\mathbf{x}_{T-1},\mathbf{y}_{T-1}) = \frac{1}{2N_T\sigma_m^2}\sum_{t=0}^{N_T-1}\sum_{i=1}^{M}\Big(x_i(t)-y_i(t)\Big)^2 + \frac{1}{2N_T\sigma_f^2}\sum_{t=0}^{N_T-2}\sum_{i=1}^{N_x}\Big(\mathrm{TRP}(x_i(t),x_i(t+1))\Big)^2 \quad . \tag{5.18}$$

This is the form that is used for calculations [3].

The value of N_T^{-1} is used to make the action intrinsic; i.e., independent of the length of the data assimilation window. This is required because the noise estimates, e.g. of the measurements σ_m^2, are obtained with respect to a single dt_{obs}, while equation (5.18) is a value with respect to the entire path.

Although different discretization methods (such as the one used in equation (5.16)) lead to different explicit formulations of the action and subsequently to the form of the path integral, it is presumed that in the appropriate limits, $dt \to 0$ and $N_T \to \infty$ such different approaches converge. In these continuum limits, the discrete summation over finite states may be formally recognized as a functional integral over the dynamics (Feynman and Hibbs (1965); Jouvet and Phythian (1979)).

5.5 Lattice Field Theory Analogy

Following the above discussion, and noting the similarities between the behavior of the synchronization error in the long time $N_T \to \infty$ limit and critical phenomena (see section 3.3 and figure 3.1), it is reasonable to look for the analogous configuration function of a classical lattice gas, and inquire about the existence of a critical phase transition at $\gamma = \gamma_c$.

Consider the analogous statistical system of a one dimensional lattice of gas particles at regularly spaced intervals (by an amount dt_{obs}), where this one dimension is the time variable of the dynamics. At each point on the lattice, the nearest-neighbor interactions between them is given by the dynamics enforced as a constraint, $\mathbf{x}(t+1) = \mathbf{F} \circ \mathbf{x}(t)$. A chaotic synchronization term in the dynamics utilizing the control parameter γ is analogous to a local field potential $h(\mathbf{t}) = \gamma \cdot (\mathbf{y}(t) - \mathbf{x}(t))$, which 'aligns' the dynamics to the data at every time point, and is measured by the synchronization error as the order parameter with $N_T \to \infty$ as a type of 'thermodynamic' limit on the required size of the system for the emergence of criticality.

In a typical lattice gas it is the interaction potential among the constituent elements that gives rise to critical phenomema, whereas in the synchronization system it is not the inter-element interactions (the dynamics) that are of interest, but the coupling to the external 'field' (i.e., the data $\mathbf{y}(t)$) from which the interesting behavior seen in section 3.3 arises; however in

[3] As a technical note, we point out that since the second term connects two locations along the path, the summation involves $N_T - 1$ terms. For the benefit of future predictions, it would perhaps be advantageous to use $\sum_{t=1}^{N_T-1} \mathrm{TRP}^2(x_i(t-1), x_i(t))$, however (5.18) accurately represents the computation that is performed, with future predictions originating from the state values at $\mathbf{x}(N_T - 2)$.

addition all of the elements in a timeseries of the dynamical system remain strongly coupled to their nearest-neighbors.

Note also that although the analogous lattice gas extends spatially in one dimension, the interactions between adjacent particles is multi-dimensional; specifically the N_x-dimensions of the dynamical model. Many systems, for example the $K > 5$ Lorenz '96 system (section 4.2.1), involve multiple dimensions of data as well. The high dimensionality of these interactions make the existence of some type of critical behaviour plausible, however it also makes this type of problem difficult to tackle analytically. We refer the interested reader to Domb and Green (1972, 1976) and Zinn-Justin (2002).

Chapter 6

Single Markov Chain Monte Carlo

In the previous chapter it is shown that if one attributes the Markov quality to the dynamics, as is also implicitly done in collocation approaches to optimization, then the problem of data assimilation can be formulated as a discrete time path integral. Additionally, one can incorporate both measurement and model error, in the form of additive gaussian noise, into the dynamics and data to achieve an approximation to this integral (Hochberg et al. (1999); Restrepo (2008)). Following this interpretation of error, the variation in the measured values of an experiment may be directly utilized to scale the fluctuations in the Markov chain through the measurement noise coefficient. In this chapter, the details of how one may use Monte Carlo methods to calculate an estimation of this integral are given.

The first piece of analysis software written by the author is a Metropolis-Hastings Markov chain simulation, implemented in parallel on a single graphics processor (GPU), to sample the distribution of state variables and parameters at a given ratio of observational error to model error.

At the end of the simulation, the estimated distribution of state variable –and therefore, also noise– information at all times $\{\mathbf{x}^*_{T-1}\}$, as well as the distribution of model parameters $\{\mathbf{p}\}$ is obtained. From these distributions, one may calculate quantities of interest as expectation values over the distribution as well as propagate the model forward to generate a prediction of the future state of the assimilated data-model system.

Twin experiment results on neuron models using this software are presented in chapter 9, after the neuron models have been described in chapter 8. Results on physical neurons from the birdsong system are given in chapter 10.

6.1 Random Search via Markov Chain

The entire process is iterative, and each iteration consists of two basic steps. Given a current point in the seach space, ω, a random move from this current point is proposed and a Metropolis-Hastings acceptance function is used as a decision process to determine if the move is to be executed or ignored. The sequence of current points comprises a *Markov chain* if the transition probability is symmetric. This guarantees that sampling from the chain appropriately estimates the desired distribution.

Let us point out that the Markov condition is used in the previous chapter as a qualification on the dynamical map advancing the dynamical state $\mathbf{x}(t)$ forward in time, while it is used here as a requirement on the sequence of candidate points $\tilde{\omega}^{(k)}$ as they are moved throughout the search space, from one iteration to the next $k \to k+1$.

6.1.1 The Search Space

The total search space for this formulation consists of the state variables at each point in time and the parameters, resulting in $\Omega = \{\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(N_T - 1), \mathbf{p}\}$, where $\Omega \subseteq \mathbb{R}^W$ and has dimension $W = N_x N_T + N_P$. This exacerbates the high dimensional sampling difficulties discussed in section 2.2.1 and 2.2.2 relative to $\Omega = \{\mathbf{x}(0), \mathbf{p}\}$. For this reason the calculation begins with a trajectory given by an optimization routine such as the one described in the following chapter, or by another method, and provides the distribution about the region of the search space that is *local* to the point ω provided.

Each timepoint of the statevector $\mathbf{x}(t)$ is given an independent *thread*, or core-process on the GPU. The individual N_x dimensions of this vector are kept together, on the same thread, so that they may be accessed more efficiently during the evaluation of the model equations.

6.1.2 The Objective Function

The 'objective function' here is essentially the action of the path integral given by (5.18), with the independent variable extended to include the full search space Ω including the model parameters,

$$\phi(\omega, \mathbf{y}_{T-1}) = \frac{1}{2N_T\sigma_m^2} \sum_{t=0}^{N_T-1} \sum_{i=1}^{M} \Big(x_i(t) - y_i(t)\Big)^2 + \frac{1}{2N_T\sigma_f^2} \sum_{t=0}^{N_T-2} \sum_{i=1}^{N_x} \Big(\mathrm{TRP}(x_i(t), x_i(t+1))\Big)^2 \,. \tag{6.1}$$

The problem here is not to minimize this function, but to evaluate it for the sequence of $\{\omega^{(k)}\}$ that constitute the Markov chain, once enough iterations have allowed this chain to reach an

equilibrium state. The distribution (5.12) being estimated by this calculation is then

$$P(\mathbf{x}_T|\mathbf{y}_T) \propto e^{-\phi(\omega,\mathbf{y}_T)} \quad . \tag{6.2}$$

Evaluating the action is the most time consuming step of the algorithm; with the calculation of the model dynamics contributing the most to the objective function. This step is implemented in parallel, so that $\mathbf{x}(t), \mathbf{x}(t+1)$ pairs of this function are evaluated (by the TRP function), and this portion of the total objective is stored. A second function is then called to sum the pieces together and obtain a value for the action over the entire path.

In this way a value for the *entire* path is obtained before being subjected to the Metropolis-Hastings acceptance function, this is distinct from other algorithms which calculate a change in the action between adjacent pairs of points in the timeseries such as in Quinn and Abarbanel (2011).

6.1.3 The Selection Function

The first important aspect to the generation of the Markov chain is the selection function, which defines how one chooses a candidate point. Formally, this amounts to a probability measure S over pairs of points in the search space,

$$\sum_{(\omega,\tilde{\omega})\in\Omega} S(\omega,\tilde{\omega}) = 1 \quad .$$

In order for the observations of Metropolis *et al.* to be valid, the samples must be generated by a Markov chain which has a reversible selection function; that is

$$S(\omega,\tilde{\omega}) = S(\tilde{\omega},\omega) \quad . \tag{6.3}$$

Where ω is the current point in the chain and $\tilde{\omega}$ is a proposed candidate point. This property of the selection function ensures that the Markov chain is *reflexive*, or k-reversible over the iterations which index the elements of the chain. Although this is not specifically required for a Markov chain calculation, it greatly simplifies the statistics provided by the chain (see for example Otten and van Ginneken (1989)).

We note that the reflexive property is often violated in attempts to locate points that contribute greatly to the distribution, such as during the burn-in phase, however it is always enforced during the gathering of samples that are used in statistics.

This function is implemented in parallel so that all N_T elements of the state-vector timeseries are perturbed simultaneously. Similarly, all dimensions of the parameter vector are also perturbed in parallel, however via a separate function call. The geometry of the graphical processor memory on which this occurs is used most efficiently when N_T is a power of 2.

6.1.4 Scaling the Stepsize α

For the implementation of MCMC used here, candidate paths are perturbed using

$$\tilde{\omega} = \omega + U(-\alpha/2, \alpha/2) \cdot (\omega_u - \omega_l) \quad ,$$

where U is a uniformly distributed random number in $[-\alpha/2, \alpha/2]$. The perturbation stepsize, α, must remain constant between subsequent iterations to satisfy the reflexive property of the transition probability.

The stepsize is the scaled quantity, $\alpha \in (0, 1)$ with ω entering the dynamical equations. All quantities are vectoral, and have the same dimensions as ω ($N_x N_T + N_p$ in this case), however the dimensions of α_i are not individually scaled. This means that resolution between different variables is uniform (with respect to their individual bounds, ω_l and ω_u), relegating the inter-variable scaling to the user who determines the variable bounds.

The equilibrium density of the system $\hat{\pi}(\omega)$ is defined to be stationary, which means that continued perturbations of the equilibrium distribution are again the equilibrium distribution. This distribution is therefore independent of the selection probability $S(\omega, \tilde{\omega})$ of candidate points. This means that the perturbation stepsize α can be given by any rule, provided that it remains constant in between successive perturbations so that equation (6.3) remains valid.

6.2 Metropolis-Hastings Algorithm

One can approximate the distribution (6.2) by randomly sampling the space (as is done via the random walking of an *unbiased* Markov chain) and then performing a weighted average over the samples, weighting the contribution from ω with $\exp(-\beta\phi(\omega))$, when computing quantities of interest such as expectation values.

Metropolis et al. (1953) pointed out that it is more computationally efficient, and numerically equivalent, to instead generate samples from the desired distribution and then weigh them uniformly when performing averages. This is more efficient because the portion of the solution space that contributes significantly to terms in these summations is quite small, and so a uniform random sampling will generate many elements that are irrelevant to the final value.

The user chooses an initial point in the search space $\omega^{(0)}$, and their bounds ω_l, ω_u, as well as the values of σ_m^2 and σ_f^2 that enter into ϕ. The algorithm then proceeds according to the steps in figure 6.1.

Metropolis Hastings Monte Carlo Algorithm

$k \leftarrow 0$

evaluate $\phi(\omega^0)$

do {

$\tilde{\omega} \leftarrow \omega^{(k)} + U(-\alpha/2, \alpha/2) \cdot (\omega_u - \omega_l)$

evaluate $\phi(\tilde{\omega})$

If $U(0,1) < \exp[-(\phi(\tilde{\omega}) - \phi(\omega^{(k)})]$ then $\omega^{(k+1)} \leftarrow \tilde{\omega}$

else $\omega^{(k+1)} \leftarrow \omega^{(k)}$

$k \leftarrow k + 1$

} while $k <$ MAX

Figure 6.1: The Metropolis-Hastings Monte Carlo Algorithm. Parallelization of both the perturbation step and the evaluation of ϕ allows for all time points of the dynamical path to be calculated simulatenously.

6.2.1 The Acceptance Function

The method proposed in Metropolis et al. (1953) and generalized in Hastings (1970) to generate a desired distribution is to bias the Markov chain with an acceptance function. As the Markov chain evolves (over iterations), this function is a rule that takes a trial move and accepts this candidate point $\tilde{\omega}$ with a probability given by

$$P_{acc} = \min\Big(1, e^{-\beta(\phi(\tilde{\omega})-\phi(\omega))}\Big) \quad . \tag{6.4}$$

If the candidate point $\tilde{\omega}$ has a lower penalty than the current point, i.e. if $\phi(\tilde{\omega}) < \phi(\omega)$, then this move is accepted with probability 1. However, if the proposed move increases the objective function, it may still be accepted but only with probability $\exp\Big(-\beta(\phi(\tilde{\omega}) - \phi(\omega))\Big) < 1$. In the limit that $\beta \to \infty$, the Markov chain is 'frozen' and only moves that reduce ϕ are accepted; we use $\beta = 1$ because the equivalent scaling is set by the noise in the measurement as described in chapter 5. The Markov chain explores the basin about the local minimum via random walk, while the Metropolis-Hastings acceptance function maintains the appropriate distribution (6.2).

A 'healthy' Markov chain, from a computational standpoint, is one where P_{acc} is far from both 0 and 1 (although not necessarily 1/2). For if P_{acc} is near zero, no new paths are

being sampled implying that the chain is not moving. While if it is near unity, it is often because the stepsize is so small that every move is accepted, again resulting in a chain that does not adequately sample the entire search space.

The relationship between the perturbation stepsize α and the average acceptance rate is generally

$$\begin{aligned} &\text{as } \alpha \to 0, && \langle P_{acc} \rangle \to 1 \\ &\text{as } \alpha \to 1, && \langle P_{acc} \rangle \to 0 \end{aligned}$$

however this may be complicated at the extremes due to numerical issues as well as the dependence of $\langle P_{acc} \rangle$ on β.

6.2.2 Sampling and Boundaries

The iterations of the Monte Carlo simulation are adjusted so that the parameters and states are sampled more uniformly, in the form of perturbation updates to the Markov chain. The parameter vector is perturbed and the changes evaluated by the Metropolis-Hastings acceptance criterion $N_x N_T / N_p$ times for every perturbation to the state vector timeseries; this distributes the sampling roughly equally between the entire timeseries $\mathbf{x}_T$ and the vector of parameters $\mathbf{p}$. This recognizes that the parameter distribution is generally of more interest, and that we do not want to under-sample it as a consequence of using a long-time data series.

The perturbation of $\mathbf{p}$ and $\mathbf{x}_T$ is accomplished by separate function calls, and both of these perturbation steps are performed in parallel, updating each dimension of $\mathbf{p}$ (or all time-points of $\mathbf{x}_T$) simultaneously.

The stochastic nature of the perturbation to the search variable ω means that it is possible to perturb a valid point across a boundary. This must be remedied, and is accomplished with pseudo-absorbing boundaries that place the perturbed variable at the boundary. Ideally, one would make the bounds large enough that they are rarely or never encountered, however in practice this cannot be realized. Caution must be taken if one reflects the perturbation in the boundary, because values are prevented from approaching the bounds arbitrarily closely without suffering from a repulsive 'force'. This may cause the appearance of a valid distribution within the bounds, when in fact the Markov chain is being biased towards the boundary.

6.2.3 Calculating Moments of the Distribution

At the termination of the Monte Carlo algorithm, one obtains a sample of points $\{\omega_i\}_{i=1}^{N_E}$ in the search space over paths and parameters where each point is $\omega = \{\mathbf{x}(0), \ldots, \mathbf{x}(N_T - 1), \mathbf{p}\}$,

and the entire sample approximates the distribution (5.12) according to

$$P(\mathbf{x}_T|\mathbf{y}_T) \approx \frac{1}{N_E}\sum_{i=1}^{N_E}\delta(\omega-\omega_i) \quad ,$$

where $\omega_i \in \{\omega_i\}_{i=1}^{N_E}$, and the dependence of the probability distribution on the parameters $\mathbf{p}$ enters through the dynamics in the model error terms of the action.

Correspondingly, expectation values of any function $\mathcal{F}(\omega)$ over the search space may be estimated using this sample set as

$$\begin{aligned}\langle\mathcal{F}(\omega)\rangle &= \int d\Omega\ \mathcal{F}(\omega)P(\mathbf{x}_T|\mathbf{y}_T) \\ &\approx \frac{1}{N_E}\sum_{i=1}^{N_E}\mathcal{F}(\omega_i) \quad . \end{aligned} \tag{6.5}$$

This is used to generate the averages and standard deviations of the parameter estimates that are reported.

Chapter 7

Stochastic Optimization with Synchronization Fiducial

In this chapter we present an optimization algorithm that relies upon synchronization to regularize chaotic instabilities, to guide the model to a solution commesurate with the data and to provide the search with approximate information of the unmeasured state variables of the model through a fiducial trajectory. The search step is provided by an ensemble of Markov chains propagating through the search space following a Metropolis-Hastings biased random walk.

The question of model error is not addressed by this formulation, which satisfies the equations of motion as delta-function equality constraints through explicit integration of initial conditions. The restriction of the search space to initial conditions and parameters $\omega = \{\mathbf{x}(0), \mathbf{p}\}$ allows each search iteration to be carried over a proper ensemble of N_E points, calculated in parallel on a graphics processor. The manor in which synchronization is used allows for the regularization of chaotic instabilities without the need to collocate the dynamical variables in time.

We demonstrate the success of this algorithm on a twin experiment using the Lorenz '63 model.

7.1 General Framework

The key computational advantage exploited by this algorithm is to take advantage of of today's modern parallel processing hardware by using a non-interacting ensemble of Markov chains to enhance the statistics capability for larger problems. For this reason, and to distinguish

it from the algorithm in the previous chapter, we refer to it as multiple Markov chain Monte Carlo (MMC). At termination the simulation provides a distribution of final states of the model at $\{\mathbf{x}^*(N_T-1)\}$ and parameters $\{p^*\}$ that can be used for subsequent forward prediction or analysis.

The strategy employed here is to solve a series of subproblems, where we create a strongly coupled fiducial trajectory (thus providing an approximation to all N_x dimensions of 'data') and search for a point that can recreate this trajectory with a lesser value of the coupling, thereby driving the coupling to its critical value.

We conjecture that, for a given coupling, all candidate points with a synchronization error (as measured from the fiducial) that is less than the fiducial's self-synchronization error (section 7.2.2) are equally likely to be an improvement over the current solution. This is reflected in the distribution generated by the Metropolis-Hastings Monte Carlo techniques that are used to solve the individual subproblems (7.5).

7.1.1 Comparison with the Marginal Distribution

The structure of this computation could loosely be seen as an estimation of the marginal distribution (5.10), as the solution depends on synchronization with the entire data timeseries $\mathbf{y}_T$, i.e. not just the data initial conditions $\mathbf{y}(0)$.

The Chapman-Kolmogorov integral over the intermediary probability distributions is collapsed to a series of delta functions, as there is no model error incorporated into this version of the calculation. Dispite the similarities due to a noise-free integral over the path, we avoid making a direct comparision to the distribution in (5.10).

7.1.2 Problem Formulation

The search is made over the space of initial conditions and parameters, $\omega = \{\mathbf{x}(0), \mathbf{p}\}$, however in practice the state variables at the end of the data assimilation window $\mathbf{x}(N_T-1)$ are of more interest, because they may be used to predict the system forward in time. They are also much easier to obtain and are outputted for exactly this purpose, however the situation we engendered is naturally phrased as an initial value problem. Formally, then, we are solving

$$\text{minimize} \quad \mu^2(\omega, \mathbf{y}_{T-1}, \gamma_c) = \frac{1}{N_T N_m} \sum_{t=0}^{N_T-1} \sum_{i=0}^{N_m} a_i(t)^2 (x_i(t) - y_i(t))^2 \quad , \tag{7.1}$$

where the dependence of the objective function on the variables ω is done through the generation of a dynamical trajectory synchronized with the data (from (3.5)),

$$\mathbf{x}_T = \text{sync}(\omega, \mathbf{y}_T, \gamma \rightarrow \gamma_c) \quad ,$$

and where γ_c is the critical value of the coupling (see section 3.3), which may be considered $\gamma_c \geq 0$ but otherwise unknown at the outset. The scaling of this objective function $\mathbf{a}(t)^2$ is discussed in detail in section 7.4.1.

The model equations are therefore treated as equality constraints between t and $t+1$ that are satisfied through the forward integration. Since there are unmeasured dimensions to the model, we cannot use the synchronization error itself as an objective function, but instead will approximate (7.1) by relying upon the synchronized fiducial as the origin of our measure. This is done so that γ_c is also acquired during the optimization.

7.2 The Synchronization Fiducial

The synchronization fiducial is a trajectory that is strongly coupled to the measured data. It serves two purposes: it functions as a reasonable guess of the unmeasured dimensions of the data, and it is used to calculate a reference value for the self-synchronization error (7.2) at reduced coupling strengths.

More precisely, it provides an approximation to the unmeasured dimensions of the proposed model while that model remains consistent with the data, such that $\mathbf{x}_m(t) \approx \mathbf{y}_m(t)$, which is the known effect of chaotic synchronization (Afraimovich et al. (1983); Fujisaka and Yamada (1983); Pecora and Carroll (1990)).

7.2.1 Creating a Fiducial

To create the synchronized fiducial trajectory $\mathbf{z}_T$, the model is synchronized with the data using strong coupling, $\gamma_{fid} >> \gamma_c$, where the initial conditions and parameters $\omega = \{\mathbf{x}(0), \mathbf{p}\}$ are provided by the current 'best' guess, denoted ω_{fid}, or by the user at the beginning of the optimization procedure. We write this as

$$\mathbf{z}_T = \text{sync}(\omega_{fid}, \mathbf{y}_T, \gamma_{fid}) \quad .$$

The stronger the coupling the more closely the fiducial will reproduce the data trajectory, keeping in mind that the $N_T \rightarrow \infty$ limit is formally required to make this statement due to finite-time transients.

7.2.2 Self-Synchronization Error Metric

If one uses the same point that created the fiducial, ω_{fid}, and instead of coupling it strongly to the data, couples it *weakly*, how much does the trajectory change? Derivative quantities with respect to γ are likely to be unstable following the same arguments as in section 2.3.4, so one should have in mind finite (and possibly large) differences between γ and γ_{fid} (while maintaining $\gamma < \gamma_{fid}$).

In practice we only evaluate the self-synchronization error for $\omega_{fid}^{(k)}$, however the idea here is to compare two trajectories generated from the same search variable $\omega = \{\mathbf{x}(0), \mathbf{p}\}$; both are coupled to the data, but with different coupling strengths. Let us define the self-synchronization error of the point ω as ϕ_0, where

$$\phi_0^2(\omega, \mathbf{y}_{T-1}, \gamma) \equiv \lim_{N_T \to \infty} \frac{1}{N_T N_x} \sum_{t=0}^{N_T-1} \sum_{i=0}^{N_x} (\tilde{z}_i(t) - z_i(t))^2 \quad , \tag{7.2}$$

and the dependence on γ and the data $y_m(t)$ is again provided through the trajectories

$$\begin{aligned} \mathbf{z}_T &= \text{sync}(\omega, \mathbf{y}_T, \gamma_{fid}) \\ \tilde{\mathbf{z}}_T &= \text{sync}(\omega, \mathbf{y}_T, \gamma) \quad \gamma < \gamma_{fid} \quad . \end{aligned}$$

The key difference between this and the synchronization error (equation (3.3)) is that the comparison trajectory is provided by a fiducial generated from the same ω; one can also see the similarities between this and the concept of generalized synchronization (Rulkov et al. (1995); Abarbanel et al. (1996)).

It should be obvious that if the self-synchronization error is evaluated with $\gamma = \gamma_{fid}$ the trajectories will be identical and $\phi_0 = 0$ for all ω, this has important consequences for the choice of the relevant sampling distribution in section 7.3.5.

7.2.3 Choice of the Fiducial Coupling, γ_{fid}

The coupling that is used to create the fiducial γ_{fid} must be strong enough to ensure that $\mathbf{z}_m(t) \approx \mathbf{y}_m(t)$ with any transients lasting for a small fraction of the data assimilation window. Numerical instabilities may occur with an integration timestep of dt_{int} if the product $dt_{int}\gamma_{fid} > 1$ due to large derivatives during integration.

Based upon our studies it is also important that γ_{fid} remain finite. Meaning that the infinite coupling limit, where $z_m(t)$ is replaced by $y_m(t)$ (see section 4.1.2), does not provide as useful metric information as a finite γ.

7.3 Solving The Subproblem

The optimization algorithm, as detailed in figure 7.2 and 7.3, consists of a series of simulated-annealing subproblems indexed by k, during which the fiducial $\mathbf{z}_T$ is held constant. Each subproblem attempts to find an improvement to $\omega_{fid}^{(k)}$ from the ensemble of candidate points $\{\tilde{\omega}_i\}_{i=1}^{N_E}$ at the current value of the coupling, $\gamma^{(k)}$. The purpose of the coupling $\gamma^{(k)}$ is to regulate the search surface, allowing for a meaningful assessment of the candidate point.

The idea is to compare the synchronized trajectory $\mathbf{x}_T$ of a candidate point $\tilde{\omega}$ with that of the synchronization fiducial, $\mathbf{z}_T$. The candidate trajectory is synchronized to the data with a smaller coupling strength $\gamma^{(k)}$ than was used to create the fiducial, i.e. $\gamma_{min} < \gamma^{(k)} < \gamma_{fid}$. If we can find a point $\tilde{\omega}$ that closely reproduces the fiducial trajectory, yet does so with a smaller coupling, then $\tilde{\omega}$ is an improvement over ω_{fid}, and will be accepted according to the Metropolis-Hastings criterion.

At the beginning of the algorithm, the user specifies the following:

$\omega_0 = \{\mathbf{x}(0), \mathbf{p}\}$	The user's initial guess for ω_{fid}.
ω_l, ω_u	Lower and upper bounds on all search variables.
$\gamma_{min}, \gamma_{max} \geq 0$	Lower and upper bounds on the coupling γ.
β_0, β_∞	Numerical limits to $\beta \in (0, \infty)$.
$\langle P_{accept} \rangle_{target}$	A target mean acceptance rate $P \in (0, 1)$.
$\mathfrak{F}$	The discretized coupled model equations of motion.
$\mathbf{y}_{T-1}$	The timeseries of measurements.

In what follows are some of the details specific to this implementation of Markov chain Monte Carlo, which primarily involves adequate estimation of the inverse generalized-temperature, β, the perturbation stepsize, α, and the proper scaling $a_i(t)^2$.

7.3.1 Subproblem Statistics via MCMC

The individual subproblems are solved using a Markov chain Monte Carlo (MCMC) procedure as outlined in the previous chapter. The statistics are gathered over a distribution of candidate solutions during a constant fiducial subproblem, after which ω_{fid} is updated with the expectation value of this distribution, i.e. the maximum likelyhood solution to the minimization problem (7.1) is used as the initial guess of the subsequent subproblem,

$$\omega_{fid}^{(k+1)} = \langle \pi(\{\omega\}^{(k)}) \rangle \quad , \tag{7.3}$$

where we denote the ensemble of search points as $\{\omega\}$.

During the kth subproblem, many Metropolis-Hastings updates may be performed for every chain in the ensemble as indexed by the l iterations shown in figure 7.1. This loop is

terminated by an equilibrium condition that may be difficult to satisfy in a meaningful way for very high dimensional systems and so contains a maximum iteration limit. For the numerical example detailed in section 7.5, quite the opposite occurrs and so a minimum number of $2W+1$ iterations is also enforced ($W = N_x + N_p$).

The Metropolis Hastings (MH) Acceptance Subroutine

$l \leftarrow 0$

do {

$\{\tilde{\omega}\} \leftarrow \{\omega\} + U(-\alpha/2, \alpha/2) \cdot (\omega_u - \omega_l)$

$\{\mathbf{x}(t)\} = \text{sync}(\{\tilde{\omega}\}, \mathbf{y}_{T-1}, \gamma^{(k)})$

$\{\phi(\tilde{\omega})^2\} = N_T^{-1} N_x^{-1} \sum_{t=0}^{N_T-1} \sum_{i=1}^{N_x} a_i(t)^2 (\{x_i(t)\} - z_i(t))^2$

$\forall_{\tilde{\omega}_i \in \{\tilde{\omega}\}}$ If $U(0,1) < \exp[-\beta^{(k)}(\phi(\tilde{\omega}_i) - \phi_0)]$ then $\omega_i \leftarrow \tilde{\omega}_i$

$l \leftarrow l + 1$; if $l > \text{MAX}_l$ then *break*;

} while $|\frac{d\langle\phi\rangle}{dk}| < \kappa\sigma(\{\phi\})$ AND $l \geq 2W + 1$

Figure 7.1: The Metropolis-Hastings Monte Carlo subroutine for use with the synchronized fiducial optimization scheme. Notice the perturbation via the fixed stepsize α and the form of the distribution. The quasi-equilibrium condition in the while loop signifies the acquisition of $\{\omega\}$ after at least $2W + 1$ iterations; $\sigma(\{\phi\}) = \sqrt{\text{var}(\{\phi\})}$.

7.3.2 Main Subproblem Algorithm

No formal annealing (increase in β) is implemented in the Markov chain loop, it simply acquires the distribution (7.5) over the ensemble $\{\tilde{\omega}\}$. However this subroutine is performed twice: at a high temperature ($\beta_0 \approx 0$) where an equilibrium is very easy to acquire, and then again at β_∞, where equilibrium is more difficult to obtain and where the chain is likely to suffer from $P_{acc} \to 0$ as the acceptance criterion can not be satisfied. To avoid this, the chain is then warmed-down ($\beta^{(k)} : \beta_\infty \to \beta_0$) until the user supplied target acceptance rate is achieved. Once the chain achieves equilibrium statistics are taken over the ensemble and the location of the optimal point is updated for the next iteration according to (7.3). The value of $\beta^{(k)}$ obtained is

used as the high temperature value for the subsequent iteration. The steps described here are presented in the overall subproblem algorithm of figure 7.3.

The Multiple-Markov Chain Initialization

$\omega_{fid} \leftarrow \omega_0$

$\{\omega\}^{(0)} \leftarrow U(\omega_l, \omega_u)$

$\{\tilde{\omega}\}^{(0)} \leftarrow U(\omega_l, \omega_u)$

$\gamma_{fid} \leftarrow \gamma_{max}$

$\gamma^{(0)} \leftarrow \gamma_{min}$

$\beta^{(0)} \leftarrow \beta_0 \approx 0$ (A high temperature β)

$k = 0$

Figure 7.2: The initialization of the synchronized fiducial multiple-Markov chain Monte Carlo algorithm. Note that the distribution is started with a small $\beta \approx 0$.

7.3.3 The Subproblem Objective Function

Although the search space only includes the initial conditions and parameters of the model, in order for the data to be relevant to the assimilation procedure a long-time trajectory of the dynamical system must enter into the objective function. This is done by integrating forward from the initial conditions $\mathbf{x}(0)$ until the end of the data assimilation window, using the parameters $\mathbf{p}$ and the coupling value of the k-th iteration, $\gamma^{(k)}$, obtaining

$$\mathbf{x}_T = \text{sync}(\omega, \mathbf{y}_T, \gamma^{(k)}) \qquad \gamma^{(k)} < \gamma_{fid}$$

Once the trajectory $\mathbf{x}_T$ is generated in this way from the search variable $\omega = \{\mathbf{x}(0), \mathbf{p}\}$, the subproblem objective function

$$\phi^2(\omega, \gamma^{(k)}) = \frac{1}{N_T N_x} \sum_{t=0}^{N_T-1} \sum_{i=1}^{N_x} a_i(t)^2 (x_i(t) - z_i(t))^2 \quad , \tag{7.4}$$

with the fiducial

$$\mathbf{z}_T = \text{sync}(\omega_{fid}, \mathbf{y}_T, \gamma_{fid}) \quad ,$$

is evaluated. This is essentially the same as the self-synchronization error (equation 7.2), except that now we allow it to measure the separation between any candidate point $\tilde{\omega}$ and the fiducial

$\mathbf{z}_T$. In other words, the self-synchronization error ϕ_0 is a special case of the subproblem objective function with $\omega = \omega_{fid}$. This allows for a direct comparison to be made between $\tilde{\omega}$ and ω_{fid} at $\gamma^{(k)}$, ascertaining the ability of a candidate point $\tilde{\omega}$ to reproduce the fiducial trajectory with $\gamma^{(k)} < \gamma_{fid}$.

Let us clearly point out that trajectories are only ever coupled to the actual data $\mathbf{y}_T$, the fiducial is not used to provide additional 'data' dimensions for the sake of synchronization; it is only ever used for additional metrical information via the objective function.

7.3.4 Determining the Subproblem Coupling, $\gamma^{(k)}$

It is crucial to determine a proper coupling $\gamma^{(k)}$ to use for evaluation of the synchronization-fiducial objective function during the k-th iteration of the subproblem. Recall from section 7.1 that we wish to drive $\gamma \to \gamma_c$ but are prevented from doing so by the need to regulate instabilities in the search surface. Choosing $\gamma^{(k)}$ too large, i.e. very close to γ_{fid}, can be problematic because the use of a synchronization fiducial creates an artificial 'solution' to the minimzation problem when $\gamma^{(k)}$ is large. To see this, simply let $\gamma^{(k)} = \gamma_{fid}$, then the point which minimizes equation (7.2) is clearly ω_{fid}. For this reason, the distribution used to estimate the subproblem solution is chosen with some care as described in the following subsection.

Alternatively, using a coupling that is close to γ_c can be detrimental when the current guess ω_{fid} is far from the solution ω^*. This is because the irregularities in the synchronization manifold begin to manifest as $\gamma \to \gamma_c$. One should think of γ_c as being the coupling that regularizes the manifold between the *single* trajectory defined by ω^* and the data, and that as one increases $\gamma \to \infty$ it regularizes an increasingly large portion of Ω about ω^*. This behavior is described in chapter 4, and illustrated in figure 4.2.

As $\gamma \to \gamma_c$ the self-synchronization error increases in a manor depicted by figure 3.1. Since the strongly coupled fiducial $\mathbf{z}_T$ is only able to reproduce the measured data approximately, with an error given by $\mu(\omega_{fid})$, any improvement that is made to ω_{fid} should be done so without deteriorating its alignment with the data $y_m(t)$. Therefore $\gamma^{(k)}$ is assigned as

$$\gamma^{(k)} = \min_{\gamma \in [\gamma_{min}, \gamma_{fid})} \quad |\phi_0(\omega_{fid}, \gamma) - \mu(\omega_{fid})| \quad ,$$

where $\mu(\omega_{fid})$ is given by equation (7.1), and measures the separation between the strongly coupled fiducial $\mathbf{z}_T$ and the measured data dimensions $y_m(t)$. Recall that γ_{min} here is set by the user as a lower bound to the expected critical coupling γ_c of the system.

The subroutine used to calculate ϕ_0 is implemented in parallel with each thread choosing an equally spaced γ for a resolution in $\gamma^{(k)}$ of $(\gamma_{fid} - \gamma_{min})/N_{THD}$, where N_{THD} is the number of

threads (i.e., independent parallel processes) available on the graphics processor, typically about $2^{10} - 2^{14}$ depending upon the problem size.

7.3.5 The Distribution of Subproblem Solutions

Motivated by the above discussion, we conjecture that the relevant distribution that we are interested in collecting with our ensemble of Markov chains is

$$\pi(\omega) \propto e^{-\beta(\phi(\omega)-\phi_0)} \quad , \tag{7.5}$$

This is the micro-canonical ensemble of a system with $\phi(\omega)$ as the configuration function and 'energy' ϕ_0, with allowed fluctuations about this energy at the inverse 'temperature' β. Again, the normalization factor of (7.5) is of little consequence as the distribution is directly sampled via the Metropolis-Hastings acceptance criterion.

To presume that the distribution (7.5) describes the desired ensemble of Markov chains the assumption is made that, for a given value of the coupling, all trajectories with the same synchronization error are *equally likely* to generate the data using the model dynamical system.

The expectation value $\langle\omega\rangle_E$ over this ensemble at the end of the kth subproblem forms the initial guess $\omega_{fid}^{(k+1)}$ of the following subproblem.

7.4 Scaling The Subproblem

Firstly, the dynamical variables in the objective function are scaled to ± 1 by the user-supplied bounds. The parameters are not scaled in this way as they enter the objective function through the dynamics, however the parameters are similarly given a uniform resolution over the search space during the Monte Carlo procedure via the perturbation stepsize α.

7.4.1 Objective Function Scaling

The subproblem objective function is scaled individually at every time point by the N_x dimensional timeseries $\mathbf{a}(t)^2$, where

$$\mathbf{a}(t)^2 = \frac{(\tilde{\mathbf{z}}(t) - \mathbf{z}(t))^2}{\mathrm{var}(\{\mathbf{z}(t)\}_0)} \quad , \tag{7.6}$$

and

$$\mathbf{z}_T = \mathrm{sync}(\omega_{fid}, \mathbf{y}_T, \gamma_{fid})$$
$$\tilde{\mathbf{z}}_T = \mathrm{sync}(\omega_{fid}, \mathbf{y}_T, \gamma) \quad \gamma < \gamma_{fid}.$$

The variance in the denominator is taken over the high entropy distribution $\{\omega\}_0 = e^{-\beta_0(\phi)}$ where $\beta_0 \approx 0^+$ and then synchronized (using γ_{fid}) with the data. This acts like a modification of the variable bounds at each time point; the user supplied bounds represent the overall dynamical range of the variables, $x_i(t) \in (x_l, x_u)$, however the variance of a synchronized high-temperature distribution provides time resolved information about a range that reasonably restricts the solution trajectory.

The numerator contains the same two trajectories that enter into the self-synchronization error; it increases when there is a discrepancy between the strongly and weakly coupled trajectories of the fiducial point ω_{fid}. This increases the relative importance of finding a candidate trajectory that matches the fiducial at this location along the timeseries.

7.4.2 Scaling of the Perturbation Stepsize α

It is likely that perturbations to the different dimensions of the search variable ω_i do not effect the objective function in uniform way. The problem of non-hyperspherically symmetric search directions is discussed as it relates to the variational problem in section 2.1.2. To address this issue for a stochastic search algorithm, the perturbation stepsizes are set according to the variance of the resultant distribution π according to

$$\alpha_i^2 = \mathrm{var}(\{\omega_i\}) \quad , \tag{7.7}$$

where α and ω are vectors of the same dimension and the index implies that the variance is taken over the ith dimension of all $\omega \in \{\omega\}$.

This scaling is done at the beginning of every subproblem, using the previously obtained distribution, so that the scaling is relevant to the evaluations of ϕ. Calculating (7.7) is fast enough that it is reasonable to perform this scaling twice in a single subproblem, once before and once after $\beta^{(k)}$ is obtained, however this was not found to be significant for the reported test problem.

7.5 Lorenz '63 Twin-Experiment

The behavior of this algorithm is tested on a twin experiment using the Lorenz '63 model (Lorenz (1963)). The equations of this model $\mathfrak{f}$ with the explicit coupling term are

$$\begin{aligned} \dot{x_1} &= p_0(x_2 - x_1) + \gamma_1(y_1 - x_1) \\ \dot{x_2} &= x_1(p_1 - x_3) - x_2 \\ \dot{x_3} &= x_1 x_2 - p_2 x_3 \quad . \end{aligned} \tag{7.8}$$

As a twin experiment, the data is generated using the uncoupled model ($\gamma_1 = 0$) with parameters and initial conditions given in table 7.1, and integrated using a fourth order explicit Runge-Kutta method with a timestep of $dt_{int} = 0.001$ reported at $dt_{obs} = 0.01$. All dimensions of the system are generated and recorded for later comparison, however $N_M = 1$, indicating only $y_1(t)$ is 'known' to the model, and coupled into the model according to (7.8).

The data assimilation window, the number of timepoints that are provided to the optimization routine, consists of $N_T = 2^{14} = 16384$ points, with an interval timestep of $dt_{int} = 0.01$, corresponding to the timestep of the measured data. However we point out that with the current implementation of this algorithm it is possible to use any dt_{int} that is an integer fraction of dt_{obs} if such precision is required for stability.

The range of the coupling allowed for the optimization is $\gamma \in (0.0, 50.0)$, or $\gamma \cdot dt_{int} \in (0.0, 0.500)$, meaning that the initial synchronization fiducial is created with the 'strong' coupling $\gamma_{fid} = 50.0$. From figure 3.1, the critical coupling for this model is $\gamma_c \approx 8$, although such information is presumed unknown at the outset.

The ensemble of Markov chains consists of $N_E = 4096$ elements, each one corresponding to an individual thread on the GPU. The averages and variances that are reported are made over this ensemble. The $\langle P_{acc} \rangle_{target} = 0.3$ and the upper and lower limits to β are $(10^{-4}, 10^{10})$.

7.5.1 Initial Conditions and Parameter Results

Table 7.1: The initial conditions and parameters used for the Lorenz '63 twin experiment, with optimization performed using the MMC method with synchronization fiducial. The results of the optimization are shown as $\langle \omega \rangle_E \pm \sigma_\omega$ (over the ensemble) along with the user-supplied bounds (ω_l, ω_u) and user-supplied initial guess ω_0.

$\omega = \{\mathbf{x}(0), \mathbf{p}\}$	'data' value	$\langle \omega \rangle$	σ_ω	ω_l	ω_0	ω_u
p_0	10.040	1.00190906524658e+01	1.238e-04	0.0	4.1	30.0
p_1	2.6767	2.67646455764771e+00	2.490e-05	0.0	1.1	7.0
p_2	28.011	2.80191955566406e+01	7.107e-05	0.0	35.0	50.0
$x_1(0)$	-4.010	-3.59577608108521e+00	3.110e-04	-30.0	-1.0	45.0
$x_2(0)$	1.1119	8.92393589019775e-01	3.132e-04	-30.0	8.0	45.0
$x_3(0)$	2.7170	2.70061039924622e+00	3.078e-04	-30.0	-5.0	45.0
$x_1(N_T - 1)$	15.6098	1.55878105163574e+01	1.996e-05	–	–	–
$x_2(N_T - 1)$	15.6777	1.64476032257080e+01	9.455e-05	–	–	–
$x_3(N_T - 1)$	36.6666	3.58734626770020e+01	1.956e-04	–	–	–

In figure 7.4 the three dimensions of the data system, including the measured dimension $x_1(t)$ and the two unmeasured dimensions are shown, along with the forward prediction using the resultant state vector distribution at $\{\mathbf{x}(N_T-1)\}$. The values of $\{\mathbf{x}(N_T-1)\}$ that are predicted forward for $t \geq N_T$ come from the final distribution of $\{\hat{\omega}\}$ (which contains only initial condition information) at the end of the optimiation process. This distribution is coupled to the data with γ_{fid} and integrated over the data assimilation window,

$$\{\mathbf{x}(N_T-1)\} = \text{sync}(\{\hat{\omega}\}, \mathbf{y}_T, \gamma_{fid}) \quad ,$$

to obtain an estimate for state at $t = N_T - 1$, again only using the N_M data dimensions.

The prediction is generated by integrating this forward, again with explicit RK4, using the uncoupled version of the model; i.e. completely independent of the data system, with $\gamma = 0$. Due to the chaos inherent in the system, this prediction will inevitably deteriorate on a time scale given by the largest Lyapunov exponent, which is $\lambda_1 \approx 0.9$ for this system.

We use the error in the long time prediction of $\pm\sigma_{x_1} \approx 7$ from figure 7.4 as $\Delta x_1(\infty)$, and the values in table 7.1 between the optimization result and the data to provide $\Delta x_1(N_T-1) = x_1(N_T-1) - y_1(N_T-1) = 0.022$. Equation (2.3) approximates the length of time for which we expect our prediction to be valid as

$$\tau_{max} = \frac{1}{0.9} \ln \frac{7}{0.022} \quad , \tag{7.9}$$

which is indicated on figure 7.4 as a vertical line at $t = N_T + \tau_{max} = 170.2$. All of this information about the quality of the prediction may be obtained from the measured variables and the optimization results and is therefore available for non-twin-experimental situations.

The initial conditions of a chaotic system are of little interest for the purposes of generating a predicition, especially when the data assimilation window is very long. Nevertheless, it is possible to improve the quality of their estimation by repeating this method with a reduced value for γ_{fid} closer to γ_c^+. This, and possibly reducing the length of the data assimilation window[1] have the effect of increasing the relevance that the transient resulting from any discrepancy, $\mathbf{x}(0) \neq \mathbf{y}(0)$, has on the objective function.

7.5.2 Equilibrium Estimation of the Critical Coupling, γ_c

Over the course of the algorithm iterations k, the subproblem coupling $\gamma^{(k)}$ reaches an equilibrium value, at which point the optimization would terminate with the ensemble distribution as the result, however it was allowed to continue for an arbitrary total of $k = 500$ subproblems

[1] After the procedure is run once, we already have an estimate for γ_c and the requirement of $N_T \to \infty$ does not have to be as strictly enforced.

to illustrate the stability of the solution $\{\hat{\omega}\}$. This is shown in figure 7.5, where the value of the coupling used for the subproblem is plotted as a function of the subproblem iterations. Recall that the value of γ_{fid} is never reduced, so that for each successive subproblem, the reported value of $\gamma^{(k)}$ represents the point where the self-synchronization error of ω_{fid} 'crosses' the value of the main problem objective function (see equation (7.1) and section 7.3.4).

The number of required Markov chain perturbations for the subproblem to determine a new ω_{fid} varies per subproblem, however the average number of Markov chain updates per subproblem over the $k = 500$ subproblems was about 280.

As can be clearly seen in figure 7.5, the equilibrium value of $\langle\gamma^{(k)}\rangle_E \cdot dt_{int} = 0.074$, is in agreement with what is seen for γ_c of this system in figure 3.1. One is therefore justified in using $\gamma^{(k)}$ at the end of the optimization as an approximation for γ_c, which is the least amount of coupling that must be supplied to the system to overcome the chaotic instabilities that will eventually occur with any prediction.

For a true experimental situtation, γ_c is not known *a priori* because it depends upon the parameters of the data system. However, once it is estimated as we have done here, it may be used to couple measurements at later times $y_m(t > N_T)$ with the assimilated model to provide high fidelity estimations of the unmeasured model variables much farther into the future than the uncoupled prediction (not shown). Importantly, if this is done with γ_c (as opposed to an arbitrarily large γ) it represents the minimal alteration of the model required to overcome the instabilities on the synchronization manifold between the data and model.

7.5.3 Additional Details

The length of the assimilation window purposefully pushes the $N_T \to \infty$ limit, as an attempt to satisify the theoretical arguments that require it, including the supposition of critical behavior for ϕ near γ_c. As can be seen in figure 3.1, $N_T = 2^{14}$ is qualitatively unsatisfactory compared to say, 2^{20}. We note that, while computation time is not irrelevant, at $N_T = 2^{16}$ the reported values for ϕ would consistently overflow the memory limitations of the GPU.

On the subject of computation time, by far the longest step in the current implementation is the ebvaluation of $\mathbf{a}(t)^2$, which takes at least half of the total time within a subproblem. The integration of the ensemble of Markov chains is performed in parallel for each timestep, however the calculation of the variance is done on the CPU, with the additional slow step of moving a large amount of data from the GPU to the CPU for every timepoint. This is done out of a desire for stability in the calculation (performed with double precision on the CPU), which is by no means a requirement, but avoids possible errors in the coding of the more complicated parallel version.

The most immediate performance increases would result from a completely parallel version of this function. The computation of the 500 subproblems used in this twin experiment took six hours to complete.

The Multiple-Markov Chain Subproblem Algorithm

do {

$k \leftarrow k + 1$

$\mathbf{z}_{T-1} = \text{sync}(\omega_{fid}, \mathbf{y}_{T-1}, \gamma_{fid})$

$\mu^2 = N_T^{-1} N_M^{-1} \sum_{t=0}^{N_T-1} \sum_{i=1}^{N_M} a_i(t)^2 (z_i(t) - y_i(t))^2$

Subroutine to Obtain $\gamma^{(k)}$, and ϕ_0 such that $\phi_0 = \phi(\omega_{fid}, \mathbf{y}_{T-1}, \gamma^{(k)}) \approx \mu$

$\tilde{\mathbf{z}}_{T-1} = \text{sync}(\omega_{fid}, \mathbf{y}_{T-1}, \gamma^{(k)})$

$\mathbf{a}(t)^2 = \frac{(\mathbf{z}(t)^{(k)} - \mathbf{z}(t))^2}{\text{var}(\{\omega\})}$

$\alpha = \sigma(\{\omega\}^{(k-1)})/(\omega_u - \omega_l)$

MH subroutine using $\beta^{(k-1)}$ to obtain 'high temp' $\{\omega\}_0^{(k)}$

MH subroutine using β_∞ to obtain 'low temp' $\{\omega\}_\infty^{(k)}$

Reduce $\beta_\infty \rightarrow \beta^{(k)}$ such that $P_{acc} \approx \langle P_{acc} \rangle_{target}$

MH subroutine using $\beta^{(k)}$ to obtain $\{\omega\}^{(k)}$

if $\langle \phi(\omega) \rangle < \phi_0$ then $\omega_{fid} \leftarrow \langle \{\omega\}^{(k)} \rangle_E$

} while $\langle \frac{d\gamma^{(k)}}{dk} \rangle_k > \epsilon_{EQ}$

Figure 7.3: The subproblem algorithm of the synchronized fiducial multiple-Markov chain Monte Carlo optimization routine.

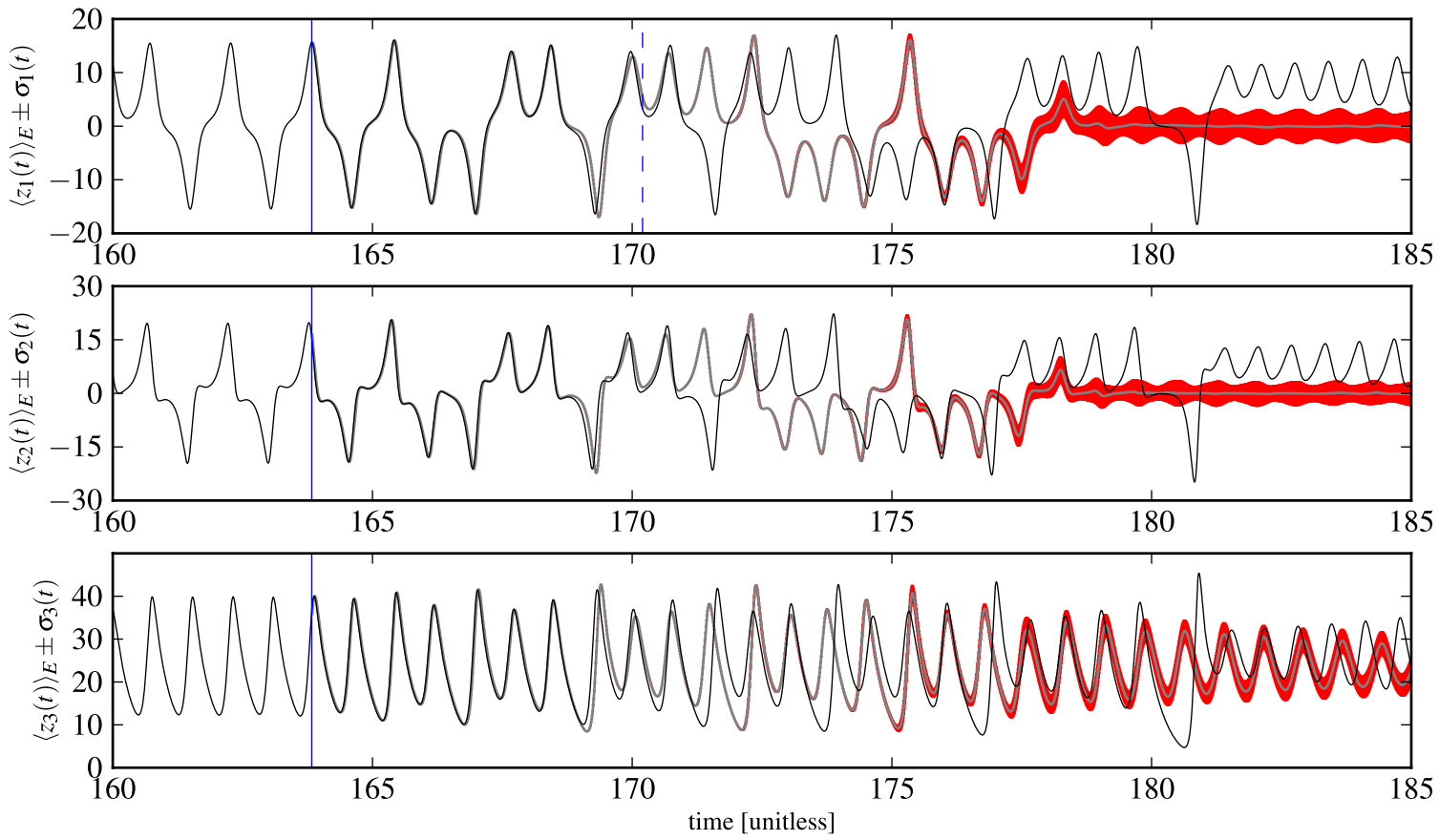

Figure 7.4: The three dimensions of the Lorenz '63 model (7.8) are potted (thin black line) as twin experiment data. The MMC optimized distribution at the end of the data assimilation window is used (along with the estimates of the parameters) to predict forward in time. The average of this distribution $\langle z_i(t)\rangle_E$ is plotted in grey, while the $\pm\sigma_i$ envelope of model error is shown in red. Only the single dimension of data $y_1(t)$ is known to x_i. The second vertical line at $t \approx 170.2$ represents the confidence window of the prediction as estimated in (7.9).

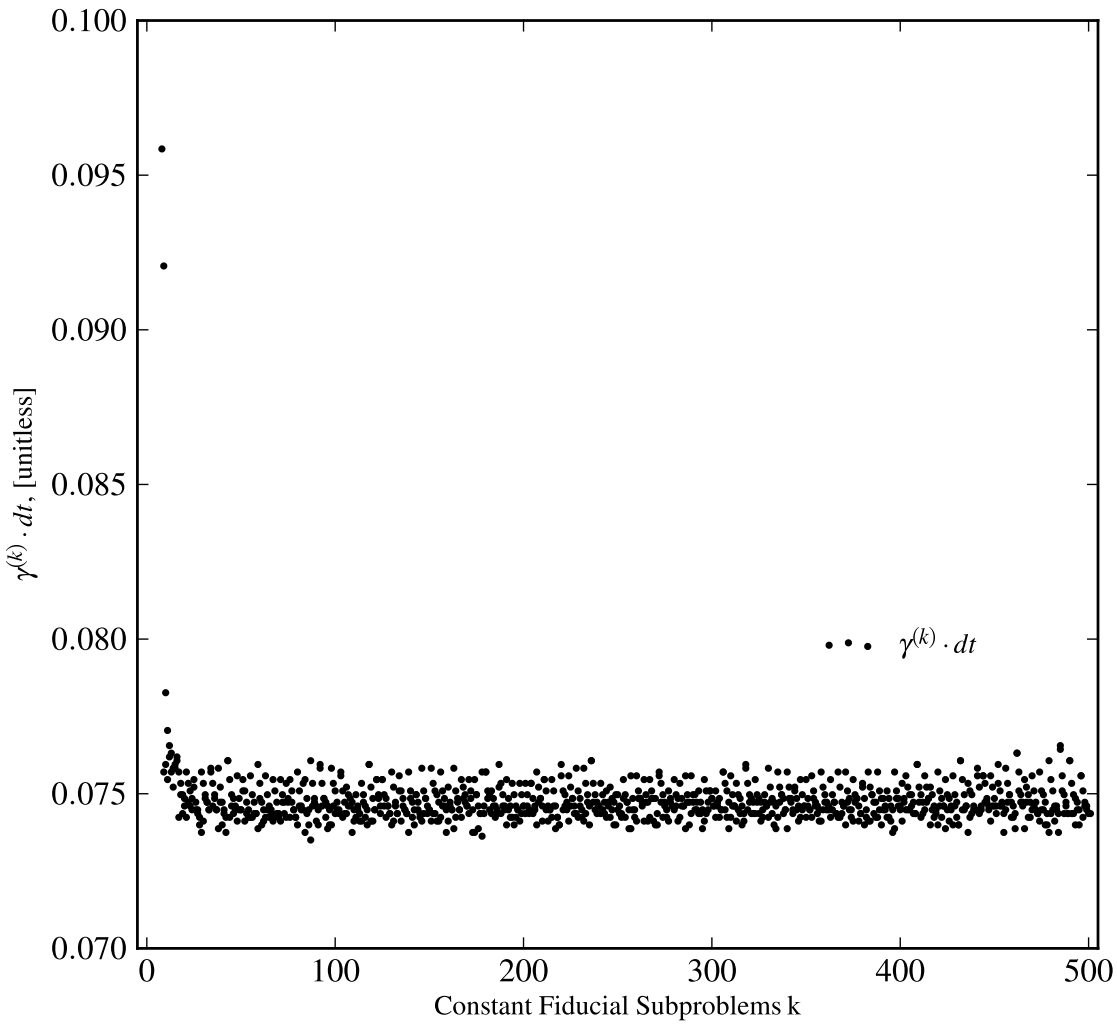

Figure 7.5: The value of $\gamma^{(k)}$ that is obtained in accordance with section 7.3.4 is plotted. An average of 280 Markov updates are performed per subproblem. Notice that an equilibrium value of ≈ 0.0745 is obtained rather quickly, after which it continually returns to the same value of $\gamma^{(k)}$.

Chapter 8

Birdsong Neuron Models

The ability of a bird to sing is a pleasantly recognized skill; the acquisition of this ability requires learning the skill, and is looked upon as an archetype for language learning in general. The effects of many pheneomena such as practice, auditory and sensory-motor feedback, sleep, and age upon the learning of and production of birdsong have been well studied. We refer the reader to the following review articles for a more complete introduction to this subject: Brainard and Doupe (2000, 2002); Margoliash (2004); Knudsen and Gentner (2010).

The zebra finch is one species of songbird that is studied due to its limited repetoire of a single song, and very distinct learning period. Physically, the nuclei comprising the pre-motor portion of the song-production system are very pronounced. Much is known about the large scale connective structure and in some cases the role of the different nuclei in the various stages of birdsong production (Long and Fee (2008); Wang et al. (2008)) Each nuclei contains approximately $10^4 - 10^5$ densely packed neurons, most likely with some level of heterogeneous characteristics amongst them.

The High Vocal center (HVc) is generally considered to be the location of pre-motor song initiation, with very sparse yet importantly-timed signals projecting to RA, the robust nucleus of the archistriatum (Hahnloser et al. (2002)). HVc is also known to project to area X, which is part of a set of delay-time feedback nuclei collectively known as the anterior forebrain pathway. Projections from RA then lead to the motor control of the syrinx and lungs (see for example Abarbanel et al. (2004); Wang et al. (2008); Gibb et al. (2009)).

Properly modelling the production, control and learning of birdsong would certainly take detailed knowledge of this network, and possibly of finer structures within each nuclei. However, with this vision in mind, the motivations for developing methods of data assimiliation

that have been discussed until now become all the more apparent; especially when one considers how it might be possible to probe such a high dimensional structure with limited measurements, or limited intrusion (Buibas and Silva (2011)). Additionally, feedback networks of neurons are known to exhibit chaotic properties, even as many of the neuron models that we construct do not have obvious chaotic parameters sets.

We set out now with a far more modest task; to probe the nodes of this network, single cells, with data assimilation attempts on an *in vitro* neuron from HVc. In this chapter we discuss the pieces and creation of these models that will be used for that neuron and for the neuron twin-experiments in the following chapter.

8.1 Gross Neuro-Anatomy

The cells that make up the nervous system of organisms utilize electrochemical gradients of various ions, such as Na^+ and K^+, to generate fast electrical signals in response to external stimuli. These signals also provide regulatory feedback, and control muscular tissue.

The propagation of electrical impulses through a neuron is generally in one direction, originating from many small branched-structures that make up a dendritic tree, through the *soma* or cell-body, and down an extended axon to another highly branching terminal region, where it connects to subsequent neurons. Between neurons, signals are propogated via direct electrical connections called *gap junctions*, or via *synapses* where chemical signals diffuse across inter-cellular space (Cowen et al. (2001)).

The voltage across the cell membrane of the neuron is governed by the flow of ions through specialized channel protiens embedded within the membrane. Like most cells, the membrane of the neuron consists of a phospho-lipid bilayer; the phospho-lipids sit tail to tail, creating a hydrophobic region in the center of the cell membrane, while the polar heads are exposed to the aqueous medium of both the interior and exterior of the cell.

A wide variety of channel protiens have been discovered through their functional effects on isolated neurons and in some cases their representation in the genome has been identified (Kew and Davies (2010)). Some of them operate passively, as pores with a varying degree of ion selectivity achieved by steric restrictions. Some 'active' channel protiens undergo conformational changes causing them to open or close in response to changes in the potential across the membrane, to the presence of ions such as calcium or magnesium, or by the attachment of chemical signaling modelcules called *ligands*. These interdependencies create nonlinearity in the behavior of the ion channel protiens and is one of the mechanisms which allows for the quick response to

stimuli and the characteristic spiking observed in neuron membrane voltage.

As inferred from the genomic encoding of these protiens, neurons that serve different specialized roles within a single organism are likely to draw from the same finite pool of channel protiens. In this way they can be seen as a finite set of building blocks from which one can account for a wide range of neural behavior, while specificity may be achieved through different densities of these channels in the membrane.

8.2 Trans-Membrane Current Flux

The voltage difference across the cell membrane is defined (Johnston and Wu (1994)) as $V_m = V_{int} - V_{ext}$, and the resting potential of typical neurons is about -65 to -75 mV. The double-wall structure of the phospho-lipid bilayer that makes up the cell membrane can store charge like a parallel plate capacitor; assuming that this membrane capacitance C_m is constant, the definition of current

$$I = C_m \frac{dV_m}{dt} \quad , \tag{8.1}$$

is used as a starting point to model the behavior of the membrane voltage. Often the 'size' of a neuron is colloquially described by its capacitance, where small interneurons may have a $C_m \approx 20 - 25pF$ while large neurons can have a C_m as high as $100pF$ (Szücs (2012)). Additionally, the specific membrane capacitance $\tilde{C}_m \equiv 1\mu F/cm^2$ is used as a unit of conversion, while biophysical values range from $0.5 - 3\mu F/cm^2$ (White (1970); Byzov (1958)).

The current flux accross the membrane occurs in different directions for different ionic species due to the gradients established by the cell. The processes that establishes these gradients such as the sodium-potassium pump operate on timescales that are much greater than present interest and are not modeled. The definition of V_m and equation (8.1) together uniquely define an orientation of the cell membrane in which the *inward* pointing normal is defined as positive[1]. Considering the typical movement of ions during an action potential, Na^+ and Ca^{+2} entering the cell is a positive current flux, and K^+ leaving the cell is negative (Cl^- tends to flow inward, as a negative current).

[1]We note that this introduces a sign difference between the membrane currents reported here and neuron models which define the outward flux as positive and write $C_m dV/dt = -I$. Ideally, one would define outward flux of positive ions as positive, and use $C_m dV/dt = I$, however this implies $V_m = V_{ext} - V_{int}$ which gives neurons a resting potential of about +70mV.

8.2.1 The Goldman-Hodgkin-Katz Equation

The trans membrane currents result from ion movement, which in turn depends upon the local electric field as well as the ionic concentration gradients. One can describe this current flux Φ_s for ion species S, with charge z_s, through a trans membrane channel using the Goldman-Hodgkin-Katz (GHK) equation (Goldman (1943); Johnston and Wu (1994)),

$$\tilde{\Phi}_s = -P_s z_s^2 V X_s(V) \frac{qF}{k_B T} \left(\frac{[S]_{int} - [S]_{ext} \exp(-z_s V q/k_B T)}{1 - \exp(-z_s V q/k_B T)} \right) \quad , \tag{8.2}$$

where $X_s(V)$ is the voltage dependent probability that the given channel is open. The functional form of $X_s(V)$, along with the ion species that passes through the pore, uniquely characterizes the channel and is discussed in the following section. The variable definitions and units used in the GHK equation are given in table 8.1.

Table 8.1: For proper numerical scaling and biophysical relevance, reasonable units for the experimental parameters in equation (8.2) are given. (Note, 1 mL is assumed to be 1 cm^3.)

$\tilde{\Phi}_s$	μA/cm^2	
P_s	μm/s	
V_m	mV	
$[S]$	mM	
z_s	[unitless]	
T	Kelvin	
q/k_B	K/mV	11.604519
F	deca-Coulomb/mmol	9.64853365

Numerical Stability

The GHK equation (8.2) requires L'Hopital's rule to evaluate $V = 0$ and therefore must be approximated for numerical stability. Since the evaluation of the model functions are often the most time consuming step of an optimization algorithm, it is important not to retain too many terms in the expansion.

To this end, a Taylor expansion is performed about $V = 0$, and the coefficients of the highest order terms (both positive and negative) are adjusted to extend the domain of voltage over which the expansion is a close approximation. Rewriting (8.2) as

$$\tilde{\Phi}_s = P_s z_s X_s(V) F \Big([S]_{ext} - [S]_{int} \exp(z_s V q/k_B T) \Big) \cdot \left(\frac{z_s V q/k_B T}{\exp(z_s V q/k_B T) - 1} \right) \quad ,$$

the expansion is performed upon the last term in parentheses.

It is also important to write them so that the terms do not combine floating point values of vastly different magnitudes, therefore the following functional form is used for optimization,

$$\frac{\xi V}{e^{\xi V}-1} \approx 1 - \xi V/A + (\xi V/B)^2\big(1-(\xi V/9)^2(1-(\xi V/13)^2)\big)/3 \quad , \tag{8.3}$$

where $\xi = z_s q/k_B T$. The coefficients were fit by eye at $T = 35C$ so that for the $z_s = 1$ case, $A = 2.000$ and $B = 2.018$ results in $< 0.5\%$ relative error over $V \in (-120, 65)$ mV. For the $z_s = 2$ case (e.g. calcium), $A = 1.950$ and $B = 1.965$ results in $< 3\%$ relative error over $V \in (-120, 45)$ mV and $< 13\%$ relative error over $V \in (-120, 65)$ mV. Note that it is more challenging to match a given voltage range for divalent ions because of z_s in the expansion.

8.2.2 Ohmic Current Flux

If the ratio of the internal to external ion concentration is approximately unity, the Nernst equation may be used to calculate a reversal potential E_{rev} for that ion species,

$$E_s = \frac{k_B T}{z_s q} \ln \frac{[S]_{ext}}{[S]_{int}} \quad . \tag{8.4}$$

The reversal potential is specific to an ion species, and so the subscript is reserved for the ion type. This is a valid approximation for sodium and potassium, however the calcium concentrations can differ by a few orders of magnitude and necessitate the use of the GHK equation for their accurate description.

The Nernst equation may then be used to define a current flux given by Ohm's law,

$$\tilde{\Phi}_s = \tilde{g}_s X_s(V) \cdot (E_s - V) \quad , \tag{8.5}$$

where again, $X_s(V)$ is the probability that that channel type will be found open, and is voltage dependent. The specific maximal conductance, $\tilde{g}$ has units of millisiemens per square centimeter, $1\text{S} = \Omega^{-1}$, giving $[\tilde{\Phi}_s] = \mu\text{A/cm}^2$ again. This form of membrane current flux was used by Hodgkin and Huxley in their studies of the giant squid axon (Hodgkin and Huxley (1952)).

Both this form for the current flux (8.5) and equation (8.2) may be used in the same neuron's voltage equation, allowing one to model the effects of calcium flux while taking numerical advantage of the Nernst simplification for sodium and potassium.

8.2.3 The Effect of Temperature

Aside from consistency, a benefit to using the GHK equation (8.2) for all ionic species is that it keeps the parameters of the model independent of temperature, whereas the reversal

potentials in the Nernst equation are temperature dependent. By separating out this dependency, one may arrive at more general characterizations of neuron populations.

The author speculates that, while the average temperature of the experimental setup is very nearly constant (especially over the $100 - 1000$ msec used for the optimization) the exponential dependence of the gating kinetics, reversal potentials (and therefore spiking threshold), on temperature likely makes the system as a whole sensitive to error in this value. The effects of such measurement error is exactly what the arguments of chapter 5 are designed to address, however to date the effects of error in temperature have not been evaluated.

8.3 Single Barrier Ion-Channel Model

The ion channels that are responsible for passive membrane currents, also known as *leak* currents, are modeled with $X(V) = 1$; these are combined into one ohmic current with a unique reversal potential E_{Leak} that takes into account multiple types of passive channels, and their lack of ion selectivity.

The voltage dependent, or *active* channels that appear in (8.2) have far more structure, and each usually consists of a few independently operating channel protiens[2] that undergo a conformational change depending with the voltage potential of the membrane (see, for example Kew and Davies (2010)). The probability that the channel is open is then the product of the probabilities that the individual gating protiens are in the open state.

The membrane-average state of each type of gating protien is given by a unique x, representing the probability that it is found open $x \in [0, 1]$. Since the protien confirmations are assumed to only depend on their current state and the transition rates α and β between states, their dynamics is governed by the first-order kinetic equation

$$\frac{dx}{dt} = \alpha(1 - x) - \beta x \qquad . \tag{8.6}$$

The opening rate is given by $\alpha(V)$ and the closing rate by $\beta(V)$.

The details of the voltage dependent rate equations α and β depend upon the mechanism of the gating proteins. For simplicity, we assume that the energy landscape of the protein's conformational change consists of a single energetic barrier. Therefore the transition rates are given by the Boltzmann factor $e^{-E/k_B T}$, where E is the energetic height of this barrier. Then

[2]These channel protiens are also often referred to as gating 'particles'.

for a voltage dependent barrier height these rates are

$$\begin{aligned} \alpha(V) &= e^{-(G_\alpha \mp z_\alpha V)q/k_B T} \\ \beta(V) &= e^{-(G_\beta \pm z_\beta V)q/k_B T} \end{aligned} \tag{8.7}$$

where G_α and G_β represent the height of the energetic barrier (in mV) of the opening and closing rates respectively. The linear voltage dependence of the channel protien's conformational change is given by z, again the individual parameters for the opening and closing transitions indicate a possible asymmetry to the confirmational change. Physically, z can be thought of as the net charge on the particle that gates the channel.

By restricting z to be positive the $\pm$ signs designate how the channel responds to changes in voltage. The top sign is used for m-particles, that open upon increasing voltage, and the bottom sign is for h-particles, those that open with decreasing voltage.

Another representation of these kinetics, that is also often used and perhaps more aligned with experiment is

$$\frac{dx(t)}{dt} = \frac{1}{\tau}(x_\infty - x(t)) \tag{8.8}$$

where $\tau(V)$ is the time constant for the channel to return to a steady-state open-fraction, given by $x_\infty(V)$. The transformation relating (8.6),(8.7) and (8.8) is

$$\begin{aligned} x_\infty &= \frac{1}{1+\beta/\alpha} \\ \tau &= \frac{1}{\alpha+\beta} \end{aligned}$$

where

$$\beta/\alpha = \exp\Big(-(G \pm zV)\frac{q}{k_B T}\Big) = \exp\Big(-\frac{V_{1/2} - V}{\kappa}\Big) \quad .$$

The top sign is again for m-particle while $G = G_\beta - G_\alpha$ and $z = z_\alpha + z_\beta$ are the same for both types of protiens. This results in

$$\begin{aligned} V_{1/2} &= \mp\frac{G}{z} \\ \kappa &= \mp\frac{k_B T}{zq} \end{aligned} \quad ,$$

where $V_{1/2}$ is the voltage of half-activation in the $x_\infty(V)$ sigmoid, and κ is the slope at this voltage.

8.3.1 Numerical Stability

The bare exponentials can become large when the membrane voltage is very large (small) for the α (β) function. To avoid this purely numerical issue, for all calculations we rewrite them

as logistic functions with a limiting rate of $1/dt$, where dt is the integration timestep; this has little effect on the performance of the gating particle. For those situations, we use the form

$$\begin{aligned} \alpha(V) &= \frac{1}{e^{(G_\alpha \mp z_\alpha V)q/kT} + \epsilon_\alpha} \\ \beta(V) &= \frac{1}{e^{(G_\beta \pm z_\beta V)q/kT} + \epsilon_\beta} \end{aligned} \tag{8.9}$$

with $\epsilon \leq dt_{int}$, which reduces to (8.7) when ϵ_α and ϵ_β are zero.

Additionally, there are instances where the time constant of the channel is found experimentally to be appreciably non-zero at extreme voltages. This indicates that there is some kind of restriction on the maximum transition rates and so the use of a logistic function with a larger constant term ϵ is more appropriate.

Equivalently, the $\alpha(V)$ and $\beta(V)$ may also be given by

$$\begin{aligned} \alpha(V) &= \frac{1}{2\epsilon}\Big(1 + \tanh\frac{-(G_\alpha \mp z_\alpha V)q/k_BT + \ln\epsilon}{2}\Big) \\ \beta(V) &= \frac{1}{2\epsilon}\Big(1 + \tanh\frac{-(G_\beta \pm z_\beta V)q/k_BT + \ln\epsilon}{2}\Big) \quad , \end{aligned} \tag{8.10}$$

which has more numerically stable derivatives than (8.9). For an m channel, that opens with increasing voltage, use the top sign; for an h channel, that closes with increasing voltage, use the lower sign. Note that regardless of the form of the sigmoid that is used, stability with respect to the bounds of the kinetic variables must be maintained; that is, if $x(t) > 1.0$ or $x(t) < 0$ the model neuron will fail.

8.4 Single Compartment Neuron Models

The voltage measurements from a whole cell (current-clamp) experiment are dependent upon the intrinsic neuron channel dynamics and an injected current I_{inj} applied by the experimenter so that (8.1) becomes the combined equation

$$C_m \frac{dV}{dt} = A_m \sum_s \tilde{\Phi}_s(V) + I_{inj} \quad ,$$

where A_m is the surface area of the cell membrane, and the sum of current fluxes is over both the channel type and active ion; e.g. one may have currents resulting from multiple channels, all with different gating kinetics, but that are permeable to the same ion.

For calculations, this is rewritten as

$$\frac{dV}{dt} = \tilde{C}_m^{-1} \sum_s \tilde{\Phi}_s(V) + C_m I_{inj} \quad ,$$

where C_m/A_m is replaced with $\tilde{C}_m \equiv 1.0\mu F/cm^2$ (note that both $\tilde{C}_m$ and C_m appear), thus absorbing any discrepancy bewteen this assumed value and the actual value into $\tilde{g}_s$ for ohmic current-fluxes described by (8.5), or into P_s for fluxes described by (8.2). Note that the units on I_{inj}/C_m here must be Amp/Farad; typically I_{inj} is measured in nA.

8.5 Ion Channel Models

There are about a dozen well identified types of sodium channels and upwards of 40-50 potassium channels (Kew and Davies (2010)). Given the ability for combinations of the gating particles to give rise to kinetically distict varieties, not to mention genetic mutations in the channel protiens, the number of possible channel types is rather large. Specific channel identification is beyond the scope of this work, although in principle data assimilation may assist with this type of problem.

Assuming that much of the channel kinetics is conserved across neurons and even to a certain extent across organisms, it is quite reasonable to consider fixing these parameters when one is concerned with whole-cell properties. The extent to which the data assimilation method can correctly estimate these values is discussed later in the context of individual experiments; in what follows the parameters for some common channels expressed using equation (8.9) or (8.7) are taken from previously completed biological studies and compiled in the tables 8.2 and 8.3. Unless otherwise noted, the ϵ's for all channel particles are negligible ($\epsilon \approx dt$ for numerical stability).

Table 8.2: Part I. Channel kinetics obtained from the literature.

Channel	Formula	G_α [mV]	z_α [·]	G_β [mV]	z_β [·]	
$Na_v1.1$	$x(V) = m^3hj$					Barela et al. (2006)
	m	-40	3.0	78	3.0	
	h	170	1.7	9.6	2.12	
	j	-	-	-	-	not modeled
$Na_v1.2$	$x(V) = m^3h$					Stühmer et al. (1987)
	m	-10	1.3	108	1.58	
	h	150	1.2	-8.9	1.28	
$Na_v1.7$	$x(V) = m^3h$					Cox (2006)
	m	-50	2.3	94.4	2.85	
	h	240	1.14	-64	3.14	

Table 8.3: Part II. Channel kinetics obtained from the literature.

Channel	Formula	G_α [mV]	z_α [·]	G_β [mV]	z_β [·]	
$K_v1.1$	$x(V) = m^4$					Grissmer et al. (1994)
	m	60	0.45	156	2.56	
$K_v1.2$	$x(V) = m^1$					Grissmer et al. (1994)
	m	140	0.30	86.8	1.67	
$K_v3.1$	$x(V) = m^4$					Grissmer et al. (1994)
	m	140	1.5	92.9	1.44	
$K_v10.x$	$x(V) = h$					Huguenard and McCormick (1992)
(I_H)	h	388	2.28	49.7	1.86	
$K_v4.x$	$x(V) = m^4h$					Huguenard and McCormick (1992)
(I_{A1})	m	-45	1.3	135	1.7	$\epsilon_\alpha = \epsilon_\beta = 0.37$
	h	162	0.68	-235	5.1	$\epsilon_\beta = 20$

8.5.1 $Na_v1.x$ (I_{NaT}) sodium channel

The I_{NaT} current is reponsible for the neuron depolarization that generates the action potential. This channel is known to have three similar activation particles, m, and a fast in-activation particle h. It is also suspected that the $Na_v1.1$ channel has a second, slower inactivation particle j, that is not modeled because complete kinetic information was not available. Parameters are found from Barela et al. (2006) for the $Na_v1.1$ channel; from Stühmer et al. (1987) for the $Na_v1.2$ type channel and from Cox (2006) for the $Na_v1.7$ type channel.

8.5.2 $K_v4.x$ (I_A) potassium channel

The I_A current is governed by a rapidly in-activating potassium channel gated by two particles in an m^4h configuration. The h particle of the I_A channel has quite a functionally interesting time constant $\tau(V)$ which is responsible for the rapid in-activation of this channel. Huguenard & McCormick (Huguenard and McCormick (1992)) use a piecewise function to describe the time constant, however we are able to use a continuous function at the expense of using more extreme values of G_α and z_α. Additionally they find that this channel requires non-zero

time constants $\epsilon_{\alpha,\beta}$ because $\tau(V)$ remains appreciable at extreme voltages.

8.5.3 $K_{\mathbf{v}}10.x$ (I_H) potassium channel

The I_H current is seen in many cell types as a sag response to a hyperpolarizing injected current. The current is found experimentally to have a single activation particle h, that opens with decreasing voltage. This channel is thought to have a less selective pore than others, allowing both sodium and potassium ions through when it is open. For GHK currents, we model the permeability of this channel by a single parameter and fix the relative amounts of sodium and potassium based upon the inverse of their relative ionic cross-sectional area. Done this way, the sodium contribution receives a prefactor of 0.631 and the potassium 0.368. For ohmic currents, we model this channel with a unique reversal potential that accounts for both sodium and potassium ions.

8.5.4 $Ca_{\mathbf{v}}1.x$ & $3.x$ (I_L & I_T) calcium channels

The I_L and I_T currents are low-threshold calcium currents, letting calcium flow into the cell similar to the action of sodium but on a much slower timescale. Following Huguenard & McCormick again, we model I_T as being gated by two types of particles in an m^2h configuration. The I_L current follows m^2 kinetics but parameter values were not obtained from the literature.

8.5.5 $K_{\mathbf{Ca}}$ ($I_{\mathbf{KCa}}$ or I_{BK}) calcium-gated potassium channel

The I_{KCa} channel is somewhat more complicated than the ones we have discussed until now. Dispite the interest in this complicated piece of biological machinery, good experimental data of the time constant is difficult to find. We follow conventional wisdom and model this channel as being selective to potassium ions, with activity composed of four gating particles that open upon increasing voltage in an m^4 configuration. As with all channel types, there are genetic variants, however it is thought that up to three calcium ions bind to the interior region of each particle affecting the voltage sensitivity. We have found that it is possible to account for the majority of the calcium binding effect through changes to the energetic barrier of the particle as a function of the internal calcium concentration. Based upon experimental studies by Cui et al. (1997) and Prakriya et al. (1996), the effect of this calcium binding upon the charge of the particle (z_α and z_β) is measureable, however minor in its effect on the overall voltage sensitivity to the

time constant. We model the kinetics of this particle with the following equations,

$$\alpha(V,[Ca]_i) = e^{-(G_\alpha([Ca]_i)-z_\alpha V)q/k_BT}$$

$$\beta(V,[Ca]_i) = e^{-(G_\beta([Ca]_i)+z_\beta V)q/k_BT} \quad ,$$

with $z_\alpha = 0.75$ and $z_\beta = 0.5$. The calcium dependent energetic barriers are given by the following power-laws

$$G_\alpha([Ca]_i) = A_\alpha [Ca]_i^p$$

$$G_\beta([Ca]_i) = A_\beta \ln^q [Ca]_i \quad ,$$

with $A_\alpha = 160$mV, $p = -0.2$, $A_\beta = 17.5$mV, $q = 0.6$ and $[Ca]_{int}$ measured in μM. Note that because of the ln[Ca] in G_β this functional form is only valid with $[Ca]_i \geq 1\mu M$, G_β is considered to be zero at lower concentrations. These parameters were obtaind by fitting time constants by eye to the data collected by Cui et al. (1997).

8.5.6 Calcium Dynamics

Typical interior concentrations of calcium range from 5-100 nM, while the exterior concentrations are more commonly in the mM range. Due to such large variations, if one uses I_{KCa} channels it is important to model $[Ca^{+2}]_{int}(t)$ as a dynamical variable. One aspect of calcium dynamics that is possibly important but not represented in this model is that interior concentrations may be much higher locally around the calcium channels, especially if the I_{KCa} are in close proximity to the I_L and I_T channels, allowing the I_{KCa} to experience a more extreme calcium gradient. The calcium enters the cell through the voltage-gated I_L and I_T channels and is sequestered by the cell in the smooth endoplasmic reticulum where it is buffered and regulated by intermediate messenger molecules (Blackwell (2005)).

The effects of calcium dynamics, and the I_{KCa} channels were not used in any of the models here.

Chapter 9

Neuron Twin Experiments

In this chapter are presented twin experiment results with various neuron models using a variety of computational techniques. The first experiment is worked through in detail to show how the algorithm described in chapter 6 may be used to ascertain the impact of model error about a locally optimized solution.

We also describe the results of some data assimilation twin experiments performed on neuron models from the previously published works of Abarbanel et al. (2011), Toth et al. (2011), and Kostuk et al. (2012). These papers address interesting questions including the effects of measurement noise on some varational optimization methods, the attempt at assimilating data from an incorrect model, and dynamical considerations that one should address when designing an input stimulus for those systems –such as a neuron– that may be driven by the experimenter.

9.1 Model Error of a GHK Twin Experiment

We performed a twin experiment on a model (GHK-1069) inspired by the discussion in the previous chapter, that uses the Goldman-Hodgkin-Katz (GHK) equation (8.2) to describe the current flux across the cell membrane. A set of permeabilities for a spiking neuron using the various ion channel kinetics that were collected from the literature in tables 8.2 and 8.3 was not found, therefore the kinetics of the twin experiment are given by the parameters that were obtained using the variational optimization on a physical HVc neuron described in detail in the next chapter.

Using these kinetics in a twin experiment has the benefit that one can see how they respond to the incorporation of model error when a reference exists (i.e. the 'unmeasured'

variables of the twin experiment 'data'). With this in mind, the purpose of this twin experiment is to demonstrate the effect of model error on a neuron model with GHK currents, and to help us better interpret the results of allowing model error on the unmeasured variables of the physical neuron in the subsequent chapter.

The data is generated using an explicit fourth order Runge-Kutta method with a timestep of $dt_{int} = 0.002$ ms and reported at $dt_{obs} = 0.20$ ms. The data assimilation window consists of $N_T = 4096$ time points spaced at dt_{obs} for a total time of 819.2 ms; the total length of the data timeseries is required to properly sample kinetics that presumably have slow time constants, while the short integration timestep is required by the explicit method to alleviate numerical instabilities in the exponentials near the kinetic boundaries. The timestep $dt_{int} = dt_{obs} = 0.2$ ms is used in the Monte Carlo evaluation, for which any error in the discretization method is still less than the level of model error being allowed by σ_m^2 and σ_f^2. We note that this is identical to the setup used to ascertain model error of the physical neuron in the next chapter.

The model error in the states and parameters is determined using the single-Markov chain Monte Carlo algorithm described in chapter 6. The measurement noise is taken from the RMS variation in the experimental voltage trace of the physical neuron (during a period of constant current) that is analyzed in the next chapter, and is

$$\sigma_m = 0.2597 \text{ mV} \quad .$$

A value for σ_f^2 is simply chosen to reflect a presumed level of model error, however the choice greatly affects the ability of the Markov chain to reach equilibrium in a reasonable amount of time. The rule of thumb that is followed for this algorithm is to use $\sigma_f = \sqrt{2}\sigma_m^2$. As such the coefficients of the conditional mutual information term and the model error term in the action are given by

$$\begin{aligned} \frac{1}{2\sigma_m^2} &= 7.41 \text{ mV}^{-2} \\ \frac{1}{2\sigma_f^2} &= 54.91 \quad , \end{aligned}$$

respectively. The units of the different dimensions of the model noise $[\sigma_f]$ are assumed to be equal to their respective variable; mV for the voltage and unitless for the gating probabilities. For simplicity the magnitude of model noise is assumed to be constant over the different dimensions of the model.

The full set of parameter bounds and the set that is used to generate the data are given in appendix C.

9.1.1 GHK-1069 Model Equations

The full model equations that are used for this twin experiment are

$$
\begin{aligned}
\frac{dV}{dt} &= (P_{NaT} m_1^3 h_2 + 0.631 P_H h_6)([Na]_{ext} - [Na]_{int}\exp(V \cdot \mathrm{QKT}))\mathrm{GHK}_{z1} \\
&+ (P_{KA1} m_3 + P_{KA2} m_4^4 h_5 + 0.368 P_H h_6)([K]_{ext} - [K]_{int}\exp(V \cdot \mathrm{QKT}))\mathrm{GHK}_{z1} \\
&+ (g_{CaL} m_7^2 + g_{CaT} m_8^2 h_9)([Ca]_{ext} - [Ca]_{int}\exp(V \cdot \mathrm{QKT}))\mathrm{GHK}_{z2} \\
&+ g_L(E_L - V) + C_m^{-1} I_{inj} \\
\frac{m_1}{dt} &= \frac{1 - m_1}{\exp[(G_{1\alpha} - z_{1\alpha}V)\mathrm{QKT}] + \epsilon_{1\alpha}} - \frac{m_1}{\exp[(G_{1\beta} + z_{1\beta}V)\mathrm{QKT}] + \epsilon_{1\beta}} \\
\frac{h_2}{dt} &= \frac{1 - h_2}{\exp[(G_{2\alpha} + z_{2\alpha}V)\mathrm{QKT}] + \epsilon_{2\alpha}} - \frac{h_2}{\exp[(G_{2\beta} - z_{2\beta}V)\mathrm{QKT}] + \epsilon_{2\beta}} \\
\frac{m_3}{dt} &= \frac{1 - m_3}{\exp[(G_{3\alpha} - z_{3\alpha}V)\mathrm{QKT}] + \epsilon_{3\alpha}} - \frac{m_3}{\exp[(G_{3\beta} + z_{3\beta}V)\mathrm{QKT}] + \epsilon_{3\beta}} \\
\frac{m_4}{dt} &= \frac{1 - m_4}{\exp[(G_{4\alpha} - z_{4\alpha}V)\mathrm{QKT}] + \epsilon_{4\alpha}} - \frac{m_4}{\exp[(G_{4\beta} + z_{4\beta}V)\mathrm{QKT}] + \epsilon_{4\beta}} \\
\frac{h_5}{dt} &= \frac{1 - h_5}{\exp[(G_{5\alpha} + z_{5\alpha}V)\mathrm{QKT}] + \epsilon_{5\alpha}} - \frac{h_5}{\exp[(G_{5\beta} - z_{5\beta}V)\mathrm{QKT}] + \epsilon_{5\beta}} \\
\frac{h_6}{dt} &= \frac{1 - h_6}{\exp[(G_{6\alpha} + z_{6\alpha}V)\mathrm{QKT}] + \epsilon_{6\alpha}} - \frac{h_6}{\exp[(G_{6\beta} - z_{6\beta}V)\mathrm{QKT}] + \epsilon_{6\beta}} \\
\frac{m_7}{dt} &= \frac{1 - m_7}{\exp[(G_{7\alpha} - z_{7\alpha}V)\mathrm{QKT}] + \epsilon_{7\alpha}} - \frac{m_7}{\exp[(G_{7\beta} + z_{7\beta}V)\mathrm{QKT}] + \epsilon_{7\beta}} \\
\frac{m_8}{dt} &= \frac{1 - m_8}{\exp[(G_{8\alpha} - z_{8\alpha}V)\mathrm{QKT}] + \epsilon_{8\alpha}} - \frac{m_8}{\exp[(G_{8\beta} + z_{8\beta}V)\mathrm{QKT}] + \epsilon_{8\beta}} \\
\frac{h_9}{dt} &= \frac{1 - h_9}{\exp[(G_{9\alpha} + z_{9\alpha}V)\mathrm{QKT}] + \epsilon_{9\alpha}} - \frac{h_9}{\exp[(G_{9\beta} - z_{9\beta}V)\mathrm{QKT}] + \epsilon_{9\beta}}
\end{aligned}
\tag{9.1}
$$

where the GHK expansions for the cations are given by the expansion in equation (8.3), and the value of QKT is fixed at 0.037756 $\mathrm{mV}^{-1}$to mimic the experimental data of chapter 10. This model has $N_x = 10$ and $N_p = 69$.

9.1.2 Equilibrium of the Markov Chain

The burn-in phase of the Monte Carlo calculation proceeds for 50,000 iterations, generating 2.97x10^7 updates to each dimension of the parameter vector $\mathbf{p}$, in accordance with the sampling adjustment[1] described in section 6.2.2. These iterations are discarded while the Markov chain settles into an equilibrium state. The statistics gathering phase involves twice as many iterations as the burn-in phase, during which time 100 uniformly-spaced (in iterations) samples are selected to represent the distribution. These samples are used to calculate the averages and standard deviations that are reported here; it is possible to calculate higher moments if one is

[1] $N_x N_T / N_p \approx 594$

interested in knowing them. These samples form the distribution that is integrated forward, individually and noise-free, to generate the prediction distributions shown in the figures.

To ensure that the statistics are valid the Markov chain should be in a steady state during the statistics collection phase. This means that the average value of the action $\langle A \rangle$ should remain roughly constant during this time, relative to the size of the fluctuations, indicating a quasi-equilibrium state. Figure 9.1 shows $A(k)$ for both the burn-in phase (top) and the statistics phase (bottom).

The large scale undulations of A are possibly indicative of poor mixing, meaning that the stepsize α may be too small to effectively sample the entire space (Rosenthal (2011)). This stepsize is not set directly, but is adjusted prior to the burn-in phase in an attempt to match the user-determined target mean acceptance rate $\langle P_{acc} \rangle_{target}$ of the chain; this observation of poor mixing would suggest reducing $\langle P_{acc} \rangle_{target}$ towards zero.

However, adequately sampling an entire space of this size is ultimately a question of patience and the overall behavior of $\langle A \rangle$ is not trending (neither increasing nor decreasing) which is the key element for us to declare this chain as sufficiently equilibrated. If the chain has not come to equilibrium the distribution will drift away from the local solution by an amount much greater than a standard deviation of the model error.

9.1.3 Voltage Assimilation and Prediction

The results of the Monte Carlo calculation are shown for the voltage variable of the model as $\langle V \rangle \pm \sigma(V)$ in figure 9.2. The segment of the timeseries that is shown straddles both the end of the data assimilation window and the beginning of the prediction period, separated by the vertical line at $t = 819.2$ms.

The 'measured' data voltage is shown for comparison; this is the data that is provided to the model and assimilated into the state variable determination via the conditional mutual information term in the action. During the assimilation window the average values of model spikes are somewhat diminished relative to the data, however the full height of the spikes remains comparable to the data during the prediction phase. The spike-timing of the prediction with model error also remains consistent with the data system. However, without the assimilated data the average sub-threshold potential of the model neuron is clearly depolarized (shifted to higher voltage) relative to the data system. The prediction may be continued for the duration of the injected current (≈ 2.5sec) without futher deterioration (not shown).

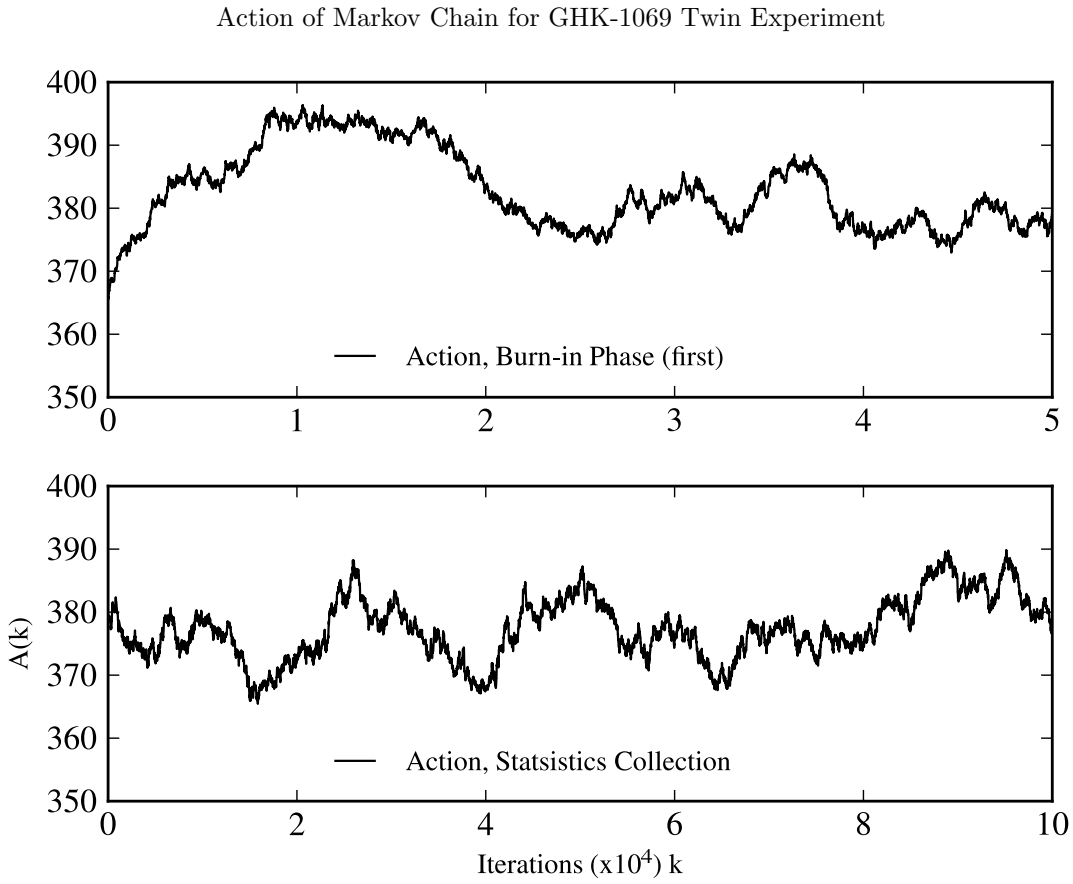

Figure 9.1: The action A as a function of the iteration of the Markov Chain Monte Carlo simulation is shown for the GHK-1069 neuron model twin experiment. A qualitiative assessment of equilibrium is provided by the overall lack of trend in $\langle A \rangle$ relative to the scale of the fluctuations in A.

9.1.4 Model Error of an Unmeasured Variable

Next, we examine the effects of model error an an unmeasured state variable in the context of a twin experiment where there is access to the 'true' value of the state. We look at one of the gating variables in the sodium current, m_1, which involves Na^+ ions entering the cell when the membrane voltage exceeds a threshold potential (of ≈ -45 mV). It is gated by two particles in an $x(V) = m^3(V)h(V)$ configuration as shown in equation (9.1).

In figure 9.3 one can see the effect that the incorporation (or lack thereof) of the measurements during the assimilation window have upon the prediction of this unmeasured state variable. Note the overall increase in the baseline of this variable's predicted average relative to the data value, as well as the relative increase in variance (shown in red).

As the kinetics of this model are adapted from the experimental problem in chapter 10, performing this twin experiment may be seen as validation of the suppositions made there. Namely, that the shift in behavior of this variable before and after the assimilation window is indicitive of it being sensitive to model error, and therefore suggestive of a focus for model improvement. This will be discussed further in section 10.3.3.

9.2 Hodgkin-Huxley Model with Noise

The results of another data assimilation twin experiment are given in Abarbanel et al. (2011), this time using the Hodgkin-Huxley neuron model (not repeated here (Hodgkin and Huxley (1952))). The twin-experiment is also setup to mimic a current-clamp protocol where a current is injected into the neuron, and its voltage response is recorded.

The related minimization problem is solved using the balanced synchronization objective function (equation (3.4)) by the nonlinear variational solver SNOPT (Gill et al. (2005)). The model has $N_x = 4$ and $N_p = 22$; the timeseries of assimilated data is short by comparison with the other results here, approxiately 100 ms, but sampled at a frequency of 100 kHz to give $N_T = 4000$. The authors suggest the use of a chaotic input current to expand the combined dimension of the data-model system, thereby allowing the optimization to better distinguish between dynamical models with similar parameter sets.

9.2.1 Observational Noise Reduction via Synchronization

Also in Abarbanel et al. (2011), observational noise that is added to the measurement (and assimilated into the model) is shown to be filtered during the optimization process through the connection to the model; a model which does not contain modes that operate as fast as the observational noise. The resultant, optimized voltage trajectory contains reduced noise relative to the input.

Limits to this filtering do exist, and if sufficient noise is present the optimization step will eventually fail. This occurs through the use of increased coupling (or constraint violations in the differential equations, if they are permitted as model error) and consequently the estimation of the unobserved state variables deteriorates.

9.3 NaKl & NaKlH Assimilation using Variational Method

A data assimilation twin experiment, between two *different* Hodgkin-Huxley like models is carried out by Toth et al. (2011). The twin experiment is setup to mimic a current-clamp protocol and the associated minimization problem is solved using a balanced synchronization objective function (equation (3.4)) by the nonlinear variational solver IPOPT (Wächter (2002)). The models equations (9.2) and (9.3) are given at the end of the chapter.

The two models that are used have a sodium, potassium and leak current, with the channel dynamics governed by hyperbolic tangent functions (as in (8.10)) and the currents given by the Nernst equation (8.4). The NaKlH (9.3) model contains an additional potassium current, the I_H current as described in section 8.5.3. In total the NaKl model has $N_x = 4$ and $N_p = 19$, while the NaKlH model has $N_x = 5$ and $N_p = 26$; typical data assimilation windows were $N_T = 1000 - 10000$ with a timestep $dt_{int} = dt_{obs} = 0.02$ ms.

9.3.1 Assessments of Incorrect-Model Assimilation

Within Toth et al. (2011) the question of assimilating data into an incorrect model is addressed in two complementary ways. Firstly, data from a NaKlH system is presented to the NaKl model for assimilation. As the NaKl model lacks the I_H current, it is incapable of reproducing the NaKlH voltage data without assistance from the synchronization coupling.

The second case used to illustrate model-incorrectness involves assimilating data from a NaKl system into a NaKlH model. In this situation, the NaKlH model contains an additional current that is not necessary to generate the data that is presented to it. The result of the optimization procedure on this experiment reported the maximal conductance of this channel to be $g_H = 10^{-9}$ mS/cm^2, indicating its irrelevance compared to the other channels with conductances ranging from $0.3 - 120$ mS/cm^2.

9.3.2 Adequate Stimulus Protocols

As the unobserved state variables must have a relevant effect on the dynamical output of the model in order to be adequately estimated, the design of the injected current is important to determining the channel properties of neuron models. In Toth et al. (2011), the influence of the injected current on the ability to estimate the maximal conductance of the additional channel g_H is also tested. The I_H current activates (opens) when the neuron is hyperpolerized (driven

to greater negative membrane voltages) with a stablizing effect, preventing further hyperpolarization; It is found that the optimization is better able to estimate the g_H parameter when the system is driven by a strongly hyperpolarizing current. This indicates that for the success of any optimization procedure it is crucial to excerise as much of the model's dynamical range as possible within the data assimilation window.

Consideration must also be paid to the frequency content of the stimulus current, as the gating dynamics operate with different refractory time constants described by (8.8). As shown in Toth et al. (2011) very high frequencies, above those that are attenuated by the nueron's RC low-pass filter properties, do not assist the parameter estimation; while it remains important to use a complicated enough stimuls to perturb, and therefore have a better chance of distinguishing, the trajectories of dynamical systems with similar parameters and state variables.

9.4 NaKl & NaKlH Model Error

The same two Hodgkin-Huxley like models are used in Kostuk et al. (2012) and examined using the Markov chain Monte Carlo techniques described in chapter 6. This approach approximates errors in the model as additive gaussian noise according to section 5.4.2, and provides statistical information about the solution provided by the variational optimization.

As the studies in Kostuk et al. (2012) are twin-experiments, the actual noise in the 'measured data' is extremely small, akin to machine precision. So to make the situation more experimentally realistic, the noise in the voltage measurements is taken to have an RMS deviation of ± 0.25 mV, which is approximately what is seen in actual experiments. The noise coefficients used in the path integral are $1/2\sigma_m^2 = 8$ mV^{-2} and $1/\sigma_f^2 = 80$.

The data assimilation window for the Monte Carlo estimations is 40.96 ms, and uses 4096 time points in the path integral approximation with an integration timestep between collocated variables of $dt_{int} = 0.01$ ms.

The error associated with the posterior distribution of the parameter estimates was typically small, with relative standard deviations $\sigma/\bar{x} \approx 0.05$, for both the NaKl and NaKlH twin experiments. However for the case of assimilating data into an incorrect model, where data is generated from the NaKl model and assimilated into the NaKlH model again results in a conductance for the I_H current (which is not present in the data system) near zero. At the end of the statistics gathering phase, $g_H = 0.019 \pm 0.015$ mS/cm^2.

NaKl Model

$$\begin{aligned}\frac{dV}{dt} &= \frac{1}{C_m}\Big\{g_{Na}m_1^3h_2(E_{Na}-V) + g_K m_3^4(E_K - V) + g_L(E_L - V) + I_{inj}\Big\}\\ \frac{dx_i}{dt} &= \frac{1}{\tau_i(V)}(x_{i,\infty}(V) - x_i)\end{aligned} \tag{9.2}$$

with

$$\begin{aligned}x_{i,\infty}(V) &= \frac{1}{2} + \frac{1}{2}\tanh\frac{V - v_i}{dv_i}\\ \tau_i(V) &= \tau_a + \tau_b\Big(1 - \tanh^2\frac{V - v_i}{dv_i}\Big)\end{aligned}$$

and $i \in \{1, 2, 3\}$; each gating particle has $\mathbf{p} = \{v_i, dv_i, \tau_a, \tau_b\}$ for a total of $N_x = 4$ and $N_p = 19$, $C_m \equiv 1\mu F/cm^2$.

NaKlH Model

$$\begin{aligned}\frac{dV}{dt} &= \frac{1}{C_m}\Big\{g_{Na}m_1^3h_2(E_{Na}-V) + g_K m_3^4(E_K - V) + g_H h_4(E_H - V) + g_L(E_L - V) + I_{inj}\Big\}\\ \frac{dx_i}{dt} &= \frac{1}{\tau_i(V)}(x_{i,\infty}(V) - x_i)\end{aligned} \tag{9.3}$$

with

$$\begin{aligned}x_{i,\infty}(V) &= \frac{1}{2} + \frac{1}{2}\tanh\frac{V - v_i}{dv_i}\\ \tau_i(V) &= \tau_a + \tau_b\Big(1 - \tanh^2\frac{V - v_i}{dv_i}\Big)\end{aligned}$$

and $i \in \{1, 2, 3\}$, as in the NaKl model. In addition, the H-current is given by

$$\begin{aligned}h_\infty(V) &= \frac{1}{2} + \frac{1}{2}\tanh\frac{V - v_a}{dv_a}\\ \tau(V) &= \tau_a + \tau_b\tanh\frac{V - v_b}{dv_b}\end{aligned}$$

which has $\mathbf{p} = \{v_a, dv_a, \tau_a, \tau_b, v_b, dv_b\}$ for a total of $N_x = 5$ and $N_p = 26$, $C_m \equiv 1\mu F/cm^2$.

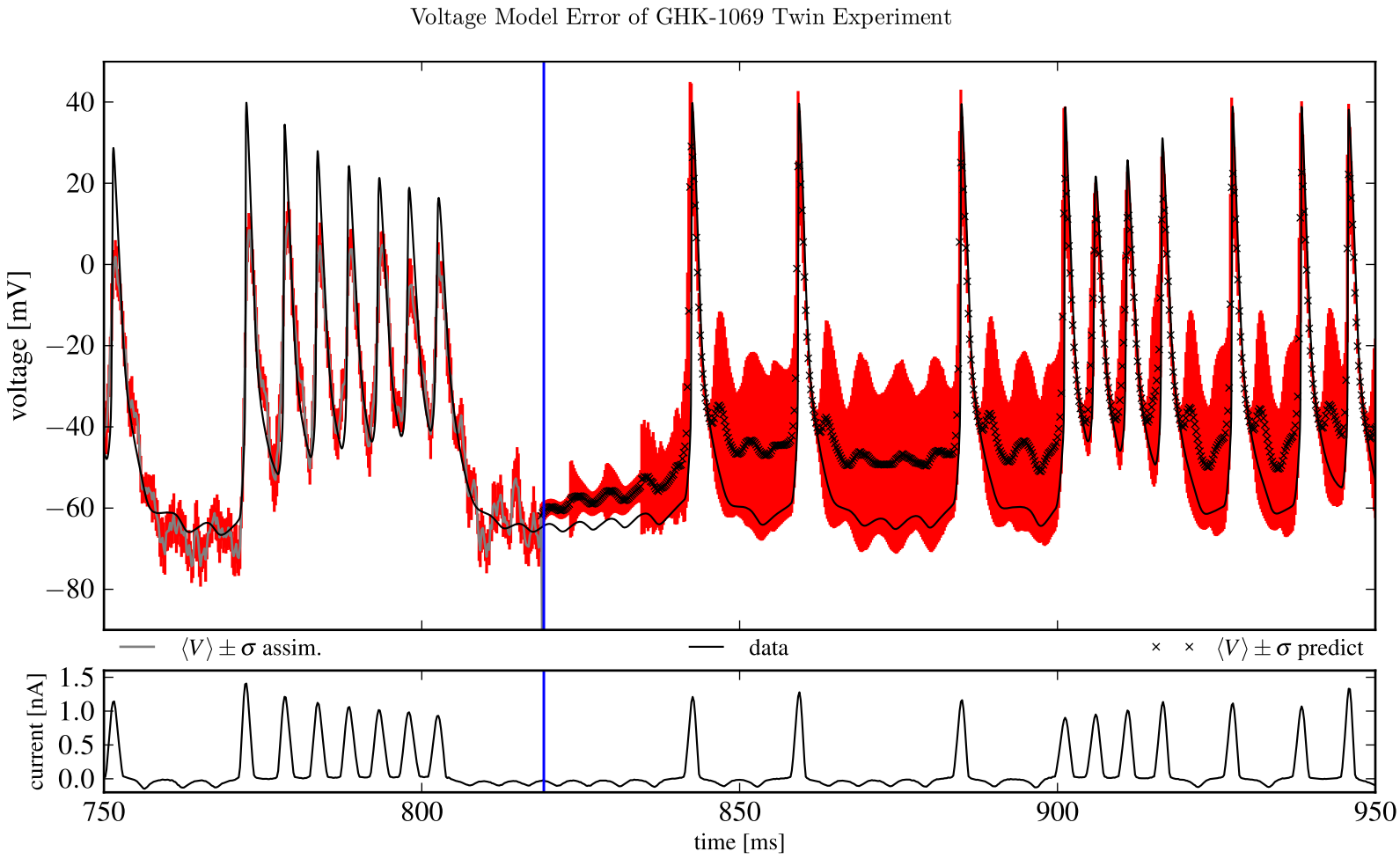

Figure 9.2: The measured 'data' voltage and the average voltage of the model are shown for the GHK-1069 model (9.1), both during a portion of the assimilation period (left) and subsequent prediction (right of the vertical line). The standard deviation of the model voltage is taken at every time point and plotted (red). Note the retention of the spike timimg, and also the shift in $\langle V_{model} \rangle$ during the prediction window relative to the data, although the data remains within approximately $\langle V_{model} \rangle \pm \sigma(V)$.

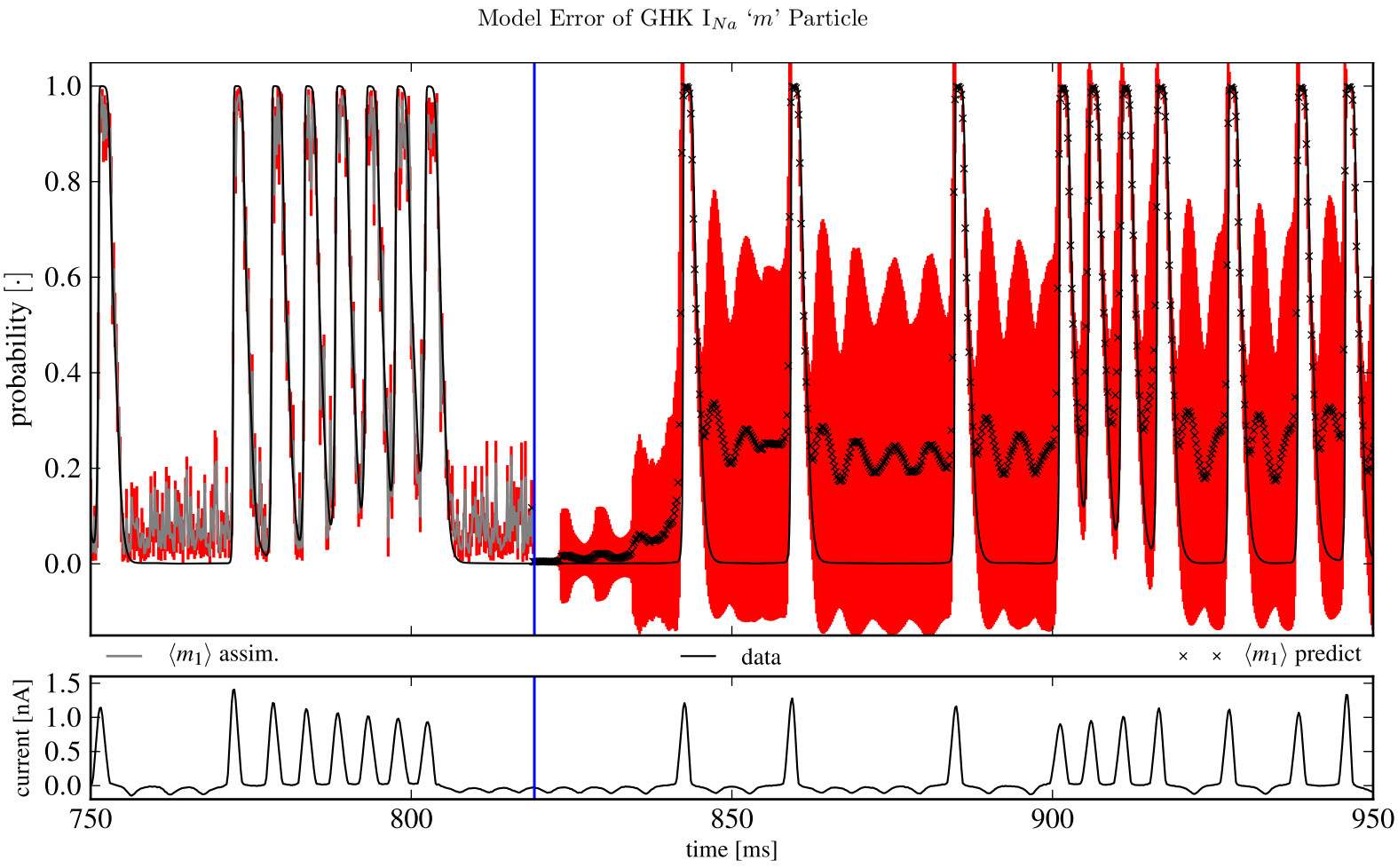

Figure 9.3: An unmeasured dimension, m_1, of the GHK-1069 model (9.1), during both the assimilation window (left) and the forward prediction of the resultant distribution. The data-system is shown for comparison as a thin black line, the model error is in red, expressed as $\pm\sigma_1$ from the mean $\langle m_1 \rangle$ which is shown in grey.

Chapter 10

High Vocal Center Neuron Results

In this chapter we report on the data assimilation results of voltage data taken from a single neuron[1], obtained from a current-clamp slice preparation of the High Vocal center of a male zebra finch, and assimilated into a Hodgkin-Huxley like model with calcium currents described by an approximation to the Goldman-Hodgkin-Katz equation.

The analysis proceeds in two parts, similar to what was done for the twin experiments of the previous chapter, where an adequate solution is found using a variational optimization method, followed by the collection of statistics about this solution using Markov Chain Monte Carlo.

10.1 Design of the Stimulus Current Protocol

Following the ideas in Abarbanel et al. (2011), the current stimulus used to drive the neuron should have enough structure to adequately excercise the full dynamical range of the neuron. A chaotic current, such as one of the dimensions of the Lorenz '63 system, has been shown to aid in the success of the associated optimization problem. However, a current that varies at too high of a frequency is lost by the RC filtering of the neuron.

In order for the channels with slower kinetics to be represented in the voltage measurement, constant step currents are useful to probe the passive membrane properties. Note that our ability to assimilate very long timeseries of data has improved since the development of the

[1] In our database it is neuron 201105_1_3 epoch 12.

current stimulus in figure 10.1, thus restrictions were placed upon the step currents to keep them each to 50 ms in duration. This severe of a restriction is no longer pertinent, and improved results on the slower channels, such as calcium and passive sodium, may be realizable with longer steps.

The scale of the injected current above and below $I_{inj} = 0$ that appropriately stimulates the neuron are different; the cell allows for a greater magnitude of current in the positive (depolarizing) direction and far less in the negative (hyperpolarizing) direction.

With these considerations in mind, the injected current stimulus in figure 10.1 is typical of the types used to stimulate neurons for data assimilation and is precisely the one used for the subsequent results.

10.2 Balanced Synchronization Results

The analysis proceeds in two steps; firstly, an optimization routine is used to find a quality solution to the data assimilation problem, and then the effects of model error and measurement noise about this solution are ascertained using the MCMC algorithm described in chapter 7. In this section we focus on the results from the first step.

The optimization of this neuron is done using the variational algorithm IPOPT (Wächter (2002)), which employs a collocated balanced synchronization objective function to solve

$$\begin{aligned} \text{minimize} \quad & \phi(\omega, \mathbf{y}_{T-1}) = \sum_{t=0}^{N_T-1} \sum_{i=1}^{N_M} \Big(x_i(t) - y_i(t) \Big)^2 + \gamma^2 \cdot \Big(1 + 5 \,\mathrm{Tanh}\Big(\frac{x_i(t)+25}{5} \Big) \Big)^2 \\ \text{subject to} \quad & \mathbf{x}(t+1) - \mathfrak{F}(\mathbf{x}(t), \mathbf{x}(t+1), \mathbf{p}, \mathbf{y}_{T-1}, \gamma) = 0 \quad . \end{aligned}$$

The Simpson rule with Hermite interpolation is used to discretize the model equations into $\mathfrak{F}$ that are presented as equality constraints. The factor multiplying the coupling γ may be interpreted as a voltage dependent penalty parameter. It has the effect of strongly weighting γ when the neuron is spiking, thus emphasizing solutions where the spiking occurs without the need for the coupling. Crafting an objective function that extracts the qualities of interest from the space of possible models remains an art form.

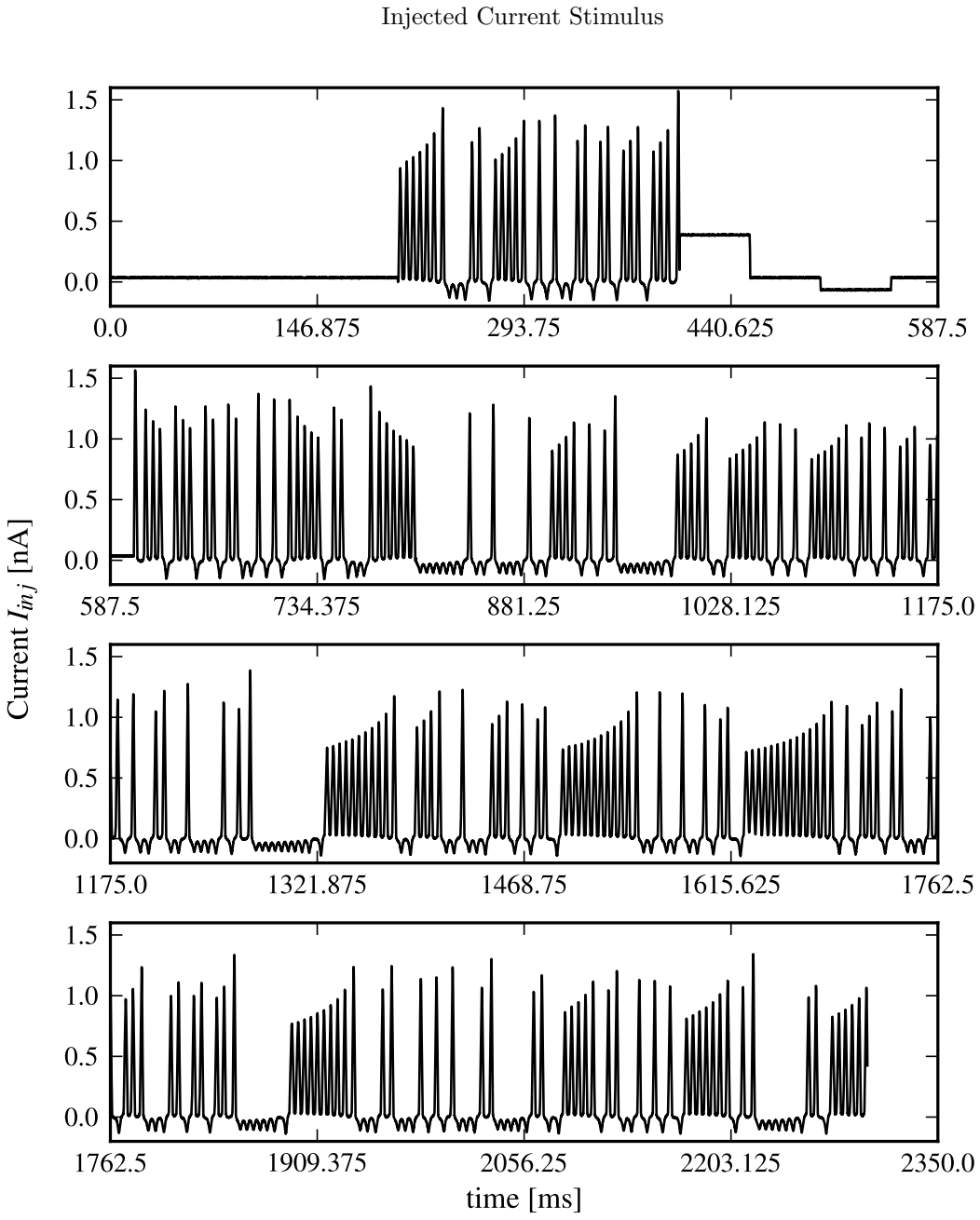

Figure 10.1: A four panel figure of a single injected current lasting approximately 2 seconds is shown. It proceeds from top to bottom, left to right. This is the current signal that is provided to the neuron that is analyzed extensively in this chapter.

10.2.1 Mixed-1280 Neuron Model

The neuron model that is used to assimilate the data contains currents based upon the Nernst approximation to the GHK equation. These Hodgkin-Huxley like currents operate according to kinetics described by the logistic function according to equation (8.9). In total, the model has $N_x = 12$ state variables and N_p=80 parameters, six for each gating particle. as well as parameters that describe the maximal conductances of each ion channel, and their reversal potentials. This accounts for two sodium channels, two potassium channels, an H current, two calcium currents and a leak current, in addition to the membrane capacitance. Not all of these channels are found to be present by the optimization routine.

We note that this model contains a completely redundant sodium channel, described by g_{Na2} and the state variables m_{10} and h_{11} in the model (10.1). It was included after noticing that many (physical) neurons show structure in their spike shape that cannot easily be accounted for by a single type of sodium channel, or that perhaps require a multi-compartmental model for their adequate description (Meliza and Margoliash (2012)). The variational optimization routine finds a solution with the maximal conductance of this channel, $g_{Na2} \approx 3\text{x}10^{-12}$, that is essentially zero, and so it appears in-consequential to the quality of the assimilated and predicted solutions, albeit necessary for the success of the optimization. An otherwise equivalent model with this channel omitted, i.e. $N_x = 10$ and $N_p = 67$, was attempted however the optimization failed to find a suitable solution. This may seem bizarre, but such is the world of nonlinear optimization.

$$
\begin{aligned}
\frac{dV}{dt} &= (g_{NaT} m_1^3 h_2 + g_{Na2} m_{10}^3 h_{11})(E_{Na} - V) \\
&\quad + (g_{KA1} m_3 + g_{KA2} m_4^4 h_5)(E_K - V) + g_L(E_L - V) + g_H h_6 (E_H - V) \\
&\quad + (g_{CaL} m_7^2 + g_{CaT} m_8^2 h_9)([Ca]_{ext} - [Ca]_{int}) 13/\mathrm{GHK}_e + C_m^{-1} I_{inj} \\
\frac{m_1}{dt} &= \frac{1 - m_1}{\exp[(G_{1\alpha} - z_{1\alpha}V)\mathrm{QKT}] + \epsilon_{1\alpha}} - \frac{m_1}{\exp[(G_{1\beta} + z_{1\beta}V)\mathrm{QKT}] + \epsilon_{1\beta}} \\
\frac{h_2}{dt} &= \frac{1 - h_2}{\exp[(G_{2\alpha} + z_{2\alpha}V)\mathrm{QKT}] + \epsilon_{2\alpha}} - \frac{h_2}{\exp[(G_{2\beta} - z_{2\beta}V)\mathrm{QKT}] + \epsilon_{2\beta}} \\
\frac{m_3}{dt} &= \frac{1 - m_3}{\exp[(G_{3\alpha} - z_{3\alpha}V)\mathrm{QKT}] + \epsilon_{3\alpha}} - \frac{m_3}{\exp[(G_{3\beta} + z_{3\beta}V)\mathrm{QKT}] + \epsilon_{3\beta}} \\
\frac{m_4}{dt} &= \frac{1 - m_4}{\exp[(G_{4\alpha} - z_{4\alpha}V)\mathrm{QKT}] + \epsilon_{4\alpha}} - \frac{m_4}{\exp[(G_{4\beta} + z_{4\beta}V)\mathrm{QKT}] + \epsilon_{4\beta}} \\
\frac{h_5}{dt} &= \frac{1 - h_5}{\exp[(G_{5\alpha} + z_{5\alpha}V)\mathrm{QKT}] + \epsilon_{5\alpha}} - \frac{h_5}{\exp[(G_{5\beta} - z_{5\beta}V)\mathrm{QKT}] + \epsilon_{5\beta}} \qquad (10.1) \\
\frac{h_6}{dt} &= \frac{1 - h_6}{\exp[(G_{6\alpha} + z_{6\alpha}V)\mathrm{QKT}] + \epsilon_{6\alpha}} - \frac{h_6}{\exp[(G_{6\beta} - z_{6\beta}V)\mathrm{QKT}] + \epsilon_{6\beta}} \\
\frac{m_7}{dt} &= \frac{1 - m_7}{\exp[(G_{7\alpha} - z_{7\alpha}V)\mathrm{QKT}] + \epsilon_{7\alpha}} - \frac{m_7}{\exp[(G_{7\beta} + z_{7\beta}V)\mathrm{QKT}] + \epsilon_{7\beta}} \\
\frac{m_8}{dt} &= \frac{1 - m_8}{\exp[(G_{8\alpha} - z_{8\alpha}V)\mathrm{QKT}] + \epsilon_{8\alpha}} - \frac{m_8}{\exp[(G_{8\beta} + z_{8\beta}V)\mathrm{QKT}] + \epsilon_{8\beta}} \\
\frac{h_9}{dt} &= \frac{1 - h_9}{\exp[(G_{9\alpha} + z_{9\alpha}V)\mathrm{QKT}] + \epsilon_{9\alpha}} - \frac{h_9}{\exp[(G_{9\beta} - z_{9\beta}V)\mathrm{QKT}] + \epsilon_{9\beta}} \\
\frac{m_{10}}{dt} &= \frac{1 - m_{10}}{\exp[(G_{10\alpha} - z_{10\alpha}V)\mathrm{QKT}] + \epsilon_{10\alpha}} - \frac{m_{10}}{\exp[(G_{10\beta} + z_{10\beta}V)\mathrm{QKT}] + \epsilon_{10\beta}} \\
\frac{h_{11}}{dt} &= \frac{1 - h_{11}}{\exp[(G_{11\alpha} + z_{11\alpha}V)\mathrm{QKT}] + \epsilon_{11\alpha}} - \frac{h_{11}}{\exp[(G_{11\beta} - z_{11\beta}V)\mathrm{QKT}] + \epsilon_{11\beta}}
\end{aligned}
$$

Note that in this model we have taken the specific membrane current $\tilde{C}_m \equiv 1\mu\mathrm{F}\cdot\mathrm{cm}^{-2}$ as a definition, relegating any physical discrepancy into the maximal conductances, g_X. State variables denoted with an m indicate gating particles that open with increasing (polarizing) voltage, while those denoted with an h open with decreasing (hyperpolarizing) voltage.

The GHK expansion for the calcium current given by

$$
\begin{aligned}
\mathrm{GHK}_e = &\Big(1 + \frac{V}{26}\Big(1 + \frac{V}{39}\Big(1 + \frac{V}{52}\Big(1 + \frac{V}{65}\Big(1 + \frac{V}{78}\Big(1 + \frac{V}{91}\Big(1 + \frac{V}{104}\Big(1 + \frac{V}{117}\Big(1 + \frac{V}{130}\Big(1 + \frac{V}{143} \\
&\times \Big(1 + \frac{V}{156}\Big(1 + \frac{V}{169}\Big(1 + \frac{V}{182}\Big(1 + \frac{V}{195}\Big(1 + \frac{V}{208}\Big(1 + \frac{V}{221}\Big(1 + \frac{V}{234}\Big(1 + \frac{V}{247}\Big(1 + \frac{V}{260} \\
&\times \Big(1 + \frac{V}{273}\Big(1 + \frac{V}{286}\Big(1 + \frac{V}{299}\Big(1 + \frac{V}{312}\Big(1 + \frac{V}{325}\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big)\Big) \; ,
\end{aligned}
$$

and was suggested by Nogaret (2011). The internal concentration of calcium does not vary in this model and is fixed at $[Ca]_{int} = 100$ nM. Also, QKT$= q/k_BT$ is not varied during the optimization procedure, but is fixed for this specific neuron according to the average temperature of the cell over the course of this recording, $\langle T \rangle = 32.87 \pm 0.06$ °C, resulting in QKT$= 0.037756$ mV^{-1}.

10.2.2 Voltage Assimilation and Prediction

The assimilation window begins at $t = 150$ ms (omitting a section of constant current) and lasts for a duration of 1500 ms; with the data sampled at a frequency of 50 kHz, this amounts to $N_T = 7.5\text{x}10^4$ data points. This is a majority of the individual recording epoch for this neuron, which lasts about 2300 ms in total. Such a long window is used to get high precision parameters, however results on twin experiments have been successfully obtained with windows as short as 50 ms. Of course, slower varying currents require longer assimilated timeseries.

After the assimilation procedure is complete[2] the state variables of all dimensions of the model are estimated, including the ones that are not measured directly such as those which describe the ion channel gating probability, as are the model parameters. The model equations are then integrated forward using the optimized parameters from the last time point of the assimilation window, $\mathbf{x}(N_T - 1)$, and with the continuation of the same injected current that was presented to the neuron to generate the prediction.

The entire voltage recording, along with the model voltage from both the assimilation windown during the optimization and the prediction after the assimilation window is shown in figure 10.2. The overlap between the two is quite remarkable, indicating that the model is capable of reproducing both the spike timing and the voltage behavior in the hyperpolarized region. The injected current is not shown, but the current that was used to stimulate this neuron is shown in figure 10.1. A detailed version of the voltage comparison between the neuron and the model is shown, accompanied by the injected current, in figure 10.3.

The parameters of this model that resulted from this optimization procedure, as well as the statistical values (average and standard deviation) that result from the subsequent MCMC calculation, are shown in three tables 10.1, 10.2 and 10.3. We will return to them after discussing the details of the Markov chain Monte Carlo estimation.

10.3 Single Markov Chain Measurement Noise and Model Error Estimates

The result of the variational optimization procedure (during the assimilation window) is used as the starting location ω_0 for the single Markov chain Monte Carlo calculation. This consists of the full $N_T N_x$ dimensional timeseries $\{\mathbf{x}(0), \ldots, \mathbf{x}(N_T)\}$ and N_p parameters $\mathbf{p}$; ω is a point in a $49,232$ dimensional search space. Due to the hardware restrictions of the GPU

[2]The optimization process took 73.3 hours on a single 3.2GHz CPU.

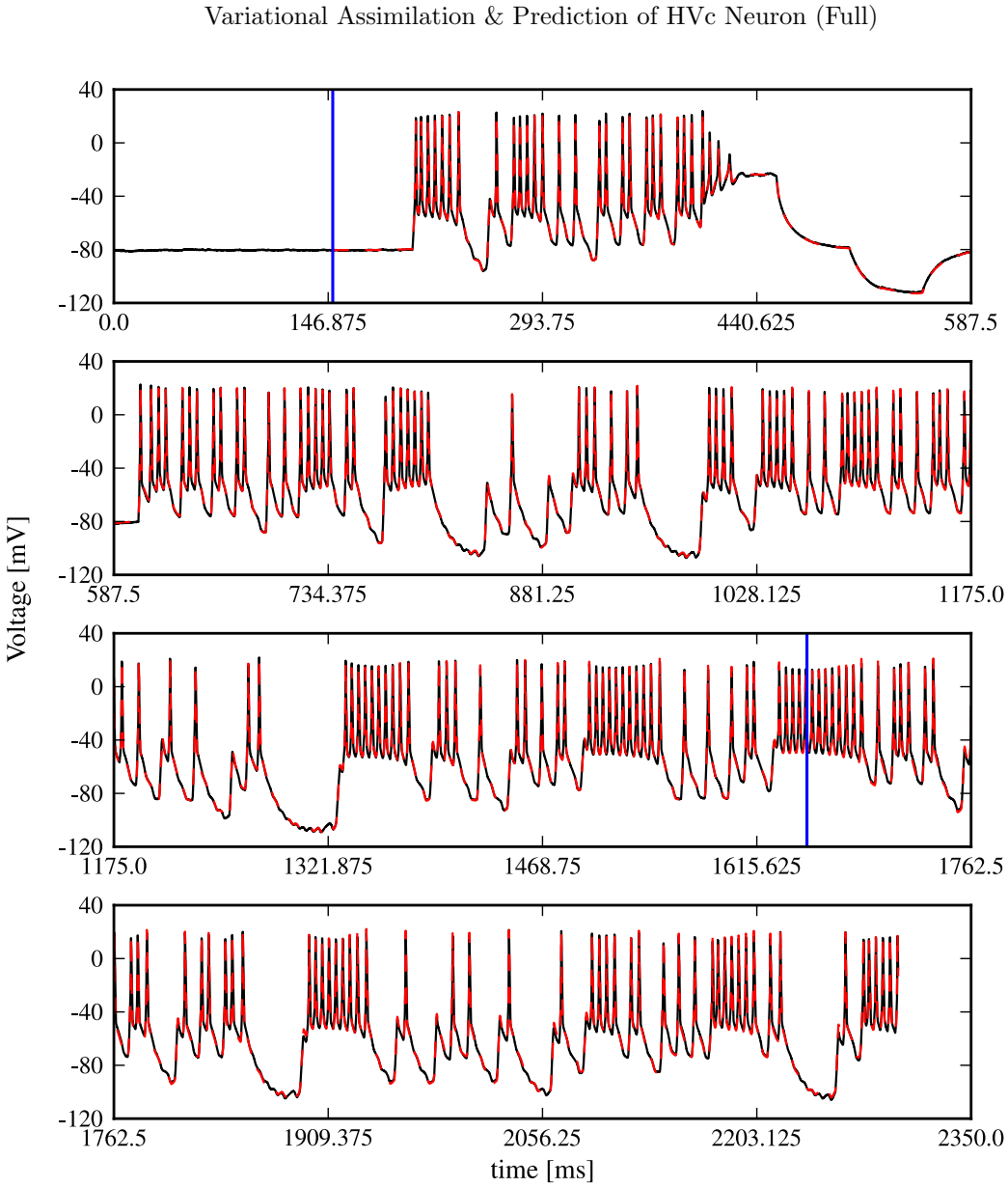

Figure 10.2: The full $\approx$ 2.3 sec of data recorded for this neuron is shown, wrapped to four consecutive panels from top to bottom. Both the measured voltage (solid black) and the model assimilation and prediction (dashed red) are given, where the vertical lines denote the assimilation window $t \in (150, 1650)$ ms and the prediction with optimized parameters (and zero synchronization coupling to the data) follows until the end of the epoch. The injected current is omitted, however the waveform used is shown in figure 10.1.

on which the MCMC calculation is run, the total number of time points that can be used is $2^{12} = 4096$. The timeseries is downsampled, so that $dt_{obs} = 0.20$ ms ($dt_{int} = 0.02$) and the full assimilation window is 819.2 ms in duration.

This starting point ω_0 is then iterated forward according to the method described in chapter 6 for 5x10^4 discared iterations, followed by 1x10^5 iterations of statistics collection, during which $N_E = 100$ samples are taken (uniformly over the statistics iterations). Note that this many iterations during the statistics collection phase involves 4.91x10^9 perturbations and acceptance evaluations to each dimension of $\mathbf{p}$.

The values for the noise allowed in the calculation is the same as was used for the GHK-1069 twin experiment,

$$\frac{1}{2\sigma_m^2} = 7.41 \text{ mV}^{-2}$$
$$\frac{1}{2\sigma_f^2} = 54.91 \quad .$$

10.3.1 Chain Equilibrium

Figure 10.4 shows the value of the action as a function of the Markov chain evolution, in terms of the algorithm interation, k. The action adjusts, during the burn-in phase, to a value commesurate with the measurement error σ_m taken from the experimental voltage trace. Afterwards it remains approximately constant (see section 9.1.2), and N_E samples are collected uniformly during this phase to estimate the distribution over the entire path $\mathbf{x}_{T-1}$ and parameters $\mathbf{p}$. The resultant ensemble of samples $\{\hat{\omega}\}$ are individually integrated forward, and variance of this integrated distribution is plotted as error bars in the figures 10.5, 10.6 and 10.7.

10.3.2 Optimized Parameters

In a three part table 10.1, 10.2 and 10.3, are listed all of the resultant parameters of the optimization. The first table collects the reversal potential for the various ionic species and the conductances of the neuron, along with the channel kinetic parameters from the Na_T channel. The following two tables contain the parameters split based upon their relevance to the model, where those channels that are *not* necessary for the fidelity of the optimized prediction are in the second table. The parameter bounds for both the IPOPT optimization and the assessment of statistics by MCMC are given in table D.1 of the appendix.

Biophysically reasonable bounds are used for the reversal potentials, as the optimization of them often proves to be problematic, resulting in a parameter that is pushed to one of the

bounds. This occurs for both the E_{Na} and E_K potentials that are strained to their extremes, which is a common occurrence with optimized neuron models, but does not occur for the I_H current reversal potential. The irrelevance of a channel current to the model voltage is declared when the maximal conductance of that channel is found by the optimization to be approximately zero, meaning that the kinetic parameters that influence these gating particles are also irrelevant.

From table 10.1 we see that the main contributions to the model's voltage response are given by the I_{A2} and I_{Na1} currents. The extreme value of the potassium current and its relative dominance over the sodium is unusual, although not completely unheard of, and is known to give similar spike timings as models with $g_{Na}/g_K \gg 1$, where subtle variations are distinguishable in the spike shape (Prinz et al. (2011)). The contributions of the additional currents that are present, such as I_{A1} and I_{CaL}, fall well within biophysically reasonable ranges as measured by their channel conductances (Golowasch et al. (1999); Goldman et al. (2000)).

10.3.3 Optimized Channel Currents

In figure 10.6, the model error of the unmeasured variable $m_1(t)$ is shown for both a portion of the data assimilation window and the prediction afterwards. This variable is the m gating variable of the sodium channel, which opens as the cell is depolarized and is reponsible for letting sodium ions flow into the cell, causing the sharp rise in voltage that is characteristic of an action potential. What can be seen in this figure is that the model error allows for a significant amount of variance in this variable, plotted as the standard deviation $\sigma_{m1}(t)$. This, and the fact that the average value $\langle m_1(t) \rangle$ is elevated (relative to its location during the assimilation time period) would have the effect of raising the overall potential of the cell. This likely contributes to the wide range of values in the predicted value of the neuron voltage.

From a data assimilation standpoint, this shows that the m_1 model variable is very sensitive to errors in the model, and in a way that has a significant impact on the predicted measurable (voltage) data. Contrast this with the m_4 model variable, which is shown in figure 10.7, and for which the same level of noise in the model dynamics does not show nearly as detrimental an impact on the predicted dynamics relative to its behavior during the assimilation period.

Comparing the dynamical trajectory of an unmeasured variable between the assimilation window and the prediction period, as done in figures 10.6 and 10.7, reveals the impact that the measurement(s) have on maintaining the Markov chain near the optimized solution. If one presumes a certain level of resilience in this biological system, or any system under study, they would likely demand similar resilience from a quality model as well. The experimenter can

therefore use this information to focus their efforts on obtaining improved models or parameter sets related to those variables that are most sensitive to noise in the model, as improving these areas are likely to have the most impact on improvements to the subsequent predictions.

10.4 Discussion of Optimized Neuron Results

This neuron is putatively an interneuron due to the relatively high frequency of spiking and small size ($20-25$ pF). We posit that the high frequency of spiking assisted the optimization process in arriving at a quality solution, as the full range of dynamics are excercised allowing for the distinction between candidate model solutions during the assimilation process.

This neuron-model assimilation result is one of the more successful ones, in terms of the fidelity of the future prediction, that has been obtained by the author to date. Collecting more quality results from a population of neurons in HVc and uncovering possible classifications between inter-neurons and projection neurons is a current work in progress (Abarbanel et al. (2012)).

As there is no measured data with which to make a comparison for the unmeasured gating particles, we show instead the resultant rate kinetics of the assimilated parameters in appendix B. To the neurobiologist, this offers insight into whether the optimized parameters plausibly represent the bio-physical channel dynamics of the neuron under study. Some of them, e.g. the I_{Na1} 'm' particle (m_1) and the I_{A2} 'h' particle (h_5), do not exhibit kinetics commesurate with known biological behavior; while others, e.g. the I_{Na1} 'h' particle (h_2), I_{CaL} 'm' (m_7), and the I_{A1} 'm' particle (m_3), do exhibit plausible behavior.

As we are not attempting to identify specific channels, futher comparison is beyond the scope of this work, however hopefully we have left the reader with the impression that such a task is answerable within the general framework of data assimilation.

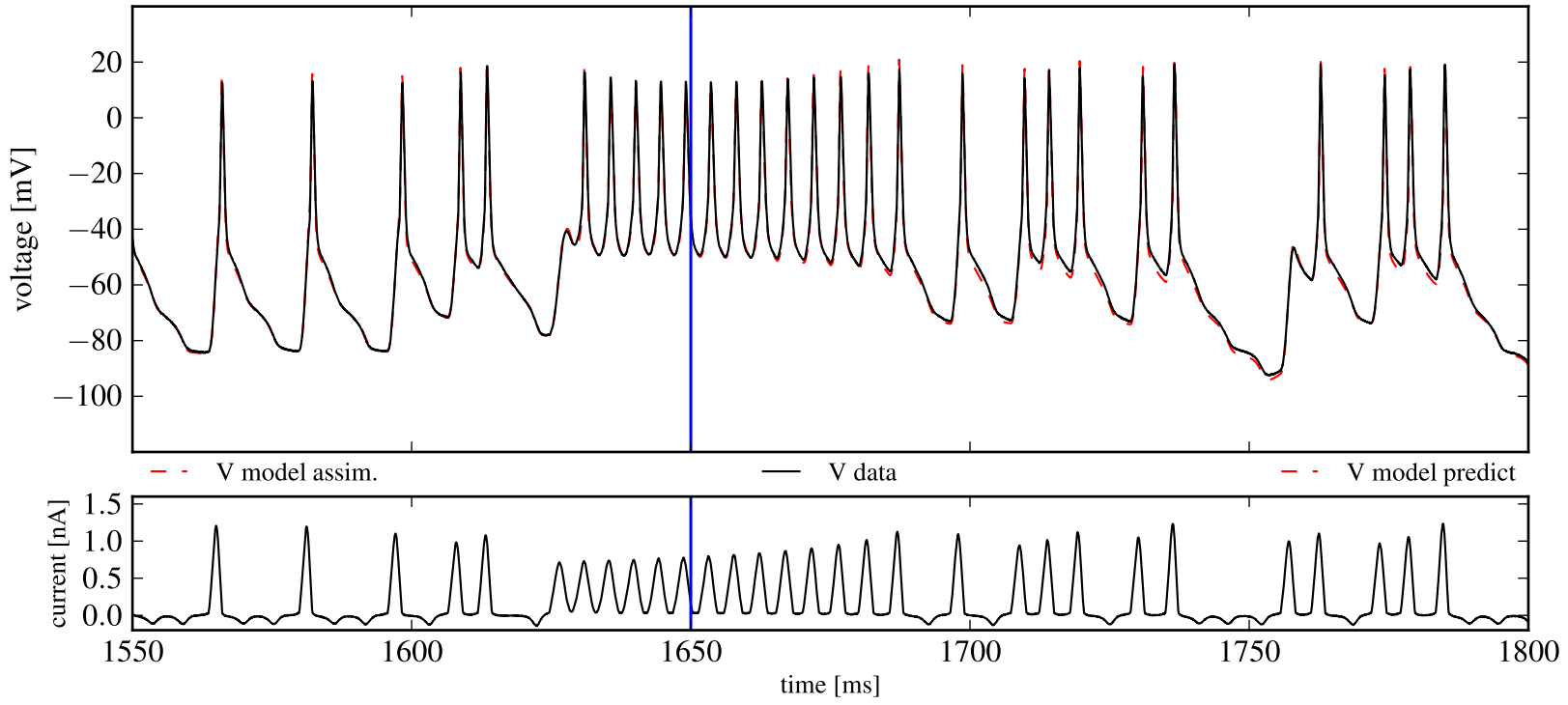

Figure 10.3: A portion of fig. 10.2, showing more detail of the measured voltage waveform of the neuron (solid black) as well as the optimized result of the model (dashed red).

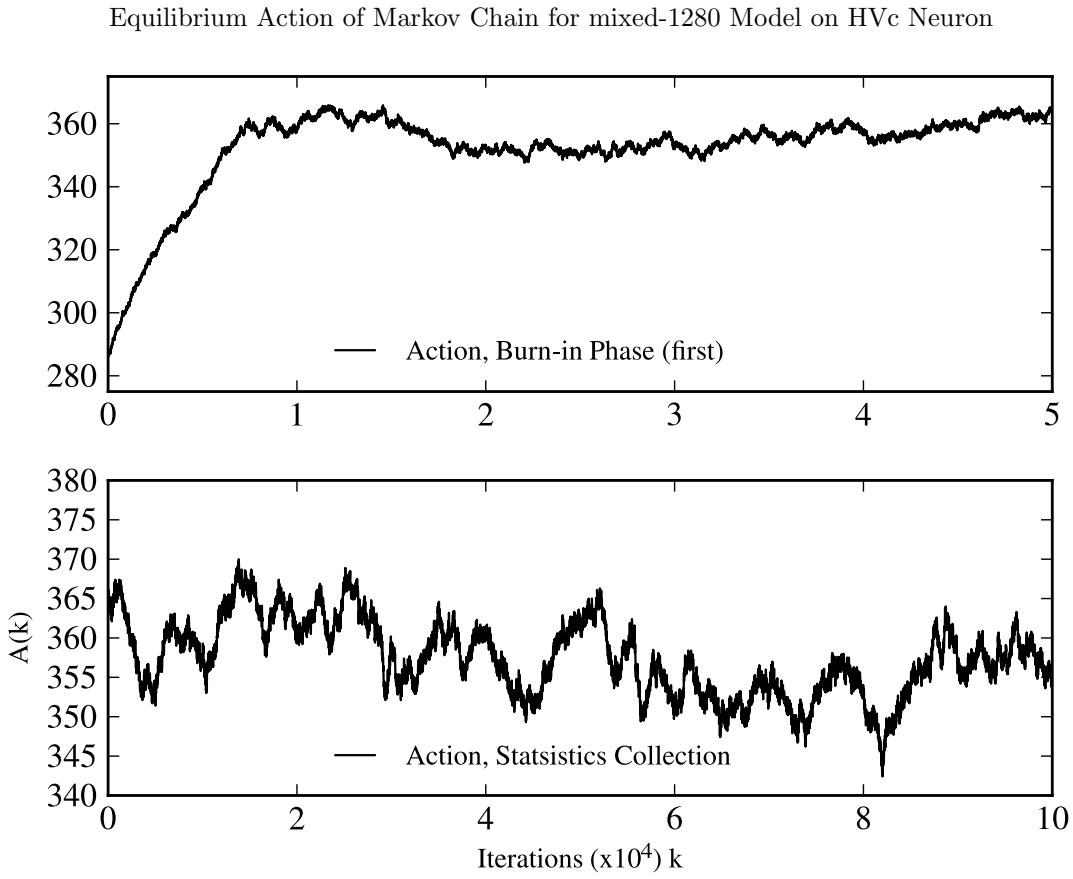

Figure 10.4: The action of the Markov chain is shown for the $5\text{x}10^4$ burn-in iterations (upper panel) along with the $1\text{x}10^5$ statistics collection iterations (lower panel). For the chain to be in equilibrium, the overall trend of $\langle A \rangle$ should not be increasing nor decreasing relative to the size of the fluctuations.

Table 10.1: Part I. The parameters that result from the variational optimization of an HVc neuron, and the assessment of noise about this solution. The bounds that were used for both procedures can be found in the Appendix. Some parameters that appear in Part I. appear again in later tables for convenience.

Parameter	Units	Variational	$\langle p \rangle$	σ_p
C_m^{-1}	nF^{-1}	3.941666e+01	50.24832	8.29715
E_{Na}	mV	4.500000e+01	59.20824	7.75353
E_K	mV	-8.500000e+01	-122.30239	4.24563
E_L	mV	-7.269787e+01	-72.75626	21.62558
E_H	mV	-4.367534e+01	-48.71512	11.54734
g_L	mS/cm^2	3.758373e-12	0.03023	0.01528
$[Ca_{ext}]$	mM	8.999548e+00	9.43534	2.65942
g_{Na1}	mS/cm^2	7.341148e+00	10.13042	0.82918
g_{Na2}	mS/cm^2	2.911103e-12	0.02138	0.01301
g_{A1}	mS/cm^2	2.470600e-01	0.21127	0.06158
g_{A2}	mS/cm^2	5.100000e+02	728.64011	34.61997
g_H	mS/cm^2	1.272465e-12	0.00927	0.00544
g_{CaL}	mS/cm^2	2.049909e-02	0.02006	0.00543
g_{CaT}	mS/cm^2	2.516841e-13	0.01953	0.01101
g_{Na1}	mS/cm^2	7.341148e+00	10.13042	0.82918
$G_{\alpha m}$	mV	-3.402891e+02	-339.01536	99.31301
$z_{\alpha m}$	$[\cdot]$	7.000000e+00	7.82044	1.88100
$\epsilon_{\alpha m}$	$[\cdot]$	4.213660e-02	0.05171	0.00670
$G_{\beta m}$	mV	4.014904e+01	41.84444	10.53611
$z_{\beta m}$	$[\cdot]$	5.000000e+00	4.95156	1.38977
$\epsilon_{\beta m}$	$[\cdot]$	4.000000e-02	0.04995	0.00602
$G_{\alpha h}$	mV	2.754432e+02	275.12590	81.13355
$z_{\alpha h}$	$[\cdot]$	5.000000e+00	5.56253	1.18198
$\epsilon_{\alpha h}$	$[\cdot]$	3.209158e-01	0.33530	0.09540
$G_{\beta h}$	mV	-2.239083e+00	-2.41987	0.63264
$z_{\beta h}$	$[\cdot]$	2.877228e+00	2.83652	0.87617
$\epsilon_{\beta h}$	$[\cdot]$	4.000000e-02	0.05041	0.00563

Table 10.2: Part II. The parameters that result from the variational optimization of an HVc neuron, and the assessment of noise about this solution. The bounds that were used for both procedures can be found in the Appendix.

Parameter	Units	Variational	$\langle p \rangle$	σ_p
g_{A1}	mS/cm^2	2.470600e-01	0.21127	0.06158
$G_{\alpha m}$	mV	-1.889764e+02	-188.69206	59.56649
$z_{\alpha m}$	[·]	7.000000e+00	7.01804	1.97803
$\epsilon_{\alpha m}$	[·]	1.739546e+00	1.76532	0.52649
$G_{\beta m}$	mV	-5.309936e+02	-560.41432	153.05653
$z_{\beta m}$	[·]	2.698601e+00	2.79542	0.80928
$\epsilon_{\beta m}$	[·]	1.438101e+02	123.75122	39.27438
g_{A2}	mS/cm^2	5.100000e+02	728.64011	34.61997
$G_{\alpha m}$	mV	-4.952332e+02	-497.24613	145.34728
$z_{\alpha m}$	[·]	5.000000e+00	4.85165	1.51840
$\epsilon_{\alpha m}$	[·]	2.294526e+00	2.59608	0.61230
$G_{\beta m}$	mV	9.801215e+00	9.90350	2.81759
$z_{\beta m}$	[·]	9.042192e-01	0.89169	0.25136
$\epsilon_{\beta m}$	[·]	4.000000e-02	0.05004	0.00600
$G_{\alpha h}$	mV	2.744450e+02	281.80443	81.41812
$z_{\alpha h}$	[·]	7.000000e+00	7.57914	2.00267
$\epsilon_{\alpha h}$	[·]	1.335403e+01	13.45143	4.46756
$G_{\beta h}$	mV	-2.781500e+02	-286.48243	75.71445
$z_{\beta h}$	[·]	7.000000e+00	6.95893	2.03511
$\epsilon_{\beta h}$	[·]	9.196644e+00	8.72900	2.62201
g_{CaL}	mS/cm^2	2.049909e-02	0.02006	0.00543
$G_{\alpha m}$	mV	-9.044624e+01	-87.86151	22.80013
$z_{\alpha m}$	[·]	1.957267e+00	1.83452	0.51937
$\epsilon_{\alpha m}$	[·]	1.289876e-01	0.12431	0.03600
$G_{\beta m}$	mV	1.631071e+02	169.90936	45.54412
$z_{\beta m}$	[·]	7.000000e+00	7.04800	1.94155
$\epsilon_{\beta m}$	[·]	4.231150e+00	3.97929	1.15302
$[Ca]_{ext}$	mM	8.999548e+00	9.43534	2.65942

Table 10.3: Part III. The parameter values of the channels with a maximal conductance ≈ 0, and therefore presumed absent from the neuron in this form.

Parameter	Units	Variational	$\langle p \rangle$	σ_p
g_{Na2}	mS/cm^2	2.911103e-12	0.02138	0.01301
$G_{\alpha m}$	mV	-2.081181e+02	-202.86855	55.12413
$z_{\alpha m}$	[·]	2.489339e+00	2.49183	0.61445
$\epsilon_{\alpha m}$	[·]	4.080179e+02	400.87347	118.25967
$G_{\beta m}$	mV	-9.230537e+00	-10.11642	2.58550
$z_{\beta m}$	[·]	3.495205e+00	3.54056	0.93217
$\epsilon_{\beta m}$	[·]	4.089970e+02	440.12758	113.74883
$G_{\alpha h}$	mV	-1.234778e+01	-12.01392	3.76722
$z_{\alpha h}$	[·]	3.493162e+00	3.86045	1.02860
$\epsilon_{\alpha h}$	[·]	4.098306e+02	416.41388	123.08780
$G_{\beta h}$	mV	-2.361097e+02	-239.61735	66.51639
$z_{\beta h}$	[·]	2.728242e+00	2.68865	0.89015
$\epsilon_{\beta h}$	[·]	4.089542e+02	398.11070	106.06683
g_H	mS/cm^2	1.272465e-12	0.00927	0.00544
$G_{\alpha h}$	mV	-1.164327e+01	-10.74471	3.31970
$z_{\alpha h}$	[·]	3.493663e+00	3.73833	0.97009
$\epsilon_{\alpha h}$	[·]	4.085324e+02	399.80118	107.57280
$G_{\beta h}$	mV	-2.128236e+02	-205.23405	60.07804
$z_{\beta h}$	[·]	2.498296e+00	2.61628	0.72541
$\epsilon_{\beta h}$	[·]	4.076516e+02	388.80110	109.39778
g_{CaT}	mS/cm^2	2.516841e-13	0.01953	0.01101
$G_{\alpha m}$	mV	-2.121085e+02	-218.07943	65.02975
$z_{\alpha m}$	[·]	2.481215e+00	2.47620	0.73729
$\epsilon_{\alpha m}$	[·]	4.117448e+02	367.89505	106.05881
$G_{\beta m}$	mV	4.745622e+01	41.22607	12.28202
$z_{\beta m}$	[·]	3.443896e+00	3.34213	1.02725
$\epsilon_{\beta m}$	[·]	4.114233e+02	388.60514	123.15097
$G_{\alpha h}$	mV	4.315966e+01	46.35745	11.61695
$z_{\alpha h}$	[·]	3.443573e+00	3.60238	1.06053
$\epsilon_{\alpha h}$	[·]	4.123252e+02	379.32965	124.03386
$G_{\beta h}$	mV	-2.375531e+02	-234.63409	66.82619
$z_{\beta h}$	[·]	2.718202e+00	2.68636	0.79592
$\epsilon_{\beta h}$	[·]	4.124946e+02	443.37927	122.06687

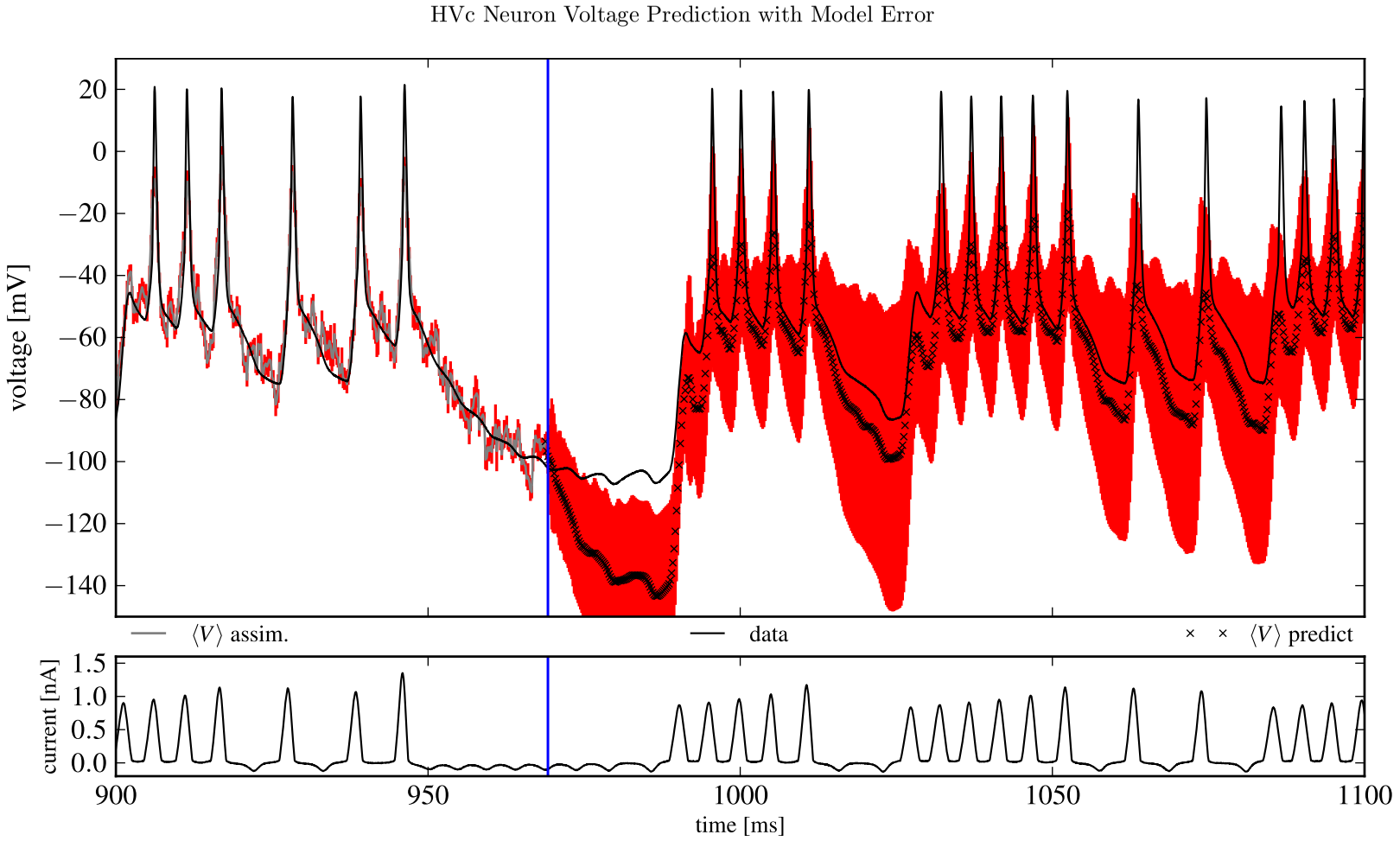

Figure 10.5: The results of the Markov Chain calculation (ch. 6) on experimental voltage data taken from a neuron of HVc. The vertical line denotes the separation between the data assimilation window (left) and the prediction of the resultant distribution (right). The optimized model parameters $\langle p \rangle \pm \sigma_p$ are given in the three tables 10.1, 10.2, and 10.3. The bounds used for the estimation procedure are given in the appendix.

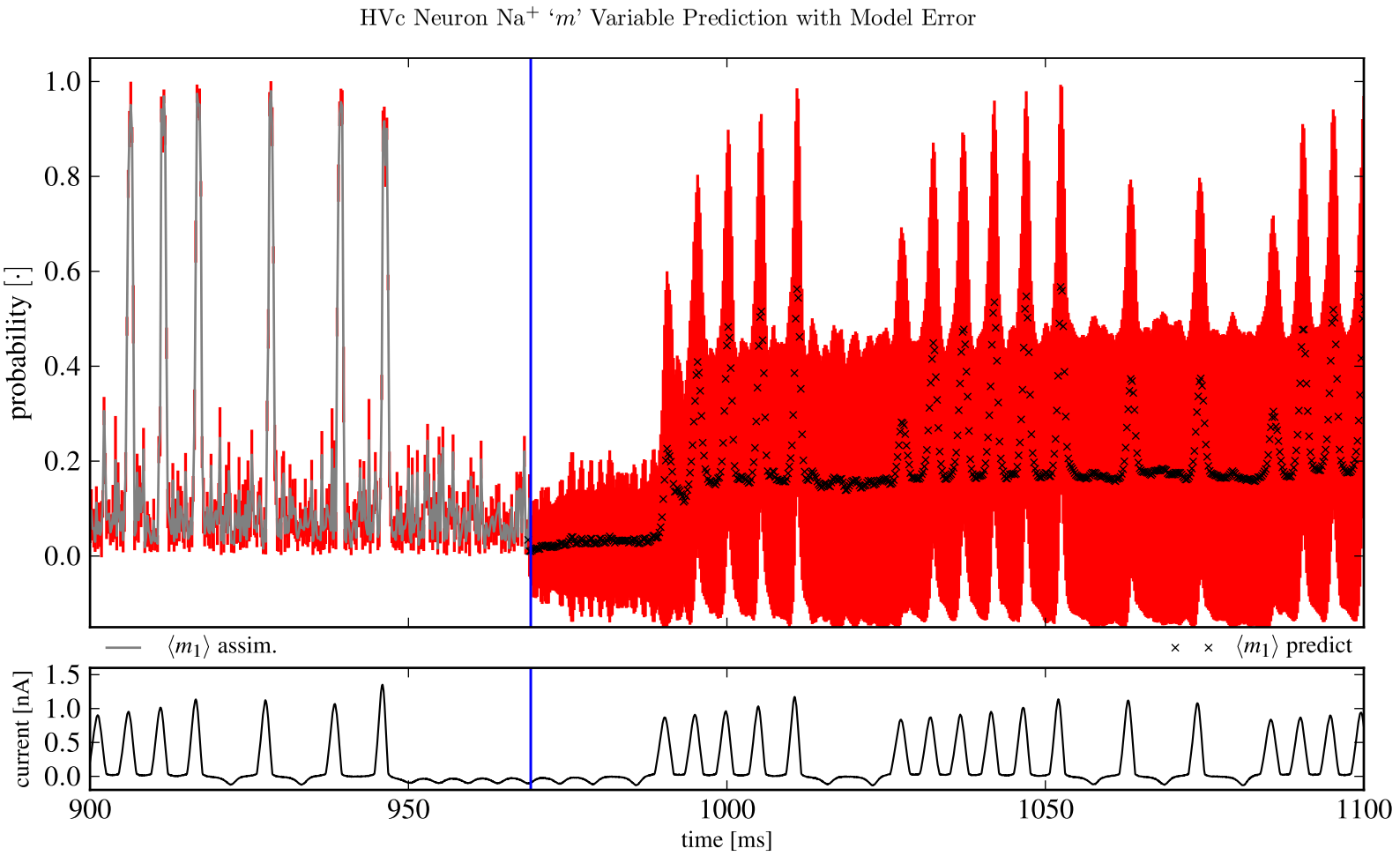

Figure 10.6: The model error in m_1 of the assimilated HVc neuron is plotted as an average (grey/black) and standard deviation (red), this is the transient sodium channel's m particle. Notice the overall shift in the average amount that this particle is open during the prediction, also the reduced dynamical range of the channel average relative to during the assimilation window. Although the standard deviation of this variable drops to negative values, this is due to it being plotted as $\langle m_1 \rangle \pm \sigma_2$; the average remains within the numerically stable region.

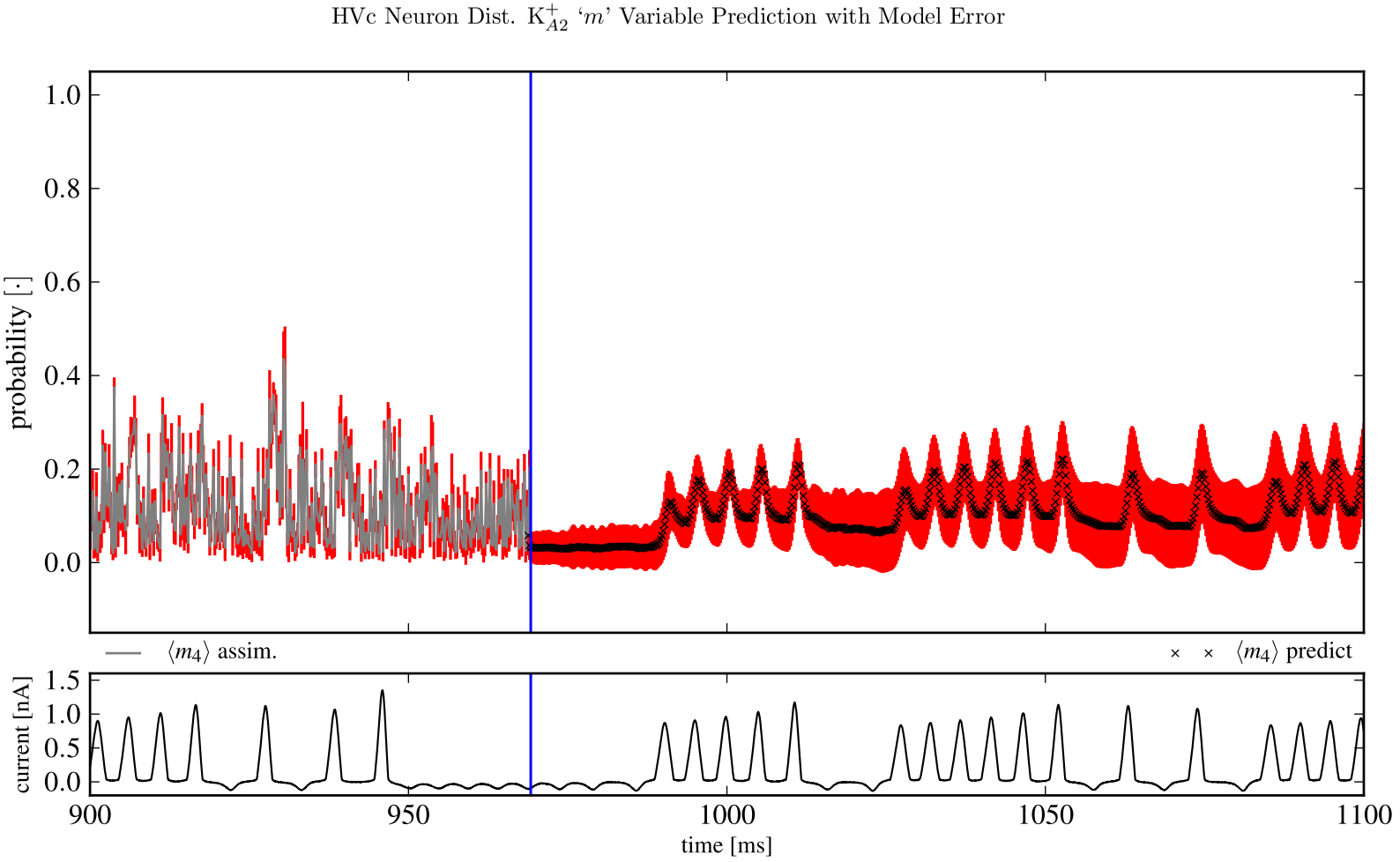

Figure 10.7: The model error in m_4 of the assimilated HVc neuron is plotted as an average (grey/black) and standard deviation (red), this is the K_{A2} channel's m particle. This channel is heavily expressed in the optimized neuron model. The average behavior of this particle during the assimilation window and prediction period are qualitatively similar, indicating relative stability of this state variable to model error.

Appendix A

Notation Index

A great effort has been made to try and keep the notation consistent and to not re-use variables whenever possible throughout this work. Collected here is a list of the most important and commonly used symbols along with their intended meaning within.

N_x The dimension of the model dynamical system.

N_T The number of time points in the data assimilation window, $t \in (0, N_T - 1)$

$\mathbf{x}(t)$ The N_x dimensional state vector at time t.

$\mathbf{x}_T = \{\mathbf{x}(t)\}\Big|_{t=0}^{t=T} = \{\mathbf{x}(0), \ldots, \mathbf{x}(N_T)\}$

The $N_T + 1$ element timeseries of the model from $t = 0$ to $t = N_T$

$\mathbf{p}$ The N_p dimensional vector of model parameters.

$\mathbf{f}(\mathbf{x}, \mathbf{p})$ The dynamical equations that describe the (uncoupled) model.

$\mathbf{F}$ The N_x dimensional map $\mathbf{F} : \mathbf{x}_t \to \mathbf{x}_{t+1}$

$\mathfrak{f} \circ (\mathbf{x}, \mathbf{p})$ The dynamical equations that describe the coupled model.

$\mathfrak{F} \circ (\mathbf{x}, \mathbf{p}, \mathbf{y}, \gamma)$ The N_x dimensional *coupled* map of the dynamics.

N_M The dimension of the measurements; $N_M \leq N_x$.

$\mathbf{y}(t)$ The N_M dimensional vector of measurements at time t.

$\mathbf{y}_T = \{\mathbf{y}(t)\}\Big|_{t=0}^{t=T} = \{\mathbf{y}(0), \ldots, \mathbf{y}(N_T)\}$

The $N_T + 1$ element timeseries of measurements from $t = 0$ to $t = N_T$.

$\mathbf{g}(\mathbf{y}, \mathbf{q})$ The dynamical equations that describe the data system, with parameters $\mathbf{q}$.

Appendix B

Ion Channel Kinetic Functions Using Parameters from the Mixed-1280 Model

Here are gathered the $\alpha(V), \beta(V), x_\infty(V)$ and $\tau(V)$ functions for the channel kinetics that were obtained from the optimization and model error estimates of neuron (20110512_1_3_12). The equations of these kinetics are all given by equation (8.9).

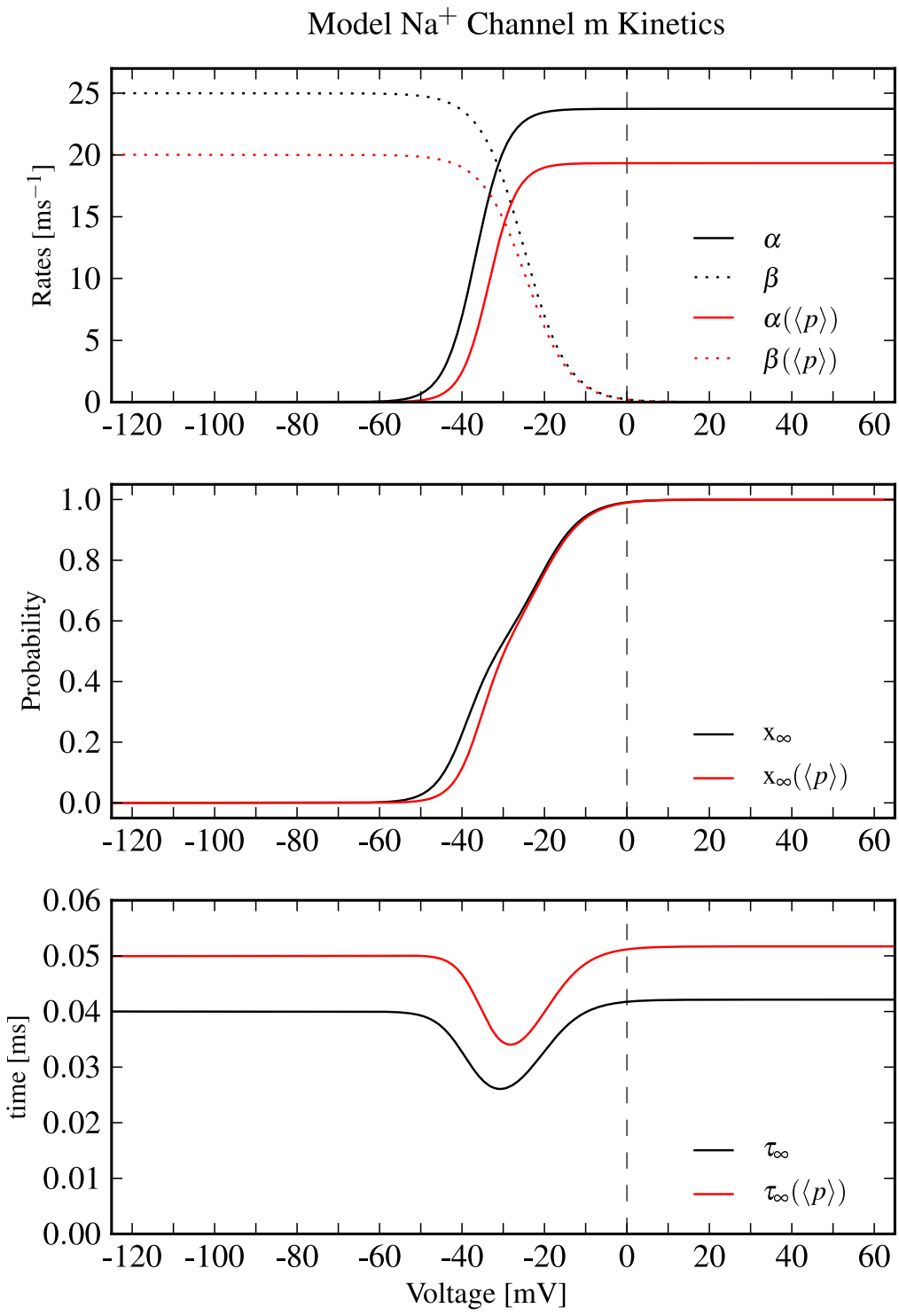

Figure B.1: Optimized Channel Kinetics of I_{NaT} 'm' Particle

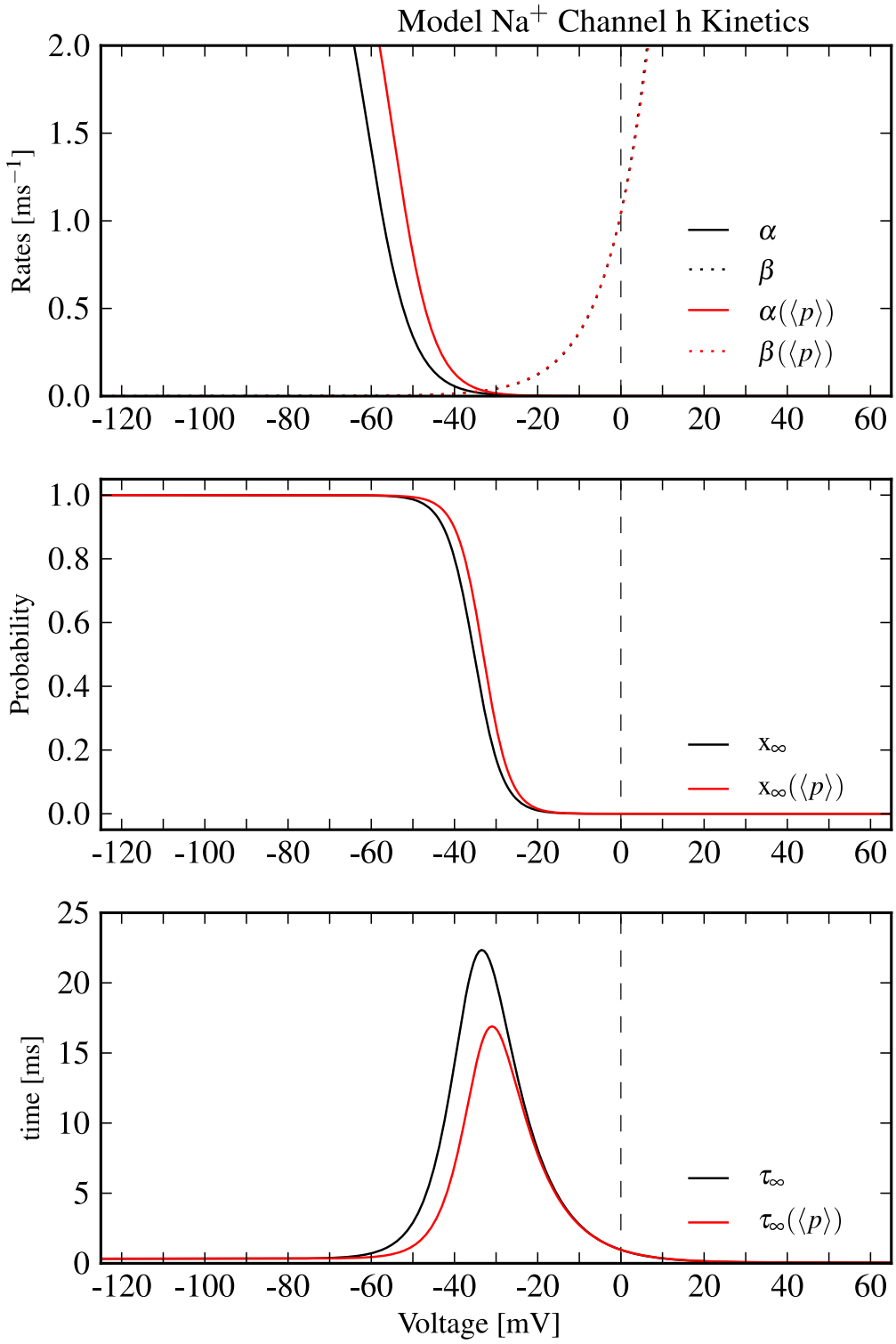

Figure B.2: Optimized Channel Kinetics of I_{NaT} 'h' Particle

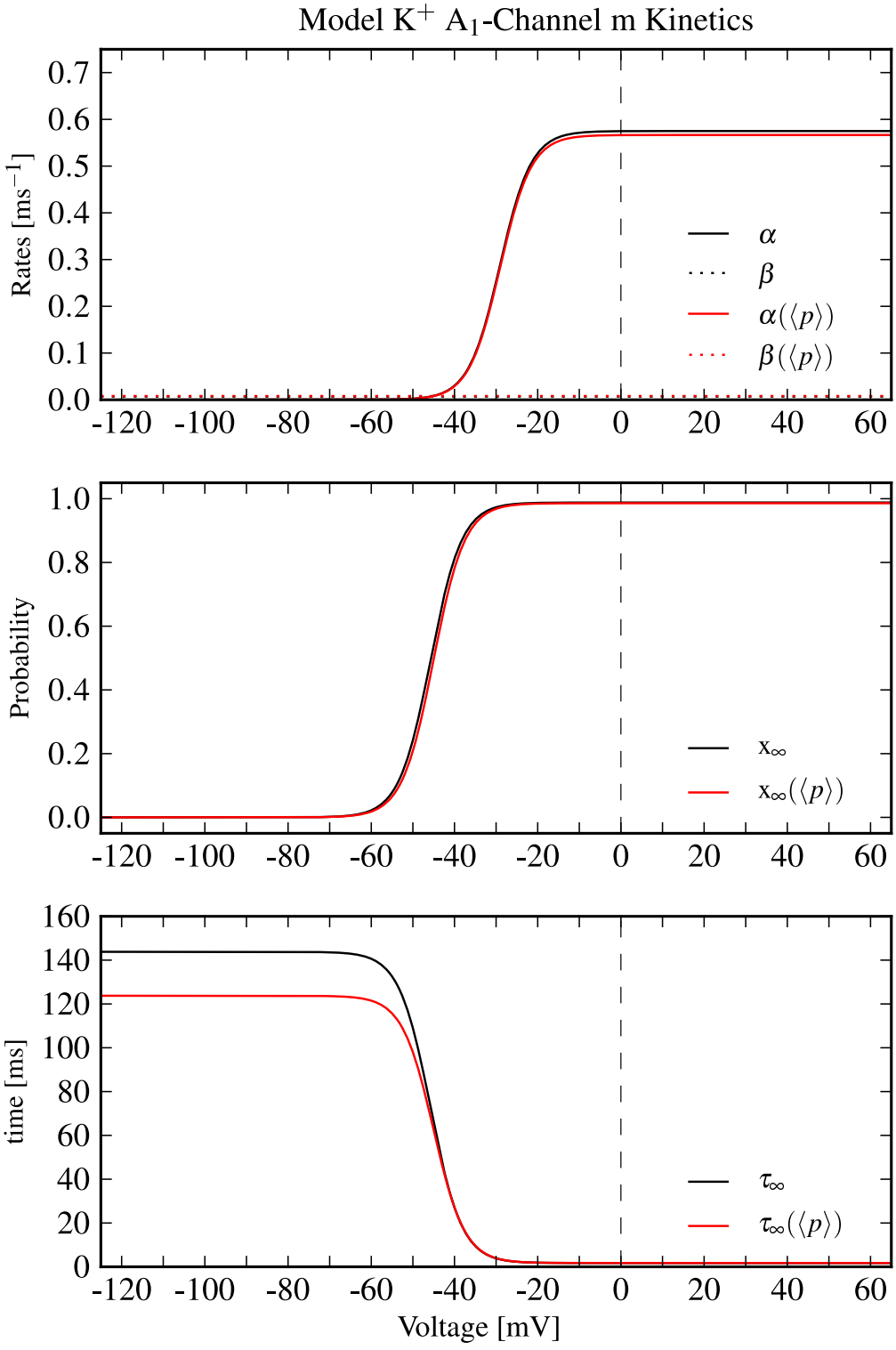

Figure B.3: Optimized Channel Kinetics of I_{KA1} 'm' Particle

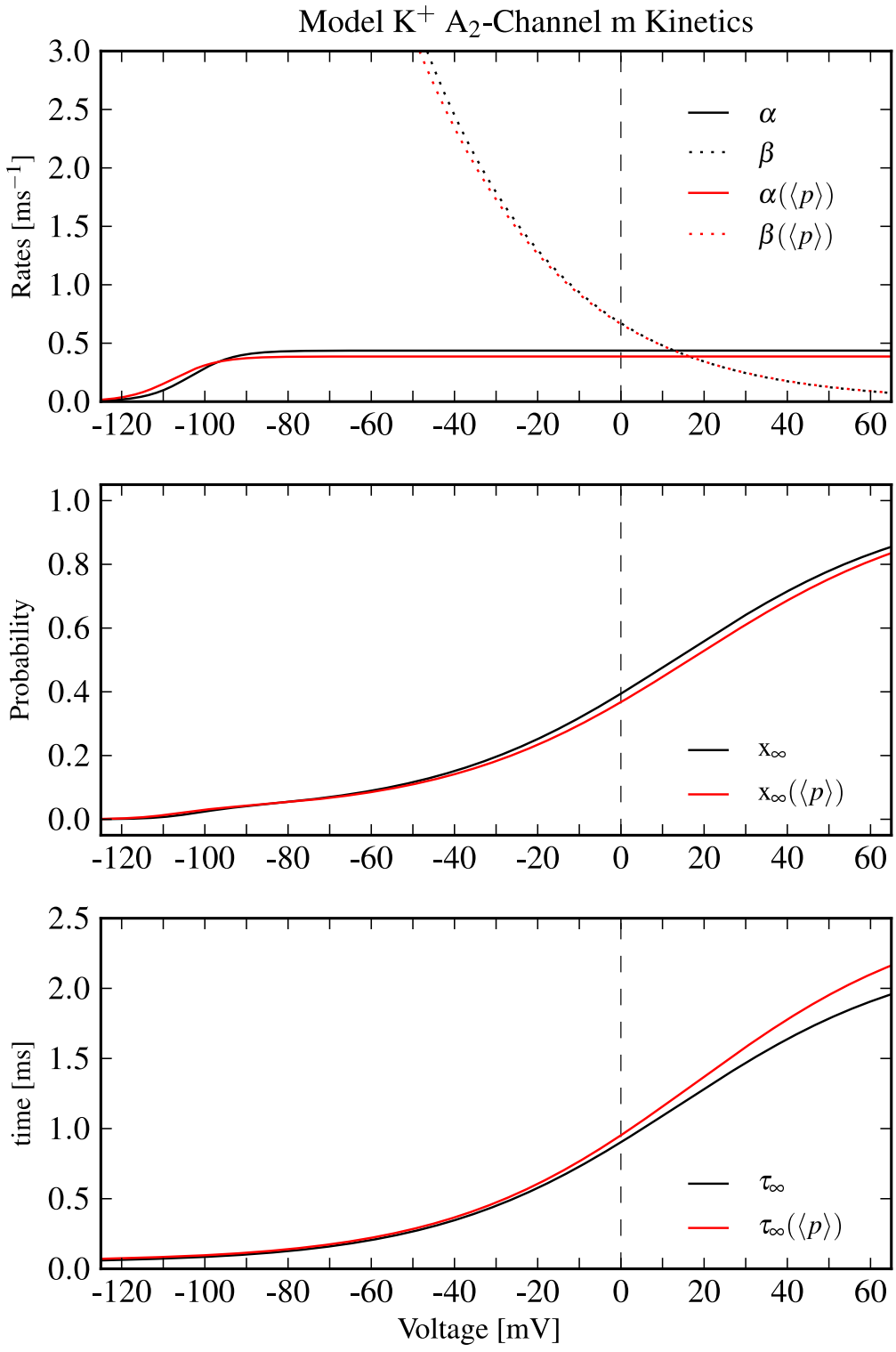

Figure B.4: Optimized Channel Kinetics of I_{KA2} 'm' Particle

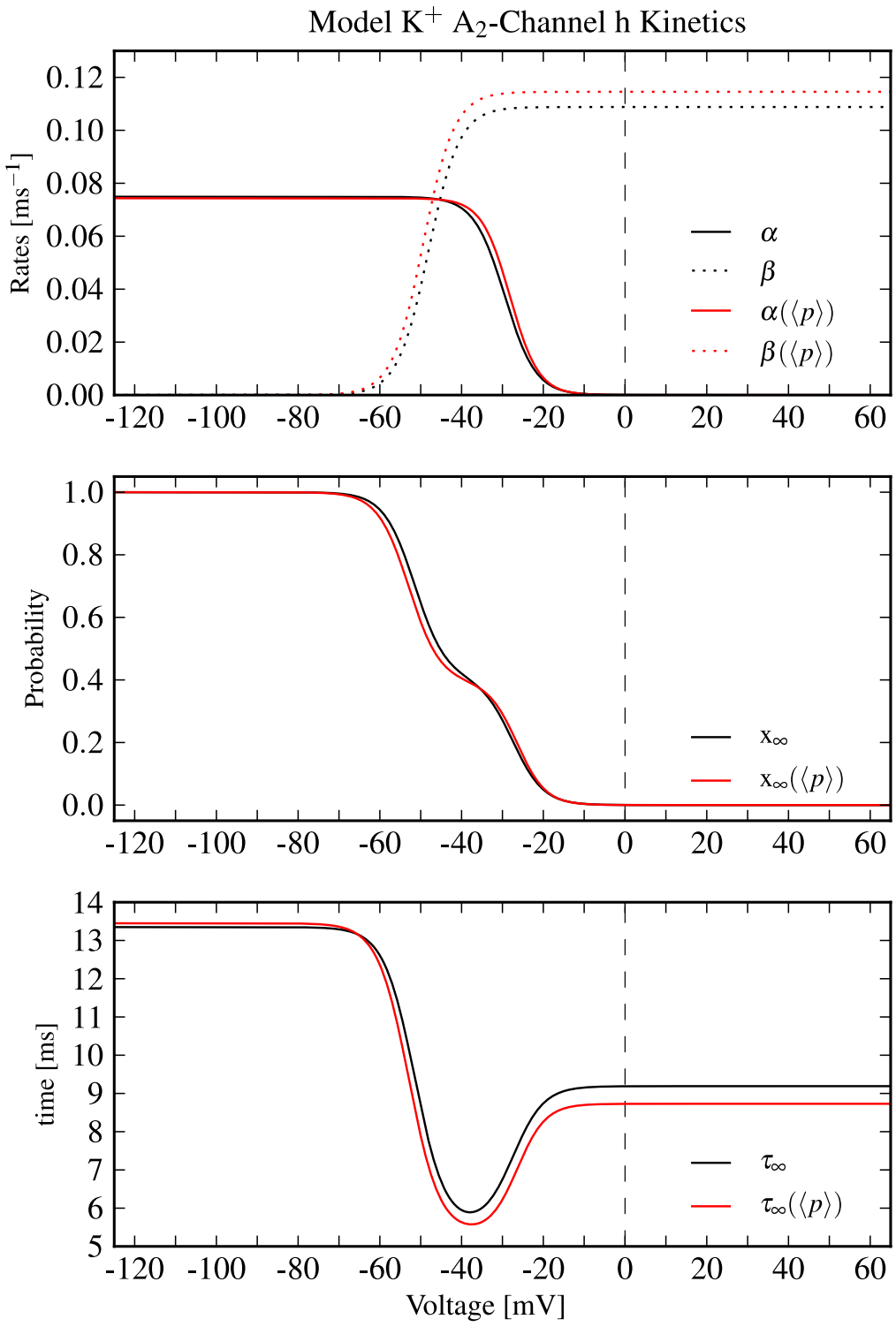

Figure B.5: Optimized Channel Kinetics of I_{KA2} 'h' Particle

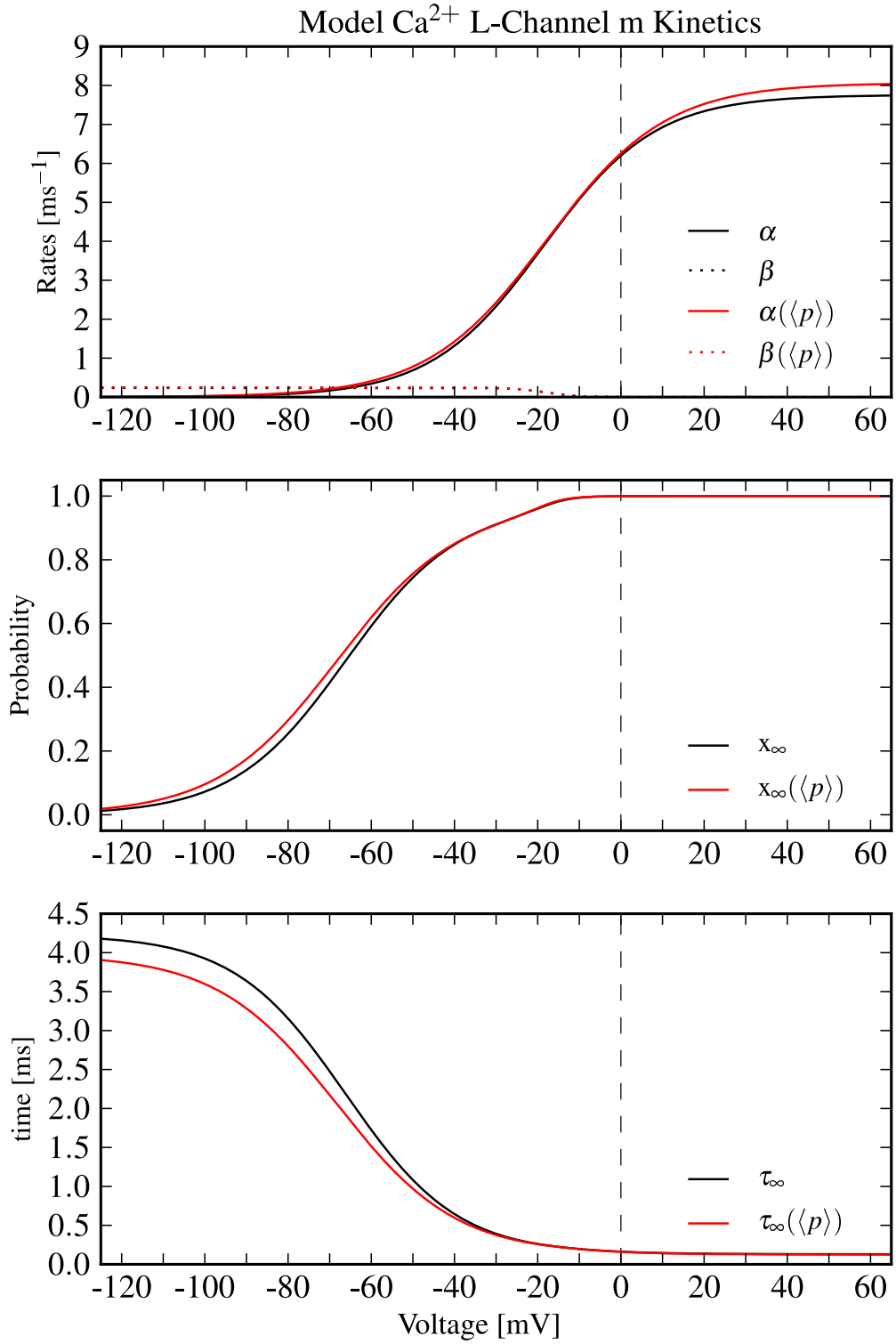

Figure B.6: Optimized Channel Kinetics of I_{CaL} 'm' Particle

Appendix C

Parameter Bounds and 'Data' Values for GHK-1069 Twin Experiment

Table C.1: Part I. Parameter Bounds and Results of GHK-1069 Model Error Analysis

		Single Markov Chain Monte Carlo				
			'Data'		Results	
Parameter	Units	Lower	Init. Cond	Upper	$\langle p \rangle$	σ_p
C_m^{-1}	nF^{-1}	35.478	39.42	43.362	39.90220	2.42565
$[Na]_{ext}$	mM	131.4	146.0	160.6	151.19007	6.51289
$[Na]_{int}$	mM	22.185	24.65	27.115	24.46308	1.53994
$[K]_{ext}$	mM	2.25	2.5	2.75	2.45049	0.13019
$[K]_{int}$	mM	64.755	71.95	79.145	71.95799	3.57256
$[Ca]_{ext}$	mM	7.92	8.8	9.68	8.69852	0.50477
$[Ca]_{int}$	mM	9e-05	1.0e-04	0.00011	0.00010	0.00001
g_L	mS/cm^2	0.0333	0.037	0.0407	0.03676	0.00176
E_L	mV	-79.97	-72.7	-65.43	-74.52784	3.64690
P_{Na1}	$\mu m/s$	1.7334	1.926	2.1186	1.98578	0.09842
$G_{\alpha m}$	mV	-374.31801	-3.402891e+02	-306.26019	-339.57535	19.22981
$z_{\alpha m}$	$[\cdot]$	6.3	7.000000e+00	7.7	6.88246	0.38713
$\epsilon_{\alpha m}$	$[\cdot]$	0.03792294	4.213660e-02	0.04635026	0.04213	0.00233
$G_{\beta m}$	mV	36.134136	4.014904e+01	44.163944	39.59822	2.31557
$z_{\beta m}$	$[\cdot]$	4.5	5.000000e+00	5.5	4.96078	0.28172
$\epsilon_{\beta m}$	$[\cdot]$	0.036	4.000000e-02	0.044	0.03975	0.00231
$G_{\alpha h}$	mV	247.89888	2.754432e+02	302.98752	275.85187	16.61517
$z_{\alpha h}$	$[\cdot]$	4.5	5.000000e+00	5.5	5.00456	0.26947
$\epsilon_{\alpha h}$	$[\cdot]$	0.28882422	3.209158e-01	0.35300738	0.31874	0.01903
$G_{\beta h}$	mV	-2.4629913	-2.239083e+00	-2.0151747	-2.26608	0.13045
$z_{\beta h}$	$[\cdot]$	2.5895052	2.877228e+00	3.1649508	2.86698	0.17590
$\epsilon_{\beta h}$	$[\cdot]$	0.036	4.000000e-02	0.044	0.04008	0.00222
P_{A1}	$\mu m/s$	0.0621	0.069	0.0759	0.07073	0.00382
$G_{\alpha m}$	mV	-207.87404	-1.889764e+02	-170.07876	-186.46662	10.95944
$z_{\alpha m}$	$[\cdot]$	6.3	7.000000e+00	7.7	6.86198	0.37611
$\epsilon_{\alpha m}$	$[\cdot]$	1.5655914	1.739546e+00	1.9135006	1.70539	0.09498
$G_{\beta m}$	mV	-584.09296	-5.309936e+02	-477.89424	-535.80879	32.69298
$z_{\beta m}$	$[\cdot]$	2.4287409	2.698601e+00	2.9684611	2.70847	0.15258
$\epsilon_{\beta m}$	$[\cdot]$	129.42909	1.438101e+02	158.19111	144.19762	8.76362
P_{A2}	$\mu m/s$	1.197	1.33	1.463	1.34163	0.06550
$G_{\alpha m}$	mV	-544.75652	-4.952332e+02	-445.70988	-510.11743	23.56007
$z_{\alpha m}$	$[\cdot]$	4.5	5.000000e+00	5.5	4.89670	0.25968
$\epsilon_{\alpha m}$	$[\cdot]$	2.0650734	2.294526e+00	2.5239786	2.34023	0.12347
$G_{\beta m}$	mV	8.8210935	9.801215e+00	10.7813365	9.63010	0.58221
$z_{\beta m}$	$[\cdot]$	0.81379728	9.042192e-01	0.99464112	0.92379	0.05107
$\epsilon_{\beta m}$	$[\cdot]$	0.036	4.000000e-02	0.044	0.03957	0.00208
$G_{\alpha h}$	mV	247.0005	2.744450e+02	301.8895	271.81162	14.80204
$z_{\alpha h}$	$[\cdot]$	6.3	7.000000e+00	7.7	7.06150	0.35962
$\epsilon_{\alpha h}$	$[\cdot]$	12.018627	1.335403e+01	14.689433	13.34129	0.72423
$G_{\beta h}$	mV	-305.965	-2.781500e+02	-250.335	-273.33965	14.16798
$z_{\beta h}$	$[\cdot]$	6.3	7.000000e+00	7.7	7.00760	0.37017
$\epsilon_{\beta h}$	$[\cdot]$	8.2769796	9.196644e+00	10.1163084	9.12169	0.54390

Table C.2: Part II. Parameter Bounds and Results of GHK-1069 Model Error Analysis

		Single Markov Chain Monte Carlo				
			'Data'		Results	
Parameter	Units	Lower	Init. Cond	Upper	$\langle p \rangle$	σ_p
P_H	$\mu m/s$	0.00558	0.0062	0.00682	0.00627	0.00032
$G_{\alpha h}$	mV	-12.807597	-1.164327e+01	-10.478943	-11.53498	0.72229
$z_{\alpha h}$	[·]	3.1442967	3.493663e+00	3.8430293	3.46242	0.19546
$\epsilon_{\alpha h}$	[·]	367.67916	4.085324e+02	449.38564	398.53929	21.98915
$G_{\beta h}$	mV	-234.10596	-2.128236e+02	-191.54124	-216.72648	10.72890
$z_{\beta h}$	[·]	2.2484664	2.498296e+00	2.7481256	2.46204	0.14631
$\epsilon_{\beta h}$	[·]	366.88644	4.076516e+02	448.41676	407.48917	24.38089
P_{CaL}	$\mu m/s$	0.01845	0.0205	0.02255	0.02063	0.00115
$G_{\alpha m}$	mV	-99.490864	-9.044624e+01	-81.401616	-91.05904	4.94502
$z_{\alpha m}$	[·]	1.7615403	1.957267e+00	2.1529937	1.94796	0.10459
$\epsilon_{\alpha m}$	[·]	0.11608884	1.289876e-01	0.14188636	0.13307	0.00639
$G_{\beta m}$	mV	146.79639	1.631071e+02	179.41781	162.22205	8.03428
$z_{\beta m}$	[·]	6.3	7.000000e+00	7.7	7.04673	0.39851
$\epsilon_{\beta m}$	[·]	3.808035	4.231150e+00	4.654265	4.28043	0.23951
P_{CaT}	$\mu m/s$	0.0459	0.051	0.0561	0.05100	0.00284
$G_{\alpha m}$	mV	-233.31935	-2.121085e+02	-190.89765	-217.66860	10.89615
$z_{\alpha m}$	[·]	2.2330935	2.481215e+00	2.7293365	2.47567	0.14446
$\epsilon_{\alpha m}$	[·]	370.57032	4.117448e+02	452.91928	414.14512	20.68319
$G_{\beta m}$	mV	42.710598	4.745622e+01	52.201842	47.31904	2.80451
$z_{\beta m}$	[·]	3.0995064	3.443896e+00	3.7882856	3.40190	0.19081
$\epsilon_{\beta m}$	[·]	370.28097	4.114233e+02	452.56563	409.36728	21.41204
$G_{\alpha h}$	mV	38.843694	4.315966e+01	47.475626	42.38841	2.52077
$z_{\alpha h}$	[·]	3.0992157	3.443573e+00	3.7879303	3.40552	0.19324
$\epsilon_{\alpha h}$	[·]	371.09268	4.123252e+02	453.55772	409.92222	22.70966
$G_{\beta h}$	mV	-261.30841	-2.375531e+02	-213.79779	-238.43556	13.44209
$z_{\beta h}$	[·]	2.4463818	2.718202e+00	2.9900222	2.70007	0.15165
$\epsilon_{\beta h}$	[·]	371.24514	4.124946e+02	453.74406	416.02345	25.07570

Appendix D

Parameter Bounds on HVc Neuron Optimization

Table D.1: Part I. Parameter Bounds of HVc Neuron Optimization

		Variational (IPOPT)			Single Markov Chain Monte Carlo	
Parameter	Units	Result	Lower	Upper	Lower	Upper
E_{Na}	mV	4.500000e+01	45	55	22.5	67.5
E_K	mV	-8.500000e+01	-85	-70	-127.5	-42.5
E_L	mV	-7.269787e+01	-77	-57	-109.046805	-36.348935
E_H	mV	-4.367534e+01	-47	-40	-65.51301	-21.83767
g_L	mS/cm^2	3.758373e-12	0.0	0.6	1.8791865e-12	5.6375595e-02
C_m^{-1}	nF^{-1}	3.941666e+01	4	90	19.70833	59.12499
g_{Na1}	mS/cm^2	7.341148e+00	0.0	190	3.670574	11.011722
$G_{\alpha m}$	mV	-3.402891e+02	-550	200	-510.43365	-170.14455
$z_{\alpha m}$	$[\cdot]$	7.000000e+00	0	7	3.5	10.5
$\epsilon_{\alpha m}$	$[\cdot]$	4.213660e-02	0.04	800	0.04	0.0632049
$G_{\beta m}$	mV	4.014904e+01	-550	550	20.07452	60.22356
$z_{\beta m}$	$[\cdot]$	5.000000e+00	0	5	2.5	7.5
$\epsilon_{\beta m}$	$[\cdot]$	4.000000e-02	0.04	800	0.04	0.06
$G_{\alpha h}$	mV	2.754432e+02	-550	550	137.7216	413.1648
$z_{\alpha h}$	$[\cdot]$	5.000000e+00	0	5	2.5	7.5
$\epsilon_{\alpha h}$	$[\cdot]$	3.209158e-01	0.04	800	0.1604579	0.4813737
$G_{\beta h}$	mV	-2.239083e+00	-550	550	-3.3586245	-1.1195415
$z_{\beta h}$	$[\cdot]$	2.877228e+00	0	7	1.438614	4.315842
$\epsilon_{\beta h}$	$[\cdot]$	4.000000e-02	0.04	800	0.04	0.06
g_{Na2}	mS/cm^2	2.911103e-12	0	150	1.4555515e-12	4.3666545e-02
$G_{\alpha m}$	mV	-2.081181e+02	-550	550	-312.17715	-104.05905
$z_{\alpha m}$	$[\cdot]$	2.489339e+00	0	7	1.2446695	3.7340085
$\epsilon_{\alpha m}$	$[\cdot]$	4.080179e+02	0.04	800	204.00895	612.02685
$G_{\beta m}$	mV	-9.230537e+00	-550	550	-13.8458055	-4.6152685
$z_{\beta m}$	$[\cdot]$	3.495205e+00	0	7	1.7476025	5.2428075
$\epsilon_{\beta m}$	$[\cdot]$	4.089970e+02	0.04	800	204.4985	613.4955
$G_{\alpha h}$	mV	-1.234778e+01	-550	550	-18.52167	-6.17389
$z_{\alpha h}$	$[\cdot]$	3.493162e+00	0	7	1.746581	5.239743
$\epsilon_{\alpha h}$	$[\cdot]$	4.098306e+02	0.04	800	204.9153	614.7459
$G_{\beta h}$	mV	-2.361097e+02	-550	550	-354.16455	-118.05485
$z_{\beta h}$	$[\cdot]$	2.728242e+00	0	9	1.364121	4.092363
$\epsilon_{\beta h}$	$[\cdot]$	4.089542e+02	0.04	800	204.4771	613.4313
g_{A1}	mS/cm^2	2.470600e-01	0	90	0.12353	0.37059
$G_{\alpha m}$	mV	-1.889764e+02	-1050	550	-283.4646	-94.4882
$z_{\alpha m}$	$[\cdot]$	7.000000e+00	0	7	3.5	10.5
$\epsilon_{\alpha m}$	$[\cdot]$	1.739546e+00	0.04	2000	0.869773	2.609319
$G_{\beta m}$	mV	-5.309936e+02	-650	550	-796.4904	-265.4968
$z_{\beta m}$	$[\cdot]$	2.698601e+00	0	7	1.3493005	4.0479015
$\epsilon_{\beta m}$	$[\cdot]$	1.438101e+02	0.04	800	71.90505	215.71515

Table D.2: Part II. Parameter Bounds of HVc Neuron Optimization

		Variational (IPOPT)			Single Markov Chain Monte Carlo	
Parameter	Units	Result	Lower	Upper	Lower	Upper
g_{A2}	mS/cm^2	5.100000e+02	0	510	255.0	765.0
$G_{\alpha m}$	mV	-4.952332e+02	-650	550	-742.8498	-247.6166
$z_{\alpha m}$	$[\cdot]$	5.000000e+00	0	5	2.5	7.5
$\epsilon_{\alpha m}$	$[\cdot]$	2.294526e+00	0.04	800	1.147263	3.441789
$G_{\beta m}$	mV	9.801215e+00	-550	550	4.9006075	14.7018225
$z_{\beta m}$	$[\cdot]$	9.042192e-01	0	7	0.4521096	1.3563288
$\epsilon_{\beta m}$	$[\cdot]$	4.000000e-02	0.04	800	0.04	0.06
$G_{\alpha h}$	mV	2.744450e+02	-550	550	137.2225	411.6675
$z_{\alpha h}$	$[\cdot]$	7.000000e+00	0	7	3.5	10.5
$\epsilon_{\alpha h}$	$[\cdot]$	1.335403e+01	0.04	800	6.677015	20.031045
$G_{\beta h}$	mV	-2.781500e+02	-1050	550	-417.225	-139.075
$z_{\beta h}$	$[\cdot]$	7.000000e+00	0	7	3.5	10.5
$\epsilon_{\beta h}$	$[\cdot]$	9.196644e+00	0.04	800	4.598322	13.794966
g_H	mS/cm^2	1.272465e-12	0	100	6.362325e-13	1.9086975e-02
$G_{\alpha h}$	mV	-1.164327e+01	-550	550	-17.464905	-5.821635
$z_{\alpha h}$	$[\cdot]$	3.493663e+00	0	7	1.7468315	5.2404945
$\epsilon_{\alpha h}$	$[\cdot]$	4.085324e+02	0.04	800	204.2662	612.7986
$G_{\beta h}$	mV	-2.128236e+02	-550	550	-319.2354	-106.4118
$z_{\beta h}$	$[\cdot]$	2.498296e+00	0	7	1.249148	3.747444
$\epsilon_{\beta h}$	$[\cdot]$	4.076516e+02	0.04	800	203.8258	611.4774
g_{CaL}	mS/cm^2	2.049909e-02	0	100	0.010249545	0.030748635
$G_{\alpha m}$	mV	-9.044624e+01	-550	550	-135.66936	-45.22312
$z_{\alpha m}$	$[\cdot]$	1.957267e+00	0	7	0.9786335	2.9359005
$\epsilon_{\alpha m}$	$[\cdot]$	1.289876e-01	0.04	800	0.0644938	0.1934814
$G_{\beta m}$	mV	1.631071e+02	-550	550	81.55355	244.66065
$z_{\beta m}$	$[\cdot]$	7.000000e+00	0	7	3.5	10.5
$\epsilon_{\beta m}$	$[\cdot]$	4.231150e+00	0.04	800	2.115575	6.346725
g_{CaT}	mS/cm^2	2.516841e-13	0	510	1.2584205e-13	3.7752615e-02
$G_{\alpha m}$	mV	-2.121085e+02	-550	550	-318.16275	-106.05425
$z_{\alpha m}$	$[\cdot]$	2.481215e+00	0	7	1.2406075	3.7218225
$\epsilon_{\alpha m}$	$[\cdot]$	4.117448e+02	0.04	800	205.8724	617.6172
$G_{\beta m}$	mV	4.745622e+01	-550	1050	23.72811	71.18433
$z_{\beta m}$	$[\cdot]$	3.443896e+00	0	7	1.721948	5.165844
$\epsilon_{\beta m}$	$[\cdot]$	4.114233e+02	0.04	800	205.71165	617.13495
$G_{\alpha h}$	mV	4.315966e+01	-550	1050	21.57983	64.73949
$z_{\alpha h}$	$[\cdot]$	3.443573e+00	0	7	1.7217865	5.1653595
$\epsilon_{\alpha h}$	$[\cdot]$	4.123252e+02	0.04	800	206.1626	618.4878
$G_{\beta h}$	mV	-2.375531e+02	-550	550	-356.32965	-118.77655
$z_{\beta h}$	$[\cdot]$	2.718202e+00	0	9	1.359101	4.077303
$\epsilon_{\beta h}$	$[\cdot]$	4.124946e+02	0.04	800	206.2473	618.7419
$[Ca]_{ext}$	mM	8.999548e+00	0.0	9	4.499774	13.499322

References

Abarbanel, H. D. I., 1996: *Analysis of Observed Chaotic Data*. Springer-Verlag, New York.

Abarbanel, H. D. I., 2009: Effective actions for statistical data assimilation. *Physics Letters A*, **373**, 4044–4048.

Abarbanel, H. D. I., Brown, R., and Kennel, M. B., 1992: Local lyapunov exponents computed from observed data. *Journal of Nonlinear Science*, **2**, 343–365.

Abarbanel, H. D. I., Bryant, P., Gill, P. E., Kostuk, M., Rofeh, J., Singer, Z., Toth, B., and Wong, E., 2011: Dynamical parameter and state estimation in neuron models. In *the dynamic brain: An Exploration of Neuronal Variability and Its Functional Significance*, editors D. Glanzman, and M. Ding, chapter 8. Oxford University Press.

Abarbanel, H. D. I., Creveling, D., and Jeanne, J., 2008: Estimation of parameters in nonlinear systems using balanced synchronization. *Physical Review E*, **77**(016208).

Abarbanel, H. D. I., Creveling, D. R., Farsian, R., and Kostuk, M., 2009: Dynamical state and parameter estimation. *SIAM Journal of Applied Dynamical Systems*, **8**(4), 1341–1381.

Abarbanel, H. D. I., Kostuk, M., Margoliash, D., Meliza, C. D., and Nogaret, A., 2012: Data-assimilated neuron models of a population of neurons in the high vocal center of the zebra finch. In progress.

Abarbanel, H. D. I., M., K., and Whartenby, W., 2010: Data assimilation with regularized nonlinear instabilities. *Quarterly Journal of the Royal Meteorological Society*, **136**(648), 769–783.

Abarbanel, H. D. I., Rulkov, N. F., and Sushchik, M. M., 1995: Blending chaotic attractors using the synchronization of chaos. *Physical Review E*, **52**(1), 214–217.

Abarbanel, H. D. I., Rulkov, N. F., and Sushchik, M. M., 1996: Generalized synchronization of chaos: The auxiliary system approach. *Physical Review E*, **53**(5), 4528–4535.

Abarbanel, H. D. I., Talathi, S. S., Mindlin, G., Rabinovich, M., and Gibb, L., 2004: Dynamical model of birdsong maintenance and control. *Physical Review E*, **70**(051911).

Afraimovich, V. S., Bykov, V. V., and Shilnikov, L. P., 1977: Appearance and structure of lorenz attractor. *Doklady Akademii Nauk SSSR*, **234**(2), 336–339.

Afraimovich, V. S., Rabinovich, M. I., and D., U. A., 1983: Critical-points and phase-transitions in the stochastic-behavior of a nonautonomous anharmonic-oscillator. *JETP Letters*, **38**(2), 72–75.

Afraimovich, V. S., and Shilnikov, L. P., 1974: Some global bifurcations connected with vanishing of a fixed-point of saddle-node type. *Doklady Akademii Nauk SSSR*, **219**(6), 1281–1284.

Ahlers, V., and Pikovsky, A., 2002: Critical properties of the synchronization transition in space-time chaos. *Physical Review Letters*, **88**(25).

Apte, A., Hairer, M., Stuart, A. M., and Voss, J., 2007: Sampling the posterior: An approach to non-gaussian data assimilation. *Physica D*, **230**, 50–64.

Arfken, G. B., and Weber, H. J., 2005: *Mathematical Methods for Physicists*. Elsevier, sixth edition edition.

Barclay, A., Gill, P. E., and Rosen, J. B., 1998: Variational calculus, optimal control, and applications. **124**, 207–222.

Barela, A. J., Waddy, S. P., Lickfett, J. G., Hunter, J., Anido, A., Helmers, S. L., Goldin, A. L., and Escayg, A., 2006: An epilepsy mutation in the sodium channel *scn1a* that decreases channel excitability. *The Journal of Neuroscience*, **26**(10), 2714–2723.

Beuter, A., Glass, L., Mackey, M. C., and Titcombe, M. S., 2003: *Nonlinear Dynamics in Physiology and Medicine*. Springer-Verlag.

Bhalla, U., and Bower, J., 1993: Exploring parameter space in detailed single neuron models: Simulations of the mitral and granule cells of the olfactory bulb. *J. Neurophys.*, **69**, 1948–1965.

Blackwell, K. T., 2005: Modeling calcium concentration and biochemical reactions. *Brains, Minds and Media*, (2).

Boltzmann, L., 1868: Studien über das gleichgewicht der lebendigen kraft zwischen bewegten materiellen punkten. *Wiener Berichte*, **58**, 517–560.

Brainard, M. S., and Doupe, A. J., 2000: Auditory feedback in learning and maintenance of vocal behavior. *Nature Reviews Neuroscience*, **1**, 31–40.

Brainard, M. S., and Doupe, A. J., 2002: What songbirds teach us about learning. *Nature*, **417**, 351–358.

Bryson, A. J., and Yu-Chi, H., 1969: *Applied Optimal Control: Optimization Estimation and Control*. Ginn and Company.

Buibas, M., and Silva, G. A., 2011: A framework for simulating and estimating the state and functional toplogy of complex dynamic geometric networks. *Neural Computation*, **23**(1), 183–214.

Byzov, A. L., 1958: Capacitance and resistance of layers in frog retina. *Biofizika [Transl]*, **3**(6), 627–638.

Cessna, J., Colburn, C., and Bewley, T. R., 2007: Multiscale retrograde estimation and forcasting of chaotic nonlinear systems. 46th IEEE Conference on Decision and Control. New Orleans, LA, USA, Dec. 12-14.

Coppersmith, S. N., 1999: A simpler derivation of feigenbaum's renormalization group equation for the period-doubling bifurcation sequence. *Am. J. Phys.*, **67**(1).

Cover, T. M., and Thomas, J. A., 2006: *Elements of Information Theory*. John Wiley & Sons, Inc., 2nd ed. edition.

Cowen, W. M., Südhof, T. C., and Stevens, C. F., editors, 2001: *Synapses*. The Johns Hopkins University Press.

Cox, J. J. *et. al.*, 2006: An SCN9A channelopathy causes congenital inability to experience pain. *Nature*, **444**(7121), 894–8.

Cui, J., Cox, D. H., and Aldrich, R. W., 1997: Intrinsic voltage dependence and Ca2+ regulation of *mslo* large conductance Ca-activated K+ channels. *J. Gen. Physiol.*, **109**, 647–673.

Domb, C., and Green, M. S., editors, 1972: *Phase Transitions and Critical Phenomena*, volume Volume 1: Exact Results. Academic Press Inc, 111 Fifth Avenue, New York, New York 10003.

Domb, C., and Green, M. S., editors, 1976: *Phase Transitions and Critical Phenomena*, volume Volume 6. Academic Press Inc, 111 Fifth Avenue, New York, New York 10003.

Eyink, G. L., Restrepo, J. M., and Alexander, F. J., 2004: A mean field approximation in data assimilation for nonlinear dynamics. *Physica D*, **195**, 347–368.

Fano, R. M., 1961: *Transmission of Information: A Statistical Theory of Communication*. The MIT Press Classics Series. MIT Press.

Feigenbaum, M. J., 1979: The universal metric properties of nonlinear transformations. *Journal of Statistical Physics*, **21**(6).

Feynman, R. P., and Hibbs, A. R., 1965: *Quantum Mechanics and Path Integrals*. McGraw-Hill, Inc.

Fraser, A. M., and Swinney, H. L., 1986: Independent coordinates for strange attractors from mutual information. *Physical Review A*, **33**(2), 1134–1140.

Friedman, J. H., 1989: Regularized discriminant analysis. *Journal of the American Statistical Association*, **84**(405), 165–175.

Fujisaka, H., and Yamada, T., 1983: *Prog. Theor. Phys.*, **69**, 32.

Gibb, L., Gentner, T. Q., and Abarbanel, H. D. I., 2009: Brain stem feedback in a computational model of birdsong sequencing. *Journal of Neurophysiology*, **102**(3), 1763–1778.

Gill, P. E., Murray, W., and Saunders, M. A., 2005: SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, **47**, 99–131.

Gill, P. E., Murray, W., and Wright, M. H., 1981: *Practical Optimization*. Academic Press, A Subsidiary of Harcourt Brace Jovanovich, London.

Goldman, D. E., 1943: Potential, impedance, and rectification in membranes. *Jour. Gen. Physiol.*, **27**(1), 37–60.

Goldman, M. S., Golowasch, J., Marder, E., and Abbott, L. F., 2000: Dependence of firing pattern on intrinsic ionic conductances: Sensitive and insensitive combinations. *Neurocomputing*, **32**, 141–146.

Golowasch, J., Abbott, L. F., and Marder, E., 1999: Activity-dependent regulation of potassium currents in an identified neuron of the stomatogastric ganglion of the crab cancer borealis. *Journal of Neuroscience*, **19**(20).

Grissmer, S., Nguyen, A., Jayashree, A., Hanson, D. C., Mather, R. J., Gutman, G. A., Karmilowicz, M. J., Auperin, D. D., and Chandy, K. G., 1994: Pharmacological characterization of five cloned voltage-gated K+ channels, types Kv1.1, 1.2, 1.3, 1.5, 3.1, stably expressed in mammalian cell lines. *Molecular Pharmacology*, **45**, 1227–1234.

Hahnloser, R. H. R., Kozhevnikov, A. A., and Fee, M. S., 2002: An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, **419**, 65–70.

Hairer, E., Lubich, C., and Wanner, G., 2002: *Geometric Numerical Integration*. Springer Series in Computational Mathematics. Springer, second edition edition.

Hastings, W. K., 1970: Monte Carlo sampling methods using markov chains and their applications. *Biometrika*, **57**(1), 97.

Hirata, Y., Suzuki, H., and Aihara, K., 2006: Reconstructing state spaces from multivariate data using variable delays. *Physical Review E*, **74**(026202).

Hochberg, D., Molina-París, C., Pérez-Mercader, J., and Visser, M., 1999: Effective action of stochastic partial differential equations. *Physical Review E*, **60**(6), 6343–6360.

Hodgkin, A. L., and Huxley, A. F., 1952: Currents carried by sodium and potassium ions through the membrane of *loligo*. *J. Physiol.*, **116**, 449–472.

Huguenard, J. R., and McCormick, D. A., 1992: Simulation of the currents involved in rhythmic oscillations in thalamic relay neurons. *Journal of Neurophysiology*, **63**(4).

Huys, Q. J. M., Ahrens, M. B., and Paninski, L., 2006: Efficient estimation of detailed single-neuron models. *J Neurophysiol*, **96**, 872–890.

Johnson, R. A., Palmer, K. J., and Sell, G. R., 1987: Ergodic properties of linear dynamical systems. *SIAM J. Math. Anal.*, **18**(1).

Johnston, D., and Wu, S., 1994: *Foundations of Cellular Neurophysiology*. MIT Press, Cambridge, MA.

Jouvet, B., and Phythian, R., 1979: Quantum aspects of classical and statistical fields. *Physical Review A*, **19**(3), 1350–1355.

Judd, K., and Stemler, T., 2009: Failures of sequential bayesian filters and the successes of shadowing filters in tracking of nonlinear deterministic and stochastic systems. *Physical Review E*, **79**(066206).

Kennel, M. B., and Isabelle, S., 1992: Method to distinguish possible chaos from colored noise and to determine embedding parameters. *Physical Review A*, **46**(6).

Kew, J. N. C., and Davies, C. H., editors, 2010: *Ion Channels: From Structure to Function*. Oxford University Press, New York.

Kirk, D. E., 1970: Optimal control theory, an introduction. Reprinted by Dover Publications, Inc. Mineola NY (2004).

Knudsen, D. P., and Gentner, T. Q., 2010: Mechanisms of song perception in oscine birds. *Brain and Language*, **115**(1), 59–68.

Kolmogorov, A. N., and Rozanov, Y. A., 1960: On strong mixing conditions for stationay gaussian procces. *Theory of Probability and its Applications*, **5**, 204–208.

Kostuk, M., Toth, B. A., Meliza, C. D., Margoliash, D., and Abarbanel, H. D. I., 2012: Dynamical estimation of neuron and network properties II: path integral Monte Carlo methods. *Biological Cybernetics*, **106**(3), 155–167.

Landau, L. D., and Lifshitz, E. M., 1980: *Statistical Physics*, volume 5 of *Course on Theoretical Physics*. Butterworth-Heinemann, third edition.

Long, M. A., and Fee, M. S., 2008: Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, **456**, 189–194.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112**(474), 1177–1194.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**(2), 130–141.

Lorenz, E. N., 1996: Predictability- a problem partly solved. In *Predictability 1995 ECMWF seminar proceedings*.

Margoliash, D., 2004: Behavioural biology: Name that tune. *Nature*, **432**, 682–683.

Meliza, C. D., and Margoliash, D., 2012: Phase plot shapes of HVc neuron spikes. Private communication.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953: Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**(6).

Nogaret, A., 2011: HVc neuron model. Intra-lab report.

Omelan, I. P., Mryglod, I. M., and Folk, R., 2003: Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations. *Computer Physics Communications*, **151**, 272–314.

Oseledec, V., 1968: A multiplicitive ergodic theorem. lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, **19**, 197–231.

Otten, R. H. J. M., and van Ginneken, L. P. P. P., 1989: *The Annealing Algorithm*. Kluwer Academic Publishers.

Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S., 1980: Geometry from a time series. *Physical Review Letters*, **45**(9), 712–716.

Pathria, R. K., 1996: *Statistical Mechanics*. Butterworth-Heinemann, second edition edition.

Pauli, W., 2000: *Statistical Mechanics*, volume 4 of *Pauli Lectures on Physics*. Dover Books.

Pecora, L. M., and Carroll, T. L., 1990: Synchronization in chaotic systems. *Physical Review Letters*, **64**, 821–824.

Powell, M. J. D., editor, 1982: *Nonlinear Optimization 1981*. NATO Conference Series. Academic Press Inc. (London) LTD.

Prakriya, M., Solaro, C. R., and Lingle, C. J., 1996: [Ca2+]i elevations detected by BK channels during Ca2+ influx and muscarine-mediated release of Ca2+ from intracellular stores in rat chromaffin cells. *J Neurosci*, **16**(14), 4344–59.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007: *Numerical Recipies*. Cambridge University Press, third edition edition.

Prinz, A. A., Smolinski, T. G., and Hudson, A. E., 2011: Understanding animal-to-animal variability in neuronal and network properties. In *the dynamic brain: An Exploration of Neuronal Variability and Its Functional Significance*, editors D. Glanzman, and M. Ding, chapter 7. Oxford University Press.

Quinn, J. C., and Abarbanel, H. D. I., 2011: Data assimilation using a gpu accelerated path integral Monte Carlo approach. *Journal of Computational Physics*. doi:10.1016/j.jcp.2011.07.015.

Restrepo, J. M., 2008: A path integral method for data assimilation. *Physica D*, **237**, 14–27.

Rosenthal, J. S., 2011: Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo*, editors S. Brooks, A. Gelman, G. Jones, and X. Meng. Chapman & Hall/CRC.

Ruelle, D., and Takens, F., 1971: On the nature of turbulence. *Commun. math. Phys.*, **20**, 167–192.

Rulkov, N. F., Sushchik, M. M., Tsimring, L. S., and Abarbanel, H. D. I., 1995: Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E*, **51**(2), 980–994.

Sakaguchi, H., 2002: Parameter evaluation from time sequences using chaotic synchronization. *Physical Review E*, **65**(027201).

So, P., Ott, E., and Dayawansa, W. P., 1994: Observing chaos: Deducing and tracking the state of a chaotic system from limited observation. *Physical Review E*, **49**(4), 2650–2660.

Sorrentino, F., and Ott, E., 2009: Using synchronization of chaos to identify the dynamics of unknown systems. *Chaos*, **19**(033108).

Strogatz, S. H., 1994: *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Perseus Books Publishing, LLC.

Stühmer, W., Methfessel, C., Sakmann, B., Noda, M., and Numa, S., 1987: Patch clamp characterization of sodium channels expressed from rat brain cDNA. *Eur Biophys J*, **14**, 131–138.

Szendro, I. G., Rodríguez, M. A., and López, J. M., 2009: On the problem of data assimilation by means of synchronization. *Journal of Geophysical Research*, **114**(D20109).

Szücs, A., 2012: The size of BNST and molluscan neurons. Private communication.

Tabak, J., Murphey, C. R., and Moore, L. E., 2000: Parameter estimation for single neuron models. *J. Comput. Neurosci.*, **9**, 215–236.

Takens, F., 1981: Detecting strange attractors in turbulence. In *Dynamic Systems and Turbulence, Warwick 1980*, editors D. A. Rand, and L. Young, volume 898 of *Lecture Notes in Mathematics*, 366–381. Springer-Verlag.

Toth, B. A., Kostuk, M., Meliza, C. D., Margoliash, D., and Abarbanel, H. D. I., 2011: Dynamical estimation of neuron and network properties I: variational methods. *Biological Cybernetics*, **105**, 217–237.

Van Kampen, N. G., 2007: *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, third ed. edition.

Vanier, M. C., and Bower, J. M., 1999: A comparative survey of automated parameter-search methods for compartmental neuron models. *J. Comput. Neurosci.*, **7**, 149–171.

Voss, H. U., Timmer, J., and Kurths, J., 2004: Nonlinear system identification from uncertain and indirect measurements. *International Journal of Bifurcation and Chaos*, **14**, 1905–1933.

Wächter, A., 2002: *An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering*. Phd thesis, Carnegie Mellon University.

Wang, C. Z. H., Herbst, J. A., Keller, G. B., and Hahnloser, R. H. R., 2008: Rapid interhemispheric switching during vocal production in a songbird. *PLoS Biol*, **6**(10).

White, S. H., 1970: A study of lipid bilayer membrane stability using precise measurements of specific capacitance. *Biophysical Journal*, **10**(12), 1127 – 1148.

Wolpert, D. H., and Macready, W. G., 1997: No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, **1**(1).

Yu, W., Guanrong, C., Jinde, C., Jinhu, L., and Parlitz, U., 2007: Parameter identification of dynamical systems from time series. *Physical Review E*, **75**(067201).

Zinn-Justin, J., 2002: *Quantum Field Theory and Critical Phenomena*. International Series of Monographs on Physics. Oxford Science Publications.