


Fall 2014

Probabilistic uncertainty quantification and experiment design for nonlinear models: Applications in systems biology

Vu Cao Duy Thien Dinh
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

 Part of the [Biomedical Engineering and Bioengineering Commons](#), [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Dinh, Vu Cao Duy Thien, "Probabilistic uncertainty quantification and experiment design for nonlinear models: Applications in systems biology" (2014). *Open Access Dissertations*. 259.
https://docs.lib.purdue.edu/open_access_dissertations/259

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Vu Cao Duy Thien Dinh

Entitled
PROBABILISTIC UNCERTAINTY QUANTIFICATION AND EXPERIMENT DESIGN FOR
NONLINEAR MODELS: APPLICATIONS IN SYSTEMS BIOLOGY

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Professor Gregory T. Buzzard

Professor Ann E. Rundell

Professor Zhilan Feng

Professor Guang Lin

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Professor Gregory T. Buzzard

Approved by Major Professor(s): _____

Approved by: Professor David Goldberg

12/04/2014

Head of the Department Graduate Program

Date

PROBABILISTIC UNCERTAINTY QUANTIFICATION AND EXPERIMENT
DESIGN FOR NONLINEAR MODELS: APPLICATIONS IN SYSTEMS
BIOLOGY

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Vu Cao Duy Thien Dinh

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

The completion of this work would have not been possible without the support of my colleagues and mentors. I would like to specifically thank Professors Gregory Buzzard and Ann Rundell for their support and advice, which have greatly aided my development as a researcher. Professors Zhilan Feng and Guang Lin, who fill out my thesis committee, have provided helpful feedback for which I am grateful. My fellow graduate students from the Rundell Lab and the Theoretical Machine Learning Group also have my gratitude for many engaging conversations.

My time at Purdue was made enjoyable in large part due to many friends and groups that have become part of my life. I would like to send special thanks to friends from the Vietnamese International Student Association at Purdue, my office-mates Alan Legg, Jonathan Montano and Mike Perlmutter, my colleagues from the Department of Mathematics, and Nam-Anh Nguyen for all the wonderful time spent.

Lastly, I owe many thanks to my family, whose support and encouragement have never faltered.

TABLE OF CONTENTS

| | Page |
|--|------|
| LIST OF FIGURES | vi |
| ABSTRACT | x |
| CHAPTER 1. INTRODUCTION | 1 |
| 1.1 Objectives | 1 |
| 1.2 Background | 2 |
| 1.3 Organization | 4 |
| 1.3.1 A probabilistic framework for uncertainty quantification and experimental design in the face of unidentifiability | 4 |
| 1.3.2 Effective sampling schemes for behavior discrimination | 5 |
| 1.3.3 Data-free identifiability analysis and data-free uncertainty quan- tification | 6 |
| 1.3.4 Convergence of perturbed Monte Carlo Markov Chains | 7 |
| 1.4 Concluding remarks | 8 |
| 1.5 References | 9 |
| CHAPTER 2. EXPERIMENTAL DESIGN FOR DYNAMICS IDENTIFICA- TION OF CELLULAR PROCESSES | 11 |
| 2.1 Preface | 11 |
| 2.2 Abstract | 11 |
| 2.3 Introduction | 12 |
| 2.4 Mathematical framework | 18 |
| 2.4.1 Model formulation | 18 |
| 2.4.2 Expected Dynamics Estimator (EDE) | 19 |
| 2.5 EDE Consistency for noise-free data | 20 |
| 2.5.1 Randomly chosen experimental design points | 21 |
| 2.5.2 Design Points Using the Maximally Informative Next Experi- ment | 25 |
| 2.6 EDE Consistency with Noisy Data | 31 |
| 2.7 EDE consistency with model mismatch | 36 |
| 2.8 Proofs of Supporting Lemmas | 39 |
| 2.8.1 Lemma 2.5.1 | 39 |
| 2.8.2 Lemma 2.5.2 | 41 |
| 2.9 Numerical examples | 42 |
| 2.9.1 A simple ODE model | 42 |
| 2.9.2 An ODE model of the T-cell signaling pathway | 45 |

| | Page |
|--|------|
| 2.10 Conclusion | 50 |
| 2.11 References | 53 |
| CHAPTER 3. EFFECTIVE SAMPLING SCHEMES FOR BEHAVIOR DIS- CRIMINATION IN NONLINEAR SYSTEMS | 55 |
| 3.1 Preface | 55 |
| 3.2 Abstract | 55 |
| 3.3 Introduction | 56 |
| 3.4 Methodology | 59 |
| 3.4.1 Description of the algorithm | 59 |
| 3.4.2 Main results | 61 |
| 3.4.3 Comparison to other approaches | 64 |
| 3.5 Convergence results | 65 |
| 3.5.1 Low-discrepancy sampling | 66 |
| 3.5.2 Sequential sampling | 73 |
| 3.6 Behavior Discrimination in enzymatic networks. | 75 |
| 3.6.1 A model of the acute inflammatory response to infection . . | 75 |
| 3.6.2 A model of collagen degradation | 78 |
| 3.6.3 A model of the T-cell signaling pathway | 79 |
| 3.7 Additional properties | 83 |
| 3.7.1 Convergence | 83 |
| 3.7.2 Dependence on the number of terms in polynomial expression | 84 |
| 3.7.3 Boundary with multiple components | 85 |
| 3.7.4 Robustness | 85 |
| 3.8 Conclusions and discussions | 87 |
| 3.9 References | 90 |
| CHAPTER 4. DATA-FREE IDENTIFIABILITY ANALYSIS OF BIOLOGI- CAL SYSTEMS | 91 |
| 4.1 Introduction | 91 |
| 4.2 Data-free identifiability | 93 |
| 4.2.1 Mathematical setting | 93 |
| 4.2.2 Uncertainty and identifiability | 94 |
| 4.3 A unifying framework for data-free identifiability analysis and a priori uncertainty quantification | 96 |
| 4.3.1 A Bayesian framework for practical data-free identifiability anal- ysis ($ \mathcal{E} < \infty$ and $\sigma > 0$) | 96 |
| 4.3.2 A reinterpretation of structural identifiability | 98 |
| 4.3.3 Dynamics identifiability and a priori uncertainty quantification | 102 |
| 4.3.4 Computational procedure | 103 |
| 4.4 Results | 104 |
| 4.4.1 An intuitive example | 105 |
| 4.4.2 A model of influenza A virus infection | 106 |

| | Page |
|--|------|
| 4.4.3 Analysis of Goodwin's model | 109 |
| 4.4.4 A model of the T-cell signalling pathway | 112 |
| 4.5 Convergence analysis | 117 |
| 4.6 Discussions and Conclusions | 120 |
| 4.7 References | 122 |
| CHAPTER 5. CONVERGENCE OF THE GRIDDY GIBBS SAMPLING AND OTHER PERTURBED MARKOV CHAINS | 123 |
| 5.1 Abstract | 123 |
| 5.2 Introduction | 124 |
| 5.3 Mathematical framework | 127 |
| 5.3.1 Gibbs transition | 128 |
| 5.3.2 Ergodic properties of the Markov Chains generated by the Gibbs sampling | 129 |
| 5.3.3 Griddy Gibbs transition | 130 |
| 5.4 Existence, uniqueness, and regularity of the invariant measure of a monte carlo markov chain generated by the griddy gibbs sampling . | 131 |
| 5.4.1 Existence and uniqueness | 132 |
| 5.4.2 Some supporting lemmas | 133 |
| 5.4.3 Regularity | 136 |
| 5.5 Sensitivity and convergence of non-uniformly ergodic Markov Chains | 137 |
| 5.5.1 Continuity of eigenspaces for eigenvalue 1 | 137 |
| 5.5.2 Convergence results | 140 |
| 5.6 Extension to non-compact support distributions | 144 |
| 5.7 Numerical examples | 147 |
| 5.7.1 A 2D example | 147 |
| 5.7.2 An example in systems biology. | 149 |
| 5.8 Conclusion | 152 |
| 5.9 References | 157 |
| CHAPTER 6. CONCLUSIONS AND FUTURE WORK | 159 |
| 6.1 Summary of work | 159 |
| 6.2 Future work: Localized analysis/uncertainty quantification and unsu- pervised behavior discrimination of biological systems | 161 |
| 6.3 Conclusion | 164 |
| 6.4 References | 165 |
| VITA | 166 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 2.1 (Two dimensional paramter space) Left: Measured time points designed by MINE criteria. The algorithm focuses on two regions in time that capture the system dynamics. Right: The L^∞ errors of EDE on log-scale in three different cases: (i) Data collected with no noise, using the original MINE criteria (2.1) (ii) Data with Gaussian noise, using the original MINE criteria (2.1), (iii) Data with noise, using criterion (2.6) on a finite set of output values and possible measurement time points. | 44 |
| 2.2 (19-dimensional paramter space) Design points in three different cases: (i) Data collected with no noise, using the original MINE criteria (2.1) (ii) Data with Gaussian noise, using the original MINE criteria 2.1, (iii) Data with noise, using criterion 2.6 on a finite set of output values and possible measurement time points. | 46 |
| 2.3 (19-dimensional paramter space). The L^∞ errors of EDE on log-scale in three different cases: (i) Data collected with no noise, using the original MINE criteria (2.1) (ii) Data with Gaussian noise, using the original MINE criteria (2.1), (iii) Data with noise, using criterion (2.6) on a finite set of output values and possible measurement time points. | 48 |
| 2.4 (19-dimensional paramter space). Left: EDEs using different sparse grid interpolators to approximate the dynamics. The EDEs are evaluated after 10 steps. Right: Expected dynamics estimator and predicted confidence intervals of the output dynamics with $\epsilon = 0.05$ | 51 |
| 3.1 (A) A model of the acute inflammatory response to infection: The predicted boundaries computed by sparse sampling at the 16-point, 36-point and 80-point sparse grid nodes. (B) A model of collagen degradation: the design points and predicted boundaries computed at the nodes of the 80-point sparse grid. In both figures, the expected boundaries are computed using a 10^5 -point Monte Carlo Markov Chain on a 13-dimensional coefficient space. | 77 |

| Figure | Page | |
|--------|---|-----|
| 3.2 | Relative contribution of MT1-MMP and MMP2 to collagen degradation. The boundary separates points for which the contribution of MT1-MMP is dominant from points for which MMP2's contribution is dominant: (A) Design points and predicted boundary derived by the sequential sampling scheme. (B) A characterization of uncertainty in discrimination by variance. Notice that the points with high variance lie around the true boundary, which explains why the data sampled on the figure on the left also tends to focus around the true boundary. | 80 |
| 3.3 | A model of the T-cell signalling pathway with discrimination based on a threshold value for pERK at the final simulation time: (A) 3-dimensional case: Design points and predicted boundary derived by the sequential sampling schemes with $N=57$. (B) 8-dimensional case with $N=321$: Error in prediction as the number of samples increase in three different scenarios: Latin hypercube sampling, sequential sampling scheme and oracular sampling | 82 |
| 3.4 | Convergence rate of prediction error for the function in equation (3.18) when number of samples increases. In both cases, when the number of sampling points increase, the error of prediction converges to zeros. However, the convergence rate of the sequential scheme is significantly faster than that of the sparse grid sampling. | 84 |
| 3.5 | (A) Decrease in prediction error when the number of terms (N) in the approximation increases for the function in equation (3.19). (B) Sample predicted boundary with different value of N . The boundaries are computed using 500 samples collected uniformly at random, while the error rates are estimated by the empirical prediction error on 10^6 uniformly distributed random points. | 86 |
| 3.6 | (A) Example of a boundary with multiple components: discrimination between the region of positive and negative values of the elliptic function in equation (3.20) (B) Performance of the algorithm when multiple assumptions of the setting are mildly violated (equation (3.21)). | 88 |
| 4.1 | Uncertainty in identifying model parameters and dynamics: Experiments are to be made for $x_2(1)$ and $x_2(3)$, where data contains noise of standard deviation $\sigma = 0.01$. (Top) Identifiability of model parameters: $g_1(q) = q_1$, $g_2(q) = q_2$, $g_3(q) = q_3$. (Bottom) Identifiability of model dynamics: $g_{k,t}(q) = x_k(t)$, $k = 1, \dots, 3$, $t \in [0, T]$. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable. | 107 |

| Figure | Page |
|---|------|
| 4.2 The dynamics of $x_2(t)$ with two different parameter configurations that have same output at $t = 1s$ and $t = 3s$ | 108 |
| 4.3 Structural identifiability analysis of Baccam's model. (Top) Uncertainty in identifying model parameters. (Bottom) Two different parameters set with indistinguishable dynamics. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable. | 110 |
| 4.4 Identifiability analysis of Goodwin's model: (Top) Uncertainty in identifying model parameters; Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable. (Bottom) Two different parameters set with indistinguishable dynamics | 113 |
| 4.5 Analysis of Lipniack's model: (Top) Variance in predicting ω^i ; (Bottom) Variance in predicting different state variables. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable. | 115 |
| 4.6 Analysis of Lipniack's model | 116 |
| 5.1 Comparison between Gibbs sampling and Griddy Gibbs sampling: Although the two transition operators P and Q are close, the Markov chain $\{Y_n\}$ is not reversible in general, so the existence and uniqueness of the invariant measure η is not guaranteed. Even when η uniquely exists, an estimate of the distance between π and η is needed to guarantee the validity of the Griddy Gibbs sampling. | 131 |
| 5.2 Left: Error of the 1D marginal empirical cumulative distribution function, and Right: error of the empirical cumulative distribution function, both as a function of the number of points used in the approximation grid. . | 148 |
| 5.3 Conditional and marginal distribution for the T-cell model. Left: The difference between the conditional distributions on the first parameters (one curve for each value of this parameter). Right: The difference between the marginal joint distributions of the first two parameters, achieved from Griddy Gibbs and Tierney's algorithm. Figure 4 shows that the differences between corresponding ECDFs are of the same magnitude as the error of the Monte Carlo method ($O(\frac{1}{\sqrt{N}})$, where N is the number of samples) . | 153 |

| Figure | Page |
|---|------|
| 5.4 The difference between the marginal distributions computed by Griddy Gibbs and Tierney's algorithm is of the same magnitude as the error of the Monte Carlo method itself ($O(\frac{1}{\sqrt{N}})$, where N is the number of samples). | 154 |
| 5.5 Left: The expected dynamics estimator based on one data point, generated by Griddy Gibbs and Tierney's samples. Right: Auto-Correlation coefficients of the Markov Chains generated by Griddy Gibbs algorithm and Tierney's algorithm. | 155 |
| 6.1 Comparison of (Top) non-smooth EMPC control surface and (Bottom) Localized EMPC controller | 162 |
| 6.2 The 2-dimensional parameter space of the Fitz Hugh-Nagumo model is partitioned according to dynamics behavior using the spectral clustering method based on the Pearson correlation distance: Red and green regions correspond to the oscillatory and transient behavior of the membrane potential. Blue and black regions both correspond to cases when the membrane potential saturates at a high level; the distinguishing feature is that in the black region, the membrane potential decreases at the beginning before being activated and saturating. | 163 |

ABSTRACT

Dinh, Vu Cao Duy Thien Ph.D., Purdue University, December 2014. Probabilistic Uncertainty Quantification and Experiment design for nonlinear models: applications in systems biology. Major Professors: Gregory T. Buzzard.

Despite the ever-increasing interest in understanding biology at the system level, there are several factors that hinder studies and analyses of biological systems. First, unlike systems from other applied fields whose parameters can be effectively identified, biological systems are usually unidentifiable, even in the ideal case when all possible system outputs are known with high accuracy. Second, the presence of multivariate bifurcations often leads the system to behaviors that are completely different in nature. In such cases, system outputs (as function of parameters/inputs) are usually discontinuous or have sharp transitions across domains with different behaviors. Finally, models from systems biology are usually strongly nonlinear with large numbers of parameters and complex interactions. This results in high computational costs of model simulations that are required to study the systems, an issue that becomes more and more problematic when the dimensionality of the system increases. Similarly, wet-lab experiments to gather information about the biological model of interest are usually strictly constrained by research budget and experimental settings. The choice of experiments/simulations for inference, therefore, needs to be carefully addressed.

The work presented in this dissertation develops strategies to address theoretical and practical limitations in uncertainty quantification and experimental design of non-linear mathematical models, applied in the context of systems biology. This work resolves those issues by focusing on three separate but related approaches:

- (i) the use of probabilistic frameworks for uncertainty quantification in the face of unidentifiability

- (ii) the use of behavior discrimination algorithms to study systems with discontinuous model responses
- and
- (iii) the use of effective sampling schemes and optimal experimental design to reduce the computational/experimental costs.

This cumulative work also places strong emphasis on providing theoretical foundations for the use of the proposed framework: theoretical properties of algorithms at each step in the process are investigated carefully to give more insights about how the algorithms perform, and in many cases, to provide feedback to improve the performance of existing approaches. Through the newly developed procedures, we successfully created a general probabilistic framework for uncertainty quantification and experiment design for non-linear models in the face of unidentifiability, sharp model responses with limited number of model simulations, constraints on experimental setting, and even in the absence of data. The proposed methods have strong theoretical foundations and have also proven to be effective in studies of expensive high-dimensional biological systems in various contexts.

CHAPTER 1. INTRODUCTION

1.1 Objectives

This dissertation addresses theoretical and practical limitations in uncertainty quantification and experimental design of non-linear systems that are prevalent in biological studies. The work herein places emphasis on applications in systems biology, most prominently intracellular signaling networks. However, it is an objective of this work to create theoretical frameworks to address those limitations in a general mathematical setting. For that reason, the strategies developed herein are not limited by either the type of model or application contexts and are applicable beyond the scope of biological studies to improve the efficiency in analyzing mathematical models in other fields of predictive science.

From a computational viewpoint, the focus of this work is to tackle several primary technical hurdles to analyses of biological systems. These include: unidentifiability of parameters, discontinuity/sharp model responses, high-dimensionality, strong non-linearity and high experimental/computational expense. While covering a wide range of topics, this work focuses on three separate but related approaches to resolve these issues, namely,

- (i) the use of probabilistic frameworks for uncertainty quantification in the face of unidentifiability
- (ii) the use of behavior discrimination algorithms to study systems with discontinuous model responses
- and
- (iii) the use of effective sampling schemes and optimal experimental design to reduce the computational/experimental costs.

This cumulative work also places strong emphasis on providing theoretical foundations for the use of the proposed framework: theoretical properties of algorithms employed at each step in the process are investigated carefully to give more insights about how the algorithms would perform, and in many cases, to provide feedback to improve the performance of existing approaches.

The remainder of this introduction provides background material to inform the motivation for the work. Section 1.2 explains the primary issues in studying biological systems, while Sections 1.3 provides summaries of several approaches that are described in detail in later chapters and gives a general picture about the organization of the dissertation. This introduction also concludes with a few remarks for further reading.

1.2 Background

Despite the ever-increasing interest in understanding biology at the system level, there are several factors that hinder studies and analyses of biological systems [1].

First, unlike system from other applied fields whose parameters can be effectively constrained by data, biological systems are usually unidentifiable, even in the ideal case when all possible system outputs are known with high accuracy [2,3]. This comes from the fact that in order to attain system robustness, which is crucial for survival, living cells usually maintain various forms of system control, such as negative-feedback and feed-forward control, and system redundancy, whereby multiple components with equivalent functions are introduced for backup. The direct consequence is that for a given cell state, the system are insensitive to perturbation on many parameters that are required for an adequate description of the system [1]. On the other hand, due to technical limitations and experimental constraints, data collected to study biological systems are often sparse and noisy: the number of data points we can collect may even be smaller than the number of model parameters and the collected data may be severely contaminated by various types of noise [4]. These two different types of

parameter unidentifiability (namely, structural unidentifiability and practical unidentifiability), along with the strong nonlinearity of the model [5] and the non-normality of the noise distribution [6], render traditional methods of parameter estimation and experimental design via optimization implausible.

Second, as a common problem with high-dimensional and complex dynamical systems, the presence of multivariate bifurcations often leads the system to behaviors that are completely different in nature [7, 8]. In such cases, system outputs (as function of parameters/inputs) are usually discontinuous or have sharp transitions across domains with different behaviors. This property makes the problem of uncertainty quantification computationally intractable: polynomial-based techniques require inherent regularities of the approximated functions while multi-element methods are too expensive [9]. As a result, analyses of high-dimensional biological systems, such as sensitivity analysis, identifiability analysis or model order reduction, are usually performed locally in a neighborhood of a nominal parameter where the system outputs can be assumed to possess certain regularity [10–12]. To make a global approach to such analyses possible, regions of the parameter/input space with different qualitative behaviors need to be treated differently and should be identified before the analyses are performed. While there are a variety of methods to address this problem for linear systems, few successful techniques have been developed for nonlinear models. Existing methods often rely on numerical simulations without rigorous bounds on the numerical errors and usually require a large number of model evaluations, rendering those methods impractical for studies of high-dimensional and expensive systems [13, 14].

Finally, models from systems biology are usually strongly nonlinear with large numbers of parameters and complex interactions. This results in high computational costs for the model simulations required to study the systems, an issue that becomes more and more problematic when the dimensionality of the system increases. Similarly, wet-lab experiments to gather information about the biological model of interest are usually strictly constrained by research budget and experimental settings. The

choice of experiments/simulations for inference, therefore, needs to be carefully addressed [15]. This leads us to the problem of experimental design/ effective sampling schemes to study high-dimensional and computationally expensive models. Such sampling schemes/designs have been widely used in the uncertainty quantification [16], experimental design [17] and machine learning [18] literatures independently. However, at least in some contexts, algorithms of this type lack theoretical support and in some cases may lead to misleading and incorrect answers due to sampling bias [18]. One goal of this dissertation is to provide a common framework with strong theoretical foundations for the problems of experimental design and the concept of effective sampling schemes to study high-dimensional and computational expensive models.

1.3 Organization

1.3.1 A probabilistic framework for uncertainty quantification and experimental design in the face of unidentifiability

To some extent, this dissertation is centered around the problem of probabilistic uncertainty quantification and experimental design for dynamics identification of biological systems. This problem is analyzed in detail in Chapter 2 and was published in the journal *Bulletin of Mathematical Biology* [19].

In this work, to resolve the problem of unidentifiability, we take a different approach toward the problem of system identification. In contrast to the tradition of using optimization techniques to achieve a "best" estimate of parameter value for inferences about the system, we focus directly on the estimation and uncertainty quantification of the system dynamics of interest. This is done within a probabilistic framework with the use of statistical inference. Specifically, available data are used to induce a probability distribution on the parameter space. The expected value and the variance in prediction with respect to this distribution then act as an estimate of the quantity of interest and a representation of uncertainty in predicting it, respectively.

Within this framework, we avoid the task of parameter fitting and enable system analysis, design and control even in the case of system unidentifiability.

Building upon this approach, we introduce the Expected Dynamics Estimator (EDE) and address the problem of using nonlinear models to design experiments to characterize the dynamics of cellular processes by using the approach of the Maximally Informative Next Experiment (MINE, which was proposed in [17] and subsequently in [4, 5]). We then prove the consistency of this estimator (uniform convergence to true dynamics) even when the chosen experiments cluster in a finite set of points. We extend this proof of consistency to various practical assumptions on noisy data and moderate levels of model mismatch. Through the derivation and proof, we develop a relaxed version of MINE that is more computationally tractable and robust than the original formulation. The results are illustrated with numerical examples on two nonlinear ordinary differential equation models of biomolecular and cellular processes.

1.3.2 Effective sampling schemes for behavior discrimination

To address the issue of uncertainty quantification/system analysis in the presence of bifurcation and discontinuous responses, we propose the construction of a map of the parameter space by different qualitative behaviors. That is, instead of performing local analysis/approximation around a nominal value, we partition the parameter space into regions of parameter values with similar qualitative behaviors (for example, regions with transient dynamics and regions with oscillatory dynamics). Within each region, the dynamics will possess certain regularity, so that smooth approximations or polynomial-based uncertainty quantification can be performed with high accuracy. The regions of the parameter/input space with different qualitative behaviors need to be treated differently and should be identified before the analyses are performed.

This idea is formalized in Chapter 3, in which the concept of behavior discrimination is defined as the problem of identifying sets of parameters for which the system

does (or does not) reach a given set of states. This chapter was published in the journal *International Journal of Uncertainty Quantification* [20].

In this work, we further developed the framework proposed in Chapter 2 to address the problem of behavior discrimination. In our approach, we directly parameterize the, yet unknown, boundary by the zero level-set of a polynomial function, then use statistical inference on available data to identify the coefficients of the polynomial. Building upon this framework, we consider the problem of choosing effective data sampling schemes for behavior discrimination of nonlinear systems in two different settings: the low-discrepancy sampling scheme, and the uncertainty-based sequential sampling scheme. In both cases, we successfully derive theoretical results about the convergence of the expected boundary to the true boundary of interest. Both methods have also proven to be effective in studies of expensive high-dimensional biological systems in various contexts.

1.3.3 Data-free identifiability analysis and data-free uncertainty quantification

Established as an extension of the probabilistic uncertainty quantification framework proposed in Chapter 2, Chapter 4 addresses the problem of quantifying the uncertainty in prediction of a quantity of interest before actual experimental observations are made. Within this new framework, we can explore the concept of *data-free identifiability*, which concerns the question of unique system identification under a given experimental setting, without actual experimental observations. As a data-independent property, data-free identifiability can be considered as a generalization of structural identifiability while at the same time addressing identifiability in the face of experimental constraints and noises.

With this novel concept, we propose a Bayesian approach to address system identifiability when data are not yet available. As we illustrate later, our approach is global, strongly theoretically supported, amenable to high-dimensional cases, can be

used to study various types of identifiability and is compatible with a large class of experimental settings. The framework is also built not only to assess parameter identifiability but also to quantify the uncertainty in prediction of any quantity of interest, and hence, can be used to address dynamics identifiability, a concept that has become of growing interest in the recent years. This also draws a direct connection between studies of identifiability and the concept of uncertainty quantification in predictive sciences. With this method, we also attempt to lay a unifying framework for the problems of structural/practical identifiability analysis, dynamics identifiability analysis and data-free uncertainty quantification.

1.3.4 Convergence of perturbed Monte Carlo Markov Chains

The probabilistic framework for identifiability analysis, uncertainty quantification and experimental design of non-linear models proposed in this dissertation is made possible by the employment of Monte Carlo Markov Chain methods to sample from a likelihood function on some high-dimensional parameter spaces. Since direct computation of the likelihood function is costly, in practice, approximation methods are usually employed to reduce some of the computational burden.

Throughout this work, we use Griddy Gibbs sampling as an effective way to sample from the likelihood functions of interest. The Griddy Gibbs sampling was proposed by Ritter and Tanner [21] as a computationally efficient approximation of the well-known Gibbs sampling method. The algorithm is simple and effective and has been used successfully to address problems in various fields of applied science. However, the approximate nature of the algorithm has prevented it from being widely used: the Markov chains generated by the Griddy Gibbs sampling method are not reversible in general, so the existence and uniqueness of its invariant measure was not guaranteed. Even when such an invariant measure uniquely exists, there was no estimate of the distance between it and the probability distribution of interest, hence no means to ensure the validity of the algorithm as a means to sample from the true distribution.

In Chapter 5, we show, subject to some fairly natural conditions, that the Griddy Gibbs method has a unique, invariant measure. Moreover, we provide L^p estimates on the distance between this invariant measure and the corresponding measure obtained from Gibbs sampling. These results provide a theoretical foundation for the use of the Griddy Gibbs sampling method. We also address a more general result about the sensitivity of invariant measures under small perturbations on the transition probability. That is, if we replace the transition probability P of any Monte Carlo Markov Chain by another transition probability Q where Q is close to P , we can still estimate the distance between the two invariant measures. The distinguishing feature between our approach and previous work on convergence of perturbed Markov Chain is that by considering the invariant measures as fixed points of linear operators on function spaces, we don't need to impose any further conditions on the rate of convergence of the Markov Chain. For example, the results we derived in this paper can address the case when the considered Monte Carlo Markov Chains are not uniformly ergodic.

1.4 Concluding remarks

This dissertation develops new procedures to address challenges in studying biological systems via probabilistic frameworks for uncertainty quantification and experimental design in the face of unidentifiability, sharp model responses with limited number of model simulations, constraints on experimental setting, and even in the absence of data. Theoretical foundations and the effectiveness of the proposed methods in studies of expensive high-dimensional biological systems will be investigated in detail in various contexts in the subsequent chapters.

1.5 References

- [1] Hiroaki Kitano, (2002), Systems Biology: a Brief Overview. Science, Vol. 295, Issue 5560.
- [2] O.T. Chis, J.R. Banga, E. Balsa-Canto, (2011), Structural identifiability of systems biology models: a critical comparison of methods. PloS One 6: e27755.
- [3] H. Miao, X. Xia, A.S. Perelson, H. Wu , (2011), On identifiability of nonlinear ode models and applications in viral dynamics. SIAM review 53: 3–39.
- [4] M. M. Donahue, G. T. Buzzard, and A. E. Rundell, (2010), Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids. IET System Biology, 4:249–262.
- [5] J. N. Bazil, G. T. Buzzard, and A. E. Rundell, (2011), A global parallel model based design of experiments method to minimize model output uncertainty. Bulletin of Mathematical Biology, 74:688–716.
- [6] Gregory T. Buzzard and Bradley J. Lucier, (2013), Optimal filters for high-speed compressive detection in spectroscopy. Proceedings of SPIE Volume 8657, Computational Imaging XI, 865707 (February 14, 2013).
- [7] David Angeli, James E. Ferrell, and Eduardo D. Sontag, (2004), Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. Proceedings of the National Academy of Sciences of the United States of America, 101.7: 1822-1827.
- [8] T. Lipniacki, B. Hat, J. R. Faeder, and W. S. Hlavacek, (2008), Stochastic effects and bistability in T-cell receptor signaling. Journal of Theoretical Biology, 254(1):110–122.
- [9] K. Sargsyan, C. Safta, B. Debusschere, and H. Najm, (2012), Uncertainty quantification given discontinuous model response and a limited number of model runs. SIAM Journal on Scientific Computing, 34(1):B44–B64.
- [10] Harvey M. Wagner, (1995), Global sensitivity analysis. Operations Research 43.6 (1995): 948-969.
- [11] Gregory T. Buzzard, (2012), Global sensitivity analysis using sparse grid interpolation and polynomial chaos. Reliability Engineering and System Safety 107 (2012): 82-89.
- [12] Sam T. Roweis, and Lawrence K. Saul, (2012), Nonlinear dimensionality reduction by locally linear embedding. Science 290.5500: 2323-2326.
- [13] A. Donzé, G. Clermont,, and C. J. Langmead, (2010), Parameter synthesis in nonlinear dynamical systems: Application to systems biology, *Journal of Computational Biology*, 17(3):325–336.
- [14] A. Donzé, E. Fanchon, L. M. Gattepaille, O. Maler, and P.Tracqui, (2011), Robustness analysis and behavior discrimination in enzymatic reaction networks. *PLoS One*, 6(9):e24246.

- [15] F. Pukelsheim, (1993), Optimal design of experiments. John Wiley and Sons: New York.
- [16] John D. Jakeman, Richard Archibald, and Dongbin Xiu, (2011), Characterization of discontinuities in high-dimensional stochastic problems on adaptive sparse grids. *Journal of Computational Physics* 230.10: 3977-3997.
- [17] W. Dong, X. Tang, Y. Yu, R. Nilsen, R. Kim, J. Griffith, J. Arnold, and H. Schuttler, (2008), Systems biology of the clock in *neurospora crassa*. *PLoS ONE*, page e3105.
- [18] Sanjoy Dasgupta, (2011) Two faces of active learning. *Theoretical computer science* 412.19 (2011): 1767-1781.
- [19] Vu Dinh, Ann E. Rundell and Gregory T. Buzzard, (2014), Experimental Design for Dynamics Identification of Cellular Processes. *Bulletin of Mathematical Biology*, 76.3: 597-626.
- [20] Vu Dinh, Ann E. Rundell and Gregory T. Buzzard, (2014), Effective sampling schemes for behavior discrimination of nonlinear models. *International Journal for Uncertainty Quantification*.
- [21] C. Ritter and M. A. Tanner, (1992), Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *J. Amer. Stat. Assoc.*, 87(419):861–868.

CHAPTER 2. EXPERIMENTAL DESIGN FOR DYNAMICS IDENTIFICATION OF CELLULAR PROCESSES

2.1 Preface

The material presented in this chapter was originally published in the Bulletin of Mathematical Biology:

Vu Dinh, Ann E. Rundell and Gregory T. Buzzard. Experimental Design for Dynamics Identification of Cellular Processes. *Bulletin of Mathematical Biology*, 76.3 (2014): 597-626.

This article has been reproduced with material omitted or summarized to befit the focus of this dissertation. It has been modified to conform to the format required.

2.2 Abstract

We address the problem of using nonlinear models to design experiments to characterize the dynamics of cellular processes by using the approach of the Maximally Informative Next Experiment (MINE), which was introduced in [W. Dong, et al. Systems biology of the clock in *neurospora crassa*. *PLoS ONE*, page e3105, 2008] and independently in [M. M. Donahue, et al. Experiment design through dynamical characterization of non-linear systems biology models utilising sparse grids. *IET System Biology*, 4:249–262, 2010]. In this approach, existing data is used to define a probability distribution on the parameters; the next measurement point is the one that yields the largest model output variance with this distribution. Building upon this approach, we introduce the Expected Dynamics Estimator (EDE), which is the expected value using this distribution of the output as a function of time. We prove the consistency of this estimator (uniform convergence to true dynamics) even when

the chosen experiments cluster in a finite set of points. We extend this proof of consistency to various practical assumptions on noisy data and moderate levels of model mismatch. Through the derivation and proof, we develop a relaxed version of MINE that is more computationally tractable and robust than the original formulation. The results are illustrated with numerical examples on two nonlinear ordinary differential equation models of biomolecular and cellular processes.

2.3 Introduction

The development and simulation of mathematical models of cellular processes can enhance our understanding of the underlying biological mechanisms ([1]). Two important components of model development are the collection of data and the tuning of parameters for a given model structure to approximate the data. In many settings, the collection of data is difficult and/or expensive, while tuning model parameters to data often involves a difficult nonlinear optimization, with potentially many local optima. Moreover, the choice of data may make this tuning more or less difficult; thus we aim to design experiments to collect the most informative data for a given model structure.

The review [8] provides a broad overview of model-based experimental design methodologies for systems biology, including methods for various optimality conditions governing unique, structural and practical parameter identification. Of course, many books and articles have been written about experimental design, both from the frequentist and the Bayesian points of view. We make no attempt to review them all here; a classic mathematical reference is [15]. Many methods for experimental design focus on identifying the parameters – designing experiments to minimize some measure of uncertainty in the parameter values given a model structure.

In contrast, we are more concerned with developing a method to explore and elucidate the observable response (which we refer to as the output dynamics) of a cellular process rather than identifying the model parameters themselves. One

motivation for this is that for a systems biology model with N parameters, the set of possible output dynamics is often contained in a space of lower dimension $l \ll N$ (or perhaps in a small neighbourhood of such a space). This feature, which is an obstacle for the problem of unique parameter identification, is an advantage for designing experiments to identify dynamics: that is, we can choose a good design with very few experiments (approximately $O(l)$), but still obtain enough information to identify the dynamics.

Methods related to approximating the observable response as a function of independent input variables fall under the broad heading of regression or response surface methodology. Once again there are many books and articles on the topic of experiment design for fitting response surfaces, and there are many approaches for representing a response surface. Most such approaches (e.g., Kriging and generalized polynomial chaos) seek to approximate the response surface with a linear combination of a fixed set of basis functions, such as polynomials or trigonometric functions. See [11] for an overview.

In this paper we focus on experiment design for accurate approximation of the response surface using a given biologically-based model. However, beyond this, as explained in [5], the method we apply acts as a kind of imaging method for understanding the behavior of a cell. Based on an initial understanding of cell behavior (which is likened to an image from a microscope), we choose the next experiment to provide as much resolving power as possible in our next measurement (which is likened to focusing the microscope to enhance a particular feature). Instead of a linear combination of basis functions, as is often assumed in the experiment design literature, we assume the model structure encodes the dominant interactions and mechanisms using a nonlinear system of differential equations. To fix ideas, suppose our model output has the form $y = f(\omega, t)$, where $\omega \in \Omega \subset \mathbb{R}^N$ is a fixed vector of unknown parameters, and where our quantity of interest is the dynamics output, which is a function of time, $t \in [0, T]$ (more generally t could be a vector of inputs to represent any independent variables such as time, voltage, etc.). Measurements of y at a given

time t can be modeled as a random variable. An estimate of ω based on these random variables is also a random variable. This estimate can be used to estimate the output $y(t)$ for any t , again giving a random variable. Classical experiment design typically seeks to minimize the variance in the estimate of ω or y . One approach to designing experiments for accurate response surface modeling is to use the condition of G-optimality. Roughly, this condition chooses an experiment design to minimize the variance in the output. In the case of a model that is linear in parameters, the Kiefer-Wolfowitz Equivalence Theorem states that this is equivalent to D-optimality, in which the design is chosen to minimize the determinant of the covariance matrix for the estimate of ω (the inverse of the Fisher Information Matrix) [15, Chapter 9]. There is an extension of this result to nonlinear models [18]; however, this result depends upon knowing the true parameters in the model, which are not known in general. In [6, Section 5.6], this problem is addressed by either (i) using a minimax approach, in which the design is chosen to minimize over experiment designs the maximum over parameter space of the determinant; or (ii) using a Bayesian approach, in which an optimality criterion for a design (such as the determinant of the dispersion matrix or the maximum output variance) is averaged using a prior distribution on parameter space, and then the design is chosen to maximize this expected criterion. A computational difficulty with both of these approaches is the need to evaluate a complex optimality criterion at many points in parameter space for each candidate design.

Alternatively, the Maximally Informative Next Experiment (MINE) algorithm proposed in [5] and later [4] uses a sequential approach to experiment design in which existing data is used to construct a probability distribution on the parameter space; this distribution is then used to calculate the variance in the output as a function of time (perhaps normalized by expected experimental variance). Based on this calculation, the next measurement will be taken at the time point with highest current (normalized) variance. That is, the next sampled time point will be chosen at the time point that has highest current uncertainty in the output. This method of design

is modified to produce a parallel (nonsequential) design in [2]. Intuitively, each new point in such a design should provide the maximum possible information about the dynamics of the system and hence lead to convergence to the true system dynamics. This approach is theoretically appealing in that it doesn't depend on an estimate of the true parameter values, and it is computationally appealing in that it requires a relatively simple sampling over the parameter space according to a specified distribution; this may be achieved reasonably efficiently with Monte Carlo methods.

However, little is known about the convergence properties of this method: Is this method sufficient to characterize the response surface (in the limit as the number of experimental points tends to ∞)? In fact, in general, this scheme will not sample densely over the interval $[0, T]$, so it is not at all clear that it is sufficient to completely characterize the dynamics over this interval. Moreover, it's not entirely clear how to use these nondense samples to estimate the dynamics.

Motivated by the MINE algorithm, we address the following two problems for the identification of systems dynamics:

- (A) Specify the Dynamics Estimator: Given a set of data $(t_1, d_1), (t_2, d_2), \dots, (t_m, d_m)$ and a model $y = f(\omega, t)$, how should we estimate the system dynamics?
- (B) Prove convergence of Dynamics Estimator: For a given sequential approach to choosing measurement points t_j and given the dynamics estimator in (A), do the estimated dynamics converge to the true dynamics?

In the derivation of the solutions to these problems, we developed variations of the MINE algorithm that are more computationally efficient than the original. We describe these variations and solutions to Problems (A) and (B) with various assumptions in the body of the paper.

Most approaches to problem (A) use the data to estimate parameter values and then use these parameters to obtain the corresponding dynamics. For a complex, non-linear model this is a difficult optimization problem with possibly many local optima and perhaps even multiple global optima. In place of using an estimated vector of

parameters to estimate the dynamics, we propose what we call the expected dynamics estimator (EDE). This uses the available data to induce a probability distribution on parameter space and then averages the dynamic output using this distribution. There are a number of advantages of this method of dynamics identification over parameter identification. First, since the dynamics for a deterministic system are unique, we don't need to worry about multiple global solutions. Second, by using the EDE, we look for the average behaviour of the system (with respect to a carefully constructed probability distribution). This task is typically much simpler than solving a nonlinear optimization problem. Furthermore, the Markov Chain Monte Carlo method can be employed to reconstruct the system's true dynamics.

Another important advantage of the probabilistic framework over parameter estimation (via optimization) is that it is a feasible approach in cases of unidentifiability. A crucial problem in parameter estimation is the calculation of confidence intervals for the estimated parameters. In the simplest scenario when the number of data points is smaller than the number of parameters, any parameter estimation (via optimization) method will fail to provide a reliable estimate of the confidence region. Such methods (that return a single parameters estimate) will never be able to predict unknown output with high confidence (or any confidence at all). In order to do so, it needs to compute all possible parameter values that are consistent with available data, which is very unlikely in practice. This also extends to the case when the model's parameters are unidentifiable, which is a common phenomenon in systems biology. Our probabilistic framework provides a feasible way to address this issue: a given set of measurements gives a probability distribution on parameters, which can be used to construct confidence interval for output dynamics in addition to the EDE.

Problem (B) is a question about the consistency of the estimator (the ability to recover the true dynamics) as a function of a particular choice of measurement points. This question highlights the fact that the ability of the EDE to recover the true dynamics (consistency) depends heavily on the experimental design algorithm. We note here that although the MINE method shares some features of a Bayesian

approach, in that a probability distribution on parameters is updated based on new data samples, it does not fall into the class of Bayesian experimental design since the design points are not chosen to maximize an expected utility function.

This paper is organized to prove and illustrate the consistency of the EDE in various situations that progressively increase in complexity towards practical applicability. In Section 2.4, we introduce the mathematical framework and the main assumptions about the behaviour of the investigated model that we use throughout the paper. We also define the EDE to address problem (A). Section 2.5 addresses problem (B) for the ideal case when the investigated model is a correct model (can reproduce the true dynamics exactly) and data are noiseless. Theorem 2.5.1 deals with the case in which the experiments are made at random time points; this result is provided primarily to illustrate the ideas to be used in later results but in a setting that avoids some technical assumptions that are needed later. Theorem 2.5.2 and Theorem 2.5.4 provide results in the case when the experiments are designed sequentially as in [4] and [5] in two different settings: when the parameter space is discretized and when the set of possible measured time points and output values are discretized. We then extend the consistency result to a larger class of designs by relaxing the choice of a point with maximal variance to a point with variance within a fixed constant multiple of maximal variance (Theorem 2.5.3). Our results imply that for these designs, we can always recover the true dynamics, even if all the measurements are made in a small portion of the time interval $[0, T]$. Section 2.6 extends the result about EDE's consistency to the case when the experimental data are subject to random noise (Theorem 2.6.2). In this section we require that the set of possible measurement points is finite in order to guarantee convergence even in the face of noisy data. This assumption is reasonable for the practical implementation of any experiment design. Section 2.7 relaxes the requirement of a correct model by allowing for a bounded error between the true dynamics and the closest approximation of the model (Theorem 2.7.1). From this result, we also justify the use of approximation methods in the algorithm to design experiments. In Section 2.9, we illustrate our

theoretical findings and demonstrate the efficacy of our method to design sequential experiments for dynamics identification with various biological models. We also give an example to show that the choice of a design point within a fixed constant of maximal variance can lead to a faster rate of convergence of the EDE relative to the original MINE algorithm. It is worth noting that although the framework we use in this paper is sequential, one can extend the result to the parallel case following the approach suggested in [2].

2.4 Mathematical framework

2.4.1 Model formulation

We assume a mathematical model of a cellular process in the form

$$\dot{x} = \alpha(\omega, x) \quad (\text{System of ODEs})$$

$$x(0) = x_0(\omega) \quad (\text{Initial conditions})$$

$$y(t) = f(\omega, t) = \beta(\omega, x(t)) \quad (\text{Output})$$

where $x = (x_1, x_2, \dots, x_{n_x}) \in M \subset \mathbb{R}^{n_x}$ is the state variable, with M a subset of \mathbb{R}^{n_x} containing the initial state, and $f(\omega, t) \in \mathbb{R}^L$ is the observable response (output dynamics) that correspond to L different experimentally observable quantities. Throughout this paper, for the sake of simplicity, we will assume that $L = 1$. However, all of the arguments can be extended to the case of multi-dimensional output without any difficulty.

It is worth noting that the set of possible outputs is not necessarily the same as the number of dynamic variables occurring in the system. An output could be any kind of prediction, e.g. also a sum or ratio or even integral of dynamic variables. However, in the case $L = 1$, there is only one observable output y . Identification of y will lead to identification of all possible outputs as well as a characterization of the uncertainty in unidentifiable outputs.

The purpose of our experimental design framework is to determine as accurately as possible the output dynamics based on measurements. This is a kind of interpolation problem. We do not address the extrapolation problem, in which measurements of one output are used to make inference about an unobservable quantity.

The vector of unknown parameters is denoted by $\omega = (\omega_1, \dots, \omega_N) \in \mathbb{R}^N$ and is assumed to belong to a subset Ω of \mathbb{R}^N . In most parts of the paper, the parameter space Ω will be assumed to be an open set along with a probability measure on Ω , or a discrete subset of \mathbb{R}^n along with a probability measure. The components of α and β are assumed to be C^1 functions of their arguments. These functions and initial conditions may depend on the parameter vector $\omega \in \Omega$.

The system will therefore be associated with the mapping $F : \Omega \rightarrow C^1([0, T], \mathbb{R})$ defined by $F(\omega) = f(\omega, \cdot)$, where $f(\omega, \cdot)$ is the observable response of the system as a function of $t \in [0, T]$ for a given ω . The image of Ω under f , $Y = f(\Omega, \cdot) \subset C^1([0, T], \mathbb{R})$ will be referred to as the dynamics space in this paper.

Throughout this paper, the true dynamics and the data values at a given time, t , will be denoted by $g(t)$ and $d(t)$, respectively. We assume that $d(t) = g(t) + \epsilon$, where ϵ is a random variable describing the noise in measurements. In Section 2.5, we assume that $\epsilon = 0$, so that the data are completely noise-free. In later sections we address the case of noisy data. In Sections 2.5 and 2.6 we assume that the model is correct; that is, there is some $\omega_0 \in \Omega$ so that $f(\omega_0, t) = g(t)$ for all $t \in [0, T]$. We relax this assumption in Section 2.7.

2.4.2 Expected Dynamics Estimator (EDE)

A given data set $(t_1, d_1), \dots, (t_n, d_n)$ will be used to induce a probability distribution on the parameter space. We do this through the normalized likelihood function,

$$p_n(\omega) = c_n \exp\left(-\sum_{i=1}^n (d_i - f(\omega, t_i))^2\right),$$

(or a variant of this expression), where c_n is a constant so that p_n is a probability distribution on Ω . (Note that if no data has been observed, the distribution p_0 is just the uniform distribution in Ω .)

The expected dynamics estimator (EDE) with respect to this probability distribution is then

$$\hat{D}_n(t) = E_{p_n(\omega)}[f(\omega, t)],$$

which we use as an estimator of the system's true dynamics. Thus, instead of trying to maximize the likelihood function in order to estimate dynamics, we average the output dynamics, weighted by the probability as determined by the likelihood function. It is also worth noting that the EDE is the natural estimator that is used frequently as a part of the ensemble method, and is usually computed by Monte Carlo Markov Chain methods.

2.5 EDE Consistency for noise-free data

In this section, we establish results about the consistency of the expected dynamics estimator, that is, the ability to recover the true dynamics under a specified experimental design. The proof will be provided for two different cases:

1. When the sampled time points $\{t_n\}$ are chosen at random from an absolutely continuous probability distribution μ on $[0, T]$.
2. When the sampled time points are chosen sequentially as in [4] and [5], where the next sampled time point will be the point with highest current uncertainty (output variance).

Before moving forward to analyze the convergence of the EDE in these two cases, it is worth mentioning the distinction between two different sources of uncertainties (in both parameters and output dynamics): noise in data (aleatoric uncertainty), and structural uncertainty (epistemic uncertainty) in the model. Given a set of noise-free data, the corresponding set of parameter values that fit the data perfectly well can

still be an infinite set (usually, is a union of low-dimensional manifolds). The simplest example for this phenomenon is when the number of data is less than the number of model parameters.

In unidentifiable nonlinear systems, this set of "fitted" parameters may not collapse to a point mass even if all measurable outputs are known completely. This uncertainty in parameters may never be eliminated. The forward propagation of this uncertainty to the output space is the target in this noise-free framework.

The likelihood proposed in the noise-free setting, therefore, is not associated with noise in data, but with the structural uncertainty in model parameters from available data (how well a parameter set fits the data). Instead of focusing on a low-dimensional set of "fitted" parameters, we use an everywhere positive likelihood function to constrain the parameter space. From a methodological point of view, the idea here is similar to those behind simulated annealing methods for optimization and multiple Monte Carlo Markov Chains method for statistical inference: since the objects of interest is difficult to identify, we relax it by heated objects that are easier to study and use our experimental design algorithm to sequentially reduce the temperature in an optimal way to identify the true output dynamics.

2.5.1 Randomly chosen experimental design points

To illustrate the ideas used in later results, we consider the case when the sampled time points $\{t_n\}$ are chosen independently at random from an absolutely continuous probability distribution, with the assumption that the data are noise free (i.e. $d(t_i) = g(t_i)$ for all i). In this setting, we have the following theorem, which says that in the limit when $n \rightarrow \infty$, the expected dynamics estimator converges to the system's true dynamics.

Theorem 2.5.1 *Suppose there exists $\omega_0 \in \Omega$ such that $f(\omega_0, t) = g(t)$ for all $t \in [0, T]$. Suppose also $\{t_n\}$ are chosen independently at random from an absolutely continuous probability distribution μ on $[0, T]$ and that $1 \leq r < \infty$. Let*

$$p_n(\omega) = c_n \exp \left(- \sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \right),$$

where c_n is the normalizing constant to ensure that p_n is a probability distribution on Ω . Then for all $t \in [0, T]$,

$$\lim_{n \rightarrow \infty} E_{p_n}[f(\omega, t)] = g(t).$$

Moreover, the convergence is uniform in t .

Before proving the theorem, we provide some intuition. Every time a new time point is sampled, the likelihood function is multiplied by a new term of the form $\exp(-|f(\omega, t_{n+1}) - g(t_{n+1})|^r)$. If ω does not correspond to the true dynamics, there must be a region of $[0, T]$ where $f(\omega, t)$ differs from $g(t)$. Since the $\{t_n\}$ are chosen independently at random from an absolutely continuous probability distribution, eventually multiple time points will be sampled in this region, causing the value of the likelihood at ω go to 0. Therefore, in the limit when $n \rightarrow \infty$, the distribution $p_n(\omega)$ will concentrate more and more on the set of ω which corresponds to the true dynamics. Hence the expected dynamics will also converge to the system's true dynamics.

We use the following two lemmas, whose proofs will be provided in Section 2.8. The first is a result on the convergence of Monte Carlo integration. The second is a result on the convergence of the EDE.

Lemma 2.5.1 *Let points t_i be chosen as in Theorem 2.5.1, and let $1 \leq r < \infty$. Define*

$$h_n(\omega) = \exp \left(- \frac{1}{n} \sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \right)$$

and

$$h(\omega) = \exp \left(- \int_0^T |f(\omega, t) - g(t)|^r d\mu(t) \right).$$

Then

$$\frac{h_n(\omega)}{h(\omega)} \rightarrow 1 \quad \text{uniformly in } \omega \in \Omega$$

and

$$\lim_{n \rightarrow \infty} \|h_n\|_n = \|h\|_\infty.$$

Lemma 2.5.2 *Let a and b be continuous functions on $\Omega \times [0, T]$ and $[0, T]$, respectively, and let $\{p_n\}$ be a sequence of probability distributions on Ω .*

a) *Define*

$$h(\omega) = \exp \left(- \int_0^T |a(\omega, t) - b(t)|^r d\mu(t) \right),$$

and suppose that

(i) *for any $\alpha < 1$, there exists $\delta < 1$ and $C > 0$ such that if $\omega \in \Omega$ with $h(\omega) \leq \alpha \|h\|_\infty$, then $p_n(\omega) < C\delta^n \forall n$;*

(ii) *there exists $\omega_0 \in \Omega$ such that $a(\omega_0, t) = b(t)$ for all $t \in [0, T]$.*

Then

$$\lim_{n \rightarrow \infty} E_{p_n} [a(\omega, t)] = b(t) \quad \forall t \in [0, T]$$

and

$$\lim_{n \rightarrow \infty} \text{Var}_{p_n} [a(\omega, t)] = 0 \quad \forall t \in [0, T].$$

Moreover, for both limits, the convergence is uniform in t .

b) *Assume that Ω is finite and that there exists a set $S \subset [0, T]$ such that*

$$\{\omega \in \Omega : p_n(\omega) \not\rightarrow 0\} \subset \{\omega \in \Omega : a(\omega, t) = b(t) \forall t \in S\}.$$

Then for all t in S ,

$$\lim_{n \rightarrow \infty} E_{p_n} [a(\omega, t)] = b(t)$$

and

$$\lim_{n \rightarrow \infty} \text{Var}_{p_n} [a(\omega, t)] = 0.$$

Moreover, for both limits, the convergence is uniform in t .

Proof [Proof of Theorem 2.5.1] Let

$$q_n(\omega) = \exp \left(- \sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \right).$$

Then $p_n = c_n q_n$, and $q_n = (h_n)^n$, where h_n is defined as in Lemma 2.5.1, so

$$p_n(\omega) = \frac{q_n(\omega)}{\int_{\Omega} q_n(\omega) d\omega} = \frac{h_n^n(\omega)}{\int_{\Omega} h_n^n(\omega) d\omega} = \left(\frac{h_n(\omega)}{\|h_n\|_n} \right)^n.$$

Let $0 < \alpha < 1$, and suppose $\omega \in \Omega$ with $h(\omega) \leq \alpha \|h\|_{\infty}$. By Lemma 2.5.1 we have $\lim_{n \rightarrow \infty} h_n(\omega) = h(\omega)$ and $\lim_{n \rightarrow \infty} \|h_n\|_n = \|h\|_{\infty}$. Let $\epsilon > 0$ and $\delta = \alpha(1 + \epsilon)^2$. For ϵ small, we have $\delta < 1$, and for this ϵ there exists N (independent of ω) large enough such that if $n > N$, then

$$h_n(\omega) \leq (1 + \epsilon)h(\omega) \leq \alpha(1 + \epsilon)\|h\|_{\infty} \leq \alpha(1 + \epsilon)^2 \|h_n\|_n.$$

Hence for all $n > N$,

$$p_n(\omega) = \left(\frac{h_n(\omega)}{\|h_n\|_n} \right)^n \leq \delta^n,$$

with $\delta < 1$. Since there exists $\omega_0 \in \Omega$ such that $f(\omega_0, t) = g(t)$ for all $t \in [0, T]$, we can apply Lemma 2.5.2 (a) with $a = f$ and $b = g$ to obtain the uniform convergence

$$\lim_{n \rightarrow \infty} \int_{\Omega} p_n(\omega) f(\omega, t) d\omega = g(t) \quad \text{for all } t \in [0, T].$$

The integral on the left is $E_{p_n}[f(\omega, t)]$, so this gives the desired equality. ■

Note that the proof depends on the sequence $\{t_i\}$ only through Lemma 2.5.1, so the result holds for any sequence that yields the conclusion in that lemma. A quasi-random sequence satisfying a low-discrepancy condition [10] is one such sequence, so we make the following remark.

Remark 2.5.1 *The conclusion of Theorem 2.5.1 is still valid if $\{t_i\}$ is a low-discrepancy sequence, i.e.*

$$D_N(\{t_1, \dots, t_N\}) := \sup_{B \subset \Omega} \left| \frac{\#\{1 \leq i \leq N : t_i \in B\}}{N} - \text{Vol}(B) \right| \rightarrow 0$$

when N approaches infinity.

The results in this section imply that if data is collected uniformly at random, we can recover the true dynamics from the sampled data. This is one example of a so-called space-filling design [3]. However, in practice, randomly chosen points do not produce an efficient experimental design, since many of the measurements will not give much information about the system; the convergence, although guaranteed, may be slow.

2.5.2 Design Points Using the Maximally Informative Next Experiment

Intuitively, we expect the sequential designs of [4] and [5], for which the next sampled time point is the one that has the highest current uncertainty (variance) to increase the convergence rate relative to randomly selected design points. On the other hand, the measured points may no longer be dense in $[0, T]$, so it's not clear that the dynamics may be recovered on the entire interval.

In the following theorem, we extend the consistency result in the previous subsection to this type of sequential design, with the additional assumption that Ω is finite. This assumption was also used in the context of parameter identification in [13] and [14]. We conclude that we can recover the entire true dynamics, even if all the measurements are made in a small subset of $[0, T]$ (in the extreme case, at one point). As in the previous subsection, we still assume that data are subject to no error.

Theorem 2.5.2 *Let ω_0 , r , p_n be as in Theorem 2.5.1 and assume that Ω has finite cardinality. Suppose that each t_{n+1} is chosen so that*

$$\text{Var}_{p_n(\omega)} [f(\omega, t)] \leq \text{Var}_{p_n(\omega)} [f(\omega, t_{n+1})] \quad \forall t \in [0, T]. \quad (2.1)$$

Then

$$\lim_{n \rightarrow \infty} E_{p_n} [f(\omega, t)] = g(t) \quad \forall t \in [0, T].$$

That is, the EDE converges to the true dynamics of the system. Moreover, the convergence is uniform in t .

By choosing the next time point to be the point with highest variance, we put a constraint on the variance of the whole dynamics: variance at other points must be smaller than variance at the measured points, which in turn converges to 0. In this case we deduce that the expected dynamics on the whole interval converges to some limit dynamics. If we can prove further that a “true” parameter vector ω_0 is still in the support of the limit distribution, then obviously this limit dynamics is equal to the true system dynamics.

As above, this is straightforward when the t_i are chosen at random from an absolutely continuous distribution. However, the case when Ω is an open set and t_i are chosen according to (2.1) is a bit different. In a continuous framework, a parameter vector has measure zero and good performance of the true parameter vector does not guarantee that it will stay in the support of the limit distribution. Such a situation can happen in the case when the model is not robust around the true parameter and at the chosen time points, the neighbourhood around true parameters in the parameter space fit the data worse than some other regions. This may cause the expected dynamics to converge to incorrect dynamics. Though this situation is perhaps unlikely to happen in practice, we cannot exclude such a possibility for a convergence result.

To resolve this issue, we assume in Theorem 2.5.2 that Ω is a finite set. This may be achieved, for example, by subdividing each coordinate axis using a fixed step size and taking the set of points in Ω that lie on the resulting grid. An alternative approach in which the outputs and the set of possible measured time points are discretized instead of Ω is also suggested in Theorem 2.5.4. Both assumptions are natural and do not hinder the applicability of the method in practice.

Proof [Proof of Theorem 2.5.2] As in the proof of Theorem 2.5.1, let

$$q_n(\omega) = \exp \left(- \sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \right),$$

and recall that $p_n = c_n q_n$. Also, let A be the set of cluster points of $\{t_n\}$: points $t \in [0, T]$ such that there exists a subsequence $\{t_{n_k}\}$ of with $t_{n_k} \rightarrow t$.

We claim first that if $p_n(w)$ does not tend to 0 with n (so that ω has probability above some fixed $\rho > 0$ for infinitely many n), then $f(\omega, t) = g(t)$ for all $t \in A$. Indeed, consider any $\omega \in \Omega$, $t \in A$ such that $|f(\omega, t) - g(t)| = c > 0$. Since A is the set of limit points of $\{t_n\}$, there exists a subsequence $\{t_{n_k}\}$ of $\{t_n\}$ such that $t_{n_k} \rightarrow t$. Since f and g are continuous, for k large enough, we have $|f(\omega, t_{n_k}) - g(t_{n_k})| \geq c/2$. Hence

$$\sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \rightarrow \infty$$

when $n \rightarrow \infty$, and so $q_n(\omega) \rightarrow 0$.

On the other hand, the assumption that some ω_0 gives the true dynamics implies that $f(\omega_0, t) - g(t) = 0$ for all t , hence $q_n(\omega_0) = 1$. Therefore, $p_n(\omega_0)/p_n(\omega) \rightarrow \infty$. Since Ω is a finite space, $p_n(\omega_0) \leq 1$, and hence $p_n(\omega) \rightarrow 0$. Hence $p_n(w) \not\rightarrow 0$ implies $f(\omega, t) = g(t)$ for all $t \in A$.

Using Lemma 2.5.2 (b) (for finite Ω) with $a = f$ and $b = g$, we deduce that

$$\lim_{n \rightarrow \infty} E_{p_n}[f(\omega, t)] = g(t) \quad \forall t \in A$$

and

$$\text{Var}_{p_n(\omega)}[f(\omega, t)] \rightarrow 0 \quad \forall t \in A.$$

On the other hand, the choice of t_{n+1} gives

$$\text{Var}_{p_n(\omega)}[f(\omega, t)] \leq \text{Var}_{p_n(\omega)}[f(\omega, t_{n+1})] \quad \forall t \in [0, T]. \quad (2.2)$$

Now we claim that

$$\text{Var}_{p_n(\omega)}[f(\omega, t_{n+1})] \rightarrow 0. \quad (2.3)$$

Indeed, by contradiction, assume that there exists a subsequence $\{t_{n_k}\}$ and a positive constant C such that

$$\text{Var}_{p_{n_k}(\omega)}[f(\omega, t_{n_k+1})] \geq C$$

for all k . Since $[0, T]$ is compact, we can drop to a subsequence to assume that t_{n_k+1} converges to some $t_0 \in A$. By the continuity of f and its derivatives on the compact set $\Omega \times [0, T]$, there is $C_0 > 0$ so that for all $k > 0$ and $\omega \in \Omega$,

$$|f(\omega, t_{n_k}) - f(\omega, t_0)| \leq C_0 |t_{n_k} - t_0|. \quad (2.4)$$

Hence by using this inequality, we have

$$\limsup_{k \rightarrow \infty} E_{p_{n_k}(\omega)} |f(\omega, t_{n_k+1}) - f(\omega, t_0)| \leq \lim_{k \rightarrow \infty} C_0 |t_{n_k+1} - t_0| = 0$$

which implies that

$$\lim_{k \rightarrow \infty} E_{p_{n_k}(\omega)} [f(\omega, t_{n_k+1})] = \lim_{k \rightarrow \infty} E_{p_{n_k}(\omega)} [f(\omega, t_0)].$$

By a similar argument, we also have

$$\lim_{k \rightarrow \infty} E_{p_{n_k}(\omega)} [f^2(\omega, t_{n_k+1})] = \lim_{k \rightarrow \infty} E_{p_{n_k}(\omega)} [f^2(\omega, t_0)].$$

Therefore

$$\lim_{k \rightarrow \infty} \text{Var}_{p_{n_k}(\omega)} [f(\omega, t_{n_k+1})] = \lim_{k \rightarrow \infty} \text{Var}_{p_{n_k}(\omega)} [f(\omega, t_0)] = 0,$$

which contradicts the choice of C .

From (2.2) and (2.3) we obtain

$$\text{Var}_{p_n(\omega)} [f(\omega, t)] \leq \text{Var}_{p_n(\omega)} [f(\omega, t_{n+1})] \rightarrow 0 \quad \forall t \in [0, T].$$

In other words, for all t in $[0, T]$,

$$\lim_{n \rightarrow \infty} \sum_{\omega \in \Omega} p_n(\omega) (f(\omega, t) - E_{p_n} [f(\omega, t)])^2 = 0. \quad (2.5)$$

The fact that ω_0 gives the true dynamics implies that ω_0 is a maximum for q_n , hence for p_n . Hence $p_n(\omega_0) \geq p_n(\omega)$ for all $\omega \in \Omega$, and from the fact that Ω is finite, we deduce that $p_n(\omega_0) \geq 1/|\Omega|$. Using this with (2.5) gives

$$\begin{aligned} (f(\omega_0, t) - E_{p_n} [f(\omega, t)])^2 &\leq |\Omega| p_n(\omega_0) (f(\omega_0, t) - E_{p_n} [f(\omega, t)])^2 \\ &\leq |\Omega| \sum_{\omega \in \Omega} p_n(\omega) (f(\omega, t) - E_{p_n} [f(\omega, t)])^2 \\ &\rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. Hence

$$E_{p_n(\omega)} [f(\omega, t)] \rightarrow f(\omega_0, t) = g(t) \quad \forall t \in [0, T].$$

■

As in the discussion before the proof, the only reason we use the variance criterion is to bound the variance of dynamics at unmeasured points by the variance at measured points. This criterion can be relaxed as follows.

Theorem 2.5.3 *The result from Theorem 2.5.2 is still valid if Condition (2.1) (that the next time point has maximum variance) is replaced by the condition that the variance at the next time point is within a fixed constant of the maximum variance. That is, there exists $C > 1$ so that for all $t \in [0, T]$,*

$$\text{Var}_{p_n(\omega)} [f(\omega, t)] \leq C \text{Var}_{p_n(\omega)} [f(\omega, t_{n+1})]. \quad (2.6)$$

There are several motivations for this relaxation of criterion (2.1). First, in practice, the optimization of the variance function (which is usually done by Markov Chain Monte Carlo methods) is subject to random effects arising in the sampling process. By using criterion (2.6), we look for a near-optimal solution of the optimization problem; this condition is stable with respect to a MCMC scheme. Second, in real experiments, some sets of measurements may be more expensive and technically difficult than the others. By looking for a near-optimal solution, we make it possible for experimenters to find an alternative measurement when the optimization problem gives rise to an optimum that is experimentally difficult to implement. Finally, as we will see in Section 2.6, the new criterion allows us to apply additional criteria for point selection in order to facilitate resampling, which will be used to establish convergence rate of the EDE in the face of noisy data.

As noted above, instead of discretizing Ω , we may discretize the output space and the measurement space. This gives the following result.

Theorem 2.5.4 *The result from Theorem 2.5.2 is still valid if we assume that Ω is open and bounded, but that the possible outputs of the system and the set of possible measured time points are both finite. In this case, we get convergence of the EDE on the full (finite) set of possible measured time points.*

Proof We denote by $\mathcal{T} = \{\tau_i\}_{i=1}^K$ the set of all possible measured time points and assume that at each time point, the output function $f(\omega, t)$, as well as the true dynamics $g(t)$, are discretized by a finite grid. That is, the continuous function $g(t)$ is approximated by values $(Rg(\tau_1), \dots, Rg(\tau_K))$, where $Rg(t)$ is obtained from $g(t)$ by rounding to the nearest of some finite set of allowable output values.

By making arbitrarily small perturbations to the possible output values if necessary, we can assume without loss of generality that $\forall \tau \in \mathcal{T}$, the true output value $g(\tau)$ does not lie midway between two allowable output values. Therefore, there exists an open neighbourhood U_{ω_0} of ω_0 such that if $\omega \in U_{\omega_0}$, then $Rf(\omega, \tau) = Rf(\omega_0, \tau) = Rg(\tau) \quad \forall \tau \in \mathcal{T}$. For the remainder of the proof we use f and g to mean Rf and Rg .

As in the proof of Theorem 2.5.2, we now consider any $\omega \in \Omega$ such that $p_n(\omega)$ does not tend to 0 with n . Assume that $f(\omega, t) \neq f(\omega_0, t)$ for some t in the cluster set, A , of $\{t_i\}$ (since \mathcal{T} is finite here, this is the set of points that are measured infinitely many times). Using the same argument as in the proof of Theorem 2.5.2, we deduce that $p_n(\omega) \leq p_n(\omega_0)$ and $p_n(\omega_0)/p_n(\omega) \rightarrow \infty$. Note that in this case, although Ω is not finite, the argument is still valid since p_n is constant on the open set U_{ω_0} , so that $p_n(\omega_0)$ is bounded above by $1/\text{Vol}(U_{\omega_0})$, where $\text{Vol}(U)$ denotes the volume of a set U .

Therefore, $p_n(\omega) \not\rightarrow 0$ implies $f(\omega, t) = g(t)$ for all $t \in A$. Since Ω may be infinite, Lemma 2.5.2 cannot be applied directly in this case. However, by denoting $U_A = \{\omega \in \Omega : f(\omega, t) = g(t) \quad \forall t \in A\}$, we have for all $t \in A$

$$\left| \int_{\Omega} p_n(\omega) f(\omega, t) d\omega - g(t) \right| \leq \int_{\Omega \setminus U_A} p_n(\omega) |f(\omega, t) - g(t)| d\omega.$$

Since f and g are bounded and $p_n(\omega_0) < 1/\text{Volume}(U_{\omega_0})$, we have $|p_n(\omega)| |f(\omega, t) - g(t)| \leq Cp_n(\omega_0) \leq C/\text{Vol}(U_{\omega_0})$. Also, $p_n(\omega) \rightarrow 0$ on $\Omega \setminus U_A$, so by the Dominated Convergence Theorem, the right hand side converges to 0 as n tends to ∞ . We deduce that $E_{p_n(\omega)}[f(\omega, t)] \rightarrow g(t) \quad \forall t \in A$. By a similar argument, we also have $\text{Var}_{p_n(\omega)}[f(\omega, t)] \rightarrow 0 \quad \forall t \in A$.

We next use the fact that the set $\mathcal{T} = \{\tau_i\}_{i=1}^K$ of possible measured time points is finite to deduce that $\text{Var}_{p_n(\omega)}[f(\omega, t_{n+1})] \rightarrow 0$. Indeed, assume that

$$\text{Var}_{p_{n_k}(\omega)}[f(\omega, t_{n_k+1})] \geq C$$

for some subsequence $\{n_k\}$ and positive constant C . Since \mathcal{T} is finite, there exists $t_0 \in A$ that appears in the subsequence $\{t_{n_k}\}$ infinitely many times; this implies that $\text{Var}_{p_{n_k}(\omega)}[f(\omega, t_0)] \not\rightarrow 0$, which is a contradiction.

Hence, $\text{Var}_{p_n(\omega)}[f(\omega, t_{n+1})] \rightarrow 0$. Combining this with (2.1), we see that in fact $\text{Var}_{p_n(\omega)}[f(\omega, t)] \rightarrow 0$ for all $t \in \mathcal{T}$. This proves that the EDE converges to the true system dynamics on \mathcal{T} . ■

Note that the condition of discrete measured time points in the previous result allows us to avoid the need for the regularity condition (2.4), which does not hold for the piecewise constant functions obtained by discretizing the system outputs. In the next section we provide further justification for a finite set of measurement points.

2.6 EDE Consistency with Noisy Data

In practice, of course, data from experiments are subject to noise. Hence in this section we extend the results from previous sections to the case of additive Gaussian noise. As is common in many settings, we assume that

$$d(t_i) = g(t_i) + \epsilon_i$$

where $g(t)$ is the true dynamics (which is unknown), $d(t_i)$ is the measured data at the sampled time point t_i , and ϵ_i are i.i.d. Gaussian random variables (see [5] for empirical support of this noise model).

The analysis in the case of noisy data is a bit different from that used in the previous section. Intuitively, if "close", but not exactly the same points in time, t_1 and t_2 , are measured, and if there is a functional relation between the output at t_1 and at t_2 , then the measurement at t_1 will also help refine the information about

data at t_2 . However, theoretically, this assertion is difficult to prove and may even be incorrect, due to nonlinearity: if the relation between output at t_1 and t_2 are nonlinear, using data at t_2 to constrain t_1 may create a bias in the fitted output.

For example, if $f_2 = f_1^2$ then

$$(f_1 + e)^2 = f_1^2 + 2ef_1 + e^2 = f_2 + ef_1 + e^2$$

When using averaging, the linear error term will go away by the strong law of large number, but the quadratic term will have positive expectation, which results in a bias in estimation of f_2 . The stronger the non-linearity is, the larger the bias and that makes it hard clarify the convergence.

In order to obtain a convergence result using noisy data, we need to be able to average over multiple trials, which makes sense only if we measure repeatedly at a given time. In the theorem below, as in Theorem 2.5.4, we discretize the time interval and allow measurements to be taken at only finitely many specified points and use a slightly different form of probability distribution. This guarantees that experiments will be replicated many times at “important” points. When data are collected multiple times, the average value is used to constrain the dynamics: the larger the number of times we make the measurement at a time point t , the more confidence we put on the average data at that point. With this framework, we again have the convergence of the EDE to the true system dynamics.

The idea of using a finite grid to replace the whole time interval to facilitate resampling is a common technique in the problem of parameter identification ([13], [14]). In studies of ODEs, under the assumption that f is analytic, this is further supported by the following theorem from [17], which guarantees that if we can identify the system dynamics on a finite grid, we can identify the dynamics on the whole interval.

Theorem 2.6.1 (*Sontag [17]*)

Assume $f(w, t)$ depends analytically on w and t , and let N be the dimension of the parameter space. Then, for Lebesgue almost every randomly chosen set of $2N + 1$

experiments, the following property holds: For any two parameters that have distinct dynamics, one of the experiments in this set will distinguish them.

We further note that the assumption of analyticity may be replaced by an assumption that Ω is a finite set. That is, with this assumption we can find a finite grid $\mathcal{T} \subset [0, T]$ that satisfies the above property: For any two parameters that have distinct dynamics on $[0, T]$, one of the experiments $t \in \mathcal{T}$ in this grid will distinguish them.

Finally, even in the case when the parameter space Ω is an open set, by choosing the discretized time points to be the nodes for an efficient interpolation scheme, then by interpolating on this finite set, convergence on the finite set of times \mathcal{T} converts to uniform approximation on the entire interval $[0, T]$.

Hence throughout this section, we discretize $[0, T]$ to a finite grid $\mathcal{T} = \{\tau_i\}_{i=1}^K$, and assume that the experiments can be made only at the nodes of this grid. We also continue to assume that Ω has finite cardinality. With these assumptions, we have the following theorem.

Theorem 2.6.2 *Let $C > 1$. Assume that Ω is finite and at step n , $t_{n+1} \in \mathcal{T}$ is chosen so that*

$$\text{Var}_{p_n(\omega)} [f(\omega, t)] \leq C \text{Var}_{p_n(\omega)} [f(\omega, t_{n+1})] \quad \forall t \in \mathcal{T}.$$

For $1 \leq k \leq K$, let $k_n(\tau_i)$ be the number of experiments made at time τ_i up through step n and $\{d_j(\tau_i)\}_{j=1}^{k_n(\tau_i)}$ be the data values from those experiments, with $d_j(\tau_i) = g(\tau_i) + \epsilon$; the ϵ are iid $N(0, \sigma^2)$. Define $B_n = \{\tau_i : k_n(\tau_i) > 0\}$ and

$$p_n(\omega) = c_n \exp \left(- \sum_{\tau_i \in B_n} k_n(\tau_i) \left(f(\omega, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right)^r \right),$$

where c_n is the normalizing constant and $r > 2$. Then

$$\lim_{n \rightarrow \infty} E_{p_n} [f(\omega, t)] = g(t) \quad \forall t \in \mathcal{T}.$$

Moreover, the convergence is uniform in $t \in \mathcal{T}$.

The same result with $r = 2$ is also valid if the following condition is satisfied

$$\lim_{n \rightarrow \infty} \frac{\log \log k_n(\tau_1)}{k_n(\tau_2)} = 0 \quad \forall \tau_1, \tau_2 \in A \quad (2.7)$$

where A is the set of all cluster points of $\{t_n\}$.

Note that if \mathcal{T} satisfies the conditions in the discussion following Theorem 2.6.1, then determining the dynamics in \mathcal{T} is sufficient to determine the dynamics on all of $[0, T]$.

Proof Let $A = \{\tau_i : \lim_{n \rightarrow \infty} k_n(\tau_i) = \infty\}$ and

$$q_n(\omega) = \exp \left(- \sum_{\tau_i \in B_n} k_n(\tau_i) \left(f(\omega, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right)^r \right).$$

We claim that

$$\{\omega : q_n(\omega) \not\rightarrow 0\} \subset \{\omega : f(\tau_i, \omega) = g(\tau_i), \forall \tau_i \in A\}.$$

Indeed, consider any $\omega \in \Omega$, $\tau_{i_0} \in A$ such that $|f(\omega, \tau_{i_0}) - g(\tau_{i_0})| = c > 0$. Let $X_j = d_j(\tau_{i_0}) = g(\tau_{i_0}) + \epsilon$ and note that $\{X_j\}$ is a Gaussian sequence of iid random variables with $E[X_j] = g(\tau_{i_0}) = f(\omega, \tau_{i_0})$. By the law of large numbers, we have with probability 1

$$g(\tau_{i_0}) = \lim_{n \rightarrow \infty} \frac{1}{k_n(\tau_{i_0})} \sum_{j=1}^{k_n(\tau_{i_0})} d_j(\tau_{i_0})$$

Hence there exists N such that for all $n > N$

$$\left| g(\tau_{i_0}) - \lim_{n \rightarrow \infty} \frac{1}{k_n(\tau_{i_0})} \sum_{j=1}^{k_n(\tau_{i_0})} d_j(\tau_{i_0}) \right| \leq c/2$$

which implies (by triangle inequality)

$$\left| f(\tau_{i_0}, \omega) - \lim_{n \rightarrow \infty} \frac{1}{k_n(\tau_{i_0})} \sum_{j=1}^{k_n(\tau_{i_0})} d_j(\tau_{i_0}) \right| \geq c/2$$

Therefore

$$\sum_{\tau_i \in B_n} k_n(\tau_i) \left(f(\omega, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right)^r \geq (c/2)^r k_n(\tau_{i_0}) \rightarrow \infty \quad (2.8)$$

as $n \rightarrow \infty$.

Now consider any $\tau_i \in A$. By the law of the iterated logarithm, we have with probability 1

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{k_n(\tau_i)}{\log \log k_n(\tau_i)}} \left(f(\omega_0, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right) = \sqrt{2}$$

If $r > 2$, there exists a constant C such that

$$k_n(\tau_i) \left(f(\omega_0, \tau_i) - \frac{1}{k} \sum_{j=1}^k d_j(\tau_i) \right)^r \leq C \frac{\log \log k_n(\tau_i)}{(k_n(\tau_i))^{(r/2-1)}} \rightarrow 0 \quad (2.9)$$

as $n \rightarrow \infty$.

Since this is true for any τ_i in the finite set A , we have

$$\lim_{n \rightarrow \infty} \sum_{\tau_i \in B_n} k_n(\tau_i) \left(f(\omega_0, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right)^r < \infty.$$

Therefore, $q_n(\omega_0)$ is bounded below, so $\frac{p_n(\omega_0)}{p_n(\omega)} = \frac{q_n(\omega_0)}{q_n(\omega)} \rightarrow \infty$. Since Ω is a finite space, $p_n(\omega_0) \leq 1$. This makes $p_n(\omega) \rightarrow 0$.

In the case when $r = 2$, we have

$$k_n(\tau_i) \left(f(\omega_0, \tau_i) - \frac{1}{k} \sum_{j=1}^k d_j(\tau_i) \right)^2 \leq C \log \log k_n(\tau_i)$$

which implies

$$\limsup_{n \rightarrow \infty} \sum_{\tau_i \in B_n} k_n(\tau_i) \left(f(\omega_0, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right)^2 \leq C_1 + \sum_{\tau_i \in A} C \log \log k_n(\tau_i).$$

Equation (2.7) implies that there exists N such that for all $n \geq N$

$$\log \log k_n(\tau_i) \leq \frac{(c/2)^2}{2C(\#A)} k_n(\tau_{i_0}) \quad \forall \tau_i \in A$$

where $\#A$ denotes the cardinality of A , and c is the constant in (2.8).

Combine this inequality and (2.8) (with $r = 2$), we have

$$\frac{p_n(\omega)}{p_n(\omega_0)} = \frac{q_n(\omega)}{q_n(\omega_0)} \leq \exp \left(C_1 - \frac{c^2}{8} k_n(\tau_{i_0}) \right) \rightarrow 0$$

as $n \rightarrow \infty$. Hence $p_n(\omega) \rightarrow 0$.

At this point, we have proved that

$$\{\omega : p_n(\omega) \not\rightarrow 0\} \subset \{\omega : f(\omega, \tau_i) = g(\tau_i), \forall \tau_i \in A\}.$$

Hence by Lemma 2.5.2 (b)

$$\lim_{n \rightarrow \infty} E_{p_n}[f(\omega, \tau_i)] = g(\tau_i) \quad \forall \tau_i \in A$$

and

$$\lim_{n \rightarrow \infty} \text{Var}_{p_n}[f(\omega, \tau_i)] = 0 \quad \forall \tau_i \in A.$$

On the other hand, we have

$$\text{Var}_{p_n(\omega)}[f(\omega, t)] \leq C \text{Var}_{p_n(\omega)}[f(\omega, t_{n+1})] \quad \forall t \in \mathcal{T}.$$

Using the same argument as in the proof of Theorem 2.5.2, we have

$$\text{Var}_{p_n(\omega)}[f(\omega, t)] \rightarrow 0 \quad \forall t \in \mathcal{T}$$

and

$$E_{p_n(\omega)}[f(\omega, t)] \rightarrow f(\omega_0, t) = g(t) \quad \forall t \in \mathcal{T}.$$

■

By an argument similar to that used in the proof of Theorem 2.5.4, we obtain the following result.

Theorem 2.6.3 *The result of Theorem 2.6.2 is still valid if we replace the condition of finite cardinality of Ω with the condition of a finite set of output values and possible measurement time points.*

2.7 EDE consistency with model mismatch

So far we have investigated various schemes to design experiments for dynamics identification, under the assumption that the investigated model is a correct model, i.e. there exists $\omega_0 \in \Omega$ such that $f(\omega_0, t) = g(t)$ for all $t \in [0, T]$. Here we relax this condition using the concept of ϵ -equivalence.

Definition 2.7.1 Let $\epsilon > 0$ and suppose g and h are continuous on $[0, T]$. Then g and h are ϵ -equivalent means that $\|g - h\|_\infty < \epsilon$.

To obtain the main result in this section, we also need to assume that the function outputs are discretized by a finite grid of resolution ϵ , similar to the discretization in Theorem 2.5.4. However, here we use an adaptive discretization in that it changes based on the measurements obtained so far and based on the time point.

Definition 2.7.2 Let h be continuous on $[0, T]$, let \mathcal{T} and $d_j(\tau_i)$ be as in Theorem 2.6.2, and let $\epsilon > 0$. Define

$$d_n^*(\tau_i) = \begin{cases} \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) & \text{if } k_n(\tau_i) > 0 \\ 0 & \text{if } k_n(\tau_i) = 0 \end{cases}$$

and

$$R_n^\epsilon h(\tau_i) = d_n^*(\tau_i) + \text{sgn}(h(\tau_i) - d_n^*(\tau_i)) \left\lfloor \frac{|h(\tau_i) - d_n^*(\tau_i)|}{\epsilon} \right\rfloor \epsilon.$$

This choice of discretization is needed to guarantee convergence of the estimated dynamics. With this setting, we have the following theorem, in which we use the framework of Theorem 2.6.2 but with the output discretization just given. The proof of this theorem is a combination of the techniques employed in Theorem 2.6.2 and Theorem 2.5.4. Here $\lfloor x \rfloor$ denotes the largest integer less than or equal to x .

Theorem 2.7.1 Let Ω , C , \mathcal{T} , B_n , k_n , $d_j(\tau_i)$ be as in Theorem 2.6.2, $\epsilon_0 > 0$ and assume that there is $\omega_0 \in \Omega$ such that $f(\omega_0, t)$ and $g(t)$ are ϵ_0 -equivalent. For $\epsilon > \epsilon_0$ define

$$p_n(\omega) = c_n \exp \left(- \sum_{\tau_i \in B_n} k_n(\tau_i) \left(R_n^\epsilon f(\omega, \tau_i) - \frac{1}{k_n(\tau_i)} \sum_{j=1}^{k_n(\tau_i)} d_j(\tau_i) \right)^2 \right),$$

where c_n is the normalizing constant, and assume that at each step, the next measurement is chosen so that

$$\text{Var}_{p_n(\omega)} [R_n^\epsilon f(\omega, t)] \leq C \text{Var}_{p_n(\omega)} [R_n^\epsilon f(\omega, t_{n+1})] \quad \forall t \in \mathcal{T}.$$

Then, for almost every $\epsilon > \epsilon_0$, the expected dynamics converges (uniformly in $t \in \mathcal{T}$) to limit dynamics that are ϵ -equivalent to $g(t)$.

Proof Denote by A the set of all $t \in \mathcal{T}$ that are measured infinitely many times. By the strong law of large numbers, we have $d_n^*(\tau) \rightarrow g(\tau)$ for all $\tau \in A$. Since Ω and \mathcal{T} are finite, there is a full measure set of $\epsilon > \epsilon_0$ such that for all $\tau \in \mathcal{T}$ and $\omega \in \Omega$, the distance between $g(t)$ and $f(\omega, \tau)$ is not a multiple of ϵ . This implies that $\lim_{n \rightarrow \infty} |f(\omega, \tau) - d_n^*(\tau)|/\epsilon$ is not an integer for any $\omega \in \Omega$ and $\tau \in A$, hence that $\lim_{n \rightarrow \infty} R_n^\epsilon f(\omega, \tau)$ exists for all such ω and τ . Also, if $\tau \notin A$, then $R_n^\epsilon f(\omega, \tau)$ is constant for n large enough. Hence $\lim_{n \rightarrow \infty} R_n f(\omega, \tau)$ exists for all $\omega \in \Omega, \tau \in \mathcal{T}$.

On the other hand, the assumption on ω_0 implies that for all $\tau \in \mathcal{T}$

$$|d_n^*(\tau_i) - f(\omega_0, \tau)| \leq |d_n^*(\tau_i) - g(t)| + \epsilon_0.$$

For n sufficiently large, the right hand side is less than ϵ for all $\tau \in \mathcal{T}$. For such n we have $R_n^\epsilon f(\omega_0, \tau) = R_n^\epsilon g(\tau)$ for all τ in \mathcal{T} .

Now consider $\omega \in \Omega$ such that $\lim_{n \rightarrow \infty} R_n^\epsilon f(\omega, \tau) \neq \lim_{n \rightarrow \infty} R_n^\epsilon f(\omega_0, \tau)$ for some τ in A . Using the same argument as in the proof of Theorem 2.6.2, we deduce that $p_n(\omega_0)/p_n(\omega) \rightarrow \infty$. Since Ω is a finite space, $p_n(\omega_0) \leq 1$. This makes $p_n(\omega) \rightarrow 0$. We have proved that

$$\{\omega : p_n(\omega) \not\rightarrow 0\} \subset \{\omega : \lim_{n \rightarrow \infty} R_n f(\omega, \tau) = \lim_{n \rightarrow \infty} R_n g(\tau), \forall \tau \in A\}.$$

Then by Lemma 2.5.2 (b)

$$\lim_{n \rightarrow \infty} E_{p_n}[R_n^\epsilon f(\omega, \tau)] = \lim_{n \rightarrow \infty} E_{p_n}[R_n^\epsilon g(\tau)] = g(\tau) \quad \forall \tau \in A$$

and

$$\lim_{n \rightarrow \infty} \text{Var}_{p_n}[R_n^\epsilon f(\omega, \tau)] = 0 \quad \forall \tau \in A.$$

On the other hand, we have

$$\text{Var}_{p_n(\omega)}[R_n^\epsilon f(\omega, t)] \leq C \text{Var}_{p_n(\omega)}[R_n^\epsilon f(\omega, t_{n+1})] \quad \forall t \in \mathcal{T}.$$

Using the same argument as in the proof of Theorem 2.5.2, we have

$$\text{Var}_{p_n(\omega)}[R_n^\epsilon f(\omega, t)] \rightarrow 0 \quad \forall t \in \mathcal{T}$$

and

$$\lim_{n \rightarrow \infty} E_{p_n(\omega)} [R_n^\epsilon f(\omega, t)] = \lim_{n \rightarrow \infty} R_n g(t) \quad \forall t \in \mathcal{T}.$$

This proves that the EDE converges to limit dynamics that are ϵ -equivalent to the true system dynamics on \mathcal{T} . ■

2.8 Proofs of Supporting Lemmas

In this section, we provide the proofs of the two lemmas that have been used throughout this paper.

2.8.1 Lemma 2.5.1

(Convergence of Monte Carlo integration)

Proof [Proof]

First, we note that

$$\frac{h_n(\omega)}{h(\omega)} = \exp \left(\int_0^T |f(\omega, t) - g(t)|^r d\mu(t) - \frac{1}{n} \sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \right).$$

Using the Koksma-Hlawka inequality for convergence of quasi-Monte Carlo integration [12], we have

$$\begin{aligned} \left| \int_0^T |f(\omega, t) - g(t)|^r d\mu(t) - \frac{1}{n} \sum_{i=1}^n |f(\omega, t_i) - g(t_i)|^r \right| & \\ \leq r D_n^* \int_0^T |f(\omega, t) - g(t)|^{r-1} \left| \frac{\partial f}{\partial t}(\omega, t) - g'(t) \right| d\mu(t), & \end{aligned} \quad (2.10)$$

where D_n^* is the discrepancy of the finite sequence $\{t_1, t_2, \dots, t_n\}$ (see [10] for more information about the discrepancy).

Since f is a C^1 function on the compact set $\Omega \times [0, T]$ and g is C^1 on $[0, T]$, there exists M independent of ω and t such that

$$\int_0^T |f(\omega, t) - g(t)|^{r-1} \left| \frac{\partial f}{\partial t}(\omega, t) - g'(t) \right| d\mu(t) \leq M. \quad (2.11)$$

Since μ is absolutely continuous with respect to Lebesgue measure, we have $D_n^* \rightarrow 0$ as $n \rightarrow \infty$.

From (2.10) and (2.11) we have for any $\omega \in \Omega$ that

$$|\log(h_n(\omega)/h(\omega))| \leq rMD_n^* \rightarrow 0.$$

Hence $h_n(\omega)/h(\omega) \rightarrow 1$ uniformly in $\omega \in \Omega$.

Also, since $h_n(\omega) \leq 1$ for all ω and $h(\omega_0) = 1$, we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \|h_n\|_n &= \limsup_{n \rightarrow \infty} \left(\int_{\Omega} h_n^n d\omega \right)^{1/n} \\ &\leq \limsup_{n \rightarrow \infty} \text{Vol}(\Omega)^{1/n} \\ &= 1 = \|h\|_{\infty}. \end{aligned} \tag{2.12}$$

To get a lower bound, let $\epsilon > 0$. Note that if $h(\omega) \geq 1 - \epsilon/2$ and $|h_n(\omega) - h(\omega)| \leq \epsilon/2$, then by the triangle inequality, we have $h_n(\omega) \geq 1 - \epsilon$. Also, since h is continuous and $\|h\|_{\infty} = 1$, we have

$$C_{\epsilon} := \text{Vol}(\{\omega : h(\omega) \geq 1 - \epsilon/2\}) > 0.$$

Since h_n/h converges uniformly on Ω to 1 and $|h| \leq 1$, there exists $N(\delta, \epsilon)$ large enough such that for all $n \geq N$ and all $\omega \in \Omega$, $|h_n(\omega) - h(\omega)| \leq \epsilon/2$. Hence

$$\text{Vol}(\{\omega : h_n(\omega) \geq 1 - \epsilon\}) \geq C_{\epsilon}.$$

So

$$\int_{\Omega} h_n^n d\omega \geq \int_{\{h_n \geq 1-\epsilon\}} h_n^n d\omega \geq C_{\epsilon}(1 - \epsilon)^n$$

and

$$\left(\int_{\Omega} h_n^n dx \right)^{1/n} \geq C_{\epsilon}^{1/n}(1 - \epsilon).$$

Taking $n \rightarrow \infty$, we deduce

$$\|h\|_{\infty} \geq \limsup_{n \rightarrow \infty} \|h_n\|_n \geq \liminf_{n \rightarrow \infty} \|h_n\|_n \geq \|h\|_{\infty} - \epsilon.$$

Since ϵ was arbitrary, $\lim_{n \rightarrow \infty} \|h_n\|_n = \|h\|_{\infty}$. ■

2.8.2 Lemma 2.5.2

(Convergence of the Expected Dynamics Estimator)

Proof [Proof] We provide the proof for part (a). The proof for part (b) uses a similar argument.

Let $\epsilon > 0$ and define

$$U = \{\omega \in \Omega : |a(\omega, t) - b(t)| < \epsilon, \forall t \in [0, T]\}.$$

Since a, b are continuous and $a(\omega_0, t) = b(t)$ for all t , we see that U is a neighborhood of ω_0 . Then for $t \in [0, T]$, we have

$$\begin{aligned} \left| \int_{\Omega} p_n(\omega) a(\omega, t) d\omega - b(t) \right| &\leq \int_{\Omega} p_n(\omega) |a(\omega, t) - b(t)| d\omega \\ &= \int_{\Omega \setminus U} p_n(\omega) |a(\omega, t) - b(t)| d\omega + \int_U p_n(\omega) |a(\omega, t) - b(t)| d\omega \\ &\leq \int_{\Omega \setminus U} p_n(\omega) |a(\omega, t) - b(t)| d\omega + \epsilon \end{aligned}$$

Now we claim that there exists $\alpha < 1$ such that $\forall \omega \in \Omega \setminus U$, $h(\omega) \leq \alpha$. Indeed, assume that $\exists \omega_n \in \Omega \setminus U$ with $h(\omega_n) \rightarrow 1$. Then for each n there is $t_n \in [0, T]$ such that $|a(\omega_n, t_n) - b(t_n)| \geq \epsilon$. Since $\Omega \times [0, T]$ is compact, without loss of generality, we can assume that $\omega_n \rightarrow \omega^* \in \Omega$, $t_n \rightarrow t^* \in [0, T]$. Since a and b are continuous, we deduce that $|a(\omega^*, t^*) - b(t^*)| \geq \epsilon$ and $h(\omega^*) = 1$.

However, $h(\omega^*) = 1$ implies that $\int_0^T |a(\omega^*, t) - b(t)|^r d\mu(t) = 0$. Since μ is absolutely continuous and a and b are continuous, this implies that $a(\omega^*, t) = b(t)$ for all $t \in [0, T]$, which contradicts $|a(\omega^*, t^*) - b(t^*)| \geq \epsilon$.

Therefore, there exists $\alpha < 1$ such that $\forall \omega \in \Omega \setminus U$, $h(\omega) \leq \alpha$. Hence, by using hypothesis (i), we have

$$\int_{\Omega \setminus U} p_n(\omega) |a(\omega, t) - b(t)| d\omega \leq \text{Vol}(\Omega) \delta^n \sup_{(\omega, t)} |a(\omega, t) - b(t)|$$

and hence

$$\left| \int_{\Omega} p_n(\omega) a(\omega, t) d\omega - b(t) \right| \leq \epsilon + C_1 \delta^n,$$

where C_1 is a constant that does not depend on t and ω . Since ϵ is arbitrary, we deduce that

$$\lim_{n \rightarrow \infty} E_{p_n}[a(\omega, t)] = \lim_{n \rightarrow \infty} \int_{\Omega} p_n(\omega) a(\omega, t) d\omega = b(t)$$

uniformly in $t \in [0, T]$. Note that this argument actually shows the somewhat stronger statement that $E_{p_n}[|a(\omega, t) - b(t)|] \rightarrow 0$ uniformly in t . The same argument shows that $E_{p_n}[|a(\omega, t) - b(t)|^2] \rightarrow 0$ uniformly in $t \in [0, T]$. Hence taking $\bar{a}(t) = E_{p_n}[a(\omega, t)]$, we have

$$\text{Var}_{p_n}[a(\omega, t)] = E_{p_n}[|a(\omega, t) - b(t)|^2] - |b - \bar{a}(t)|^2$$

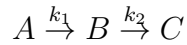
which converges to 0 uniformly in $t \in [0, T]$. ■

2.9 Numerical examples

In this section we provide numerical examples to illustrate our theoretical findings and demonstrate the efficacy of our variations on the MINE method to design experiments for dynamics identification.

2.9.1 A simple ODE model

We consider a simple biochemical system that contains 3 chemicals:



where k_1 and k_2 are the (unknown) degradation rates of A and B , respectively. We also assume that at the beginning, the system only contains A .

We model this system using

$$\frac{dA}{dt} = -k_1 A, \quad \frac{dB}{dt} = k_1 A - k_2 B, \quad \frac{dC}{dt} = k_2 B,$$

$$(A(0), B(0), C(0)) = (1, 0, 0).$$

In this particular example, we are interested in the dynamics of B . The parameter space is $[0.1, 10] \times [0.1, 10]$, the time interval is $[0, 180]$ (seconds) and the “true”

dynamics of the system will correspond to a fixed value ω_0 that is chosen randomly from the uniform distribution on the parameter space.

The experiments are designed sequentially using criteria (2.1) or (2.6), depending on the assumptions for a given example. In all cases, at step $n + 1$, the expected dynamics and the corresponding variance function are calculated using the Markov Chain Monte Carlo method. A Markov chain of length 10000 with respect to the invariant measure p_n is sampled on the parameter space using Griddy-Gibbs sampling [16]. To speed up the sampling process, a sparse grid interpolant [7] is used to approximate the model output. At each point of the chain, the corresponding dynamics is evaluated using the polynomial interpolant. The average of these sampled dynamics is computed to approximate the EDE, and the variance is approximated in a similar manner. The interpolant we used in this example has an estimated L^∞ error of order 10^{-4} , which is small in comparison to the experiment error and therefore is negligible. The error of the interpolant is estimated by the difference between the interpolated dynamics and the exact dynamics evaluated using the MatLab solver ode15s at 1000 parameter vectors chosen at random from the uniform distribution on the parameter space.

First, we use the framework of Theorem 2.5.2, in which the data is collected with no noise, the time interval is not discretized, and the experiments are designed using condition (2.1). The selected sampling times are shown in Figure 2.1(left panel). We see that even without the discretization of the possible sampled time points, the algorithm focuses on two regions in time that are sufficient to capture the system dynamics. This is consistent with the fact that the system is controlled by two parameters. Figure 2.1 (right panel (i), solid curve) shows how rapidly the EDE approximates the actual response. After 5 experiments, the EDE has converged to the true system dynamics within a negligible error.

Next, we consider the case when the data are subject to Gaussian noise with $\sigma^2 = 0.01$. The dashed curve in Figure 2.1 (right panel, (ii)) represents the error of the EDE as described in the original algorithm. The dash-dot curve (iii) corresponds

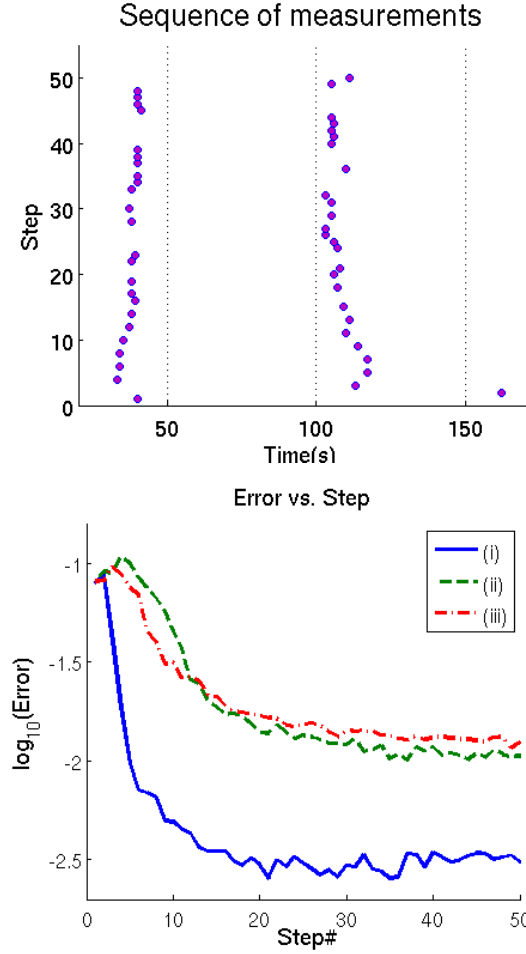


Figure 2.1. (Two dimensional parameter space) Left: Measured time points designed by MINE criteria. The algorithm focuses on two regions in time that capture the system dynamics. Right: The L^∞ errors of EDE on log-scale in three different cases: (i) Data collected with no noise, using the original MINE criteria (2.1) (ii) Data with Gaussian noise, using the original MINE criteria (2.1), (iii) Data with noise, using criterion (2.6) on a finite set of output values and possible measurement time points.

to the assumptions of Theorem 2.6.2. In this case, the experiments are designed sequentially using criteria (2.6) with $C = 2$, and the time interval is discretized by a uniform grid whose distance between neighbour points is equal to 20. In either case, the algorithm provides a good approximation of the true dynamics after just a few sequential experiments.

2.9.2 An ODE model of the T-cell signaling pathway

In this example, we consider a mathematical model of the T-cell signaling pathway proposed by Lipniacki et al. in [9]. This is a system of ODEs with 37 state variables, 19 parameters, and fixed initial conditions. We seek to design experiments to identify the dynamics of pZAP, one of the state variables of the system.

In this example, the parameter space is defined relative to a nominal parameter vector. That is, for each component of the nominal vector, we define a range of five times smaller to five times larger than this component. The whole parameter space is the 19-dimensional set formed by the product of these 19 intervals. The time interval is $[0, 201]$ (seconds). The true dynamics of the system are given by a fixed choice of ω_0 that is chosen randomly from the uniform distribution on the parameter space. The expected dynamics and the corresponding variance function are calculated as described in the previous example. To reduce the computational cost, we also construct a sparse grid interpolant to approximate the output of the ODE system. We use a sparse grid with 50,000 points to construct the interpolant. Even so the interpolant has an L^∞ error of order 10^{-2} , so that there is some mismatch in the model.

Figure 2.2 shows the sequence of design points created by the algorithm in 3 different cases: (i) Data collected with no noise, using the original MINE criteria (2.1); (ii) Data with Gaussian noise, using the original MINE criteria (2.1); (iii) Data with noise, using criterion (2.6) on a finite set of output values and possible measurement time points. In all three cases, the design algorithm focuses on two distinct regions, one of which is precisely defined, the other of which is somewhat nebulous and may perhaps be considered as two regions, particularly in case (iii). This result suggests that although the ODE system is controlled by 19 different parameters, the set of possible system dynamics is contained (at least approximately) in a space of dimension 3. It is worth noting that in [5], the authors also predicted a pile-up of data points under MINE criteria in the open-loop setting (where multiple measurements are chosen in

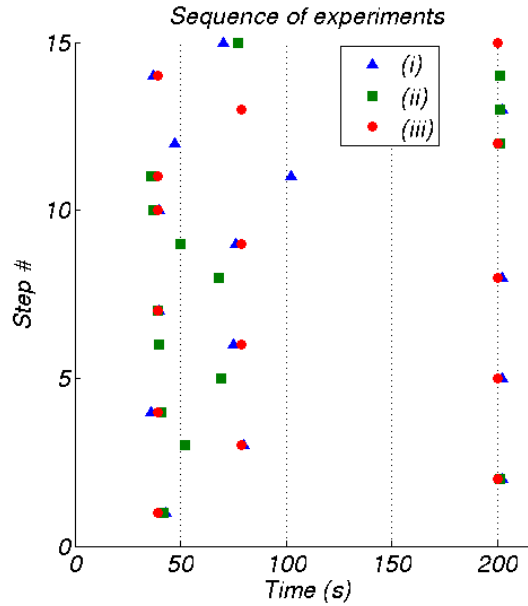


Figure 2.2. (19-dimensional parameter space) Design points in three different cases: (i) Data collected with no noise, using the original MINE criteria (2.1) (ii) Data with Gaussian noise, using the original MINE criteria 2.1, (iii) Data with noise, using criterion 2.6 on a finite set of output values and possible measurement time points.

one step). Our result confirms the same property in the closed-loop case, where the measurements are chosen sequentially with updated probability distributions.

Figure 2.3 shows the approximation error of the EDE in these three cases. As in the previous example, the error in case (i) decreases quickly to a level consistent with the error in the interpolant and the error in MCMC sampling. This supports the assertion that if we know the exact values of the dynamics at three important points, we are able to recover the whole time course of the dynamics.

Since our algorithms in case (ii) and (iii) are data-dependent, we present two different realizations of the performance.

In the first case (left panel of Figure 2.3), the original algorithm using the MINE criteria with noisy data does not do very well in recovering the dynamics after the first 15 experiments. The problem here is that a measurement with significant noise, especially in the first few steps, can cause the estimator to shift toward a region

of parameter space in which the dynamics do not agree with the true dynamics. Moreover, if the output function at this point of measurement is relatively insensitive to parameter changes in this region, it may take many additional measurements to overcome this initial misestimation.

In our example, we encounter this issue: the second measurement made at $t = 201$ gives a data value of nearly 1 when the true value is approximately 0.75. This measurement shifts the probability distribution toward a broad region in the parameter space where the corresponding dynamics saturate to the maximum value 1. This reduces the system variance at time 201 to a relatively small value in comparison to that of other time points. A direct consequence is that in the next eight experiments, no measurement is made around time 200, and the EDE's error does not improve. However, during this process, the parameters that correspond to the true dynamics gain weight, causing the variance around time 200 to increase. Finally, a measurement at time 201 is made in step 11, which significantly decreases the error of the expected dynamics estimator.

This example illustrates the fact that although the convergence of the original algorithm is guaranteed, the convergence may be slow. Some drawbacks of the original algorithm are removed in case (iii) by replacing criteria (2.1) by criteria (2.6) and by restricting the set of possible time points to be finite. By making the set of possible measurement points finite, we collapse the important regions in the time interval to single points and facilitate resampling to get more accurate data at these important points. Also, by using criteria (2.6), we obtain the freedom to select the next measurement point subject to multiple criteria, as described next.

For Figure 2.3, as in the previous example, the set of all possible measurement time points is restricted to a uniform grid of resolution 20, starting from 1. To design experiments, we used the following ranking: among time points with relatively high variance (specifically, that have variance larger than half of the maximum), the time points that have already been measured are given more priority (to promote resampling); among time points that have been measured, the points with fewer

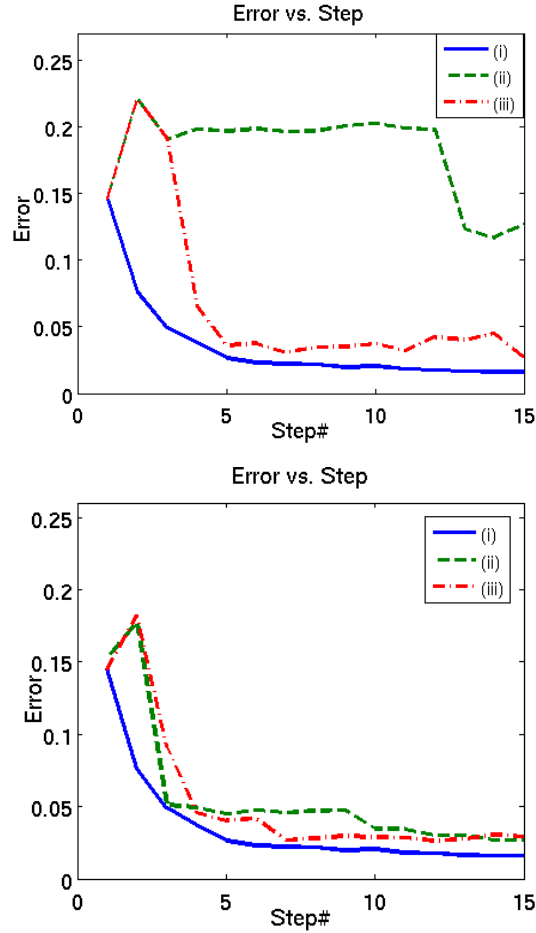


Figure 2.3. (19-dimensional parameter space). The L^∞ errors of EDE on log-scale in three different cases: (i) Data collected with no noise, using the original MINE criteria (2.1) (ii) Data with Gaussian noise, using the original MINE criteria (2.1), (iii) Data with noise, using criterion (2.6) on a finite set of output values and possible measurement time points.

measurements have more priority; among time points that have the same number of measurements, the ones with higher variance have more priority.

The advantages of this variation of the algorithm are illustrated in Figure 2.3 (left panel). After 2 measurements that coincide with those of the previous case (including the point with large measurement error), the algorithm then selects a different measurement point that leads the error to drop quickly. After 6 experiments, the ex-

pected dynamics estimator has converged to the true dynamics within an acceptable error.

On the right panel of Figure 3, we consider a different realization of data on the first measurement. In this case, the random data is obtained with small error and leads to quick convergence of the EDEs corresponding to both criteria.

This example also illustrates the fact that the probabilistic framework in experimental design works well in the case when the number of data is less than the number of parameters, or when the model is unidentifiable: our examples couldn't have been done using a method of parameter estimation via optimization. Assume that in example 2, we can make measurements with high accuracy at 3 time points 50, 100 and 200 and want to know the value at time 150. The number of data points in this case is less than the number of parameters and any method that returns a single parameter estimate will never be able to predict with high confidence (or any confidence at all) the output value at 150. In order to do so, it needs to compute every possible parameter values that fit the data, which is very unlikely in practice. Our probabilistic framework provides a feasible way to address the issue: we considered such an example in Figure 2.4, in which we quantify the uncertainty of the dynamics with only 10 noisy measurements ($\sigma = 0.1$) that accumulate at 3 time points (see also Figure 2.2), which is much less than the number of model parameters (19).

Model mismatch: Finally, we illustrate the effect of model mismatch on the convergence of the EDE by using different sparse grid interpolants to approximate the system output. In this particular example, we run the algorithm with the relaxed MINE criteria on a finite set of measured time points and output values with three different sparse grids of 1000, 2000, 9000 grid points, respectively. The estimated L^∞ -errors of the three interpolants are 0.2, 0.1 and 0.05. As in the previous example, the errors of the interpolants are estimated by the difference between the interpolated dynamics and the exact dynamics evaluated using the MatLab solver ode15s at 1000 parameter vectors chosen uniformly at random from the parameter space. These interpolants are considered as approximate models with varying degrees of mismatch.

In each case, the EDE is evaluated after 10 points selected according to criteria (2.6) with $C = 2$.

The results of this example are given in Figure 2.3. It is not surprising that all three cases give good estimates of the true dynamics: since we are not concerned with the identification of parameters, as long as the dynamics space of the approximate models are close to the dynamics space of the true model, the algorithm will work well. Although the sparse grid interpolant with 1000 grid points is not a good approximation of the system output, it has enough degrees of freedom to capture the behaviour of the system so that a weighted average over parameter space gives a good estimation of the true dynamics.

2.10 Conclusion

Building upon the Maximally Informative Next Experiment algorithm, we have developed several variants of a model-based experiment design algorithm. This algorithm uses existing data to produce a probability distribution on parameter space and then identifies possible measurement points whose output values have large variance under this distribution. We have also proven the convergence of the associated EDE (expected dynamics estimator) to the true system dynamics under a variety of assumptions on the model and data, even when the chosen experiments cluster in a small finite set of points. This approach provides an effective way to incorporate the knowledge arising from nonlinear models into the experiment design process. We illustrated our results with numerical examples on various models of cellular processes.

There are several avenues for future work. First, in [5], the authors proposed several MINE criteria for experimental design. In this work, we establish the theoretical foundations for one of them. The next step would be validating other MINE criteria within a more general model setting.

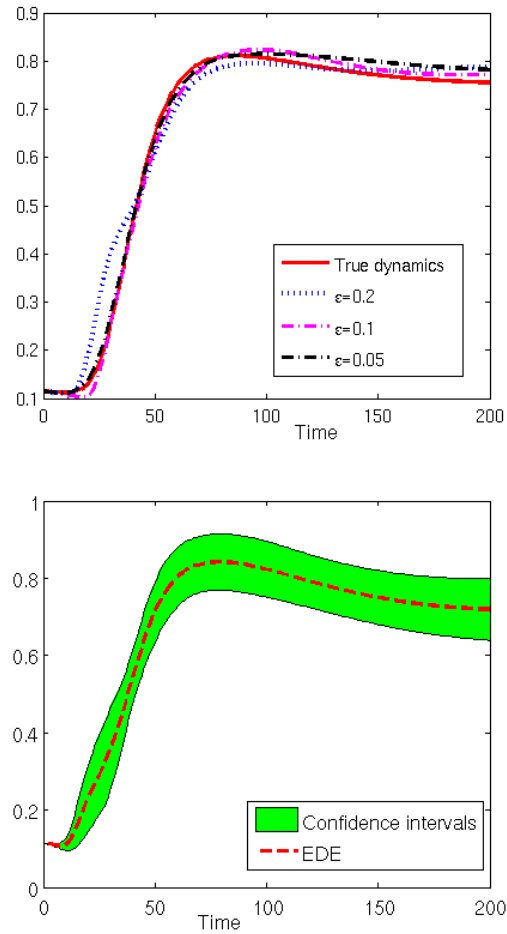


Figure 2.4. (19-dimensional parameter space). Left: EDEs using different sparse grid interpolators to approximate the dynamics. The EDEs are evaluated after 10 steps. Right: Expected dynamics estimator and predicted confidence intervals of the output dynamics with $\epsilon = 0.05$

Second, in this paper, we focused on the identification of observable outputs and did not attempt to address the extrapolation problem, in which measurements of one output are used to make inference about an unobservable output. However, it is worth noting that our framework can be naturally extended to identify unobservable outputs that are theoretically identifiable given that all information about possible observable outputs is known. The problem of determining which unobservable outputs are identifiable in a given experimental setting will be addressed in one of our independent but related works.

2.11 References

- [1] Hiroaki Kitano, (2002), Systems Biology: a Brief Overview. *Science*, Vol. 295, Issue 5560.
- [2] J. N. Bazil, G. T. Buzzard, and A. E. Rundell, (2011), A global parallel model based design of experiments method to minimize model output uncertainty. *Bull. Math. Biol.*, 74:688–716.
- [3] D. Cox and N. Reid. *The Theory of the Design of Experiments* (2000) Monographs on Statistics and Applied Probability. Chapman & Hall/CRC.
- [4] M. M. Donahue, G. T. Buzzard, and A. E. Rundell, (2010), Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids. *IET System Biology*, 4:249–262.
- [5] W. Dong, X. Tang, Y. Yu, R. Nilsen, R. Kim, J. Griffith, J. Arnold, and H. Schuttler, (2008), Systems biology of the clock in *neurospora crassa*. *PLoS ONE*, page e3105.
- [6] V. V. Fedorov and P. Hackl, (1997), *Model-oriented design of experiments*. Springer-Verlag: New York.
- [7] A. Klimke and B. Wohlmuth, (2005), Algorithm 847: Spinterp: piecewise multi-linear hierarchical sparse grid interpolation in matlab. *ACM Trans. Math. Softw.*, 31(4):561–579.
- [8] C. Kreutz and J. Timmer, (2009), Systems biology: experimental design. *FEBS Journal*, 276(4):923–942.
- [9] T. Lipniacki, B. Hat, J. R. Faeder, and W. S. Hlavacek, (2008), Stochastic effects and bistability in T-cell receptor signaling. *Journal of Theoretical Biology*, 254(1):110–122.
- [10] W. J. Morokoff and R. E. Caflisch, (1994), Quasi-random sequences and their discrepancies. *SIAM J. Sci. Comput*, 15:1251–1279.
- [11] R. H. Myers and D. C. Montgomery (2002) *Response surface methodology: Process and product optimization using designed experiments*. John Wiley and Sons: New York.
- [12] H. Niederreiter, (1978), Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.*, 84(6):957–1041.
- [13] L. Pronzato, (2009), Asymptotic properties of nonlinear estimates in stochastic models with finite design space. *Statistics and Probability Letters* 79(21):2307–2313.
- [14] L. Pronzato, (2010), One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal. *Metrika* 71(2):219–238.
- [15] F. Pukelsheim (1993) *Optimal design of experiments*. John Wiley and Sons: New York.
- [16] C. Ritter and M. A. Tanner, (1992), Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *J. Amer. Stat. Assoc.*, 87(419):861–868.

- [17] E. D. Sontag, (2002), For differential equations with r parameters, $2r+1$ experiments are enough for identification. *J. Nonlin. Sci*, 12:553–583.
- [18] L. V. White, (1973), An extension of the general equivalence theorem to nonlinear models. *Biometrika*, 60(2):pp. 345–348.

CHAPTER 3. EFFECTIVE SAMPLING SCHEMES FOR BEHAVIOR DISCRIMINATION IN NONLINEAR SYSTEMS

3.1 Preface

The material presented in this chapter was originally published in the International Journal for Uncertainty Quantification:

Vu Dinh, Ann E. Rundell and Gregory T. Buzzard. Effective sampling schemes for Behavior Discrimination in nonlinear systems. *International Journal for Uncertainty Quantification*.

This article has been reproduced with material omitted or summarized to befit the focus of this dissertation. It has been modified to conform to the format required.

3.2 Abstract

Behavior discrimination is the problem of identifying sets of parameters for which the system does (or does not) reach a given set of states. While there are a variety of methods to address this problem for linear systems, few successful techniques have been developed for nonlinear models. Existing methods often rely on numerical simulations without rigorous bounds on the numerical errors and usually require a large number of model evaluations, rendering those methods impractical for studies of high-dimensional and expensive systems.

In this work, we describe a probabilistic framework to estimate the boundary that separates contrasting behaviors and to quantify the uncertainty in this estimation. In our approach, we directly parameterize the, yet unknown, boundary by the zero level-set of a polynomial function, then use statistical inference on available data to identify the coefficients of the polynomial. Building upon this framework,

we consider the problem of choosing effective data sampling schemes for behavior discrimination of nonlinear systems in two different settings: the low-discrepancy sampling scheme, and the uncertainty-based sequential sampling scheme. In both cases, we successfully derive theoretical results about the convergence of the expected boundary to the true boundary of interest.

We then demonstrate the efficacy of the method in several application contexts with a focus on biological models. Our method outperforms previous approaches to this problem in several ways and proves to be effective to study high-dimensional and expensive systems.

Keywords: Representation of uncertainty, Variance Reduction methods, High-dimensional methods, Classification, Sequential Data, Probabilistic inference, Biological modeling.

3.3 Introduction

Behavior discrimination, or parameter synthesis, is the problem of identifying sets of parameters for which the system does (or does not) reach a given set of states. This problem appears in science and engineering contexts in various forms. For examples, in studies of biological systems, regions of the parameter/input space with different qualitative behaviors need to be treated differently and should be identified before other tools, such as sensitivity analysis, identifiability analysis or model order reduction can be performed. Similarly, in optimal control theory, the constrained optimization problem is well-defined only on the feasible region of the parameter space and the task of computing that region is crucial in designing the control [1]. In uncertainty quantification of discontinuous model response with limited model runs, a preliminary step to identify the structure of the discontinuity also needs to be done before the reconstruction of the model response can be established [3].

While there are a variety of methods to address this problem for linear systems, few techniques have been developed for nonlinear models. One of the recent successful

methods for behavior discrimination of enzymatic reaction networks is proposed by Donze et.al. in [4] and later [5], based on a system's reachability and robustness analysis. This algorithm employs Monte Carlo sampling and sensitivity analysis to explore the space P of possible parameter values and identify subsets of P which robustly satisfy a property ϕ . If a subset of P does not satisfy/violate that property, the algorithm iteratively subdivides it until each subdivisions entirely satisfies or entirely violates ϕ , or which are of insignificant size. The expected result of the overall procedure is a partition of P into small subsets around the boundary between satisfaction and violation of ϕ and larger regions where the satisfaction or violation is robust.

As discussed in [4] and [5], this approach has a few limitations. First, the refinement process implies that the number of partitions increases exponentially with the number of unknown parameters. Thus, in practice, some variables must be held fixed while analyzing the behavior of the model. Second, a common issue in studying complex ODE systems is that the cost for evaluation of system output for a particular set of parameter values can be very expensive. For systems with a large number of equations and unknown parameters, it is usually not computationally feasible to sample enough points by Monte Carlo sampling to explore the parameter space. Third, the method relies on numerical simulations, without rigorous bounds on the numerical errors. Finally, although the method generates a map that partitions the parameter space into regions of the same behavior, no analytical formula to describe the boundary is derived.

In this paper we introduce an alternative way to identify the boundary between the satisfaction and violation regions of a property ϕ . In our approach, we directly parameterize the, yet unknown, boundary by the zero level-set of a polynomial function, then use statistical inference on available data to identify the coefficients of the polynomial.

This idea of using statistical inference to locate the surface of discontinuity was proposed by Sargsyan et. al. [3] for uncertainty quantification of mathematical models

with discontinuity and limited model runs. It is worth noting that [3] focused on finding a probabilistic description of the boundary between the "good" and "bad" regions based on given data, and did not specify an effective way to choose the points where data should be collected for inference. In most examples, Monte Carlo sampling (which can be expensive in complex systems) was employed to choose the points of inference, with an implicit assumption that the selected data contain enough information to constrain the distribution on the set of possible boundaries effectively. Since no estimator of the true boundary was specified, no theoretical result about the convergence or rigorous bound on the numerical errors was provided.

To resolve the issue of computational cost, as well as to provide a theoretical foundation for the usage of statistical inference to locate the surface of discontinuity, in this paper, we investigate two classes of sampling methods to choose the points that need to be evaluated for inference: the low-discrepancy sampling method and the sequential sampling method. The former is a well-known and effective class of sampling schemes to study high dimensional structure, while the latter selects points where current understanding about the location of the boundary is most uncertain. As we will show later in our computational results, the low-discrepancy sampling method can quickly identify the general structure of the boundary, while after a burn-in period, the sequential sampling only samples at points near the boundary of interest. With these effective sampling schemes, the number of model evaluations needed to produce a fair approximation of the boundary is significantly smaller than that required by the methods discussed above.

Our method has several advantages over previous approaches. First, the number of points to be sampled is relatively insensitive to the number of unknown parameters. Thus, the algorithm is a practical method to study high dimensional/computationally expensive systems. Second, by parameterizing the boundary as a zero level set of a polynomial function, we successfully explore a large class of boundary curves, including the case when the boundary has multiple components, which has not been addressed and analyzed in the literature. Third, for both classes of sampling schemes,

we are able to provide theoretical results about the convergence of the estimated surface to the boundary of interest. Finally, by employing a probabilistic framework, our method provides a feasible way to quantify uncertainty in the discriminations. This enables further uncertainty analysis of the system of interest.

The paper is organized as follows. Section 2 provides the mathematical framework and describes our algorithm of behavior discrimination, as well as compares the method with other approaches to the problem. Section 3 establishes the convergence of the estimated surface to the boundary of interest for the two mentioned classes of sampling schemes. In Section 4, we investigate various mathematical models of biological systems to illustrate the efficacy of the method in applications. Further properties about the performance of the algorithm is provided by simulation in Section 5. Finally, in section 6, we conclude the paper with discussions about the method and some descriptions of future work.

3.4 Methodology

3.4.1 Description of the algorithm

In this work, we consider a continuous nonlinear system that can be described by a functional relation between a parameter vector ω and the system output y

$$y = F(\omega)$$

where $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega \subset \mathbb{R}^n$ is the parameter vector; F is a continuously differentiable function of its arguments and $y = (y_1, y_2, \dots, y_{n_y})$ is the vector of system outputs.

For a given set of parameter values, our algorithms first solve the system, then decide whether or not the corresponding trajectory satisfies the property of interest. Hereafter, we will use $G(\omega)$ to denote the response of the system with parameter ω , in which $G(\omega) = 1$ if the system satisfies the property of interest, and $G(\omega) = -1$ otherwise. Since the closed form of G is unknown and evaluating G might be

expensive, we seek to approximate G by a simple rule of discrimination with small error using as few evaluations of G as possible.

Probabilistic representation and uncertainty quantification of the discrimination

To identify the boundary that separates the "good" and "bad" region, we assume without loss of generality that the boundary of interest $\partial\Gamma$ is the zero level set of a smooth function $\Gamma(\omega)$ that can be well approximated by polynomials. Moreover, the polynomial approximations are expressed in the Legendre basis, instead of the monomial basis:

$$f(c, \omega) = \sum_{i=0}^N c_i \eta_i(\omega)$$

with the vector c contained in some coefficient space \mathcal{C} . In other words, the boundary of interest will be modeled as the set $\{\omega \in \Omega : f(c, \omega) = 0\}$, where the coefficient c needs to be inferred from available data.

In our method, based on the collected data $(\omega_i, G(\omega_i))$, $1 \leq i \leq m$, a probability distribution π_m is generated on \mathcal{C} , where $\pi_m(c)$ corresponds to the likelihood that the zero level set of the polynomial $f(c, \cdot)$ is the boundary that separates the two regions

$$\pi_m(c) \propto \exp \left(- \sum_{i=0}^m |G(\omega_i) - \phi(c, \omega_i)| \right) \quad (3.1)$$

where $\phi(c, \omega) = \text{sign}(f(c, \omega))$. Note that this distribution may be generalized (see Remark 3.4.1).

This distribution, when propagated to the space of all possible boundary curves, induces a probabilistic representation of the boundary. The expected prediction function with respect to the distribution π_m is defined as:

$$\bar{\phi}_m(\omega) = E_{\pi_m}[\phi(c, \omega)] \quad (3.2)$$

while the uncertainty in the discrimination at a point ω can partially be represented by the variance in prediction

$$\text{Var}_{\pi_m}[\phi(c, \omega)] \quad (3.3)$$

Effective sampling schemes for behavior discrimination

We will illustrate in this paper that when the collected data ω_i , $1 \leq i \leq m$ satisfies certain patterns, the zero level set of $\bar{\phi}_m(\omega)$ will converge to the true boundary. Specifically, we investigate the convergence of the algorithm in two separate settings:

1. Low-discrepancy sampling method: ω_i , $1 \leq i \leq m$ is a sequence with low discrepancy.
2. Sequential sampling method: data is collected sequentially, where the next data point ω_{m+1} is taken at the point where the maximum of variance in prediction with respect to π_m is achieved.

$$\omega_{m+1} = \arg \max_{\omega \in \Omega} \text{Var}_{\pi_m}[\phi(c, \omega)] \quad (3.4)$$

3.4.2 Main results

The intuition behind the low-discrepancy sampling schemes is simple: if the data we sample on the parameter space is dense enough, and assuming that the boundary between the contrasting behaviors is smooth, we will have enough information to recover the true boundary that separates the two regions. Since low-discrepancy sampling is a well-known and effective scheme to unravel high-dimensional structure (see, for example, [6]), it is natural to use such sampling schemes for discrimination. This intuitive idea is supported by the following result:

Theorem 3.4.1 *Assume that the approximate model is correct, i.e.*

$$\exists c_0 \in \mathcal{C} : \quad G(\omega) = \phi(c_0, \omega) \quad \forall \omega \in \Omega$$

and $\{\omega_i\}$ has discrepancy D_m tending to 0 when $m \rightarrow \infty$.

Then with π_m and $\bar{\phi}_m$ defined as in (3.1) and (3.2), we have

$$\lim_{m \rightarrow \infty} \bar{\phi}_m(\omega) = G(\omega) \quad \forall \omega \in \Omega$$

That is, the predicted classification converges pointwise to the true classification.

Moreover, for all $\epsilon > 0$, if we denote

$$D^{-1}(\epsilon) = \sup\{m + 1 : D_m \geq \epsilon\}$$

Then for $m = \Theta(D^{-1}(\epsilon) + N \frac{1}{\epsilon} \log \frac{1}{\epsilon})$, where N is the number of terms in the polynomial expansion, we have

$$\int_{\Omega} |\bar{\phi}_m(\omega) - G(\omega)| d\omega \leq \epsilon$$

This theorem guarantees that when data is collected with a low-discrepancy scheme, the predicted curves converge to the true boundary that separates two regions. The theorem also provides a numerical bound for the number of points to be sampled to achieve a given level of accuracy. Notice that the number of evaluations needed to approximate the boundary within a given accuracy ϵ does not directly depend on the dimension d of the parameter space, but on the dimension of the coefficient space N . For a fixed degree of smoothness of the true boundary surface, N will increase as a polynomial in d when d becomes larger.

However, by exploring the whole parameter space by a low-discrepancy sequence, we also collect data at insignificant points that do not give much information about the location of the boundary. For complicated systems where the cost for each evaluation of data is high, such a strategy may not be practical. The sequential sampling scheme is proposed to address the issue. By choosing to observe the response at the point with highest uncertainty, the method introduces an effective way to refine the structure of the boundary. While at first sight this method may appear to be heuristic, the convergence of the method is also guaranteed by the following result:

Theorem 3.4.2 *Assume that the approximate model is correct, \mathcal{C} has finite cardinality, and data is collected sequentially as in (3.4).*

$$\omega_{m+1} = \arg \max_{\omega \in \Omega} \text{Var}_{\pi_m}[\phi(c, \omega)]$$

Then

$$\lim_{m \rightarrow \infty} \bar{\phi}_m(\omega) = G(\omega) \quad \forall \omega \in \Omega$$

That is, the predicted classification converges pointwise to the true classification.

In the rest of the paper, the algorithm will be analyzed under the setting specified above. However, we make the following remarks about how the results can be generalized and varied to adapt to different applications. The needed modifications to the proofs are straightforward but are omitted for clarity.

First, we notice that the form of the distribution π_m can be generalized to weighted forms. This accounts for the fact that in some applications, some samples are given more weight than others. For example, discriminations between oscillatory and non-oscillatory dynamics in the face of noise can be made only with certain degree of confidence, and those with higher confidence should be given more weight in the analysis. In feasible analysis of optimal control, one would prefer an under-approximation of the feasible set over an over-approximation one, and two types of misclassification should be weighted differently.

Remark 3.4.1 (Generality of the distribution form) *Theorem 3.4.1 and Theorem 3.4.2 are still valid when the probability distribution in (3.1) is of the form:*

$$\pi_m(c) \propto h(c) \exp \left(- \sum_{i=0}^m k(\omega_i) D(G(\omega_i), \phi(c, \omega_i)) \right)$$

where h, k are arbitrary positive weight functions on (C) and Ω , D is a metric on the set of real numbers. Note that the choice of D does not have much influence on the distribution since the values of G and ϕ are restricted to ± 1 .

Second, criteria (3.4) corresponds to an optimization problem, which may be difficult in some situations. In Remark 3.4.2, we relaxed this condition to a sub-optimization problem that can be solved easily in most cases. Not only does this mean that we don't need to find the optimal sample with high accuracy, it also implies that as long as we take into account the uncertainty in prediction and make an effort to reduce it, a quick convergence toward the true boundary is to be expected.

Remark 3.4.2 (Generality of the sequential sampling criteria) *The convergence result from Theorem 3.4.2 are still valid when criteria (3.4) is replaced by the condition that the variance at the next evaluation point is within a fixed constant of the maximum variance. That is, there exists $C > 1$ so that for all $\omega \in \Omega$ and $m \geq 1$*

$$\text{Var}_{\pi_m(c)} [\phi(c, \omega)] \leq C \text{Var}_{\pi_m(\omega)} [\phi(c, \omega_{m+1})]. \quad (3.5)$$

3.4.3 Comparison to other approaches

One main difference between our method and other algorithms of boundary detection and behavior discrimination comes from our choice of estimator: instead of trying to maximize the likelihood function in order to estimate the boundary, we average the possible curves of discontinuity, weighted by the likelihood function. Our expected estimator has several advantages over the maximum likelihood estimator. First, a prediction using the averaging method is more stable than the prediction of an algorithm based on the maximum likelihood estimator [7]. Secondly, the problem of identifying the maximum likelihood estimator is a global optimization problem, which is more difficult in both theoretical and computational aspects, while the expected estimator can be computed easily by employing the Monte Carlo Markov chain method (whose convergence is relatively insensitive to dimension [8]). Finally, the probabilistic framework we propose in this paper provides a feasible way to quantify uncertainty in the discriminations, which is not available in other techniques.

Another key difference is the choice of design points where data is collected. Designed to address the problem of behavior discrimination in high-dimensional and computationally expensive systems, the sequential design takes into account the information that one gains (or alternatively, the uncertainty one reduces) by collecting data. Not only does this provide a theoretical foundation for the convergence of the methods but also the number of model evaluations needed to produce a fair ap-

proximation of the boundary using our method is smaller than other methods in the literature.

It is worth noting that although our method shares the same idea of using statistical inference for boundary detection employed in [3], the method of inference here is quite different from the Bayesian framework suggested in that paper: instead of using the posterior distribution for inference, our computations are based on a prior distribution, which represents current understanding about the system. By this, we simplify the computing process and relax further assumption on the achieved values and the expected error in the next evaluation. The choice of basis functions (tensorized Legendre polynomials) also helps produce more stable results than the monomial basis employed in [3].

The mathematical formulation of the problem of behavior discrimination also shares a lot of similarity with the problem of classification from machine learning. In fact, to some extent, our algorithm can be framed as an active Bayesian learning algorithm for classification. The distinguishing feature is the choice of data collected for learning: data for statistical learning (either in passive or active learning settings) are generally sampled from some underlying natural unknown distribution and learning this distribution is also a part of the process [9]. Our problem setting, however, allows us to get the response at any point on the parameter space, which provides us more freedom in data sampling. To the best of our knowledge, no analysis of a similar algorithm with either low-discrepancy or sequential data exists in the machine learning literature. From an application point of view, the ability to quantify and reduce uncertainty in inferences also distinguishes the method from other machine learning based classification techniques.

3.5 Convergence results

In this section, we establish results about the convergence of the expected prediction function, that is, in the limit when the number of samples m approaches infinity

(following the algorithms described in the previous section), the approximation zero level set converges to the boundary between the two regions.

For simplicity, we define for every $c \in \mathcal{C}$, the binary classifications

$$\phi(c, \omega) = \text{sign}(f(c, \omega))$$

and

$$G(\omega) = \text{sign}(\Gamma(\omega))$$

(We recall that Γ is the smooth function whose zero level set is assumed to be the boundary that separates regions with different behaviors. Hence, G is the classification function of the parameter space by behaviors.)

3.5.1 Low-discrepancy sampling

We first consider the case when the sequence of samples $\{\omega_m\}$ has discrepancy approaching 0 when $m \rightarrow \infty$. More precisely, for a sequence of points $\{\omega_i\} \subset \Omega$ and a subset B of Ω , let $\#B_m$ be the number of points of $\omega_1, \dots, \omega_m$ contained in B . Then the discrepancy of the first m points is

$$D_m(\{\omega_1, \dots, \omega_m\}) = \sup_{B \subset \Omega} \left| \frac{\#B_m}{m} - \text{Vol}(B) \right|, \quad (3.6)$$

and the low-discrepancy condition means that D_m tends to 0 as m tends to infinity.

This condition covers a large class of sampling schemes, for example: (1) when the set of $\{\omega_i\}$ is grid points of a multi-level sparse grid in parameter space (2) $\{\omega_i\}$ is quasi-random, such as those that are collected by Latin hypercube sampling (3) when $\{\omega_i\}$ is sampled independently from an absolutely continuous distribution, or a Markov Chain whose invariant measure is such a distribution.

Theorem 3.5.1 *Assume that the approximate model is correct, i.e.*

$$\exists c_0 \in \mathcal{C} : \quad G(\omega) = \phi(c_0, \omega) \quad \forall \omega \in \Omega$$

and $\{\omega_i\}$ has discrepancy D_m tending to 0 when $m \rightarrow \infty$.

Then with π_m and $\bar{\phi}_m$ defined as in (3.1) and (3.2), we have

$$\lim_{m \rightarrow \infty} \bar{\phi}_m(\omega) = G(\omega) \quad \forall \omega \in \Omega.$$

That is, the predicted classification converges pointwise to the true classification.

Moreover, for all $\epsilon > 0$, if we denote

$$D^{-1}(\epsilon) = \sup\{m + 1 : D_m \geq \epsilon\},$$

then for $m = \Theta(D^{-1}(\epsilon) + N^{\frac{1}{\epsilon}} \log \frac{1}{\epsilon})$, where N is the number of terms in the polynomial expansion, we have

$$\int_{\Omega} |\bar{\phi}_m(\omega) - G(\omega)| d\omega \leq \epsilon.$$

Theorem 3.4.1 guarantees that when data is collected by one of the schemes discussed above, the predicted curves converge to the true boundary that separates two regions. The theorem also provides a numerical bound for the number of points to be sampled to achieve a given level of accuracy.

Preliminary lemmas

Before moving forward to provide the proof for Theorem 3.4.1, we first establish the following two lemmas

Lemma 3.5.1 *Let g be a non constant polynomial on a compact subset $\mathcal{P} \subset \mathcal{R}^n$. Denote $V = \{p \in \mathcal{P} : g(p) = 0\}$.*

Then there exists k and C_1 depending only on g such that

$$\text{Vol}(\{p \in \mathcal{P} : |g(p)| \leq \epsilon\}) \leq \text{Vol}(\{p \in \mathcal{P} : \text{dist}(p, V) \leq (C_1 \epsilon)^{1/k}\})$$

for all $\epsilon > 0$.

Proof By Lojasiewicz inequality (see, for example, [10]), there exists k and C_1 depends only on g such that

$$\text{dist}(p, V)^k \leq C_1 |g(p)|.$$

Hence, for every p that satisfies $|g(p)| \leq \epsilon$, we have $\text{dist}(p, V) \leq (C_1\epsilon)^{1/k}$. This deduces

$$\{p \in \mathcal{P} : |g(p)| \leq \epsilon\} \subset \{p \in \mathcal{P} : \text{dist}(p, V) \leq (C_1\epsilon)^{1/k}\}.$$

■

Lemma 3.5.2 *Let V be an algebraic surface with Hausdorff dimension $n - 1$ on a compact subset $\mathcal{P} \subset \mathcal{R}^n$. Then*

$$\text{Vol}(\{p \in \mathcal{P} : \text{dist}(\omega, V) \leq \epsilon\}) \leq C_2\epsilon \quad \forall \epsilon > 0,$$

where C_2 depends only on V and the volume constant in n dimensions.

Proof Since V has Hausdorff dimension $n - 1$, there exist C depending only on V such that for all $\epsilon > 0$, the number of balls with radius ϵ needed to cover V ($K(\epsilon)$) satisfies

$$K(\epsilon) \leq \frac{C}{\epsilon^{n-1}}.$$

It is worth noting that such cover of V will also contain the set

$$\{p \in \mathcal{P} : \text{dist}(\omega, V) < \epsilon\}$$

as a subset. Since the volume of a n -dimensional ball with radius r is equal to $C_n r^n$, we deduce that

$$\text{Vol}(\{p \in \mathcal{P} : \text{dist}(p, V) \leq \epsilon\}) \leq \frac{C}{\epsilon^{n-1}} C_n \epsilon^n = C C_n \epsilon$$

which completes the proof. ■

Proof of Theorem 3.4.1

Throughout this section, we denote

$$e_m(c) = \frac{1}{m} \sum_{i=1}^m |\phi(c, \omega_i) - G(\omega_i)|, \quad (3.7)$$

$$e(c) = \int_{\Omega} |\phi(c, \omega) - G(\omega)| d\omega \quad (3.8)$$

and

$$\mathcal{C}_\epsilon = \{c \in \mathcal{C} : e(c) \leq \epsilon\} \quad (3.9)$$

The proof for Theorem 3.4.1 can be summarized as follows: In Lemma 3.5.3, we prove that outside the set of "good" candidates \mathcal{C}_ϵ (on which the error of prediction is less than ϵ), the distribution π_m converges to zero exponentially at a rate depending on $\text{Vol}(\mathcal{C}_\epsilon)$. Lemma 3.5.5 and 3.5.6 establish a lower bound on $\text{Vol}(\mathcal{C}_\epsilon)$ in term of ϵ and N (the number of terms in the polynomial expansion). The combination of those results completes the proof of Theorem 3.4.1.

Lemma 3.5.3 *Let $\epsilon > 0$ and $m \geq D^{-1}(\epsilon/8)$. For all $c \in \mathcal{C} \setminus \mathcal{C}_\epsilon$, we have*

$$\pi_m(c) \leq \frac{\exp(-m\epsilon/4)}{\text{Vol}(\mathcal{C}_{\epsilon/4})}$$

Proof Denote $r_m(c) = \exp(-e_m(c))$ and

$$q_m(c) = \exp\left(-\sum_{i=1}^m |\phi(c, \omega_i) - G(\omega_i)|\right).$$

Consider $\epsilon > 0$ and $m \geq D^{-1}(\epsilon/8)$, if we denote $B = \{\omega : \phi(c, \omega) \neq G(\omega)\}$, then

$$|e(c) - e_m(c)| = 2 \left| \frac{\#\{\omega_i \in B\}}{m} - \text{Vol}(B) \right| \leq 2D^{-1}(\epsilon/8) = \epsilon/4$$

for all $c \in \mathcal{C}$.

Hence, for $c \in \mathcal{C}_{\epsilon/4}$, we deduce that $e_m(c) \leq |e(c) - e_m(c)| + e(c) \leq \epsilon/2$.

Therefore

$$\|r_m\|_m^m = \int_{\mathcal{C}} |r_m(c)|^m \geq \int_{\mathcal{C}_{\epsilon/4}} |r_m(c)|^m \geq \exp(-m\epsilon/2) \text{Vol}(\mathcal{C}_{\epsilon/4}).$$

Now for $c \in \mathcal{C} \setminus \mathcal{C}_\epsilon$, since $|e(c) - e_m(c)| \leq \epsilon/4$, we deduce that $e_m(c) > 3\epsilon/4$. Hence $r_m(c) \leq \exp(-3\epsilon/4)$ and

$$\begin{aligned} \pi_m(c) &= \frac{q_m(c)}{\int_{\mathcal{C}} q_m(c) dc} = \left(\frac{r_m}{\|r_m\|_m} \right)^m \\ &\leq \frac{\exp(-m\epsilon/4)}{\text{Vol}(\mathcal{C}_{\epsilon/4})} \end{aligned}$$

■

Lemma 3.5.4 *For $c_1, c_2 \in \mathcal{C}$, we have*

$$|e(c_1) - e(c_2)| \leq 2 \text{Vol}(\{\omega : |f(c_1, \omega)| \leq |f(c_2, \omega) - f(c_1, \omega)|\}).$$

Proof For any $c \in \mathcal{C}$, denote

$$\Omega_c = \{\omega \in \Omega : \phi(c, \omega) \neq G(\omega)\}$$

We have

$$\begin{aligned} e(c_2) - e(c_1) &= 2 \left(\int_{\Omega_1} d\omega - \int_{\Omega_2} d\omega \right) \\ &= 2 \left(\int_{\Omega_1 \setminus \Omega_2} d\omega - \int_{\Omega_2 \setminus \Omega_1} d\omega \right) \\ &\leq 2 \text{Vol}(\{\omega : \text{sign}(|f(c_1, \omega)|) \neq \text{sign}(|f(c_2, \omega)|)\}). \end{aligned}$$

Note that for $a, b \in \mathcal{R}$ and $b \neq 0$, $\text{sign}(a) \neq \text{sign}(b)$ implies $|a| < |a - b|$. Then

$$|e(c_1) - e(c_2)| \leq 2 \text{Vol}(\{\omega : |f(c_1, \omega)| \leq |f(c_2, \omega) - f(c_1, \omega)|\}).$$

■

Lemma 3.5.5 *There exists $C > 0, k \geq 1$ depending only on \mathcal{C} and Ω such that for all $c_1, c_2 \in \mathcal{C}$ and $\epsilon > 0$*

$$|e(c_1) - e(c_2)| \leq C|c_1 - c_2|^{1/k}.$$

Proof Let $c_1, c_2 \in \mathcal{C}$. Since \mathcal{C} and Ω are compact and f is smooth, there exist C_3 depending only on \mathcal{C} and Ω such that

$$|f(c_2, \omega) - f(c_1, \omega)| \leq C_3|c_2 - c_1|$$

We deduce

$$\begin{aligned} \text{Vol}(\{\omega \in \Omega : |f(c_1, \omega)| \leq |f(c_2, \omega) - f(c_1, \omega)|\}) \\ \leq \text{Vol}(\{\omega \in \Omega : |f(c_1, \omega)| \leq C_3|c_2 - c_1|\}). \end{aligned} \tag{3.10}$$

Using this and Lemma 3.5.4, we have

$$|e(c_1) - e(c_2)| \leq 2 \text{Vol}(\{\omega \in \Omega : |f(c_1, \omega)| \leq C_3|c_2 - c_1|\}). \quad (3.11)$$

Applying Lemma 3.5.1 with $\epsilon = C_3|c_2 - c_1|$ and $V = \{\omega \in \Omega : f(c_1, \omega) = 0\}$, we deduce that there exist C_1 and k depending only on c_1 such that

$$\begin{aligned} \text{Vol}(\{\omega : |f(c_1, \omega)| \leq C_3|c_2 - c_1|\}) \\ \leq \text{Vol}(\{\omega \in \Omega : \text{dist}(\omega, V) \leq (C_1 C_3|c_2 - c_1|)^{1/k}\}). \end{aligned} \quad (3.12)$$

On the other hand, by Lemma 3.5.2 with $\epsilon = (C_1 C_3|c_2 - c_1|)^{1/k}$, we have

$$\begin{aligned} \text{Vol}(\{\omega \in \Omega : \text{dist}(\omega, V) \leq (C_1 C_3|c_2 - c_1|)^{1/k}\}) \\ \leq C_2(C_1 C_3|c_2 - c_1|)^{1/k}. \end{aligned} \quad (3.13)$$

Combining (3.10), (3.11), (3.12) and (3.13), we deduce

$$|e(c_1) - e(c_2)| \leq 2C_2(C_1 C_3)^{1/k}|c_2 - c_1|^{1/k}.$$

Since \mathcal{C} is compact, the constants can also be chosen independent of c_1 . ■

Lemma 3.5.6 *For \mathcal{C}_ϵ defined as in (3.7), k the constant defined in Lemma 3.5.5 and N the dimension of the coefficient space, there exists C depending only on \mathcal{C} and Ω such that for all $\epsilon > 0$*

$$\text{Vol}(\mathcal{C}_\epsilon) \geq C\epsilon^{Nk}$$

Proof Recall that c_0 is the true vector of coefficients (hence $e(c_0) = 0$). Then for $c \in \mathcal{C}$ such that $|c - c_0| \leq (\frac{\epsilon}{C})^k$, where k and C are defined as in Lemma 3.5.5, we have

$$|e(c) - e(c_0)| \leq C|c - c_0|^{1/k} \leq \epsilon$$

which implies that $c \in \mathcal{C}_\epsilon$.

Therefore

$$B = \{c \in \mathcal{C} : |c - c_0| \leq \left(\frac{\epsilon}{C}\right)^k\} \subset \mathcal{C}_\epsilon$$

and

$$\text{Vol}(\mathcal{C}_\epsilon) \geq \text{Vol}(B) = C_1 \left(\frac{\epsilon}{C}\right)^{Nk}$$

where C_1 is the volume constant in N -dimensional space. ■

Proof of Theorem 3.4.1

Proof We have

$$\begin{aligned} \int_{\Omega} |G(\omega) - \bar{\phi}_m(\omega)| d\omega &= \int_{\Omega} |E_{\pi_m}[G(\omega) - \phi(c, \omega)]| d\omega \\ &\leq \int_{\Omega} E_{\pi_m}[|G(\omega) - \phi(c, \omega)|] d\omega = E_{\pi_m} \left[\int_{\Omega} |G(\omega) - \phi(c, \omega)| d\omega \right] \\ &= E_{\pi_m}[e(c)] = \int_{e(c) > \epsilon/2} e(c) \pi_m(c) dc + \int_{e(c) \leq \epsilon/2} e(c) \pi_m(c) dc \end{aligned} \quad (3.14)$$

$$\leq \text{Vol}(\mathcal{C}) \exp(-m\epsilon/8) \frac{1}{\text{Vol}(\mathcal{C}_{\epsilon/8})} + \epsilon/2 \quad (3.15)$$

with the inequality obtained by using Lemma 3.5.3 for $m \geq D^{-1}(\epsilon/16)$. Hence if we further choose m that satisfies

$$\text{Vol}(\mathcal{C}) \exp(-m\epsilon/8) \frac{1}{\text{Vol}(\mathcal{C}_{\epsilon/8})} \leq \epsilon/2$$

or equivalently, in Θ -notation (using Lemma 3.5.6)

$$m = \Theta \left(D^{-1}(\epsilon/16) + Nk \frac{1}{\epsilon} \log \frac{1}{\epsilon} \right)$$

then

$$\int_{\Omega} |G(\omega) - \bar{\phi}_m(\omega)| d\omega \leq \epsilon,$$

which completes the proof of Theorem 3.4.1. ■

3.5.2 Sequential sampling

The sequential sampling scheme is proposed to study high-dimensional and expensive systems. By choosing to observe the response at the point with highest uncertainty, the method introduces an effective way to reduce the uncertainty and refine the structure of the boundary. While at first sight this method may appear to be heuristic, the convergence of the method is also guaranteed by the following result:

Theorem 3.5.1 *Assume that the approximate model is correct, \mathcal{C} has finite cardinality, and data is collected sequentially as in (3.4).*

$$\omega_{m+1} = \arg \max_{\omega \in \Omega} \text{Var}_{\pi_m}[\phi(c, \omega)]$$

Then

$$\lim_{m \rightarrow \infty} \bar{\phi}_m(\omega) = G(\omega) \quad \forall \omega \in \Omega$$

That is, the predicted classification converges pointwise to the true classification.

Before moving forward to provide the proof of the theorem, it is worth noting that in the sequential setting, an additional condition is imposed on the coefficient space: the space is supposed to have finite cardinality. This condition comes from the fact that in a continuous framework, a coefficient vector has measure zero and good performance of the true coefficient vector does not always guarantee that it will stay in the support of the limit distribution. Though this situation is perhaps unlikely to happen in practice, we cannot exclude such a possibility for a convergence result. We also want to note that this assumption was also commonly used in the context of parameter identification [11,12]. In practice this condition may be achieved without affecting the model's ability to approximate the true boundary, for example, by subdividing each coordinate axis using a fixed step size and taking the set of points in that lie on the resulting grid.

Proof Denote

$$q_m(c) = \exp \left(- \sum_{i=1}^m |\phi(c, \omega_i) - G(\omega_i)| \right),$$

Then $\pi_m = C_m q_m$. Also, let A be the set of cluster points of $\{\omega_m\}$: points $\omega \in \Omega$ such that there exists a subsequence $\{\omega_{m_k}\}$ of with $\omega_{m_k} \rightarrow \omega$.

Step 1: We claim first that if $\pi_m(c)$ does not tend to 0 with m (so that c has probability above some fixed $\rho > 0$ for infinitely many m), then $f(c, \omega) = G(\omega)$ for all $\omega \in A$.

Proof:

Consider any $c \in \mathcal{C}$, $\omega \in A$ such that $|\phi(c, \omega) - G(\omega)| > 0$. Then there exists a subsequence $\{\omega_{m_k}\}$ of $\{\omega_m\}$ such that $\omega_{m_k} \rightarrow \omega$ and $|\phi(c, \omega_{m_k}) - G(\omega_{m_k})| \geq 1$. Hence

$$\sum_{i=1}^m |\phi(c, \omega_i) - G(\omega_i)| \rightarrow \infty$$

when $m \rightarrow \infty$, and so $q_m(c) \rightarrow 0$.

On the other hand, the assumption that there exists c_0 such that $\phi(c_0, \omega) = G(\omega)$ for all ω implies that $q_m(c_0) = 1$. Therefore, $\pi_m(c_0)/\pi_m(c) \rightarrow \infty$. Since \mathcal{C} is a finite space, $\pi_m(c_0) \leq 1$, and hence $\pi_m(c) \rightarrow 0$. Hence $\pi_m(c) \not\rightarrow 0$ implies $\phi(c, \omega) = G(\omega)$ for all $\omega \in A$.

We deduce that

$$\lim_{m \rightarrow \infty} E_{\pi_m}[\phi(c, \omega)] = G(\omega) \quad \forall \omega \in A$$

and

$$\text{Var}_{\pi_m}[\phi(c, \omega)] \rightarrow 0 \quad \forall \omega \in A.$$

Step 2: Now we claim that

$$\text{Var}_{\pi_m}[\phi(c, \omega_{m+1})] \rightarrow 0. \tag{3.16}$$

Proof:

We have

$$\begin{aligned} & E_{\pi_m} [|\phi(c, \omega_{m+1}) - G(\omega_{m+1})|^2] \\ &= \sum_{c: \pi_m(c) \rightarrow 0} |\phi(c, \omega_{m+1}) - G(\omega_{m+1})|^2 \pi_m(c) + \sum_{c: \pi_m(c) \not\rightarrow 0} |\phi(c, \omega_{m+1}) - G(\omega_{m+1})|^2 \pi_m(c). \end{aligned}$$

However, by the same argument as above, $\pi_m(c) \not\rightarrow 0$ implies that c makes only a finite number of mistakes in prediction, which implies that $\phi(c, \omega_{m+1}) = G(\omega_{m+1})$ for m large enough. Therefore

$$E_{\pi_m} [|\phi(c, \omega_{m+1}) - G(\omega_{m+1})|^2] \rightarrow 0$$

and

$$\begin{aligned} \text{Var}_{\pi_m} [\phi(c, \omega_{m+1})] &= E_{\pi_m} [|\phi(c, \omega_{m+1}) - G(\omega_{m+1})|^2] - (G(\omega_{m+1}) - E_{\pi_m} \phi(c, \omega_{m+1}))^2 \\ &\leq E_{\pi_m} [|\phi(c, \omega_{m+1}) - G(\omega_{m+1})|^2] \rightarrow 0. \end{aligned}$$

Step 3: The choice of ω_{m+1} gives

$$\text{Var}_{\pi_m} [\phi(c, \omega)] \leq \text{Var}_{\pi_m} [\phi(c, \omega_{m+1})] \quad \forall \omega \in \Omega, \quad (3.17)$$

and the right hand side tends to 0 as $m \rightarrow \infty$ by step 2.

Then for all $\omega \in \Omega$

$$\lim_{m \rightarrow \infty} \sum_{\pi_m(c) \neq 0} |\phi(c, \omega) - E_{\pi_m} \phi(c, \omega)|^2 = 0.$$

From the fact that $\pi_m(c_0) \geq \pi_m(c) \forall c \in C$ and $\phi(c_0, \omega) = G(\omega) \forall \omega \in \Omega$, we have

$$\begin{aligned} |G(\omega) - E_{\pi_m} \phi(c, \omega)|^2 &= |\phi(c_0, \omega) - E_{\pi_m} \phi(c, \omega)|^2 \\ &\leq \#C \sum |\phi(c, \omega) - E_{\pi_m} \phi(c, \omega)|^2 \pi_m(c) = \#C \text{Var}_{\pi_m} [\phi(c, \omega)] \rightarrow 0 \end{aligned}$$

as $m \rightarrow \infty$. Hence

$$\lim_{m \rightarrow \infty} E_{\pi_m} \phi(c, \omega) = G(\omega) \quad \forall \omega \in \Omega.$$

■

3.6 Behavior Discrimination in enzymatic networks.

3.6.1 A model of the acute inflammatory response to infection

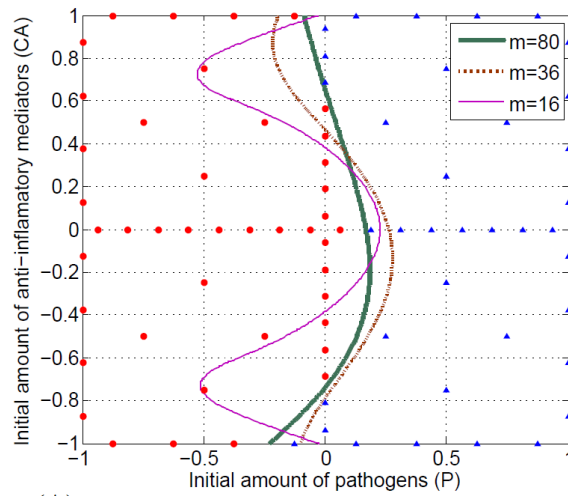
We applied our method to a model of the acute inflammatory response to infection. This 4-equation, 22-parameter model was presented in [13], where the state variables

P , N_A , D , and C_A , correspond to the amounts of pathogen, pro-inflammatory mediators (e.g., activated neutrophils), tissue damage, and anti-inflammatory mediators (e.g., cortisol and interleukin-10), respectively.

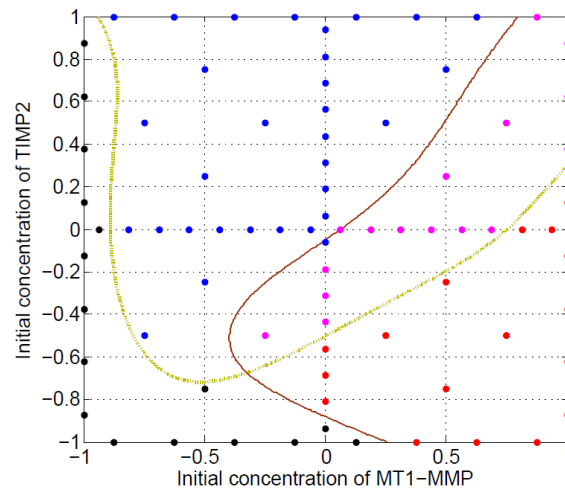
The response is a complex process that exhibits both pro and anti-inflammatory behaviors. The pro-inflammatory elements are responsible for eliminating the pathogen, but pathogen killing can cause collateral tissue damage. This in turn triggers an escalation in the pro-inflammatory response, creating a positive feedback cycle. The anti-inflammatory elements counteract this cycle, minimizing tissue damage and promoting healing. However, in cases of extreme infection, this delicate balance is destroyed, which leads to a potentially lethal amount of tissue damage.

In this example, we validated our method by reproducing results previously obtained in [13], and later in [4]. Death is defined as a sustained amount of tissue damage (D) above a specified threshold value and constitutes the undesirable outcome we wish to avoid. The two unknown parameters considered are the initial amount of pathogen, P_0 , and the initial amount of anti-inflammatory mediators, CA_0 . The parameter space Ω is defined relative to a nominal parameter vector; that is, for each component of the nominal vector, we define a range of 10 times smaller to 10 times larger than this component. The growth rate of pathogen was set to 0.3 and other parameters to their nominal values as in [4].

For boundary detection, we evaluate the model on the grid points of the 16-point, 36-point and 80-point two dimensional uniform sparse grid. The result is presented in Figure 3.1A, where the parameter space is scaled to the unit square. The expected boundary is computed using a 10^5 -point Monte Carlo Markov Chain on a 13-dimensional coefficient space (that is, $N=13$). The result is similar to what achieved in [4], but the number of model evaluations needed to produce this result is significantly lower.



(A)



(B)

Figure 3.1. (A) A model of the acute inflammatory response to infection: The predicted boundaries computed by sparse sampling at the 16-point, 36-point and 80-point sparse grid nodes. (B) A model of collagen degradation: the design points and predicted boundaries computed at the nodes of the 80-point sparse grid. In both figures, the expected boundaries are computed using a 10^5 -point Monte Carlo Markov Chain on a 13-dimensional coefficient space.

3.6.2 A model of collagen degradation

Our second example is performed on the biochemical network adapted from [14] and later extended in [5], which models the loosening of the extra-cellular matrix, a crucial process in angiogenesis, the sprouting of new blood vessels as a reaction to signals that indicate the need for additional oxygen in certain tissues. The system consists of 12 differential equations that integrates on a long time scale, which makes model evaluations become very expensive.

We investigate the relative contribution of MT1-MMP and MMP2 on collagen proteolysis using a combination of constraints imposed on (1) the amount of collagen that has been degraded after a given time and (2) the respective contribution of any of the two enzymes onto collagen degradation. Studying those two properties, we are able to create the division maps in Figure 3.1B. The region above the solid curve corresponds to the case when the amount of collagen degraded by MT1-MMP is greater than that degraded by MMP2; the region on the right of the other curve corresponds to the case when the system does (does not) manage to degrade 90% of the collagen before 12 hours. Similar to the previous example, the boundaries in Figure 3.1B were computed using data at the first 80 grid points of a uniform sparse grid. This division map replicates the result about collagen proteolysis previously achieved in [5] with a lower number of evaluations.

In Figure 3.2A, we employ the sequential sampling scheme to approximate the blue boundary in Figure 3.1B with higher accuracy using the same number of model evaluations. Starting with a prior data set collected at the first 16 points of a uniform sparse grid, we use the relaxed variance criteria to choose the next sample point until 80 data points are obtained. We can see that the boundary generated by the sequential sampling method converges quickly to the true boundary in less than a hundred model evaluations. This advantage comes from the fact that after a burn-in period, the algorithm selects points near the true boundary and focuses the probability distribution on \mathcal{C} on a few regions that contain potential candidates.

As we emphasized earlier in the introduction, the framework proposed in this method provides a natural probabilistic representation of the classifying boundary, which enables further uncertainty analysis of the system. This is a feature that distinguishes the algorithm from other approaches. In Figure 3.2B, we provide a contour map of the variance in prediction with respect to the data-dependent probability distribution π_m constructed on the coefficient space. The variance map can be considered as a representation of the uncertainty in discrimination, or a measure of confidence one has in classification. From the figure, we also notice that the region with high variance encloses around the true boundary, illustrating the fact that after a burn-in period, the sequential sampling scheme only selects points that are close to the boundary.

3.6.3 A model of the T-cell signaling pathway

Our next example is a mathematical model of the T-cell signaling pathway proposed by Lipniacki et al. in [15]. This is a system of ODEs with 37 state variables, 19 parameters, and fixed initial conditions. We seek to design a sampling scheme to identify the boundary between the region of the parameter space where pERK, a state variable of the system, stabilizes at a high level of concentration, and the region at which pERK's concentration is less than a threshold level. In this example, the parameter space is defined relative to a nominal parameter vector. That is, for each component of the nominal vector, we define a range of 5 times smaller to 5 times larger than this component. We focus our attention on the 8 most sensitive parameters determined by a global sensitivity analysis algorithm. The whole parameter space is the 8-dimensional set formed by the product of these 8 intervals. The time interval for analysis is $[0, 6000]$ (seconds).

For the sake of clarity in illustrations, we first investigate the performance of the algorithm in the case when all but the three most sensitive parameters are fixed. Starting with a prior data set collected at the first 44 points of a three-dimensional

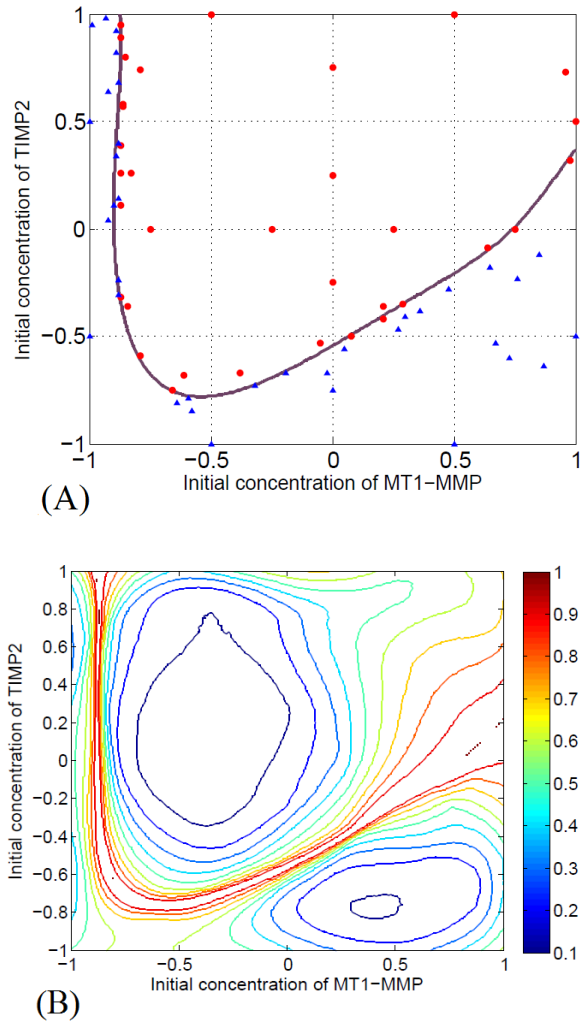


Figure 3.2. Relative contribution of MT1-MMP and MMP2 to collagen degradation. The boundary separates points for which the contribution of MT1-MMP is dominant from points for which MMP2's contribution is dominant: (A) Design points and predicted boundary derived by the sequential sampling scheme. (B) A characterization of uncertainty in discrimination by variance. Notice that the points with high variance lie around the true boundary, which explains why the data sampled on the figure on the left also tends to focus around the true boundary.

uniform sparse grid, we use the relaxed variance criteria to choose the next sample point until 144 data points are obtained. Figure 3.3A shows the boundary surface where the region above the surface corresponds to parameter values that will stabilize the concentration of pERK at a level higher than a fixed threshold. Similar to the previous example, samples selected by the sequential scheme tends to focus more and more along the boundary surface.

We then consider the case when 8 most sensitive parameters are varied in the interval mentioned above. To evaluate the performance of different sampling schemes in learning the structure of the boundary, we look at three different scenarios. In the first scenario, the Latin hypercube sampling is employed to collect 400 data points for inference. In the second scenario, starting with 100 data points chosen uniformly at random on the parameter space, the sequential sampling scheme is performed until 400 samples are obtained. Finally, to compare with the best possible performance, in the last scenario, we consider the ideal case when the sampling scheme is designed by an omniscient oracle that at each step knows all the points for which current prediction is incorrect. Starting with 100 data points chosen uniformly at random (the same as in the second scenario), the oracle uses the estimated boundary as above to derive a expected prediction function. Upon testing the expected prediction function on thousands of samples, the oracle would add the misclassified samples into the training data set and continue the process.

Since the same algorithm of boundary computation is employed in all three scenarios, this examples provides a fair evaluation of the performance of these sampling schemes. The results are provided in Figure 3.3B. We can see that the sequential sampling scheme significantly outperforms the Latin hypercube sampling, and has comparable performance to the ideal case.

This result indicates that the sequential sampling scheme can effectively approximate a complex surface in 8-dimensional space with reasonable accuracy using only 400-500 samples. We want to note that this result is not likely to be produced by traditional methods for behavior discrimination. For example, if the discrimination

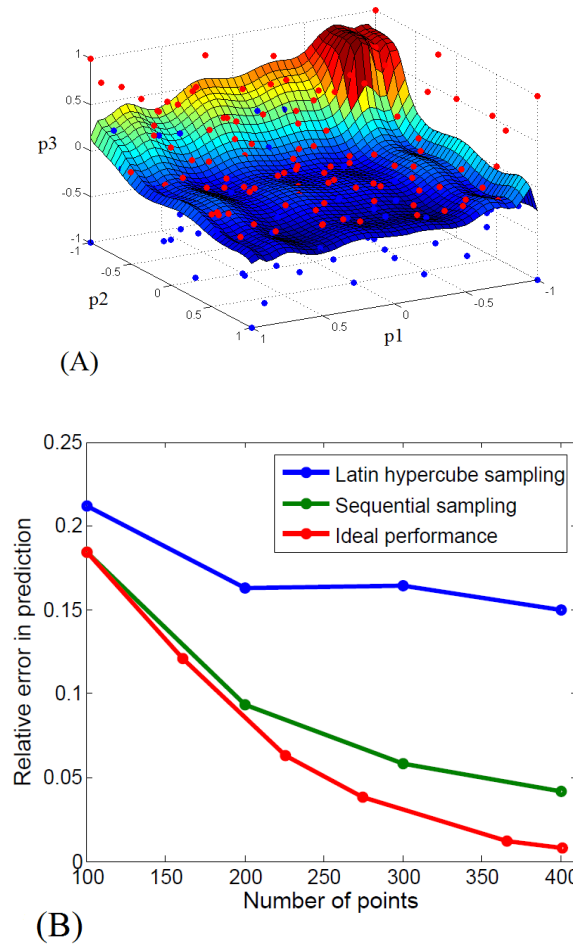


Figure 3.3. A model of the T-cell signalling pathway with discrimination based on a threshold value for pERK at the final simulation time: (A) 3-dimensional case: Design points and predicted boundary derived by the sequential sampling schemes with $N=57$. (B) 8-dimensional case with $N=321$: Error in prediction as the number of samples increase in three different scenarios: Latin hypercube sampling, sequential sampling scheme and oracular sampling

is done by iteratively subdividing cubes that neither satisfy nor violate the property of interest into smaller hypercubes, every time a subdivision is performed, $2^8 - 1$ new elements are created and the number of samples needed to check if each of those elements satisfies/violates the property is almost as large as the number of samples we used in this example.

3.7 Additional properties

In this section, through 2D examples, we validate the theoretical results established in the Section 3, as well as illustrate other properties of our algorithm. In most examples (except example 5.3), the synthetic discontinuous output u_1, u_2, \dots, u_m are generated by evaluating the signum function along some discontinuity curve $r(x, y) = 0$:

$$u = \text{sign}(r(x, y))$$

Throughout this section, except example 5.2, the boundaries are computed by Griddy Gibbs Markov Chains of 10^5 points on the coefficient space, while the number of terms in the approximation is set at 13 (i.e., we consider the first 13 terms in the sparse grid expansion of Γ).

3.7.1 Convergence

Figure 3.4 demonstrates the convergence of the algorithm in both low-discrepancy and sequential settings. In this experiment, data u_1, u_2, \dots, u_m are generated by evaluating the signum function along the discontinuity curve $r(x, y) = 0$:

$$u = \text{sign}(r(x, y)) \tag{3.18}$$

The parameter space in this case is $[-1, 1] \times [-1, 1]$, whereas the discontinuity function is described by $r(x, y) = y - (x - \frac{1}{4})^2$.

To compute the prediction error of each estimated classification, we evaluate the model at 10^6 points collected at random (uniformly) and compute the predicted results using one of the two sampling schemes. The blue curve corresponds to the case when data is sampled at the grid points of the sparse grid with uniform grid points, while for the red curve, data are collected using the sequential scheme with prior data given by the first 16 sparse grid points (low discrepancy case). In both cases, we can see that when the number of sampling points increases, the error of prediction converges

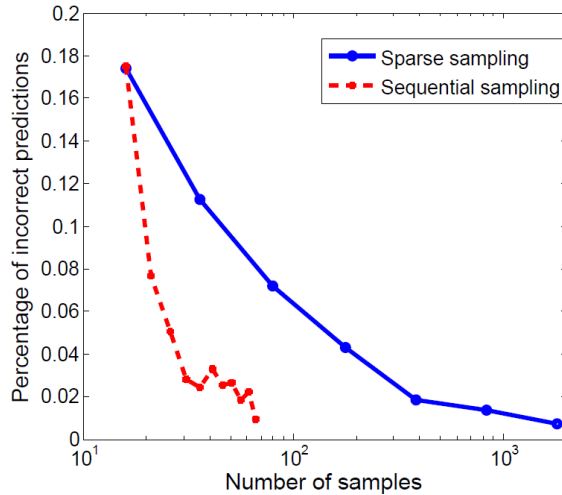


Figure 3.4. Convergence rate of prediction error for the function in equation (3.18) when number of samples increases. In both cases, when the number of sampling points increase, the error of prediction converges to zeros. However, the convergence rate of the sequential scheme is significantly faster than that of the sparse grid sampling.

to zero. However, the convergence rate of the sequential scheme is significantly faster than that of the sparse grid sampling.

3.7.2 Dependence on the number of terms in polynomial expression

In Figure 3.5A, we consider a case when the discontinuity curve cannot be expressed as the level set of a polynomial function, in which

$$r(x, y) = y^2 - (x^3 - 2x - 1 - 3e^x) \quad (3.19)$$

and the input space is $[-6, 6] \times [-6, 6]$.

In this example, to reconstruct the boundary with high accuracy, the model is evaluated at 500 points, which are chosen uniformly at random in the input space. The algorithm to identify the boundary is employed with increasing number of terms in the polynomial approximation expressions. The expected boundaries were computed by a Markov Chain of 10^5 points using the Griddy Gibbs sampling method. The

error rates were derived by the empirical prediction error on 10^6 points collected at random (uniformly). On the left panel, we plot the error rates in terms of the number of basis functions used in the approximation. On the right panel, the approximated boundaries with various degree of approximation were illustrated. As expected, when the number of terms used in the approximation increases, the predicted boundary converges to the true boundary of discrimination.

3.7.3 Boundary with multiple components

We illustrate the fact that our algorithm can deal with cases when the boundary of interest has disconnected components. In this example, the model response is computed by the elliptic function

$$r(x, y) = \frac{y^2}{4} - \left(\frac{x^3}{8} - \frac{x}{2} \right) \quad (3.20)$$

on $[-2, 2] \times [-2, 2]$.

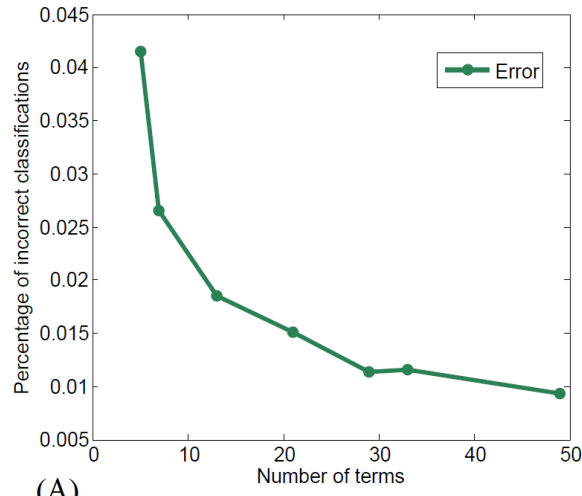
The constructed boundary is plotted in Figure 3.6A, using 200 sample chosen uniformly at random.

3.7.4 Robustness

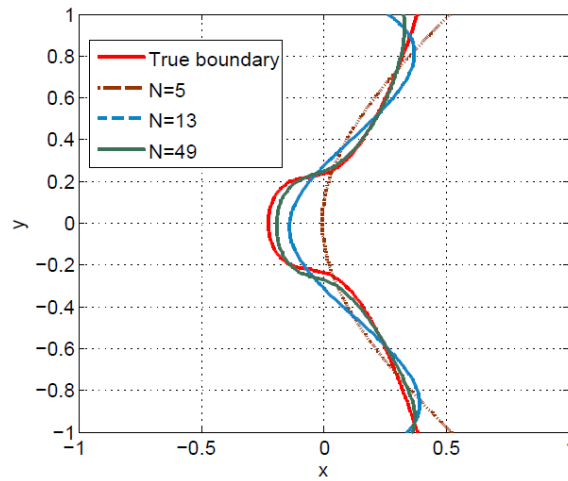
In this particular example, we consider the behavior of our algorithm when some of the assumptions are not met. In this case, the synthetic discontinuous data u_1, u_2, \dots, u_m are generated by evaluating a bivariate error function with discontinuity strength parameter γ , discontinuity curve $r(x, y)$, and an additional global oscillatory structure with amplitude δ :

$$u_i = \text{erf}(\gamma(r(x_i, y_i))) + \delta \sin\left(\frac{\pi}{3}(y_i + x_i)\right) \quad (3.21)$$

It is worth noting that this example violates (mildly) some conditions of our algorithm: (1) the response function is not discontinuous, but changes sharply across a curve, (2) the curve of discontinuity is not a perfect zero level set of any polynomial



(A)



(B)

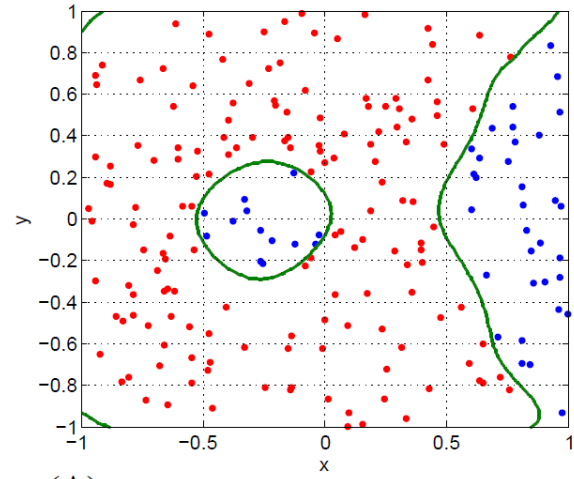
Figure 3.5. (A) Decrease in prediction error when the number of terms (N) in the approximation increases for the function in equation (3.19). (B) Sample predicted boundary with different value of N . The boundaries are computed using 500 samples collected uniformly at random, while the error rates are estimated by the empirical prediction error on 10^6 uniformly distributed random points.

functions (but approximately is) and (3) the responses on both sides of the discontinuity curve are not flat, but have some additional oscillatory structure. Despite those violations, the function itself still resembles behavior of a yes/no response, and hence would work well with our algorithm for behavior discrimination.

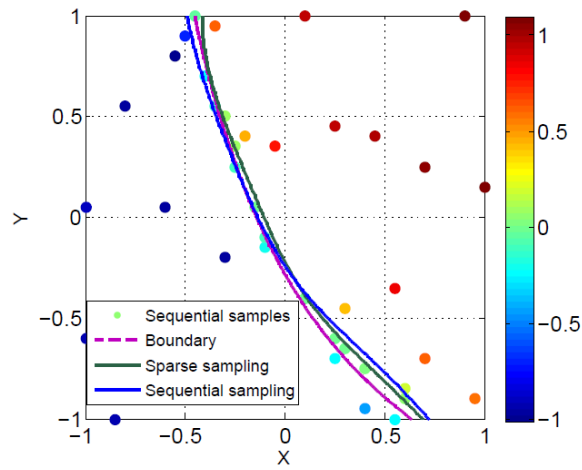
In Figure 3.6B, we replicate this example as used previously in [3], where the input parameter space is $[0.5, 6] \times [0, 2]$, while the discontinuity curve is a shifted and rescaled inverse function that goes through $(2, 2)$ and $(5, 0)$, i.e., $r(x, y) = y - \frac{20}{3} \left(\frac{1}{x} - \frac{1}{5} \right)$. The amplitude of the oscillatory structure is set to $\delta = 0.1$, and the steepness parameter is $\gamma = 2$. In the figure, the first 40 points generated by the sequential sampling are shown along with a expected boundary using 40 samples. We also computed the expected boundary derived by using first 80 nodes of a uniform sparse grid to make it convenient to compare the two methods. We can see that both methods of sampling provide good estimates of the boundary. The boundary is reconstructed with better quality than in [3], where data are collected at random locations. This example confirms the robustness of our method under perturbation of the system of interest. This feature is important for studies of reaction networks, where in some cases, the qualitative behavior of interest can only be determined within a certain level of confidence.

3.8 Conclusions and discussions

In this work, we investigated the problem of choosing effective data sampling schemes for behavior discrimination of nonlinear systems. Using a probabilistic framework to estimate the boundary and quantify the uncertainty in discrimination, we give results about two classes of sampling schemes: the low-discrepancy sampling, and the sequential sampling. In both cases, we successfully derived theoretical results about the convergence of the expected boundary to the true boundary of interest. We then demonstrate the efficacy of the method in different application contexts. The method proves to be effective to study high-dimensional and expensive systems.



(A)



(B)

Figure 3.6. (A) Example of a boundary with multiple components: discrimination between the region of positive and negative values of the elliptic function in equation (3.20) (B) Performance of the algorithm when multiple assumptions of the setting are mildly violated (equation (3.21)).

Nevertheless, there are some limitations to this method that are worth mentioning. First, although we are able to give an estimate of the convergence rate for the low-discrepancy setting, no such estimate is provided for the sequential setting (although we did provide a consistency result in that case). Second, one of our main assumptions in the paper is that the boundary of interest is well approximated by a finite polynomial expansion. In practice, however, the number of terms needed for approximation is difficult to be determined in advance. A more practical extension of the method should include an adaptive scheme to determine the number of terms in an effective way. These will be subjects of future work.

3.9 References

- [1] Summers, S., Jones, C. N., Lygeros, J., and Morari, M., A multiresolution approximation method for fast explicit model predictive control, *Automatic Control, IEEE Transactions on*, 56(11):2530–2541, 2011.
- [2] Chakrabarty, A., Dinh, V., Buzzard, G. T., Zak, S. H., and Rundell, A. E., Robust explicit nonlinear model predictive control with integral sliding mode, In *American Control Conference (ACC), 2014*, pp. 2851–2856. IEEE, 2014.
- [3] Sargsyan, K., Safta, C., Debusschere, B., and Najm, H., Uncertainty quantification given discontinuous model response and a limited number of model runs, *SIAM Journal on Scientific Computing*, 34(1):B44–B64, 2012.
- [4] Donzé, A., Clermont, G., and Langmead, C. J., Parameter synthesis in nonlinear dynamical systems: Application to systems biology, *Journal of Computational Biology*, 17(3):325–336, 2010.
- [5] Donzé, A., Fanchon, E., Gattepaille, L. M., Maler, O., and Tracqui, P., Robustness analysis and behavior discrimination in enzymatic reaction networks, *PLoS One*, 6(9):e24246, 2011.
- [6] Niederreiter, H., Quasi-Monte Carlo methods and pseudo-random numbers, *Bulletin of the American mathematical society*, 84(6):957–1041, 1978.
- [7] Freund, Y., Mansour, Y., and Schapire, R. E., Generalization bounds for averaged classifiers, *Annals of Statistics*, pp. 1698–1722, 2004.
- [8] MacKay, D. J. Introduction to Monte Carlo methods. In *Learning in graphical models*, pp. 175–204. Springer, 1998.
- [9] Settles, B., Active learning literature survey, *Computer Sciences Technical Report 1648*, University of Wisconsin, Madison, 2010.
- [10] Bierstone, E. and Milman, P. D., Semianalytic and subanalytic sets, *Publications Mathématiques de l’IHÉS*, 67(1):5–42, 1988.
- [11] Pronzato, L., Asymptotic properties of nonlinear estimates in stochastic models with finite design space, *Statistics & Probability Letters*, 79(21):2307–2313, 2009.
- [12] Pronzato, L., One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal, *Metrika*, 71(2):219–238, 2010.
- [13] Reynolds, A., Rubin, J., Clermont, G., Day, J., Vodovotz, Y., and Bard Ermentrout, G., A reduced mathematical model of the acute inflammatory response: I. derivation of model and analysis of anti-inflammation, *Journal of theoretical biology*, 242(1):220–236, 2006.
- [14] Karagiannis, E. D. and Popel, A. S., A theoretical model of type I collagen proteolysis by matrix metalloproteinase (MMP)2 and membrane type 1 MMP in the presence of tissue inhibitor of metalloproteinase 2, *Journal of Biological Chemistry*, 279(37):39105–39114, 2004.
- [15] Lipniacki, T., Hat, B., Faeder, J. R., and Hlavacek, W. S., Stochastic effects and bistability in T cell receptor signaling, *Journal of theoretical biology*, 254(1):110–122, 2008.

CHAPTER 4. DATA-FREE IDENTIFIABILITY ANALYSIS OF BIOLOGICAL SYSTEMS

4.1 Introduction

Parameter estimation by means of data fitting has always been a critical step in the model building process. Even before applying rigorous estimation methods to a model to estimate the model parameters based on experimental data, one needs to verify whether the model parameters, or a subset of parameters, or some given quantities of interest are identifiable or can be constrained based on the measurements of output variables [1]. This step is usually referred to as system identifiability analysis.

The literature on identifiability analysis makes a distinction between two main types of analysis: *structural* and *practical identifiability*. Assuming an ideal context of error-free model structure and noise-free measurements, a system is said to be structurally identifiable if there do not exist two different parameter values that give rise to the same system outputs. Practical identifiability analysis, on the other hand, considers the issue of accurately estimating parameters from noisy data.

Since structural analysis is done without any actual experimental observation and addresses the question of determining a priori whether there is any chance of uniquely estimating model unknown parameters, it is also called prior identifiability analysis. Structural identifiability has generated huge interest in various fields of applied science, especially in the context of experimental design. Reviews about methods for structural identifiability analysis can be found in [1,2]. The development of widely-applicable methods for structural identifiability analysis for general non-linear dynamic models is still an active research problem. In general, there is no method amenable to every model, thus at some point we have to face the selection of one of several possible methods [1].

The limitations of the concept of structural identifiability come from its core assumptions that all system outputs are fully observed with no error. Naturally, such ideal assumptions cannot hold true in reality: in most applications in systems biology, not all state variables incorporated in a model can be measured directly; moreover, experimental data are usually noisy and insufficient considering the size of the model, and measurements might be available only at some specific region or time/length scales that are subject to randomness. Structural identifiability fails to capture the constraints one has on the experimental setting as well as the structure of the predictive noise, thus sometimes gives correct but misleading answers. As we will illustrate through examples, it may happen that although the system is theoretically identifiable, the features that distinguish different dynamics are not detectable due to technical limitations, or arise in a time/spatial scale that can not be captured by experiment. Since one of the main ideas behind structural analysis is to determine if a system is possibly identifiable by careful experimental design, pure theoretical identifiability analyses loses some utility when disassociated from experimental settings.

As an effort to address this issue, we explore the concept of *data-free identifiability*, which concerns the question of unique system identification under a given experimental setting, without actual experimental observation. As a data-independent property, data-free identifiability can be considered as a generalization of structural identifiability while at the same time addressing identifiability in the face of experimental constraints and noise.

With this novel concept, we propose a Bayesian approach to address system identifiability when data are not yet available. As we illustrate throughout this paper, our approach is global, strongly theoretically supported, amenable to high-dimensional cases, can be used to study various types of identifiability and is compatible with a large class of experimental settings. The framework is also built not only to assess parameter identifiability but also to quantify the uncertainty in prediction of any quantity of interest, and hence, can be used to address dynamics identifiability, a concept that has become of growing interest in the recent years [3, 4, 6]. This also

draws a direct connection between studies of identifiability and the concept of uncertainty quantification in predictive sciences. With this method, we also attempt to lay a unifying framework for the problems of structural/practical identifiability analysis, dynamics identifiability analysis and a priori uncertainty quantification.

The chapter is organized as follows. Section 4.2 provides the detailed algorithm for data-free identifiability analysis and introduces the method in various settings. We then investigate its performance and compare the results with those of previous approaches in the literature in Section 4.3. Theoretical results on the algorithm's convergence is provided in Section 4.4. Finally, we conclude the paper with some remarks, discussions and description of future work.

4.2 Data-free identifiability

4.2.1 Mathematical setting

To fix ideas, we consider biological systems that can be described by a set of ordinary differential equations (ODEs), although we note that the method can be extended to any continuous parametric system. The considered model is of the following form

$$\begin{aligned}\dot{x} &= \alpha(\omega, x, t) \\ x(0) &= x_0(\omega) \\ g &= g(\omega)\end{aligned}$$

where $t \in [0, T]$, the time interval where the system is observed; $x = (x_1, x_2, \dots, x_{n_x}) \in M$, a subset of \mathbb{R}^n containing the initial state; $\omega = (\omega^1, \omega^2, \dots, \omega^{n_\omega}) \in \Omega \subset \mathbb{R}^{n_\omega}$ is the parameter vector; $g = (g_1, g_2, \dots, g_{n_g}) \in \mathbb{R}^{n_g}$ is the vector of quantities of interest that needs to be predicted but cannot be accessed by experiments; α is a known continuously differentiable functions of its arguments.

In our framework, the quantity of interest g is either a known function of parameter ω or a black-box type function that can be evaluated point-wise. For the most part,

we focus on two special types of quantities of interest, namely, the parameter of the system and the dynamics of certain unobserved state variables. For convenience, we also denote by f_i the functions that map an parameter vector ω to its corresponding output x_i , i.e. for all $\omega \in \Omega$ and $t \in [0, T]$, $x_i(t) = f_i(\omega, t)$ will be its corresponding system output at the given time.

To investigate the experimental setting, we define the set of all theoretically possible measurements that we can make for inference as

$$\mathcal{A} = \{(i, t) : 1 \leq i \leq n_x, t \in [0, T]\}$$

where the pair (i, t) corresponds to a measurement collected of $x_i(t)$, the value of i^{th} -state variable x_i at time t . In practice, however, full observation of the system is not practical; for example, not all state variables incorporated in a model can be measured directly, or some may be available only at some specific region or time scales. In this paper, we denote by \mathcal{E} the set of all practically possible measurements that we can make for inference and refer to it as the *experimental constraint*.

Since data in practice is usually contaminated by noise and other type of uncertainties, throughout this paper, we denote the data obtained by measurements by $d(t) = x(t) + \varepsilon$ where ε is a random variable describing the noise in measurements. For simplicity, we assume that the noise in measurements are identically independently distributed Gaussian noise with known variance σ^2 . We note that other models of noise in measurements can also be considered without significant changes in either theoretical or computational aspect.

4.2.2 Uncertainty and identifiability

The major limitation of structural identifiability analysis is that normally, the analysis is done under the assumption that the full time course of certain (or more often, all) output dynamics can be obtained precisely ($\mathcal{E} = \mathcal{A}$ and $\sigma = 0$). While this assumption is general in theoretical studies of dynamical systems, it may not be considered as a natural assumption in practice.

To tackle this limitation, we need to consider the fact that in actuality, the number of data we can collect is finite and may even be smaller than the number of model parameters; moreover, the collected data may be severely contaminated by noise. However, by doing so, we introduce two types of uncertainty in our analysis, namely, the uncertainty that comes from lack of information (since we could not make enough measurements) and the uncertainty that comes from random noise. With the new uncertainties, we are no longer able to define “identifiability” as a yes/no property, and we need to define continuous quantities to quantify/measure system identifiability. As we will illustrate in later sections, one such measure is the uncertainty in prediction of the quantity of interest. This draws direct connections between studies of identifiability, the concept of uncertainty quantification in predictive sciences, and the theory of design of optimal experiments.

In the problem of data-free identifiability analysis, we ask ourselves the following question: if we are to make measurements under the practical experimental constraint \mathcal{E} and the noise level σ , i.e., $x_i(t)$ is known up to an amount of Gaussian noise with variance σ^2 for all $(i, t) \in \mathcal{E}$, can we identify the system parameters/quantities of interest? More precisely, as we discussed above, we aim to quantify the uncertainty in prediction of the quantities of interest, and check if any of the quantities can be estimated with low uncertainty.

Since generally we have various types of experimental constraints and can choose whether to include noise in the analysis, we end up with different types of experimental settings, each of which is interesting in its own way and has its own challenges. In the scope of our paper, we only focus on three types of identifiability:

1. $|\mathcal{E}| < \infty$ and $\sigma > 0$ (practical data-free identifiability)
2. $\mathcal{E} \neq \mathcal{A}$ and $\sigma = 0$ (constrained structural identifiability)
3. $\mathcal{E} = \mathcal{A}$ and $\sigma = 0$ (structural identifiability)

In the next section, we introduce a Bayesian framework for practical data-free identifiability analysis. We then consider the limit of that formulation when $|\mathcal{E}| \rightarrow \infty$ and $\sigma \rightarrow 0$ and achieve, separately, the other two types of identifiability.

In noise-free cases ($\sigma = 0$) where the concept of identifiability can be interpreted as a yes/no property, we have the following formal definition of constrained structural identifiability.

Definition 4.2.1 (Constrained structural identifiability ($\sigma = 0$)) *Given $\mathcal{E} \subset \mathcal{A}$, the system is said to be \mathcal{E} -identifiable if for any two parameter vectors ω^1 and ω^2 in the parameter space Ω , $f_i(\omega_1, t) = f_i(\omega_2, t), \forall (i, t) \in \mathcal{E}$ holds if and only if $\omega_1 = \omega_2$.*

We note that when $\mathcal{E} = \mathcal{A}$, we achieve the conventional concept of structural identifiability.

4.3 A unifying framework for data-free identifiability analysis and a priori uncertainty quantification

4.3.1 A Bayesian framework for practical data-free identifiability analysis ($|\mathcal{E}| < \infty$ and $\sigma > 0$)

Before moving forward to introduce our framework for identifiability analysis, we want to note that the problem of estimating uncertainty in model prediction in the presence of noise has been studied intensively in the context of experimental design of linear models [5]. The main assumptions of such analysis are that the model is linear and the random added noise is normal and independently identically distributed (i.i.d.). Another implicit assumption that comes with the condition of linearity is that when there are enough measurements, one can estimate uniquely the system parameters for every realization of data (i.e., the model is identifiable).

For linear models with additive i.i.d Gaussian noise, classical result implies that the uncertainty in estimation depends only on the experimental setting and not on data. Therefore, experiments can be designed to reduce uncertainty before measurements

are made. Such ideal conditions are no longer valid either when the model is non-linear [6] or the normal assumption is violated [7]. For a nonlinear model, such as those arising in systems biology, different realizations of data may lead to different values of uncertainty in prediction, and one needs to take into account the effect of data on uncertainty.

Another issue with biological systems is non-identifiability [8]: given a realization of data, there might be several or even an infinite set of parameter values that can fit the data equally well. In order to make reasonable predictions of the quantity of interest and address identifiability, one needs to either keep track of all parameter configurations that are consistent with data or use a probabilistic framework to assign the likelihood of each parameter configuration conditioned on the given realization of data.

In our framework, given an experimental constraint $\mathcal{E} = \{(i_1, t_1), (i_2, t_2), \dots, (i_k, t_k)\}$ and level of noise σ we define the probability distribution $\pi(\omega_1, \omega_2, \varepsilon)$ as the likelihood of ω_1 being the estimated parameter values if data is generated by the "true" parameter ω_2 and contaminated by the random noise ε , i.e. $d_{i_j}(t_i) = f_{i_j}(\omega_2, t_i) + \varepsilon_j$. While this probability cannot be written analytically, its conditional distributions with respect to each variable ω_1 , ω_2 and ε can be described by

$$\begin{aligned} \varepsilon | \omega_1, \omega_2 &\propto \mathcal{N}(0, \sigma^2) \\ \omega_1 | \omega_2, \varepsilon &\propto \exp \left(-\frac{1}{2\sigma^2} \sum_{j=1}^k |f_{i_j}(t_j, \omega_1) - (f_{i_j}(t_j, \omega_2) + \varepsilon_j)|^2 \right) \\ \omega_2 | \omega_1, \varepsilon &\propto \exp \left(-\frac{1}{2\sigma^2} \sum_{j=1}^k |f_{i_j}(t_j, \omega_1) - (f_{i_j}(t_j, \omega_2) + \varepsilon_j)|^2 \right) \pi_0(\omega_2) \end{aligned}$$

where $\pi_0(\omega_2)$ is the prior distribution of ω_2 .

Here, we note that the formula for $\pi(\omega_2 | \omega_1, \varepsilon)$ is derived by the Bayes formula

$$\pi(\omega_2 | \omega_1, \varepsilon) \propto \pi(\omega_1 | \omega_2, \varepsilon) \pi_0(\omega_2 | \varepsilon)$$

and the fact that ε is independent of ω_2 .

The expected uncertainty in prediction of the quantity of interest g is represented by the variance-covariance matrix

$$U_\pi(g) = \text{Cov}_{\pi(\omega_1, \omega_2)}[g_1(\omega_1) - g_1(\omega_2), g_2(\omega_1) - g_2(\omega_2), \dots, g_{n_g}(\omega_1) - g_{n_g}(\omega_2)]$$

while the identifiability index of g is defined as the inverse of the variance-covariance matrix

$$I_\pi(g) = (U_\pi(g))^{-1}.$$

It is worth noting that when the f 's are linear functions and the quantities of interest are model parameters ($g(\omega) = \omega$), $U(g)$ coincides with the well-known variance-covariance matrix of the maximum likelihood estimator in the context of experimental design. In this case, $I(g)$ also acts effectively as an upper bound for and is asymptotically (when number of data points approach infinity) equal to the Fisher Information.

4.3.2 A reinterpretation of structural identifiability

As a data-independent property, data-free identifiability can be considered as a generalization of structural identifiability while at the same time addressing identifiability in the face of experimental constraints and noise. In this section, we illustrate that by using the framework for practical data-free identifiability proposed in the previous section, we can prove that in the limit when the number of data goes to infinity and when the measurement error approaches 0, we achieve the traditional structural identifiability and a more general type of structural identifiability with constrained experimental setting that we dub *constrained structural identifiability*. This provides a new interpretation of traditional structural identifiability.

Structural identifiability($\mathcal{E} = \mathcal{A}$ and $\sigma = 0$)

We first recall the concept of structural identifiability.

Definition 4.3.1 *Structural identifiability: A system structure is said to be identifiable if for any two parameter vectors ω^1 and ω^2 in the parameter space Ω , $f_i(\omega_1, t) = f_i(\omega_2, t), \forall t \in [0, T], 1 \leq i \leq n_x$ holds if and only if $\omega_1 = \omega_2$.*

We can see from the definition that structural identifiability are just data-free identifiability when $\mathcal{E} = \mathcal{A}$ and $\sigma = 0$. Moreover, as we illustrate below, structural identifiability can be regarded as the limit of practical data-free identifiability when the number of data points goes to infinity.

Indeed, consider a hypothetical infinite experiment $\{(i_1, t_1), (i_2, t_2), \dots, (i_k, t_k), \dots\} \in \mathcal{A}$ such that the sequence becomes dense in \mathcal{A} in the limit as $k \rightarrow \infty$. Following the framework proposed previously for fixed k and $\sigma = 0$, the posterior distribution with uniform prior of ω_2 given data is

$$\pi_k(\omega^1, \omega^2) \propto \prod_{j=1}^k 1_{f_{i_j}(t_j, \omega_1) = f_{i_j}(t_j, \omega_2)}.$$

We then have the following theorem

Theorem 4.3.1 *Denote $\Theta = \Omega \times \Omega = \{(\omega^1, \omega^2) : \omega^1, \omega^2 \in \Omega\}$ and consider the following probability distributions on Θ :*

$$\pi_k(\omega^1, \omega^2) \propto \prod_{j=1}^k 1_{f_{i_j}(t_j, \omega_1) = f_{i_j}(t_j, \omega_2)}$$

Then i^{th} -parameter of the model is identifiable if and only if

$$\lim_{k \rightarrow \infty} U_{\pi_k}(\omega^i) = 0$$

Proof: We note that the support of π_k is just the set of all (ω_1, ω_2) that have the same output dynamics at $\{t_1, t_2, \dots, t_k\}$. By assumption, $\{t_1, t_2, \dots, t_k, \dots\}$ is dense in $[0, T]$, which implies that the support of π_k is contained the set of all pairs of parameters (ω_1, ω_2) with $f_{i_j}(t_j, \omega_1) = f_{i_j}(t_j, \omega_2) \forall j$. Intuitively, if there is such a pair (ω_1, ω_2) with $\omega_1 \neq \omega_2$ that satisfies this condition, then the variance with respect to

π_k of $\|\omega^1 - \omega^2\|$ will be bounded away from zero. Similarly, if the i^{th} parameter is unidentifiable, the variance of $|\omega_1^i - \omega_2^i|$ will also be bounded away from zero.

This theorem, however, has practically no use, since the computation of π_k corresponds to the task of identifying all pairs of parameter with identical dynamics, and the form of the distribution makes it impossible to use any practical method to sample from π_k .

We then propose an alternative form of the distribution of π_k as follows

$$\hat{\pi}_k(\omega^1, \omega^2) \propto \exp \left(-\frac{1}{\sqrt{k}} \sum_{j=1}^k |f_{i_j}(t_j, \omega_1) - f_{i_j}(t_j, \omega_2)|^2 \right)$$

In some sense, our method is directly related to the idea behind simulated annealing and multiple heated chains: the object $\{(\omega_1, \omega_2) : f_{i_j}(t_j, \omega_1) = f_{i_j}(t_j, \omega_2) \ \forall j\}$ is difficult to identify, so we relax it by a heated object (characterized by the probability distribution $\hat{\pi}_k$) that is easier to study.

It is worth noting that since the support of $\hat{\pi}_k$ is no longer the set of all pairs of parameter with identical dynamics, the uncertainty estimated using this approximated distribution is an overestimation of the uncertainty. However, despite the approximate nature of this alternative distribution, we still have the following theorem which guarantees that this approximation of uncertainty is equivalent to the actual model uncertainty.

Theorem 4.3.2 (Structural identifiability (or equivalently, \mathcal{A} -identifiability))

Consider the following probability distributions on $\Omega \times \Omega$:

$$\hat{\pi}_k(\omega^1, \omega^2) \propto \exp \left(-\frac{1}{\sqrt{k}} \sum_{j=1}^k |f_{i_j}(t_j, \omega_1) - f_{i_j}(t_j, \omega_2)|^2 \right)$$

and approximate $U_{\pi_k}(\omega^i)$ by

$$U_{\hat{\pi}_k}(\omega^i) = \lim_{n \rightarrow \infty} \text{Var}_{\pi_n(\omega^1, \omega^2)}[\omega_i^1 - \omega_i^2]$$

Suppose the i^{th} parameter of the model is identifiable. Then

$$\lim_{k \rightarrow \infty} U_{\hat{\pi}_k}(\omega^i) = 0$$

Moreover, if Ω is finite, the converse is also true.

Theorem 4.3.2 provides a criteria for inference on the identifiability of each parameter of the system and provides a probabilistic computational method to address the problem of a priori identifiability analysis. The main part of the computational process will be evaluating the variance of $(\omega_1^i - \omega_2^i)$ with respect to π_k , which can be done by generating a Monte-Carlo Markov Chain with π_k as the invariant measure. The convergence rate of MCMC is relatively insensitive to dimension, and the method can be employed to study high-dimensional systems.

Constrained structural identifiability($\mathcal{E} \neq \mathcal{A}$ and $\sigma = 0$)

The arguments in the previous subsection also extend to the case when $\mathcal{E} \neq \mathcal{A}$. For a given experiment setting \mathcal{E} , we define a (theoretical) continuous distinguishability function $D_{\mathcal{E}}(\omega_1, \omega_2)$ with respect to \mathcal{E} and assume that there exists a sequence of functions $D_k(\omega_1, \omega_2)$ such that

1.

$$D_{\mathcal{E}}(\omega_1, \omega_2) = D_{\mathcal{E}}(\omega_2, \omega_1), \quad \forall (\omega_1, \omega_2) \in \Omega \times \Omega \quad (4.1)$$

2.

$$D_{\mathcal{E}}(\omega_1, \omega_2) \geq 0, \quad \forall (\omega_1, \omega_2) \in \Omega \times \Omega \quad (4.2)$$

3.

$$D_{\mathcal{E}}(\omega_1, \omega_2) = 0 \quad \Leftrightarrow \quad f_i(\omega_1, t) = f_i(\omega_2, t) \quad \forall (i, t) \in \mathcal{E} \quad (4.3)$$

4.

$$\left| \frac{D_k(\omega_1, \omega_2)}{k} - D_{\mathcal{E}}(\omega_1, \omega_2) \right| = O\left(\frac{1}{\sqrt{k}}\right) \quad \text{uniformly in } (\omega_1, \omega_2) \in \Omega \times \Omega \quad (4.4)$$

We note that the existence of the continuous distinguishability function for any experimental constraint is guaranteed by the following theorem

Theorem 4.3.3 *Assume that \mathcal{E} is a non-empty subset of \mathcal{A} , then there exist continuous functions $D_{\mathcal{E}}$ satisfies (4.1)-(4.3).*

We then have the following result

Theorem 4.3.4 (Constrained data-free identifiability) *Given the experimental constraint \mathcal{E} and the distinguishability function $D_{\mathcal{E}}(\omega^1, \omega^2)$. Consider the following distribution*

$$\hat{\pi}_k(\omega^1, \omega^2) \propto \exp\left(-\frac{D_k}{\sqrt{k}}\right)$$

where D_k satisfies (4.4).

Suppose the i^{th} parameter of the model is \mathcal{E} -identifiable. Then

$$\lim_{k \rightarrow \infty} U_{\hat{\pi}_k}(\omega^i) = 0$$

If Ω is finite, the converse is also true.

The proofs of Theorem 4.3.4 and Theorem 4.3.3 are given in section 5.

4.3.3 Dynamics identifiability and a priori uncertainty quantification

For high-dimensional and complex biological systems, unidentifiability is somewhat expected. Therefore, some recent research on biological systems has shifted from parameter identification to identifying a quantity of interest (in most cases, some observable/ unobservable output dynamics). The most straightforward example of this phenomenon is when the model parameterization is redundant: some parameters are totally insensitive to the system output and measurements may not contain enough information to effectively identify the parameter. Despite this, one does not really need to estimate such redundant parameters to gain understanding about some output dynamics which might be well-constrained by the data.

This gives rise to the problem of *a priori* uncertainty quantification, that is, by inferences on available data, one wishes to forwardly propagate the uncertainty in parameters to the output space to test if certain quantities of interest can be effectively

constrained. This leads us to the concept of the \mathcal{E} -identifiability of a given quantity of interest.

Definition 4.3.2 *Given an experimental constraint \mathcal{E} , a quantity of interest $g(\omega)$ is said to be \mathcal{E} -identifiable if for any two parameter vectors ω_1 and ω_2 in the parameter space Ω , $g(\omega_1) = g(\omega_2)$ holds if and only if $D_{\mathcal{E}}(\omega_1, \omega_2) = 0$.*

We then have the following theorem

Theorem 4.3.5 *Given the experimental constraint \mathcal{E} and the corresponding distinguishability function $D_{\mathcal{E}}$. Consider the following distribution*

$$\pi_k(\omega_1, \omega_2) \propto \exp\left(-\frac{D_k}{\sqrt{k}}\right)$$

where D_k satisfies (4.2).

Suppose the quantity of interest $g(\omega)$ is identifiable then we have

$$\lim_{n \rightarrow \infty} U_{\pi_n}(g(\omega)) = 0$$

If Ω is finite, the converse is also true.

We note that the framework built in this section does not require the quantities of interest to be the dynamics outputs of an ODEs systems, and in fact can be used to quantify the uncertainty in prediction of any quantities of interest.

4.3.4 Computational procedure

In summary, the computational procedure to investigate all types of identifiability can be outlined as follows:

Step 1. Define problem: determine the parameter space (Ω) and type of analysis

- The quantity of interest g (e.g, parameter/dynamics)
- The experimental constraints \mathcal{E}

– The noise structure ϵ (in the normal cases, the standard deviation σ)

- Step 2.** Design the distinguishability function $D_{\mathcal{E}}$, as well as the mathematical formulation of the distribution π_n on Θ .
- Step 3.** Generate a Monte Carlo Markov Chain $\{\theta_k\}$ with $\pi_k(\omega_1, \omega_2, \epsilon)$ as the invariant distribution.
- Step 4.** Compute the variance with respect to π_k of $(g(\omega_1) - g(\omega_2))$ and use this as the criteria to assess \mathcal{E} -identifiability (via the uncertainty in prediction $U_{\pi}(g)$) of the quantity of interest.

The probabilistic framework for identifiability analysis proposed in this chapter is made possible by the employment of Monte Carlo Markov Chains method to sample from a likelihood function on some high-dimensional parameter spaces. Since direct computation of the likelihood function are costly, in practice, approximation methods are usually employed to reduce some of the computational burden.

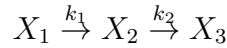
Throughout this work, we use the Griddy Gibbs sampling [12] as an effective way to sample from the likelihood functions of interest.

4.4 Results

This section proceeds in an increasing order of complexity. In the first step, we provide insight into the method via studies of practical data-free identifiability and structural identifiability. Then the method is twisted to tackle the problem of identifiability in more general experiment settings. Finally, the question of dynamics identifiability and a priori uncertainty quantification are also investigated.

4.4.1 An intuitive example

To illustrate ideas, we consider the following toy example of a simple biochemical system that contains 3 chemicals:



where k_1 and k_2 are the (unknown) degradation rates of X_1 and X_2 , respectively. We also assume that at the beginning, the system contains only X_1 .

We model this system using

$$\begin{aligned} \frac{dx_1}{dt} &= -\omega_1 x_1, & \frac{dx_2}{dt} &= \omega_1 x_1 - (\omega_2 + \omega_3) x_2, & \frac{dx_3}{dt} &= (\omega_2 + \omega_3) x_2, \\ (x_1(0), x_2(0), x_3(0)) &= (1, 0, 0). \end{aligned}$$

Note that the parameterization of the system is designed to be redundant since ω_2 and ω_3 appear only as $\omega_2 + \omega_3$; this makes the system theoretically unidentifiable.

In this particular example, the parameter space is $[0.1, 10] \times [0.1, 10] \times [0.1, 10]$, the time interval is $[0, 10]$ (seconds) and is converted to log space for convenience. The measurements of the concentration of X_2 are to be made at two different time points $t = 1s$ and $t = 3s$, contaminated by independent Gaussian noise of standard deviation $\sigma = 0.01$. We assume that at the time the analysis is performed, measurement at $t = 1s$ and $t = 3s$ are not yet available. The analysis in this case, therefore, is a data-free identifiability analysis.

We then employ the framework proposed above to assess the identifiability of (i) the model parameters $\omega^1, \omega^2, \omega^3$, and (ii) the concentration of $x_1(t)$, $x_2(t)$, $x_3(t)$ for all $t \in [0, T]$. For simplicity, we consider g as separate one-dimensional quantities of interest and ignore the correlation between them. The result is provided in Figure 4.1, where the individual uncertainty in prediction of ω_i 's and $x_i(t)$'s are plotted. The uncertainty is presented by variance in prediction of parameter/dynamics by a sample of 10^5 -point Monte Carlo Markov chain from the joint distribution of ω_1

and ω_2 as in [citation] by the Griddy Gibbs sampling. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable.

In this example, all system parameters are practically unidentifiable. The systems dynamics also could not be predicted with high accuracy, except for the part of the second state variable X_2 after 1s. (Note that the identifiability of X_2 after 1s is somewhat expected, since X_2 is the observable output of the system). We also provide in Figure 4.2 the dynamics corresponding to very different parameter configurations that give rise to indistinguishable outputs at times $t = 1s$ and $t = 3s$. It is also worth noting that with this system, some data realizations of the outputs at $t = 1s$ and $t = 3s$ do lead to parameter identifiability, while some data realizations do not. This also highlights the fact that different realizations of data in a nonlinear model may lead to different values of uncertainty in prediction.

4.4.2 A model of influenza A virus infection

We perform structural identifiability (\mathcal{A} -identifiability) analysis on a model of influenza A virus infection, proposed by Baccam et al. in [9].

$$\begin{aligned}\frac{\partial T}{\partial t} &= -\beta TV \\ \frac{\partial I}{\partial t} &= \beta TV - \delta I \\ \frac{\partial V}{\partial t} &= cI - pV\end{aligned}$$

where T is the number of uninfected target cells (epithelial cells), I is the number of productively infected cells, and V is the infectious viral titer expressed in TCID50/ml which is the only state variable to be measured.

This model was previously analyzed in [2] using the implicit function method. In the case when only V is measured, it is reported that β, δ, c are identifiable with

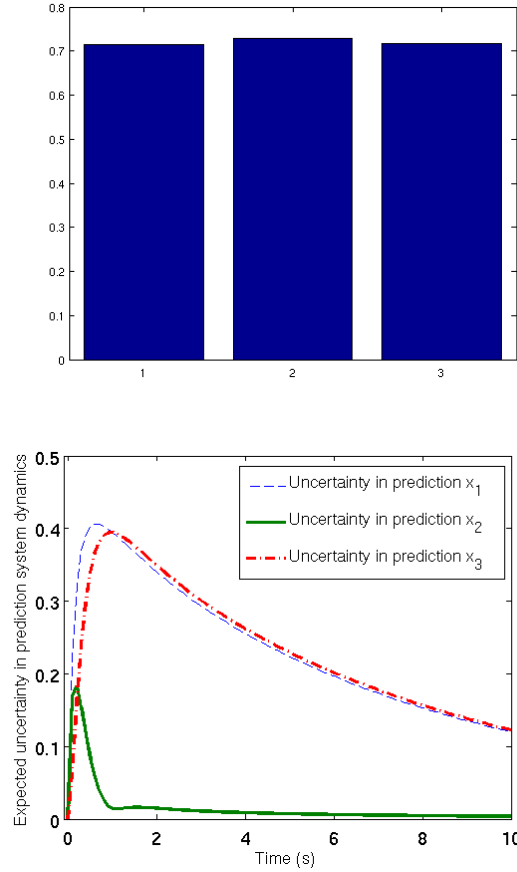


Figure 4.1. Uncertainty in identifying model parameters and dynamics: Experiments are to be made for $x_2(1)$ and $x_2(3)$, where data contains noise of standard deviation $\sigma = 0.01$. (Top) Identifiability of model parameters: $g_1(q) = q_1$, $g_2(q) = q_2$, $g_3(q) = q_3$. (Bottom) Identifiability of model dynamics: $g_{k,t}(q) = x_k(t)$, $k = 1, \dots, 3$, $t \in [0, T]$. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable.

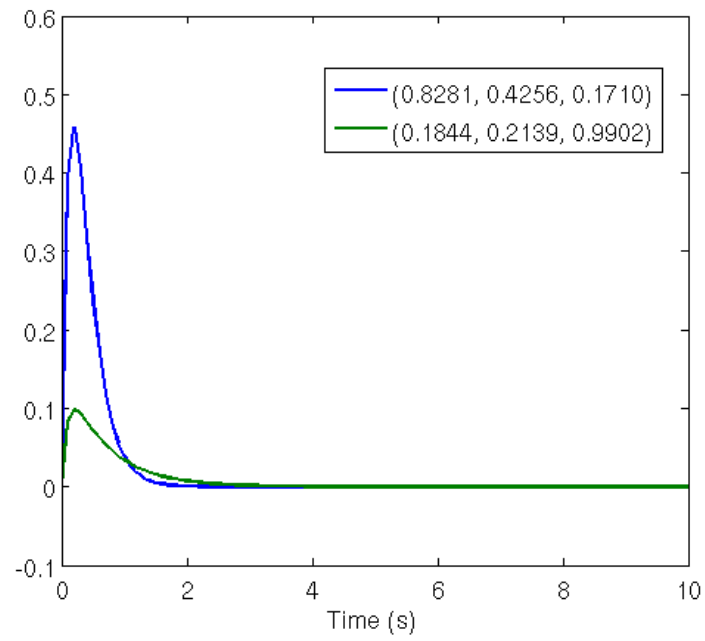


Figure 4.2. The dynamics of $x_2(t)$ with two different parameter configurations that have same output at $t = 1s$ and $t = 3s$.

a minimum of 6 measurements (that is, V need to be measured at at least 6 time points), and p is unidentifiable.

We perform our identifiability analysis on this model, using the L^2 -distance between the output V 's as the distinguishability measure, and $D_n = nD$. $U_\pi(\beta)$, $U_\pi(\delta)$, $U_\pi(c)$ and $U_\pi(p)$ are presented in Figure 4.3A. Notice that since the range of parameter in log-scale is $[-1, 1]$, our analysis indicates that with just V being measured, only c is identifiable while the others are not. Moreover, since this is a low-dimensional nonlinear dynamical system, the generated MCMC also proposes good candidates for pairs of parameter sets with similar dynamics. In Figure 4.3B, we plotted the dynamics corresponding to two different parameter sets with indistinguishable dynamics. The figure clearly indicates that β is unidentifiable. Similar examples with indistinguishable dynamics for significantly different δ can also be obtained.

It is worth noting that there are many possible explanations of the discrepancy between our analysis and those of [2], the most likely of which is that the features that distinguish different dynamics of the system are not detectable due to technical limitations, or arise in a different time scale that can not be captured by experiment. From a theoretical point of view, the structural identifiability analysis in [2] is more rigorous; while in terms of experimental design, our a priori identifiability analysis provide more accurate insight about the system.

4.4.3 Analysis of Goodwin's model

In this example, we study the structural identifiability of Goodwin's model [10]. The state variable x_1 represents an enzyme concentration whose rate of synthesis is regulated by feedback control via a metabolite x_3 while x_2 regulates the synthesis of x_1 . The model includes rational kinetics consisting of a Hill-like term, and is given by the following system of ODEs:

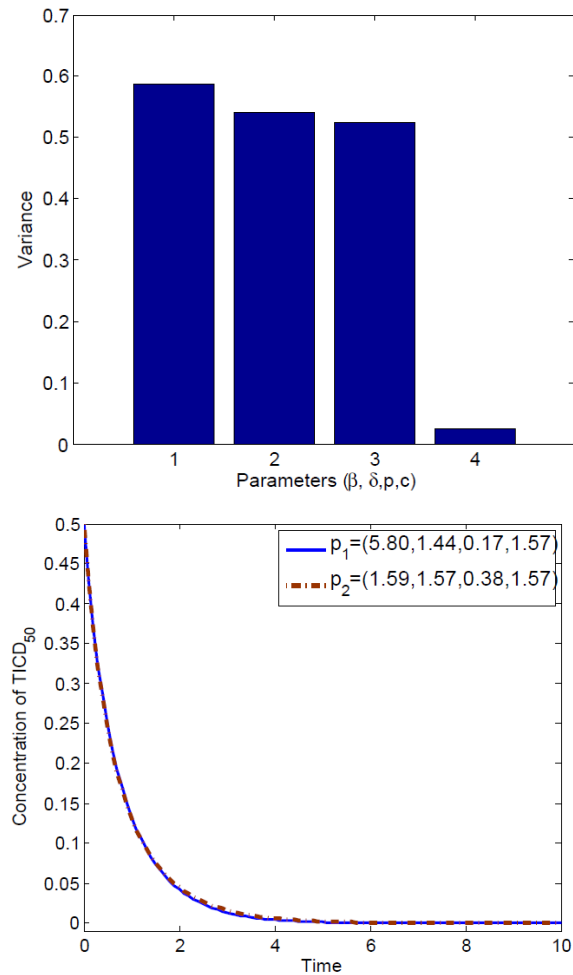


Figure 4.3. Structural identifiability analysis of Baccam's model. (Top) Uncertainty in identifying model parameters. (Bottom) Two different parameters set with indistinguishable dynamics. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable.

$$\begin{aligned}
\dot{x}_1 &= -bx_1 + \frac{a}{A + x_3^\eta} \\
\dot{x}_2 &= \alpha x_1 - \beta x_2 \\
\dot{x}_3 &= \gamma x_2 - \delta x_3 \\
x_1(0) &= 0.3617; \quad x_2(0) = 0.9137 \quad x_3(0) = 1.3934
\end{aligned}$$

As in the previous example, the parameter space is defined relative to a nominal parameter vector: for each component of the nominal vector, we define an uncertainty range of 10 times smaller to 10 times larger than this component. The time interval is $[0, 100]$ (minutes).

In the situation for which all states x_1, x_2, x_3 can be measured, a review on performances of current structural identifiability analysis methods applied to Goodwin's model can be found in [1]. The results can be summarized as follows:

1. The similarity transformation approach could not be applied.
2. The Taylor and generating series approaches suggest that Goodwin model is structurally locally identifiable.
3. The differential algebra approach, as implemented in DAISY, results in the non-identifiability of the model. No results about local identifiability were reported.
4. The method based on the implicit function theorem with fixed η indicates that the remaining parameters are structurally locally identifiable provided $\eta > 2$.
5. The dynamic reaction networks that fixes both η and A derives the structural local identifiability of the remaining parameters.

In the analysis of the model, we use the summation of the L^2 -distance between the output x_i 's as the distinguishability measure, with the empirical discrepancy $D_n = nD$. The parameter space is defined relative to a nominal parameter vector

selected from [10] . That is, for each component of the nominal vector, we define a range of 10 times smaller to 10 times larger than this component.

The uncertainty in prediction of ω^i is presented by variance in prediction of parameter/dynamics by a sample of 10^5 -point Monte Carlo Markov chain from the joint distribution of ω_1 and ω_2 as in [citation] by the Griddy Gibbs sampling. The uncertainty in prediction of ω^i is presented in Figure 4.4A. Since the range of parameter in log-scale is $[-1, 1]$, the result indicates that with the full system outputs measured, β and δ are identifiable, but σ is not. Similar to the previous example, we use the generated MCMC to propose candidates for pair of parameter sets with similar dynamics, one of which is presented in Figure 4.4B. We also perform the same analysis for the case $\sigma = 3$. The results still suggests that α and δ are identifiable, while the rest are not.

This example also highlights a difference between our approach and other current methods of a priori identifiability analysis. While most methods focus on the local properties of the system, our probabilistic approaches concentrate more on global identifiability within a certain parameter space.

4.4.4 A model of the T-cell signalling pathway

Our next example is a mathematical model of the T-cell signaling pathway proposed by Lipniacki et al. in [11]. This is a system of ODEs with 37 state variables, 19 parameters, and fixed initial conditions. As in the previous example, the parameter space is defined relative to a nominal parameter vector: for each component of the nominal vector, we define a range of 10 times smaller to 10 times larger than this component. The time interval is $[0, 100]$ (minutes).

To simplify the analysis, we first restrict our attention to the identifiability of the five most sensitive parameters: pSHP binding rate to TCR complex (ly_1), LCK dephosphorylation rate (ls_1), TCR phosphorylation rate (t_p), ZAP spontaneous phosphorylation rate (z_0) and ERK phosphorylation rate (e_1). The entire time course of

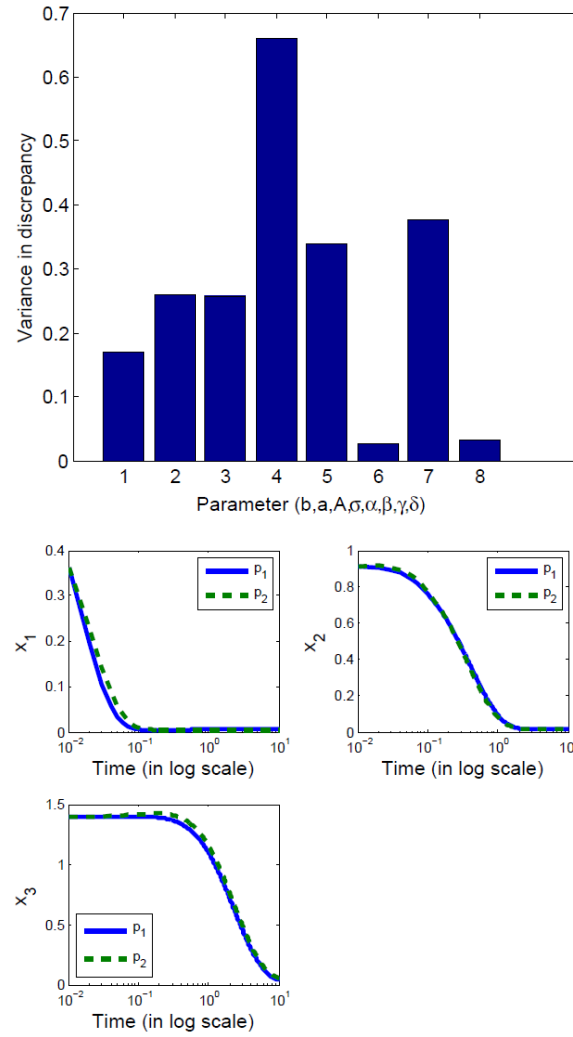


Figure 4.4. Identifiability analysis of Goodwin's model: (Top) Uncertainty in identifying model parameters; Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable. (Bottom) Two different parameters set with indistinguishable dynamics

the concentration of pERK ($x_{36} + x_{37}$) and pZAP (x_{31}) is used as the experimental constraint \mathcal{E} for constrained structural identifiability analysis of the five parameters. The identifiability of ω^i is presented in Figure 4.5A, which implies the identifiability of the third and the fifth parameters and the unidentifiability of other parameters.

The uncertainty in prediction different state variables ($U_\pi(x_i(t))$) are plotted in 4.5B. The result is consistent with biological insights about the model. Since the downstream component of the signaling pathway starts with pZap (x_{34}) and ends with pERK and ppERK (x_{36}, x_{37}) and the phosphorylation/dephosphorylation rate of MEK/pMEK are assumed to be known, understanding about the time courses of those pZap and pERK+ppERK helps identify the dynamics of the substrates in between, including ppMEK x_{34} . LCK (x_{27}) can be partially identified because of two reasons: (1) the dynamics of pZap (x_{31}) is completely determined by the LCK-membrane complexes x_9, x_{10} and the interactions with the downstream component (2) the dynamics of LCK is controlled by the creation and degradation of LCK-membrane complexes, which in turn, are strongly influenced by a feedback from ppERK. The dynamics of free TCR, however, also depends strongly on the dynamics of another protein, SHP, and cannot be identified by our experiment setting.

The figure also indicates that the early time scale of the system is more sensitive to uncertainty in parameters. This suggests that further and more careful experiment design in this early time scale can help identify the parameter with more accuracy, yet inference using the estimation of the dynamics at this time should be done with more caution.

In the second instance of this example, we consider a more practical setting when the level of phosphorylated ERK ($x_{36} + x_{37}$) is to be measured at 10 minute, 50 minute and 100 minute while a single measurement at 10 minute of pZap is to be obtained. Since this experiment setting is strongly under-determined, it is not surprising that all model parameters are somewhat unidentifiable (Figure 4.6A). Similarly, due to lack of knowledge of other important parameters, the dynamics of pMEK, LCK and TCR could not be constrained (Figure 4.6B). However, the figure suggests that the

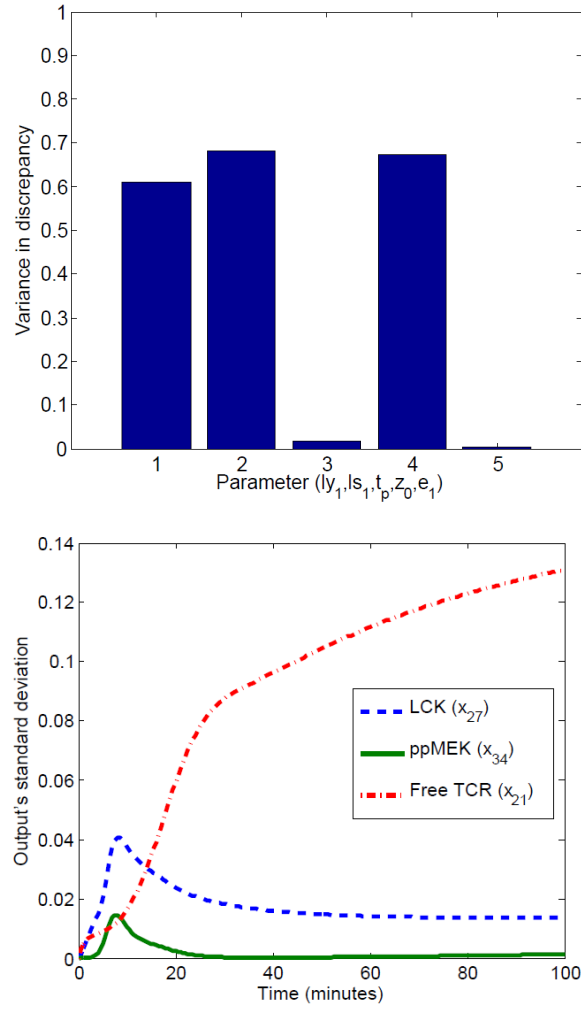


Figure 4.5. Analysis of Lipniack's model: (Top) Variance in predicting ω^i ; (Bottom) Variance in predicting different state variables. Notice that since the range of parameter in log-scale is $[-1, 1]$, an uncertainty index around 0.5 corresponds to highly unidentifiability, while a quantity with uncertainty index of order 10^{-2} is considered to be highly identifiable.

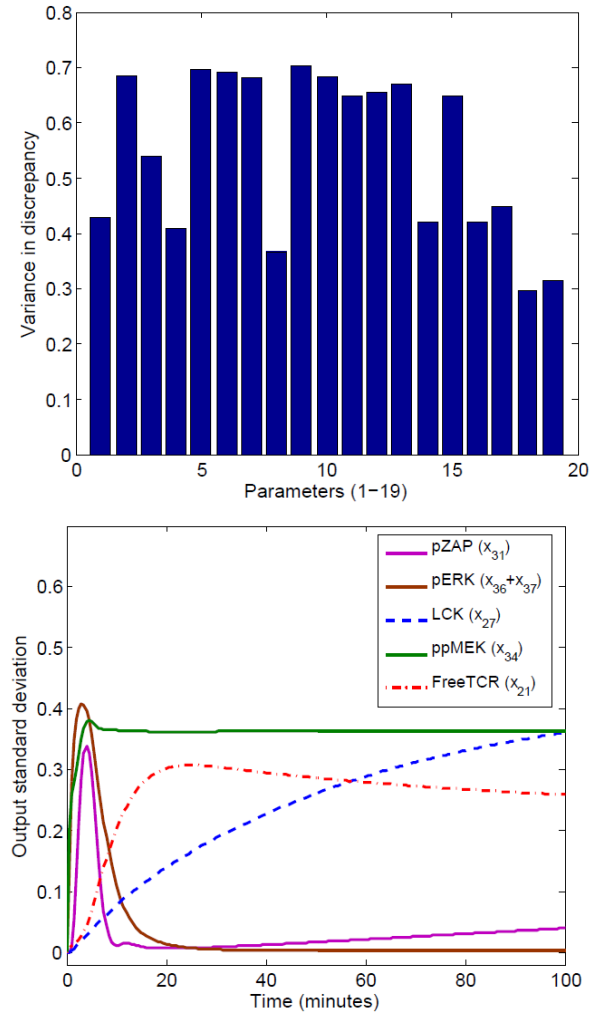


Figure 4.6. Analysis of Lipniack's model

combination of a few measurements of phosphorylated ERK and pZap can help identify the dynamics of *both* of them. From an experimental design point of view, an additional measurement of either phosphorylated ERK or pZap at around 5 minutes needs to be made for a complete study of the time courses of both substrates.

To the best of our knowledge, such an analysis of under-determined system in a general experiment setting can not be obtained by any previous approach to identifiability analysis.

4.5 Convergence analysis

Proof of the Theorem 4.3.3

Proof We consider two separate cases

1. \mathcal{E} is finite.

Assume that $\mathcal{E} = \{(i_1, t_1), (i_2, t_2), \dots, (i_N, t_N)\}$, then by defining

$$D_{\mathcal{E}} = \sum_{j=1}^N |f_{i_j}(\omega_1, t_j) - f_{i_j}(\omega_2, t_j)|^2$$

and $D_n = nD_{\mathcal{E}}$, we can easily check that $D_{\mathcal{E}}$ and D_k satisfy all the required condition.

2. \mathcal{E} is infinite.

Since \mathcal{E} is infinite, there exist an infinite sequence $E = \{(i_1, t_1), (i_2, t_2), \dots, (i_n, t_n), \dots\} \in \mathcal{E}$ that is dense in \mathcal{E} . Denote

$$D_{\mathcal{E}}(\omega_1, \omega_2) = \sum_{j=1}^{\infty} \frac{1}{2^j} |f_{i_j}(\omega_1, t_j) - f_{i_j}(\omega_2, t_j)|^2$$

Since f 's are continuous, we deduce that $D_{\mathcal{E}}$ is well-defined, symmetric and non-negative. Moreover, $D_{\mathcal{E}}(\omega_1, \omega_2) = 0$ if and only if $f_{i_j}(\omega_1, t_j) = f_{i_j}(\omega_2, t_j) \forall j$. Since E is dense in \mathcal{E} and f 's are continuous, this is equivalent to $f_i(\omega_1, t) = f_i(\omega_2, t) \forall (t, i) \in \mathcal{E}$.

■

Proof of the Theorem 4.3.4

Proof Denote $q_n(\omega_1, \omega_2) = \exp\left(-\frac{1}{\sqrt{n}}D_n(\omega_1, \omega_2)\right)$ and

$$r_n(\omega_1, \omega_2) = \exp\left(-\frac{1}{\sqrt{n}}\frac{D_n(\omega_1, \omega_2)}{n}\right)$$

For any $\epsilon > 0$, there exists $C = C(\epsilon/4)$ such that if $n \geq C(\epsilon/4)$, we have

$$\left| \frac{D_n(\omega_1, \omega_2)}{n} - D_{\mathcal{E}}(\omega_1, \omega_2) \right| \leq \epsilon/4 \quad \forall \omega_1, \omega_2 \in \Omega. \quad (4.5)$$

Denote

$$V_\epsilon = \{(\omega_1, \omega_2) \in \Omega \times \Omega : D_{\mathcal{E}}(\omega_1, \omega_2) \leq \epsilon\}$$

Note that since $(\omega, \omega) \in V_\epsilon \forall \omega \in \Omega, \epsilon > 0$, we deduce that $V_\epsilon \neq \emptyset$ for all $\epsilon > 0$.

For $(\omega_1, \omega_2) \in V_{\epsilon/4}$, we deduce from (4.5) that $\frac{D_n(\omega_1, \omega_2)}{n} \leq \epsilon/2$ if $n \geq C(\epsilon/4)$.

Therefore, for each n

$$\begin{aligned} \|r_n\|_n &= \left(\int_{\Omega \times \Omega} |r_n(\omega_1, \omega_2)|^n \right)^{1/n} \geq \left(\int_{V_{\epsilon/4}} |r_n(\omega_1, \omega_2)|^n \right)^{1/n} \\ &\geq \exp\left(-\frac{\epsilon}{2\sqrt{n}}\right) [\text{Vol}(V_{\epsilon/4})]^{1/n} \end{aligned} \quad (4.6)$$

Now consider $(\omega_1, \omega_2) \notin V_\epsilon$ with $n \geq C(\epsilon/4)$ and note that in this case $\frac{D_n(\omega_1, \omega_2)}{n} > 3\epsilon/4$, we have

$$\pi_n(\omega_1, \omega_2) = \frac{q_n(\omega_1, \omega_2)}{\int_{\Omega \times \Omega} q_n(u_1, u_2) du_1 du_2} = \left(\frac{r_n}{\|r_n\|_n} \right)^n \quad (4.7)$$

On the other hand, from (4.6) and the fact that $r_n(\omega_1, \omega_2) \leq \exp(-\frac{3\epsilon}{4\sqrt{n}})$, we get

$$\frac{r_n}{\|r_n\|_n} \leq \frac{\exp(-\frac{3\epsilon}{4\sqrt{n}})}{\exp\left(-\frac{\epsilon}{2\sqrt{n}}\right) [\text{Vol}(V_{\epsilon/4})]^{1/n}} \leq \frac{\exp(-\frac{\epsilon}{4\sqrt{n}})}{[\text{Vol}(V_{\epsilon/4})]^{1/n}} \quad (4.8)$$

Therefore

$$\pi_n(\omega_1, \omega_2) \leq \frac{\exp(-\sqrt{n}\epsilon/4)}{\text{Vol}(V_{\epsilon/4})} \quad (4.9)$$

On the other hand, by the symmetry of the distinguishability function, we see that π_n is also symmetric

$$E_{\pi_n(\omega^1, \omega^2)}[\omega_i^1 - \omega_i^2] = 0.$$

We conclude that

$$\begin{aligned} \text{Var}_{\pi_n(\omega^1, \omega^2)}[\omega_i^1 - \omega_i^2] &= E_{\pi_n(\omega^1, \omega^2)}[\omega_i^1 - \omega_i^2]^2 \\ &= \int_{V_\epsilon} |\omega_1^i - \omega_2^i|^2 \pi_n(\omega_1, \omega_2) + \int_{\Omega \times \Omega - V_\epsilon} |\omega_1^i - \omega_2^i|^2 \pi_n(\omega_1, \omega_2) \end{aligned}$$

$$\leq \max_{V_\epsilon} |\omega_i^1 - \omega_i^2|^2 + \frac{\exp(-\sqrt{n}\epsilon/4)}{\text{Vol}(V_{\epsilon/4})} \int_{\Omega \times \Omega - V_\epsilon} |\omega_1^i - \omega_2^i|^2$$

Fixing ϵ and taking $n \rightarrow \infty$ gives

$$\lim_{n \rightarrow \infty} \text{Var}_{\pi_n(\omega_1, \omega_2)} [\omega_1^i - \omega_2^i] \leq \max_{V_\epsilon} |\omega_1^i - \omega_2^i|^2$$

Since this is true for every $\epsilon > 0$, we deduce

$$\lim_{n \rightarrow \infty} \text{Var}_{\pi_n(\omega_1, \omega_2)} [\omega_1^i - \omega_2^i] \leq \max_{D_{\mathcal{E}}(\omega_1, \omega_2)=0} |\omega_1^i - \omega_2^i|^2$$

Therefore, if ω_i is identifiable, we have $\lim_{n \rightarrow \infty} \text{Var}_{\pi_n(\omega_1, \omega_2)} [\omega_1^i - \omega_2^i] = 0$.

Moreover, if Ω is finite, assume that there exists ω_1^*, ω_2^* such that $\omega_1^* \neq \omega_2^*$ and $D(\omega_1^*, \omega_2^*) = 0$. Denote

$$V_0 = \{(\omega^1, \omega^2) : D(\omega_1, \omega_2) = 0\}$$

and let $\#(V_0)$ be the cardinality of V_0 .

Since

$$\left| \frac{D_n(\omega_1, \omega_2)}{n} - D(\omega_1, \omega_2) \right| = O\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in $\Omega \times \Omega$, there exists C such that

$$\left| \frac{D_n(\omega_1^*, \omega_2^*)}{n} - \frac{D_n(\omega_1, \omega_2)}{n} \right| \leq \frac{C}{\sqrt{n}}$$

for all $(\omega_1, \omega_2) \in V_0$.

Hence,

$$\begin{aligned} \frac{\pi_n(\omega_1^*, \omega_2^*)}{\pi_n(\omega_1, \omega_2)} &= \frac{q_n(\omega_1^*, \omega_1^*)}{q_n(\omega_1, \omega_2)} \\ &= \exp\left(-\frac{D_n(\omega_1^*, \omega_1^*)}{\sqrt{n}} + \frac{D_n(\omega_1, \omega_2)}{\sqrt{n}}\right) \geq \exp(-C) > 0 \end{aligned}$$

for all $(\omega_1, \omega_2) \in V_0$.

Therefore

$$\pi_n(\omega^{1*}, \omega^{2*}) \geq \frac{\exp(-C)}{\#(V_0)} \sum_{(\omega^1, \omega^2) \in V_0} \pi_n(\omega^1, \omega^2)$$

However, since π_n converges to 0 outside V_0 (by (4.9)), we deduce that $\pi_n(\omega^1, \omega^2)$ is bounded away from 0 and

$$\text{Var}_{\pi_n(\omega^1, \omega^2)} [\omega_i^1 - \omega_i^2] \geq \pi_n((\omega^{1*}, \omega^{2*})) |\omega_i^{1*} - \omega_i^{2*}|^2 \geq C' > 0$$

which completes the proof. ■

Proof of the Theorem 4.3.5

Proof By the same argument as in the proof of Theorem 4.3.4, we have

$$\begin{aligned} \text{Var}_{\pi_n(\omega_1, \omega_2)}[g(\omega_1) - g(\omega_2)] &= E_{\pi_n(\omega_1, \omega_2)}[g(\omega_1) - g(\omega_2)]^2 \\ &= \int_{V_\epsilon} |g(\omega_1) - g(\omega_2)|^2 \pi_n(\omega_1, \omega_2) + \int_{\Omega \times \Omega - V_\epsilon} |g(\omega_1) - g(\omega_2)|^2 \pi_n(\omega_1, \omega_2) \\ &\leq \max_{V_\epsilon} |g(\omega_1) - g(\omega_2)|^2 + C \exp(-n\epsilon/4) \text{Vol}(V_{\epsilon/4}) \end{aligned}$$

which implies

$$\lim_{n \rightarrow \infty} \text{Var}_{\pi_n(\omega_1, \omega_2)}[g(\omega_1) - g(\omega_2)] \leq \max_{D(\omega_1, \omega_2)=0} |g(\omega_1) - g(\omega_2)|^2$$

Therefore, if $g(\omega)$ is \mathcal{E} -identifiable, we have $\lim_{n \rightarrow \infty} \text{Var}_{\pi_n(\omega_1, \omega_2)}[g(\omega_1) - g(\omega_2)] = 0$.

Similarly, when Ω is finite, assume that there exists ω_1^*, ω_2^* such that $\omega_1^* \neq \omega_2^*$ and $D(\omega_1^*, \omega_2^*) = 0$, we can deduce that

$$\text{Var}_{\pi_n(\omega_1, \omega_2)}[g(\omega_1) - g(\omega_2)] \geq \pi_n((\omega_1^*, \omega_2^*)) |g(\omega_1^*) - g(\omega_2^*)|^2 \geq C' > 0$$

which completes the proof. ■

4.6 Discussions and Conclusions

In this paper, we introduce the concept of a *data-free identifiability analysis*, which concerns with the question of system identification under a given experimental setting, without actual experimental observation. Building upon this concept, we suggest a probabilistic approach to address the problem of data-free identifiability analysis. Our approach is global, amenable to high-dimensional cases, can be used to study various types of identifiability and is compatible with a large class of experimental settings. With this method, we attempt to lay a unifying framework to the problems of structural identifiability analysis, dynamics identifiability analysis and a priori uncertainty

quantification in the case of unidentifiability. We then perform the analysis for different biological systems in various experiment settings, and compare the performance with those of other methods of identifiability analysis. The method is proved to be a reliable analysis and is able to provide different and unique insights about the studied systems.

One interesting question about the method is: in case of unidentifiable systems, is it possible to use the Markov Chain with invariant measure π_n to find pairs of parameter sets with indistinguishable outputs, as we did in some of the examples? The answer is, in general, no. As we emphasized in earlier parts of the report, the number of model evaluations for a pure probabilistic method to find an optimum (even local) of an objective function with a given accuracy will increase exponentially in the dimension. Even for a system of 5 parameters, the space $\Omega \times \Omega$ has dimension 10, and finding such pairs of parameter sets is not really practical. Our method focus more on the practical existence of an off-diagonal solution to the problem, without estimating such a solution.

In the paper, we characterize the parameter uncertainty by variance. It is worth noting that similar theoretical results (and variations of the method) using other characterizations of uncertainty, such as entropy measure, can also be derived.

Acknowledgments

This research was supported by the NSF grant DMS-0900277.

4.7 References

- [1] Chis OT, Banga JR, Balsa-Canto E (2011) Structural identifiability of systems biology models: a critical comparison of methods. *PloS One* 6: e27755.
- [2] Miao H, Xia X, Perelson AS, Wu H (2011) On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM Review* 53: 3–39.
- [3] Vu Dinh, Ann E. Rundell and Gregory T. Buzzard, (2014), Experimental Design for Dynamics Identification of Cellular Processes. *Bulletin of Mathematical Biology*, 76.3: 597-626.
- [4] M. M. Donahue, G. T. Buzzard, and A. E. Rundell, (2010), Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids. *IET System Biology*, 4:249–262.
- [5] F. Pukelsheim, (1993), Optimal design of experiments. John Wiley and Sons: New York.
- [6] J. N. Bazil, G. T. Buzzard, and A. E. Rundell, (2011), A global parallel model based design of experiments method to minimize model output uncertainty. *Bulletin of Mathematical Biology*, 74:688–716.
- [7] Gregory T. Buzzard and Bradley J. Lucier, (2013), Optimal filters for high-speed compressive detection in spectroscopy. *Proceedings of SPIE Volume 8657, Computational Imaging XI*, 865707 (February 14, 2013).
- [8] Hiroaki Kitano, (2002), Systems Biology: a Brief Overview. *Science*, Vol. 295, Issue 5560.
- [9] Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS (2006) Kinetics of influenza a virus infection in humans. *Journal of Virology* 80: 7590–7599.
- [10] Goodwin BC (1965) Oscillatory behavior in enzymatic control processes. *Advances in enzyme regulation* 3: 425–437.
- [11] Lipniacki T, Hat B, Faeder JR, Hlavacek WS (2008) Stochastic effects and bistability in t cell receptor signaling. *Journal of theoretical biology* 254: 110–122.
- [12] C. Ritter and M. A. Tanner, (1992), Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *J. Amer. Stat. Assoc.*, 87(419):861–868.

CHAPTER 5. CONVERGENCE OF THE GRIDDY GIBBS SAMPLING AND OTHER PERTURBED MARKOV CHAINS

5.1 Abstract

The Griddy Gibbs sampling was proposed by Ritter and Tanner (1992) as a computationally efficient approximation of the well-known Gibbs sampling method. The algorithm is simple and effective and has been used successfully to address problems in various fields of applied science. However, the approximate nature of the algorithm has prevented it from being widely used: the Markov chains generated by the Griddy Gibbs sampling method are not reversible in general, so the existence and uniqueness of its invariant measure is not guaranteed. Even when such an invariant measure uniquely exists, there was no estimate of the distance between it and the probability distribution of interest, hence no means to ensure the validity of the algorithm as a means to sample from the true distribution.

In this paper, we show, subject to some fairly natural conditions, that the Griddy Gibbs method has a unique, invariant measure. Moreover, we provide L^p estimates on the distance between this invariant measure and the corresponding measure obtained from Gibbs sampling. These results provide a theoretical foundation for the use of the Griddy Gibbs sampling method. We also address a more general result about the sensitivity of invariant measures under small perturbations on the transition probability. That is, if we replace the transition probability P of any Monte Carlo Markov Chain by another transition probability Q where Q is close to P , we can still estimate the distance between the two invariant measures. The distinguishing feature between our approach and previous work on convergence of perturbed Markov Chain is that by considering the invariant measures as fixed points of linear operators on function spaces, we don't need to impose any further conditions on the

rate of convergence of the Markov Chain. For example, the results we derived in this paper can address the case when the considered Monte Carlo Markov Chains are not uniformly ergodic.

Keywords: Griddy Gibbs, nonreversible Markov Chain, perturbed Markov kernel, non-uniformly ergodic Markov Chain

5.2 Introduction

The need to generate samples from a probability function or estimate moments of such a distribution arises in many fields of applied science, including Bayesian statistics, computational physics, computational biology and computer science. A common difficulty in generating such samples is that the distribution (hereafter denoted by π) may be high-dimensional and computationally intractable. To resolve this problem, many sampling-based approaches have been proposed: the basic idea is to construct a Markov chain with a tractable transition mechanism that has π as its invariant distribution.

One of the most widely applicable methods to construct such a Markov chain is the method of Gibbs sampling. This algorithm generates an instance from the distribution of each variable in turn, conditional on the current values of the other variables. This reduces the sampling problem to a series of one-dimensional problems. The method of Gibbs sampling is very computationally effective, especially in the case when π is high-dimensional. Gibbs sampling applies even in the case that the distribution is known only up to a normalizing constant, which occurs commonly in fitting models to data.

However, the use of the Gibbs sampling method is hindered by several factors. First, the method requires the one-dimensional conditional densities to be known, or at least to be easy to sample directly. In most contexts, such knowledge about the conditional densities is usually not available. Second, in many fields of applied sciences, sampling from the conditional distributions is computationally expensive,

despite the fact that they are one-dimensional. For instance, in systems biology, evaluating (up to a normalizing factor) the value of the distribution function π at one point might be equivalent to solving a high-dimensional system of differential equations.

To address these issues, Ritter and Tanner (1992) proposed in [3] an approximate method – the Griddy Gibbs method – as an alternative. The Griddy Gibbs sampling method evaluates the conditional density on a grid and uses piecewise linear or piecewise constant functions to approximate the cumulative distribution function of the conditional distributions based on these grid values. The resulting distribution is used to generate random variables with approximately the right distribution.

This method has been used successfully to address problems in various fields of applied science: statistical inference ([15] [16]), chemical analysis ([20]), systems biology ([21], [22]), medical science([7]), statistical computing and data analysis([12] [14]), economics([8] [17] [19]), ecological modelling([18]), acoustics ([9]), and time series analysis ([10] [11] [13]). However, the approximate nature of the algorithm still prevents it from being widely used. The approximation by linear or constant functions leads to theoretical questions about the ergodic properties of the constructed Markov chains and about the validity of the algorithm as a means to sample from the true distribution.

Many adjustments to overcome the approximate nature of the algorithm have been proposed. In [1], a Metropolis chain is embedded in the algorithm to ensure that the equilibrium distribution is exactly π even on a coarse grid. In [2], a similar strategy is proposed, in which the Multiple-try Metropolis algorithm is embedded in the sampling process. In both approaches, the convergence of the algorithms are guaranteed, but the computational costs increase considerably, the algorithms are more difficult to set up, and the approximations are restricted to piecewise linear and piecewise constant functions.

In this paper, we show, assuming that the approximations to the distribution are bounded from above and bounded away from zero, that the Griddy Gibbs method

has a unique, invariant measure. Moreover, we provide L^p estimates on the distance between this invariant measure and the corresponding measure obtained from Gibbs sampling. Subject to appropriate hypotheses, our main results about Griddy Gibbs are the following.

1. Although the Markov chains generated by the Griddy Gibbs sampler are not reversible in general, they admit unique invariant measures.
2. For $2 \leq p \leq \infty$, there is an L^p -estimate of the distance between the limit invariant measure and the correct distribution π , which guarantees the L^p -convergence of the algorithm.

The first result is obtained using tools from the theory of Markov processes. We then extend the Markov chain transition operator to L^p -spaces and use techniques from the theory of functional analysis on Hilbert vector spaces to prove the second result. These results provide a theoretical foundation for the use of the Griddy Gibbs sampling method with some guarantees of convergence.

In this paper, we go beyond these results for Griddy Gibbs sampling in two ways. First, the approximation scheme does not need to be piecewise linear or piecewise constant: any reasonable approximation scheme can be employed to obtain Griddy Gibbs sampling. In fact, in the case that the distribution is smooth but the computational cost of determining the value of the conditional distribution is much greater than the cost for approximation, high order polynomial interpolations are preferred since they increase the accuracy of the sampling process and reduce the number of function evaluations.

Second, we generalize our method to give results about the sensitivity of invariant measures under small perturbations on the transition probability. That is, if we replace the transition probability P of any Monte Carlo Markov Chain by another transition probability Q where Q is close to P , can we still estimate the distance between the two invariant measures? Our paper provides a positive answer to this question, given some mild conditions imposed on Q . The distinguishing feature be-

tween our approach and other work [29–31] on convergence of perturbed Markov Chain is that by considering the invariant measures as fixed points of linear operators on function spaces, we don't need to impose any further conditions on the rate of convergence of the Markov Chain. For example, the results we derive in this paper can address the case when the considered Monte Carlo Markov Chains are not uniformly ergodic.

The paper is organized as follows. Section 2 provides the mathematical framework used in the paper, as well as descriptions of the Gibbs and Griddy Gibbs sampling methods. Section 3 discusses the existence, uniqueness and regularity results for the invariant measure. We develop in Section 4 results about the sensitivity of invariant measures under small perturbations on the transition probability for general non-uniformly ergodic Monte Carlo Markov Chains. The estimates are then extended to the case when the distribution of interest has non-compact support in Section 5. Finally, we provide in Section 6 numerical examples to illustrate our theoretical findings and demonstrate the utility of the Griddy Gibbs sampling method.

5.3 Mathematical framework

The problem addressed by the Gibbs algorithm is the following (see [26] for reference). We are given a density function $\hat{\pi}$, on a state space with bounded Lebesgue measure $D \subset \mathbb{R}^d$. This density gives rise to an absolutely continuous probability measure π on D , by

$$\pi(A) = \int_A \hat{\pi}(x) dx, \quad \forall A \in \mathcal{B}$$

where \mathcal{B} denotes the σ -algebra of Borel sets on D . Without loss of generality, we assume throughout this paper that the the distribution π has finite variance. In other words, the density function $\hat{\pi} \in L^2(D)$.

In many applications, we want to estimate the expectations of functions $\phi : D \rightarrow \mathbb{R}$ with respect to π , i.e. we want to estimate

$$\pi(\phi) = E_{\pi}[\phi(X)] = \int_D \phi(x) \hat{\pi}(x) dx.$$

If D is high-dimensional, and π_u is a complicated function, then direct integration (either analytic or numerical) of these integrals is infeasible.

The classical Monte Carlo solution to this problem is to simulate independent and identically distributed random variables $X_1, X_2, \dots, X_N \approx \pi(\cdot)$ and then estimate $\pi(\phi)$ by

$$\pi_N(\phi) = \frac{1}{N} \sum_{i=1}^N \phi(X_i). \quad (5.1)$$

This gives an unbiased estimate with standard deviation of order $O(1/\sqrt{N})$. However, if π_u is complicated, it is difficult to directly simulate i.i.d. random variables from π . The Markov chain Monte Carlo (MCMC) approach is introduced instead to construct on D a Markov chain that is computationally efficient and that has π as a stationary distribution. That is, we want to define easily-simulated Markov chain transition probabilities $P(x, y)$ for $x, y \in D$ such that

$$\int_D \hat{\pi}(y) P(x, y) dy = \hat{\pi}(x), \quad \text{for a.e. } x \in D.$$

In principle, if we run the Markov chain (started from anywhere) to obtain samples X_n , then for large n the distribution of X_n will be approximately stationary, and the sequence $\{X_n\}$ can be used to estimate $\pi(\phi)$ as in equation (5.1).

5.3.1 Gibbs transition

The Gibbs transition is a transition probability on D defined as follows:

1. The i^{th} component Gibbs transition P_i leaves all components except the i^{th} component unchanged and replaces the i^{th} component by a draw from the full distribution π conditional on all other components:

$$P_i(x_1, \dots, x_i, \dots, x_d) = \frac{\pi_u(x_1, \dots, x_i, \dots, x_d)}{\int_{-1}^1 \pi_u(x_1, \dots, t, \dots, x_d) dt},$$

where t appears in the i^{th} position.

2. The Gibbs sampler is defined as

$$P(x, y) = P_1(y_1, x_2, \dots, x_d) P_2(y_1, y_2, x_3, \dots, x_d) \cdots P_d(y_1, y_2, \dots, y_d)$$

where $x = (x_1, x_2, \dots, x_d)$ and $y = (y_1, y_2, \dots, y_d)$.

Now let $\{X_n\}, n \geq 0$ be a time-homogeneous Markov process generated by the Gibbs sampling algorithm with transition probability P . Then

$$P(X_n \in A | X_0 = a) = P^n(a, A), \quad \forall A \in \mathcal{B}$$

where P^n is defined recursively by

$$P^1 = P, \quad P^n(a, y) = \int_D P(x, y) P^{n-1}(a, x) dx.$$

We also define the transition operator T on $\mathcal{P}(D)$, the space of probability measures on D , by

$$T\mu(A) = \int_D P(x, A) \mu(dx). \quad (5.2)$$

This transition operator can also be considered as a linear operator on $L^p(D)$, $1 \leq p \leq \infty$, by defining

$$Tf(y) = \int_D P(x, y) f(x) dx$$

Moreover, the operator T^n obtained by replacing P by P^n in (5.2) is equal to the operator obtained by applying T n times, $T^n = T \circ T \circ \dots \circ T$.

5.3.2 Ergodic properties of the Markov Chains generated by the Gibbs sampling

By standard results about ergodicity of Gibbs sampling method, we know that under rather general conditions, T admits a unique invariant measure, which is the distribution π that we want to sample, i.e. $T\pi = \pi$. Moreover, the distribution of X_n converges in total variation norm to π . We state here Theorem 6 from [4] that justifies the convergence of the Gibbs sampling method:

Theorem 5.3.1 ([4]) *Assume that for each $1 \leq i \leq d$, the conditional distributions $\pi(X_i | X_j, j \neq i)$ have densities, say p_i , with respect to some dominating measure ρ_i . Suppose further that for each $1 \leq i \leq d$, there is a set A_i with $\rho_i(A_i) > 0$, and a $\delta > 0$ such that for each $1 \leq i \leq d$*

1. $\pi(X_i = x_i | X_j = x_j, j \neq i) > 0$ whenever $x_k \in A_k$ for $k \leq i$ and x_{i+1}, \dots, x_d arbitrary.
2. $\pi(X_i = x_i | X_j = x_j, j \neq i) > \delta$ whenever $x_k \in A_k$ for $k \leq d$.

Then for π -a.e. $x \in D$:

$$\sup_{C \in \mathcal{B}} |P^n(x, C) - \pi(C)| \rightarrow 0$$

In the rest of the paper, we will assume that the distribution of interest π satisfies conditions of Theorem 5.3.1 and has finite variance.

5.3.3 Griddy Gibbs transition

Now, in the Griddy Gibbs sampling method, at each point in the sampling space and on each dimension, we use some approximation scheme to approximate P_i . The i^{th} component Griddy Gibbs transition leaves all components except the i^{th} component unchanged and replaces the i^{th} component by a draw from Q_i that approximates the conditional expectation on all other components. I.e.,

$$Q_i(x_1, \dots, y_i, \dots, x_d) \approx \frac{\pi_u(x_1, \dots, y_i, \dots, x_d)}{\int_{-1}^1 \pi_u(x_1, \dots, t, \dots, x_d) dt} \quad (5.3)$$

The transition probability and transition operator of the new Markov chain are defined similarly:

$$Q(x, y) = Q_1(y_1, x_2, \dots, x_d) Q_2(y_1, y_2, \dots, x_d) \dots Q_d(y_1, y_2, \dots, y_d) \quad (5.4)$$

$$Q^1 = Q, \quad Q^n(a, y) = \int_D Q(x, y) Q^{n-1}(a, x) dx$$

$$S\mu(A) = \int_D Q(x, A) \mu(dx) \quad (5.5)$$

We note that since the approximations on each dimension are different, the Markov chain $\{Y_n\}$ generated by Griddy Gibbs algorithm is not reversible in general.

$$\begin{array}{ll}
\text{Gibbs:} & \{X_n\} \xrightarrow{T,P} \pi \\
\text{Griddy Gibbs:} & \{Y_n\} \xrightarrow{S,Q} \eta
\end{array}$$

Figure 5.1. Comparison between Gibbs sampling and Griddy Gibbs sampling: Although the two transition operators P and Q are close, the Markov chain $\{Y_n\}$ is not reversible in general, so the existence and uniqueness of the invariant measure η is not guaranteed. Even when η uniquely exists, an estimate of the distance between π and η is needed to guarantee the validity of the Griddy Gibbs sampling.

Throughout this paper, we will use the notation $\{X_n\}$, T , P to describe a Markov chain generated by Gibbs sampling, its transition operator and its transition probability, respectively. The corresponding notations for Griddy Gibbs are $\{Y_n\}$, S , Q . A comparison between the notations used for the Gibbs sampling and Griddy Gibbs sampling is provided in Figure 5.1.

5.4 Existence, uniqueness, and regularity of the invariant measure of a monte carlo markov chain generated by the griddy gibbs sampling

In this section, we will prove that the transition operator S (obtained from the Griddy Gibbs algorithm as in (5.3), (5.4), (5.5)) admits a unique invariant measure η , assuming that the approximations Q_i are uniformly bounded above and away from 0:

$$\exists M, \epsilon > 0, \text{ such that } \epsilon \leq Q_i(x) \leq M, \forall 1 \leq i \leq d, \forall x \in D. \quad (5.6)$$

We also prove that under this condition, η is absolutely continuous with respect to Lebesgue measure and admits a bounded density function.

We note that condition (5.6) is general and does not hinder the application of the Griddy Gibbs sampling method, since we can always use additional cutoff functions

on the approximation scheme to guarantee the boundedness from above and below of f_i , without significantly affecting the accuracy of the approximation scheme.

The outline of the proof is as follow. By verifying Doeblin's condition (see Theorem 5.4.1), we can prove the existence and uniqueness of the invariant measure η ; moreover, the distribution of $\{Y_n\}$ (obtained by the Griddy Gibbs algorithm) converges to η in total variation norm. Using this and Lemma 5.4.1, we deduce that η is absolutely continuous with respect to Lebesgue measure. Finally, using Lemma 5.4.2, we can prove that the density function of η is bounded.

5.4.1 Existence and uniqueness

To verify the existence and uniqueness of the invariant measure, we use the following result from [6] on the convergence of transition probabilities. As before, we will denote by \mathcal{B} the σ -algebra of Borel sets on D .

Theorem 5.4.1 ([6]) *Suppose that the Markov chain Z_n with transition probability $K(x, \cdot)$ satisfies the Doeblin condition:*

$\exists k \in N, \epsilon > 0$, and a probability measure ϕ on (D, \mathcal{B}) such that

$$K^k(x, C) \geq \epsilon \phi(C), \forall x \in D, \forall C \in \mathcal{B}.$$

Then there exists a unique invariant probability measure ξ such that for all $n \in N$ and all $x \in D$,

$$\sup_{C \in \mathcal{B}} |K^n(x, C) - \xi(C)| \leq (1 - \epsilon)^{((n/k)-1)}.$$

Using this result, we can prove that under condition (5.6), the distribution of the Markov chain $\{Y_n\}$ generated by the Griddy Gibbs sampling method converges to a stationary distribution η in total variation norm. This is a direct analog of the convergence given in Theorem 5.3.1 above (although we still have to show that η is near π).

Theorem 5.4.2 (*Existence and uniqueness of the invariant measure for S*) Assume that the approximation scheme $\{f_i\}_{i=1}^d$ satisfies condition (5.6). Then there exists a unique probability measure η that is invariant under S , and this η satisfies

$$\sup_{C \in \mathcal{B}} |Q^n(x, C) - \eta(C)| \rightarrow 0$$

for all $x \in D$.

In other words, $\forall x \in D, Q^n(x, \cdot) \rightarrow \eta(\cdot)$ in total variation norm.

Proof We will prove that the transition probability Q constructed in the Griddy Gibbs sampling algorithm satisfies Doeblin's condition of Theorem 5.4.1. Recall that the transition probability in the Griddy Gibbs algorithm is given by

$$Q(x, C) = \int_C f_1(y_1, x_2, \dots, x_d) f_2(y_1, y_2, \dots, x_d) \cdots f_d(y_1, y_2, \dots, y_d) dy_1 dy_2 \dots dy_d.$$

Recall that $f_i \geq \epsilon$ on D from (5.6). Hence with $\text{Vol}(C)$ denoting the Lebesgue measure of C , we have

$$Q(x, C) \geq \epsilon^d \text{Vol}(C), \forall x \in D, \forall C \in \mathcal{B}.$$

This is Doeblin's condition with $k = 1$, ϕ is the Lebesgue measure on D , so applying Theorem 5.4.1, we have

$$\sup_{C \in \mathcal{B}} |Q^n(x, C) - \eta(C)| \rightarrow 0$$

■

5.4.2 Some supporting lemmas

To establish results about regularity of the invariant measure η of Markov chains generated by Griddy Gibbs sampling, we need the following two lemmas. The first result is about the absolute continuity of η , while the second result provides a basic inequality for the transition operator as a linear operator on L^p space. The proofs are standard, but we sketch them for completeness.

Lemma 5.4.1 Let μ_n be a sequence of probability measures on (D, \mathcal{B}) that converges in total variation norm to a measure μ . Assume further that each μ_n is absolutely

continuous with respect to Lebesgue measure. Then μ is also absolutely continuous w.r.t Lebesgue measure and admits a non-negative density function.

Proof Consider any Borel measurable set A with $|A| = 0$. By the assumption of absolute continuity, $\mu_n(A) = 0$, hence $\mu(A) = \lim \mu_n(A) = 0$. Since A was arbitrary, μ is absolutely continuous w.r.t. Lebesgue measure. ■

Lemma 5.4.2 For $1 \leq p \leq \infty$, let $K(x, y)$ be a bounded function on $D \times D$, and let

$$Lg(y) = \int K(x, y)g(x)dx$$

for $g \in L^p(D)$. Then

a) $L: L^2(D) \rightarrow L^2(D)$ is a compact linear operator. Moreover

$$\|L\|_{L(L^2, L^2)} = \|K\|_{L^2(D \times D)}.$$

b) $L: L^1(D) \rightarrow L^1(D)$ is a bounded linear operator. Moreover, if $K(x, y)$ is a transition probability function, then

$$\|L\|_{L(L^1, L^1)} \leq 1.$$

c) L maps $L^1(D)$ to $L^\infty(D)$, and

$$\|L\|_{L(L^1, L^\infty)} \leq \|K\|_\infty.$$

d) If $g \in L^2(D)$ and $2 \leq p \leq \infty$ then

$$\|Lg\|_p \leq \|K\|_p \max\{\|g\|_1, \|g\|_2\}.$$

Proof Part a) of the Lemma is a well-known result about Hilbert-Schmidt integral operators. For reference, cf. [5].

For b) and c), let $M = \sup_{D \times D} |K(x, y)| = \|K\|_\infty$. Then for all $y \in D$ we have

$$|Lg(y)| = \left| \int K(x, y)g(x)dx \right| \leq M \int |g(x)|dx.$$

In other words,

$$\|Lg\|_\infty \leq \|K\|_\infty \|g\|_1.$$

Integrating over D gives

$$\|Lg\|_1 \leq \text{Vol}(D) \|Lg\|_\infty \leq \text{Vol}(D) \|K\|_\infty \|g\|_1,$$

which proves b) and c). Finally, if K is a transition probability, we have

$$\int |Lg(y)| dy = \int \left| \int K(x, y) g(x) dx \right| dy \leq \int \int K(x, y) dy |g(x)| dx = \int |g(x)| dx$$

which implies $\|L\|_{L(L^1, L^1)} \leq 1$.

For d), consider the linear operator W defined on $L^2(D \times D)$ and on $L^\infty(D \times D)$ as

$$W\phi = \int_D \phi(x, y) g(x) dx$$

Then from part a) and c), we have W is a bounded linear operator that maps $L^2(D \times D)$ to $L^2(D)$, and maps $L^\infty(D \times D)$ to $L^\infty(D)$. Moreover, the following inequalities are satisfied:

$$\|W\phi\|_{L^2(D)} \leq \|g\|_{L^2(D)} \|\phi\|_{L^2(D \times D)}$$

$$\|W(\phi)\|_{L^\infty(D)} \leq \|g\|_{L^1(D)} \|\phi\|_{L^\infty(D \times D)}$$

Using Riesz-Thorin interpolation theorem (see [28]), we deduce that W also maps $L^p(D \times D)$ to $L^p(D)$, and

$$\|W\phi\|_{L^p(D)} \leq \max\{\|g\|_{L^2(D)}, \|g\|_{L^1(D)}\} \|\phi\|_{L^p(D \times D)}$$

Replace ϕ by K , noticing that $Lg = W(K)i$, we deduce

$$\|Lg\|_p \leq \|K\|_p \max\{\|g\|_1, \|g\|_2\}.$$

■

5.4.3 Regularity

These two previous lemmas allow us to prove the following result.

Theorem 5.4.3 (*Regularity of invariant measure*) *The invariant measure η of S is absolutely continuous w.r.t Lebesgue measure on D . Moreover, there exists $\hat{\eta} \in L^\infty(D)$ so that for each $C \in \mathcal{B}$,*

$$\eta(C) = \int_C \hat{\eta}(x) dx.$$

Also, $\hat{\eta}$ is invariant under S : $S\hat{\eta} = \hat{\eta}$.

Proof The proof of this theorem is straightforward from the previous theorems and lemma. From Theorem 5.4.2 and Lemma 5.4.1, we know that η is absolutely continuous and admits a density function:

$$\eta(dx) = \hat{\eta}(x) dx$$

with $\hat{\eta} \in L^1(D)$.

Now considering S as a bounded linear operator on $L^1(D)$, we have

$$\begin{aligned} \int_A \hat{\eta}(x) dx &= \eta(A) = S\eta(A) = \int_D Q(x, A) \eta(dx) \\ &= \int_D \int_A Q(x, y) dy \hat{\eta}(x) dx = \int_A \left(\int_D Q(x, y) \hat{\eta}(x) dx \right) dy. \end{aligned}$$

Since A was arbitrary, we deduce that

$$\hat{\eta}(x) = \int Q(x, y) \hat{\eta}(x) dx$$

or $\hat{\eta} = S\hat{\eta}$. From Lemma 5.4.2, S maps $L^1(D)$ to $L^\infty(D)$. Hence $\hat{\eta} = S\hat{\eta} \in L^\infty(D)$, so $\hat{\eta}$ is a bounded function. ■

Remark 5.4.1 *Since D is a subset with bounded measure of \mathbb{R}^d , $\hat{\eta}$ also belongs to $L^p(D)$, for all $1 \leq p \leq \infty$.*

5.5 Sensitivity and convergence of non-uniformly ergodic Markov Chains

Before proceeding to give result about the sensitivity of the invariant measures under perturbation, we want to make a remark that the assumption of uniformly boundedness away from 0 of the approximations Q_i was introduced only to guarantee the existence and uniqueness of an absolutely continuous invariant measure η .

As we mentioned before, we can always use additional cutoff functions on the approximation scheme to guarantee the boundedness from below of Q_i , without significantly affecting the accuracy of the approximation scheme. However, as an analysis of convergence of perturbed Monte Carlo Markov Chains, condition (5.6) is replaced by any condition that guarantees the existence and uniqueness of the invariant measure η and the ergodicity of the Markov chain $\{Y_n\}$. In a similar manner, the assumptions of Theorem 2.1 can be replaced by the existence and uniqueness of the invariant measure π and the ergodicity of the Markov chain $\{X_n\}$.

In short, we will assume the following conditions in the subsequent analyses

1. the invariant measures π, η of the Markov Chain exists and are unique.
2. the Markov Chain $\{X_n\}, \{Y_n\}$ are ergodic (not necessarily uniformly ergodic).
3. the distributions π, η have finite second moments.

The distinguishing feature between our approach and other work [29–31] on convergence of perturbed Markov Chain is that by considering the invariant measures as fixed points of linear operators on function spaces, we don't need to impose any further conditions on the rate of convergence of the Markov Chain. For that reason, the results we derived in this paper can address the case when the considered Monte Carlo Markov Chains are not uniformly ergodic.

5.5.1 Continuity of eigenspaces for eigenvalue 1

We recall from the previous part of the paper that the two transition operators T and S admit unique absolutely continuous invariant measures π and η , respectively.

Before proceeding to derive estimates of the distance between $\hat{\pi}$ and $\hat{\eta}$, we provide in this section two key lemmas to further investigate properties of the transition operators T and S as operators on $L^2(D)$.

In Lemma 5.5.1, we will prove that the eigenspaces correspond to eigenvalue $\lambda = 1$ of T and S are one-dimensional subspaces spanned by π and η , respectively. Lemma 5.5.2 investigates a special case when it is possible to estimate the distance between the positive invariant eigenvectors of two close operators.

Lemma 5.5.1 *Using the same notation as in the section 2.3 and consider T, S as operators on $H = L^2(D)$, we have*

- $\{v \in H : Tv = v\} = \langle \hat{\pi} \rangle$
- $\{v \in H : Sv = v\} = \langle \hat{\eta} \rangle,$

where $\langle \hat{\pi} \rangle$ denotes the span of $\hat{\pi}$.

Proof Consider any $w \in H - \{0\}$ such that $Tw = w$. Then

$$\int |Tw(y)|dy = \int \left| \int P(x, y)w(x)dx \right| dy \leq \int \int P(x, y)dy |w(x)|dx = \int |w(x)|dx.$$

Equality happens only when

$$\left| \int P(x, y)w(x)dx \right| = \int |P(x, y)w(x)|dx$$

for a.e. $y \in D$.

Since $P(x, y) > 0$, this happens only if w does not change sign on D . Therefore, if we define

$$w^* = \frac{w}{\|w\|_{L^1(D)}}$$

then w^* is the density function of a probability measure on D . Moreover, we also have $Tw^* = w^*$. Since π is the unique invariant measure that is also a fixed point of T , we deduce that $w^* = \hat{\pi}$. Hence, $w \in \langle \hat{\pi} \rangle$. ■

Lemma 5.5.2 *Let M and N be Hilbert-Schmidt integral operators on $H = L^2(D)$. Assume further that $u, v \in H$ such that*

- (i) $\|u\|_H = \|v\|_H = 1$
- (ii) $\{w \in H : Mw = w\} = \langle u \rangle$
- (iii) $\{w \in H : Nw = w\} = \langle v \rangle$
- (iv) u, v are positive functions.

Then there exists $\alpha > 0$ depends only on M such that

$$\|v - u\|_H \leq C(\alpha) \|M - N\|_{L(H,H)}.$$

Proof Since H is a Hilbert space, we can write

$$H = \langle u \rangle \oplus K$$

where K is the orthogonal complement of the linear space spanned by u . For the sake of convenience, in the rest of the proof, we will denote $\|\cdot\|_H$ simply by $\|\cdot\|$.

First we show that there exists $\alpha > 0$ such that

$$\|(M - I)k\| \geq \alpha \|k\| \quad \forall k \in K.$$

By way of contradiction, suppose that $\exists \alpha_n \rightarrow 0, \|k_n\| = 1, k_n \in K$ such that $\|Mk_n - k_n\| = \alpha_n$. Since M is a compact operator on H , by extracting a subsequence, we can assume that $Mk_n \rightarrow k_\infty \in H$. On the other hand, we have $\|Mk_n - k_n\| = \alpha_n \rightarrow 0$. By the triangle inequality, we have

$$\|k_n - k_\infty\| \leq \|k_n - Mk_n\| + \|Mk_n - k_\infty\| \rightarrow 0.$$

We deduce that $k_n \rightarrow k_\infty$, and hence that $\|k_\infty\| = 1$. Since K is closed we have $k_\infty \in K$, and since M is continuous we have $Mk_\infty = k_\infty$. By (ii), $Mu = u$ has no nontrivial solution in K , so we deduce that $k_\infty = 0$, which contradicts $\|k_\infty\| = 1$.

On the other hand, we can uniquely decompose

$$v = \lambda u + k \tag{5.7}$$

for some $\lambda \in \mathbb{R}, k \in K$. Since u and v are fixed points of M and N , respectively, we deduce that

$$Mv = M(\lambda u + k) = \lambda Mu + Mk = \lambda u + Mk$$

and

$$Nv = v = \lambda u + k.$$

Therefore

$$\|M - N\|_{L(H,H)} \geq \|Mv - Nv\| = \|(\lambda u + Mk) - (\lambda u + k)\| = \|Mk - k\| \geq \alpha \|k\|.$$

The orthogonal decomposition in (5.7) gives

$$1 = \|v\|^2 = \lambda^2 \|u\|^2 + \|k\|^2 = \lambda^2 + \|k\|^2,$$

so

$$\lambda^2 = 1 - \|k\|^2 \geq 1 - \left(\frac{\|M - N\|_{L(H,H)}}{\alpha} \right)^2.$$

This plus the same decomposition also gives

$$\begin{aligned} \|v - u\|^2 &= (\lambda - 1)^2 \|u\|^2 + \|k\|^2 = \lambda^2 - 2\lambda + 1 + \|k\|^2 \\ &= 2(1 - \lambda) = 2 \frac{1 - \lambda^2}{1 + \lambda}. \end{aligned}$$

On the other hand, from the facts that u, v are positive functions (by (iv)) with $\|u\| = 1$ (by (i)) and the orthogonal decomposition of v , we have $\lambda = \langle u, v \rangle = \int_D uv \, dx \geq 0$.

Hence

$$\|v - u\|^2 \leq 2(1 - \lambda^2) = 2\|k\|^2 \leq 2 \frac{\|M - N\|_{L(H,H)}^2}{\alpha^2}$$

or

$$\|v - u\| \leq \sqrt{2} \frac{\|M - N\|_{L(H,H)}}{\alpha}.$$

■

5.5.2 Convergence results

In this section, we answer the question about the sensitivity of the invariant of a Monte Carlo Markov Chain under kernel perturbations: given that $\|P - Q\| < \epsilon$ (or

equivalently, given a small perturbation on the transition operator), can we estimate the distance $\|\pi - \eta\|$ between the two invariant measures?

The outline of this section is as follows. Using Lemma 5.5.1 and 5.5.2, we derive the L^2 -estimate of the distance between $\hat{\eta}$ and $\hat{\pi}$.

Then, knowing that S maps $L^1(D)$ to $L^\infty(D)$, we bound the L^∞ -norm by L^2 -norm to produce an L^∞ -estimate, and then apply Lemma 5.4.2 to derive the L^p estimate for $2 \leq p \leq \infty$.

Since the proofs require us to switch back and forth between norms, let us recall that if $f \in L^\infty(D)$ then

$$\|f\|_1 \leq C\|f\|_2 \text{ and } \|f\|_2 \leq C\|f\|_\infty$$

where $C = \sqrt{\text{Vol}(D)}$.

Theorem 5.5.1 (*L^2 -estimate*)

There exists $\delta(\pi), C(\pi) > 0$ such that for $\|P - Q\|_2 < \delta(\pi)$, we have

$$\|\hat{\pi} - \hat{\eta}\|_2 \leq C(\pi) \|P - Q\|_2$$

Proof For clarity, we replace $\hat{\pi}$ and $\hat{\eta}$ with π and η respectively. Applying the previous theorem with $u = \frac{\pi}{\|\pi\|_2}$, $v = \frac{\eta}{\|\eta\|_2}$, we have

$$\left\| \frac{\pi}{\|\pi\|_2} - \frac{\eta}{\|\eta\|_2} \right\|_2 \leq \sqrt{2} \frac{\|T - S\|}{\alpha}. \quad (5.8)$$

Then

$$\left\| \frac{\pi}{\|\pi\|_2} - \frac{\eta}{\|\eta\|_2} \right\|_1 \leq C \left\| \frac{\pi}{\|\pi\|} - \frac{\eta}{\|\eta\|} \right\|_2 \leq C\sqrt{2} \frac{\|T - S\|}{\alpha}$$

with $C = \sqrt{\text{Vol}(D)}$. By the triangle inequality

$$\left| \left\| \frac{\pi}{\|\pi\|_2} \right\|_1 - \left\| \frac{\eta}{\|\eta\|_2} \right\|_1 \right| \leq C\sqrt{2} \frac{\|T - S\|}{\alpha}.$$

Since π and η are probability measures, we have $\|\pi\|_1 = \|\eta\|_1 = 1$, and hence

$$\left| \frac{1}{\|\pi\|_2} - \frac{1}{\|\eta\|_2} \right| \leq C\sqrt{2} \frac{\|T - S\|}{\alpha}.$$

This leads to

$$1 - \frac{\|\pi\|_2}{\|\eta\|_2} \leq C\sqrt{2} \frac{\|T - S\|}{\alpha} \|\pi\|_2. \quad (5.9)$$

If we assume further that the right hand side is less than 1, then

$$\|\eta\|_2 < \frac{\|\pi\|_2}{1 - C\sqrt{2} \frac{\|T - S\|}{\alpha} \|\pi\|_2}. \quad (5.10)$$

The triangle inequality plus (5.8) and (5.9) give

$$\begin{aligned} \|\pi - \eta\|_2 &\leq \left\| \pi - \frac{\|\pi\|_2 \eta}{\|\eta\|_2} \right\|_2 + \left\| \eta - \frac{\|\pi\|_2 \eta}{\|\eta\|_2} \right\|_2 \\ &\leq \sqrt{2} \frac{\|T - S\|}{\alpha} \|\pi\|_2 + C\sqrt{2} \frac{\|T - S\|}{\alpha} \|\pi\|_2 \|\eta\|_2 \end{aligned}$$

and then (5.10) gives

$$\|\pi - \eta\|_2 \leq \sqrt{2} \frac{\|T - S\|}{\alpha} \|\pi\|_2 \left(1 + C \frac{\|\pi\|_2}{1 - C\sqrt{2} \frac{\|T - S\|}{\alpha} \|\pi\|_2} \right). \quad (5.11)$$

Since T is defined by π , we can consider α as a function of π only. Moreover, with $\delta(\pi) = \alpha/(2C\sqrt{2} \|\pi\|_2)$ and $\|T - S\| \leq \delta(\pi)$, the right hand side of (5.9) is at most $1/2$, and the constant in parentheses in (5.11) is at most $1 + 2C \|\pi\|_2$. Hence we define

$$C(\pi) = \frac{\sqrt{2} \|\pi\|_2 (1 + 2C \|\pi\|_2)}{\alpha}$$

and note that from Lemma 5.4.2,

$$\|T - S\|_{L(L^2, L^2)} = \|P - Q\|_{L^2(D \times D)}.$$

Hence for $\|P - Q\|_2 < \delta(\pi)$, changing back to the original notations, we have the desired estimate

$$\|\hat{\pi} - \hat{\eta}\|_2 \leq C(\pi) \|P - Q\|_2.$$

■

Remark 5.5.1 From (5.10) and the choice of $\delta(\pi)$, we see that if $\|P - Q\|_2 < \delta(\pi)$, then $\|\hat{\eta}\|_2 < 2\|\hat{\pi}\|_2$.

Theorem 5.5.2 (L^∞ -estimate)

There exists $\delta'(\pi), C'(\pi) > 0$ such that if $P, Q \in L^\infty(D \times D)$ and $\|P - Q\|_\infty < \delta'(\pi)$ then

$$\|\hat{\pi} - \hat{\eta}\|_\infty \leq C'(\pi) \|P - Q\|_\infty.$$

Proof As in the proof of the previous theorem, we replace $\hat{\pi}$ and $\hat{\eta}$ with π and η respectively. Using part the fact that π and η are fixed by T and S , respectively, plus the triangle inequality and c) of Lemma 5.4.2, we have

$$\begin{aligned} \|\eta - \pi\|_\infty = \|S\eta - T\pi\|_\infty &\leq \|S\eta - T\eta\|_\infty + \|T\eta - T\pi\|_\infty \\ &\leq \|P - Q\|_\infty \|\eta\|_1 + \|P\|_\infty \|\eta - \pi\|_1 \\ &\leq C\|P - Q\|_\infty \|\eta\|_2 + C\|P\|_\infty \|\eta - \pi\|_2, \end{aligned} \quad (5.12)$$

with $C = \sqrt{\text{Vol}(D)}$. With $\delta(\pi)$ and $C(\pi)$ as in the previous theorem, define

$$\delta'(\pi) = \frac{\delta(\pi)}{C} \quad \text{and} \quad C'(\pi) = 2C\|\pi\|_2 + C^2\|P\|_\infty C(\pi).$$

If $\|P - Q\|_\infty < \delta'(\pi)$, then as mentioned previously,

$$\|P - Q\|_2 \leq C\|P - Q\|_\infty < \delta(\pi).$$

We start with 5.12 and then use $\|\eta\|_2 < 2\|\pi\|_2$ and $\|\pi - \eta\|_2 \leq C(\pi) \|P - Q\|_2$ from remark 5.5.1 and theorem 5.5.1 to get

$$\begin{aligned} \|\eta - \pi\|_\infty &\leq C\|P - Q\|_\infty \|\eta\|_2 + C\|P\|_\infty \|\eta - \pi\|_2 \\ &\leq 2C\|P - Q\|_\infty \|\pi\|_2 + C\|P\|_\infty C(\pi) \|P - Q\|_2 \end{aligned}$$

By collecting terms and noticing that $\|P - Q\|_2 \leq C\|P - Q\|_\infty$, we deduce that

$$\begin{aligned} \|\eta - \pi\|_\infty &\leq (2C\|\pi\|_2 + C^2\|P\|_\infty C(\pi)) \|P - Q\|_\infty \\ &= C'(\pi) \|P - Q\|_\infty \end{aligned}$$

■

Theorem 5.5.3 (L^p -estimate, $2 \leq p \leq \infty$)

Let $2 \leq p \leq \infty$, there exists $\delta'(\pi), C'(\pi) > 0$ such that if $P, Q \in L^p(D \times D)$ and $\|P - Q\|_p < \delta'(\pi)$ then

$$\|\hat{\pi} - \hat{\eta}\|_p \leq C'(\pi) \|P - Q\|_p.$$

Proof As before, we replace $\hat{\pi}$ and $\hat{\eta}$ with π and η respectively. Applying Lemma 5.4.2 (part d), noticing that η and π belong to $L^2(D)$, we have:

$$\|S\eta - T\eta\|_p \leq \|P - Q\|_p \max\{\|\eta\|_1, \|\eta\|_2\}$$

and

$$\|T\eta - T\pi\|_p \leq \|P\|_p \max\{\|\eta - \pi\|_1, \|\eta - \pi\|_2\}.$$

Using the fact that π and η are fixed by T and S , respectively, plus the triangle inequality and c) of Lemma 5.4.2, we have

$$\begin{aligned} \|\eta - \pi\|_p = \|S\eta - T\pi\|_p &\leq \|S\eta - T\eta\|_p + \|T\eta - T\pi\|_p \\ &\leq \|P - Q\|_p \max\{\|\eta\|_1, \|\eta\|_2\} + \|P\|_p \max\{\|\eta - \pi\|_1, \|\eta - \pi\|_2\} \\ &\leq C\|P - Q\|_p \|\eta\|_2 + C\|P\|_p \|\eta - \pi\|_2, \end{aligned} \tag{5.13}$$

with $C = \sqrt{\text{Vol}(D)}$. The rest of the proof concludes as in the proof of the previous theorem. ■

5.6 Extension to non-compact support distributions

While most of the assumption of the method on the ergodicity of the Markov Chains are quite general, one restriction of the method comes from the assumption of bounded parameter space D . Since the key ideas of our analysis of sensitivity of the invariant measures rely on moving back and forth between the L^p -norms, this condition could not be easily removed from the framework.

However, it is worth noting that for distributions with non-compact support, a variation of the Griddy Gibbs sampling method can be developed as followed: first, a

rectangular domain D is chosen by prior knowledge about π , then the Griddy Gibbs sampling with $\pi_{new} = \pi|_D$ (normalized by a constant) is proceeded as usual. By our previous analyses, the Monte Carlo Markov Chains generated by this process will have a unique invariant measure η whose distance to π can be estimated by the following theorem

Theorem 5.6.1 *Let $2 \leq p \leq \infty$. Assume that π has non-compact support on \mathcal{R}^d and that there exists $C > 0$ so that:*

$$\int_{\mathcal{R}^d} \|x\|_1 \hat{\pi}(x)^p dx \leq C_1 < \infty \quad (5.14)$$

and

$$\int_{\mathcal{R}^d} \|x\|_1 \hat{\pi}(x) dx \leq C_2 < \infty \quad (5.15)$$

where $\|x\|_1 = |x_1| + \dots + |x_d|$. Denote

$$D_t = \{x \in \mathcal{R}^d : \|x\|_\infty > t\}$$

where $\|x\|_\infty = \max_i |x_i|$.

Then there exists $\delta'(\pi), C'(\pi) > 0$ such that if $P, Q \in L^p(D \times D)$, $\|P - Q\|_p < \delta'(\pi)$ and $t \geq C_2/2$ then

$$\|\hat{\pi} - \hat{\eta}\|_p \leq C'(\pi, D_t) \|P - Q\|_p + \frac{C_2}{2t} \|\hat{\pi}\|_p + \frac{C_1}{t \|\hat{\pi}\|_p} \quad (5.16)$$

Proof We denote

$$f = \frac{\hat{\pi}^p}{\|\hat{\pi}\|_p}$$

and X is a random variable whose density function is f .

By the Chebyshev's inequality, we have for $i = 1, 2, \dots, d$

$$\begin{aligned} \mathbb{P}[|X_i| > t] &\leq \frac{1}{t} \int_{\mathcal{R}^d} |x_i| f(x) dx \\ &= \frac{1}{t \|\hat{\pi}\|_p} \int_{\mathcal{R}^d} |x_i| \hat{\pi}^p(x) dx \end{aligned}$$

Hence

$$\|\hat{\pi}\|_{L^p(\mathcal{R}^d \setminus D_t)} = \mathbb{P}[\|X\|_\infty > t] \leq \frac{\int_{\mathcal{R}^d} \|x\|_1 \hat{\pi}^p(x) dx}{t \|\hat{\pi}\|_p}$$

By a similar argument, we have

$$\|\hat{\pi}\|_{L^1(\mathcal{R}^d \setminus D_t)} \leq \frac{1}{t} \int_{\mathcal{R}^d} \|x\|_1 \hat{\pi}(x) dx \leq \frac{C_2}{t}$$

On the other hand, results for distribution with compact support in D implies

$$\left\| \hat{\eta} - \frac{\hat{\pi}}{\int_{D_t} \hat{\pi}} \right\|_{L^p(D)} \leq C'(\pi, D_t) \|P - Q\|_p,$$

We deduce that, for $t \geq 2C_2$, we have

$$\begin{aligned} \|\hat{\pi} - \hat{\eta}\|_{L^p(\mathcal{R}^d)} &\leq \|\hat{\pi} - \hat{\eta}\|_{L^p(D)} + \|\hat{\pi}\|_{L^p(\mathcal{R}^d \setminus D_t)} \\ &\leq \left\| \hat{\eta} - \frac{\hat{\pi}}{\int_{D_t} \hat{\pi}(x) dx} \right\|_p + \frac{\int_{\mathcal{R}^d \setminus D_t} \hat{\pi}(x) dx}{\int_{D_t} \hat{\pi}} \|\hat{\pi}\|_p + \|\hat{\pi}\|_{L^p(\mathcal{R}^d \setminus D_t)} \\ &\leq C'(\pi, D_t) \|P - Q\|_p + \frac{C_2}{2t} \|\hat{\pi}\|_p + \|\hat{\pi}\|_{L^p(\mathcal{R}^d \setminus D_t)} \\ &\leq C'(\pi, D_t) \|P - Q\|_p + \frac{C_2}{2t} \|\hat{\pi}\|_p + \frac{C_1}{t \|\hat{\pi}\|_p} \end{aligned}$$

■

Corollary 5.6.1 *Let $2 \leq p \leq \infty$ and assume that π has non-compact support on \mathcal{R}^d and that there exists $C_3, C_4 > 0$ so that:*

$$1. \int_{\mathcal{R}^d} |x|^2 \hat{\pi}(x) dx \leq C_3 < \infty$$

and

$$2. \|\hat{\pi}\|_{L^{2p-1}} \leq C_4 < \infty$$

Then result (5.16) is true with $C_1 = \sqrt{C_3 C_4^p}$ and $C_2 = 1 + C_4$.

That is, if prior estimates on the second moment of π and the L^{2p-1} -norm of $\hat{\pi}$ are available, the Griddy Gibbs algorithm can be adjusted accordingly to produce a good estimate on the distance between the two invariant measures.

Proof By Holder's inequality with

$$u(x) = \|x\|_1 \hat{\pi}(x)^{1/2}, \quad v(x) = \hat{\pi}^{p-1/2}$$

we have

$$\begin{aligned} \int_{\mathcal{R}^d} \|x\|_1 \hat{\pi}^p(x) \, dx &\leq \left(\int_{\mathcal{R}^d} \|x\|_1^2 \hat{\pi}(x) \, dx \right)^{1/2} \left(\int_{\mathcal{R}^d} \hat{\pi}^{2p-1}(x) \, dx \right)^{1/2} \\ &\leq \sqrt{C_1 C_2^p}. \end{aligned}$$

and

$$\int_{\mathcal{R}^d} \|x\|_1 \hat{\pi}(x) \, dx \leq \int_{\|x\|_1 \leq 1} \hat{\pi}(x) \, dx + \int_{\|x\|_1 > 1} \|x\|_1^2 \hat{\pi}(x) \, dx \leq 1 + C_4$$

■

5.7 Numerical examples

In this section, we provide numerical examples to illustrate our theoretical findings and demonstrate the utility of the Griddy Gibbs sampling method. First, we validate the estimates derived in previous sections in a simple 2D example. We then proceed to investigate the performance of the Griddy Gibbs sampling in a practical example arising from systems biology, in which it is necessary to employ the Griddy Gibbs sampling method, and demonstrate the use of the method in making inferences about the system.

5.7.1 A 2D example

In this example, we investigate the performance of the Griddy Gibbs sampling algorithm on grids of various resolutions in a simple 2D example. The chosen distribution has the following density function

$$\begin{aligned} \pi(x, y) &= \frac{1}{2} \text{Beta}\left(\frac{x+1}{2}, 2, 5\right) * \text{Beta}\left(\frac{y+1}{2}, 2, 5\right) \\ &\quad + \frac{1}{2} \text{Beta}\left(\frac{x+1}{2}, 2, 2\right) * \text{Beta}\left(\frac{y+1}{2}, 2, 2\right) \end{aligned}$$

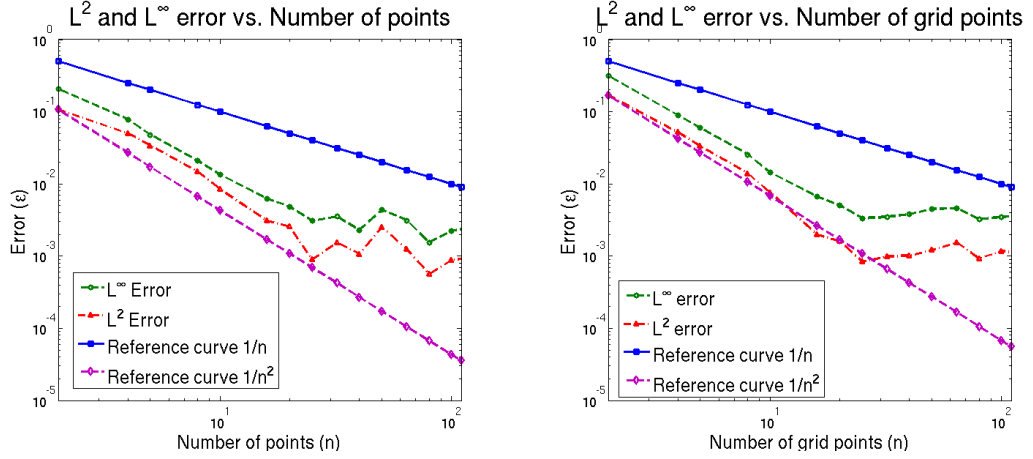


Figure 5.2. Left: Error of the 1D marginal empirical cumulative distribution function, and Right: error of the empirical cumulative distribution function, both as a function of the number of points used in the approximation grid.

where $Beta(x, \alpha, \beta)$ is the one-dimensional Beta distribution with parameter α and β .

This distribution was chosen specifically to illustrate the developed framework in the case of compact support: it has compact support in its domain $[-1, 1] \times [-1, 1]$ and has non-independent components, but the 1D marginal density functions can be obtained in simple form.

Using this probability distribution, we illustrate the estimates provided in previous sections, by expressing the L^2 and L^∞ distance between the estimator (using Griddy Gibbs) and the true distribution of interest in terms of the number of points used in the grid of approximation. For a fixed grid, a Griddy Gibbs chain of length 10^5 is generated, using standard linear interpolation as the approximation scheme for 1-dimensional distributions. We then use the sampled points to estimate the empirical cumulative distribution function (ECDF) and the 1D marginal ECDF of the invariant distribution of the chains. Finally, the L^2 and L^∞ distance between the estimated ECDFs with different number of grid points and the true CDF are calculated. We note that in the context of our example, it is more convenient to work with CDFs

rather than with PDFs for two main reasons: (i) CDFs can be approximated using nonparametric estimators; and (ii) there is a well-developed theoretical machinery for the comparison of CDFs using such estimators. Moreover, it is well-known that the ECDF is a non-parametric, unbiased estimator that converges uniformly to the true CDF (a result known as the Glivenko - Cantelli theorem [25]).

The results are illustrated in Figure 2. The error of both ECDF and the marginal ECDF of the first variable decrease faster than $O(\frac{1}{n})$ and approximately as fast as $O(\frac{1}{n^2})$ when the number of the grids point n increases, until it reaches a level at which the error of the Griddy Gibbs sampling is dominated by the error of the Monte Carlo simulation. Since the accuracy of standard 1D linear approximation method is bounded by $O(\frac{1}{n})$, and can be as fast as $O(\frac{1}{n^2})$ if the function has bounded second derivative, this confirms our theoretical results about linear dependency between error of the 1D approximation, and the distance from the estimated distribution to the true distribution of interest.

5.7.2 An example in systems biology.

In this example, we consider a mathematical model of the T-cell signaling pathway proposed by Lipniacki et al. in [23]. The behaviour of the system is modelled as an ODE system controlled by 19 different parameters with 37 state variables and fixed initial conditions:

$$\dot{x} = \alpha(\omega, x) \quad (\text{System of ODEs})$$

$$x(0) = x_0(\omega) \quad (\text{Initial conditions})$$

$$y(t) = f(\omega, t) = \beta(\omega, x(t)) \quad (\text{Output})$$

Here $x = (x_1, x_2, \dots, x_{n_x}) \in M \subset \mathbb{R}^{n_x}$ is the state variable, with M a subset of \mathbb{R}^{n_x} containing the initial state, and $f(\omega, t) \in \mathbb{R}$ is the output response (system dynamics). In the scope of this paper, we are interested in the dynamics of pZap, one of the state variables of the system. The vector of unknown parameters is denoted

by $\omega = (\omega_1, \dots, \omega_N) \in \mathbb{R}^N$ and is assumed to belong to a subset Ω of \mathbb{R}^N . These functions and initial conditions depend on the parameter vector $\omega \in \Omega$.

The traditional approach to study such a system is to estimate values of the parameters from observations. However, in the field of systems biology, usually it is not possible to estimate all parameters in a given model, in particular if the model is complex and the data is sparse and noisy. Thus, to represent explicitly the state of knowledge, it is best to consider not a single parameter valuation but the whole space of uncertain parameters. The uncertainty in parameter values is often characterized by a probability distribution $\pi(\omega)$ on the set of all possible parameter values, based on how the output of the system driven by a particular parameter valuation fits previous data. This gives a distribution with density

$$\hat{\pi}(\omega) = c_n \exp \left(- \sum_{i=1}^n |f(\omega, t_i) - d(t_i)|^2 \right), \quad (5.17)$$

where c_n is a normalizing constant, $(t_1, d_1), \dots, (t_n, d_n)$ is the set of previous data.

Inference about the system will be made based on π . For example, in [21] [24], the optimal experiment is chosen at the time point where the maximum value of the normalized variance of the outputs with respect to π is achieved. Another example was given in [22] where the expected dynamics estimator to recover the correct system dynamics is defined as the expected value of the system dynamics with respect to the distribution π .

This motivates the problem of sampling with respect to the distribution π . As noted in the introduction, the use of the standard Gibbs sampling method is hindered by two factors: first, there is no closed-form formula for the distribution π or for the corresponding one-dimensional conditional distributions; second, the evaluation of the unnormalized distribution at one point is computationally expensive (it is equivalent to solving a high dimensional system of differential equations). It is then necessary to approximate the conditional distribution by functions of simpler forms. The Griddy Gibbs method therefore is a suitable choice for this sampling process.

In this particular example, we restrict the analysis to the five most sensitive parameters with respect to perturbation. This choice is based on previous knowledge about the dynamics of the system and on the result of a global sensitivity analysis using sparse grid interpolation([27]).

To further reduce the computational cost, we also employ a sparse grid interpolant to approximate the output of the ODE system. That is, the output functions of the system of ODEs are evaluated on a sparse grids of 10^5 points on the parameter space, then the method of sparse grid interpolation is employed to approximate the outputs at other sets of parameter values. Moreover, the one-dimensional conditional distributions are then approximated by piecewise linear functions on grids of fineness $\delta = 0.2$ (which corresponds to a grid with 11 equally spaced points). It is worth noting that although this is a two-leveled approximation, it still fits into the framework developed in previous sections.

We will compare the performance of the Griddy Gibbs sampling with the variation of Gibbs sampling suggested by Tierney et al. in [1]. In Tierney's algorithm, a Metropolis chain is embedded to ensure that the equilibrium distribution is exactly π even on a coarse grid. The drawback is that the computational cost is at least twice as much as Griddy Gibbs sampling using the same grid. Moreover, the algorithm is more difficult to set up and is restricted to piecewise linear and piecewise constant approximations.

In Figure 5.3, we use samples from Griddy Gibbs and from Tierney's algorithm to compare the conditional and marginal distribution derived from the ECDFs. In the left panel, we compare the conditional distributions (using samples from Griddy Gibbs and Tierney's algorithm) of the second parameter on the first parameter, for various values of this parameter. In the right panel, the difference between the two marginal joint distributions of the first and the second variable are computed. We also compare the difference between the two marginal joint distributions of the first and the second variable while using various numbers of samples in Figure 5.4. The results from Figure 5.3 and Figure 5.4 suggest that the Griddy Gibbs sampling method is as

effective as Tierney's algorithm (whose convergence is also guaranteed theoretically) in generating Markov Chains with respect to a given invariant measure: the difference between the two marginal distributions is of the same magnitude as the error of the Monte Carlo method itself.

We then investigate the performance of the Griddy Gibbs sampling in making inferences about dynamics. For this we consider the Expected Dynamics Estimator based on one single simulated data point. This generates a distribution π_1 as in (5.17) with $n = 1$, and we then use this distribution to estimate the system dynamics by

$$\hat{D}_1(t) = E_{\pi_1(\omega)}[f(\omega, t)].$$

The results are provided in Figure 5.5 (Left). The expected dynamics are calculated using the empirical mean of the output values on the previous two sets of samples. Once again, the performance of the Griddy Gibbs sampling is as good as Tierney's algorithm in computing the expected dynamics.

Finally, Figure 5.5 (Right) compares the auto-correlation coefficients of the Monte Carlo Markov Chains generated by the two algorithms. To compute the auto-correlation coefficients, two Monte-Carlo Markov Chains of length 10^5 were generated by the two algorithms, respectively. The figure illustrates the fact that not only is the computational cost of Tierney's algorithm (to generate one instance of the chain) higher, but also its auto-correlation function converges (to zero) at a much lower rate. In this particular example, if one wants to get two sets of i.i.d samples with the same number of points by both algorithms, the computational cost for Tierney's algorithm is at least ten times that of the cost for Griddy Gibbs.

5.8 Conclusion

We have shown, subject to some fairly natural conditions, that the Griddy Gibbs method has a unique, invariant measure. Moreover, we gave L^p estimates on the distance between this invariant measure and the corresponding measure obtained

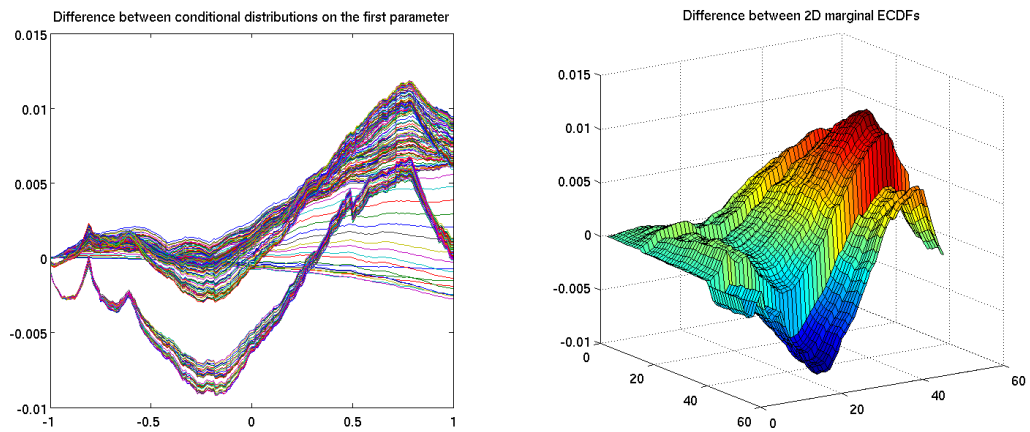


Figure 5.3. Conditional and marginal distribution for the T-cell model. Left: The difference between the conditional distributions on the first parameters (one curve for each value of this parameter). Right: The difference between the marginal joint distributions of the first two parameters, achieved from Griddy Gibbs and Tierney's algorithm. Figure 4 shows that the differences between corresponding ECDFs are of the same magnitude as the error of the Monte Carlo method ($O(\frac{1}{\sqrt{N}})$, where N is the number of samples)

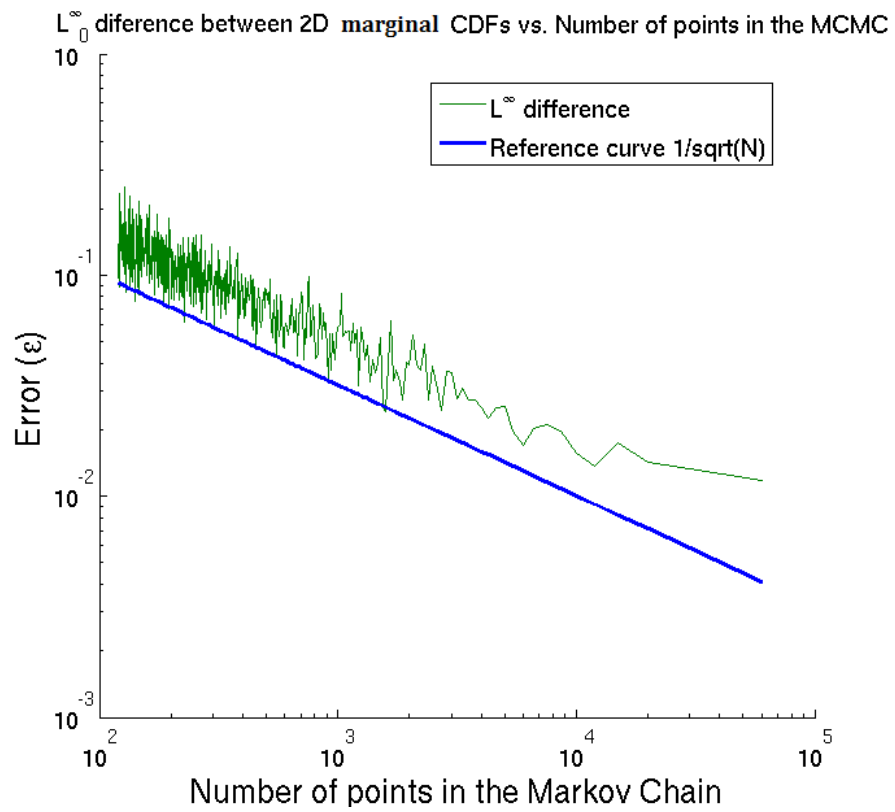


Figure 5.4. The difference between the marginal distributions computed by Griddy Gibbs and Tierney's algorithm is of the same magnitude as the error of the Monte Carlo method itself ($O(\frac{1}{\sqrt{N}})$, where N is the number of samples).

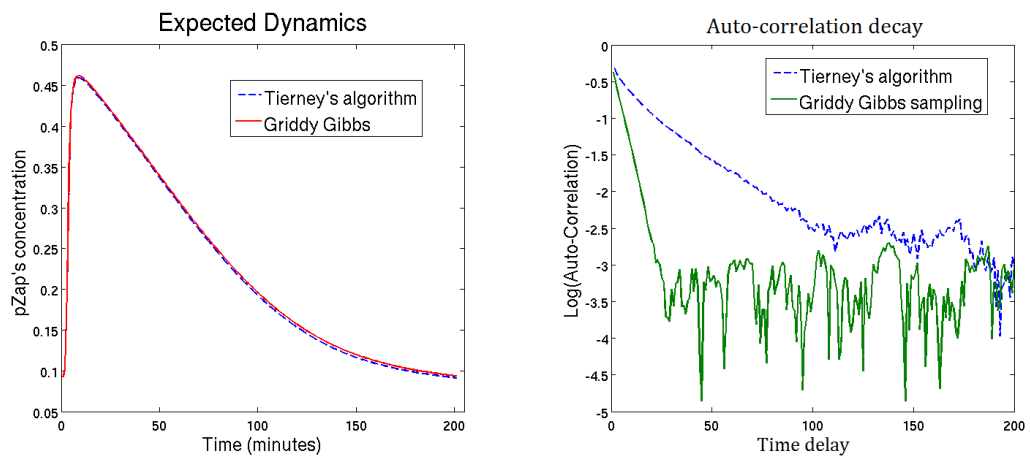


Figure 5.5. Left: The expected dynamics estimator based on one data point, generated by Griddy Gibbs and Tierney's samples. Right: Auto-Correlation coefficients of the Markov Chains generated by Griddy Gibbs algorithm and Tierney's algorithm.

from Gibbs sampling. These results provide a theoretical foundation for the use of the Griddy Gibbs sampling method.

Moreover, using the theoretical framework developed to validate the Griddy Gibbs sampling method, we also successfully provided a more general result about the sensitivity of invariant measures under small perturbations on the transition probability. Our results imply that if we replace the transitional probability P of a Monte Carlo Markov Chain by a different transitional probability Q that is close to P in L^p norm ($2 \leq p \leq \infty$), the distance between the two invariant measures (in L^p) is bounded by a constant times the L^p -distance between P and Q , provided that the approximation schemes satisfy a mild condition provided in the paper. This condition is very general and does not hinder the application of the Griddy Gibbs sampling method, since it can always be guaranteed simply by using additional cutoff functions on the approximation scheme, without significantly affecting its accuracy. The method can be generalized to validate other Monte Carlo Markov Chain sampling methods that involve approximation.

We also gave numerical examples to illustrate our theoretical findings and demonstrate the utility of the method in different applications. The numerical results confirm the linear relation between the distance between the invariant measures and the accuracy of the approximation scheme derived in theory. Moreover, our examples illustrate that Griddy Gibbs performs as well as its variants in applications and that the algorithm is simpler to implement and less computationally expensive. Additionally, the Markov Chains generated by this algorithm have significantly smaller auto-correlation coefficients than those of other variant algorithms. These features demonstrate that Griddy Gibbs is a simple and effective sampling method that can be employed in applications with confidence in its validity.

5.9 References

- [1] TIERNEY, L. (1994). Markov Chains for exploring posterior distributions. *Ann. Stats.* **22**, 1701–1728.
- [2] LIU, S.J. ET. AL. (2000). The Multiple-Try Method and local optimization in Metropolis sampling. *J. Am. Stat. Assoc.* **95**, 121–134.
- [3] RITTER, C. AND TANNER, M.A. (1992). Facilitating the Gibbs Sampler: the Gibbs Stopper and the Griddy-Gibbs Sampler. *J. Am. Stat. Assoc.* **87**, 861–868.
- [4] ATHREYA, K. ET. AL. (1996). On the convergence of the Markov Chain Simulation Method. *Ann. Stats.* **24**, 69–100.
- [5] PEDERSEN, M. (1999). Functional analysis in applied mathematics and engineering. CRC Press.
- [6] KORALOV, L. AND SINAI, Y. (2007). Theory of probability and random processes. Springer.
- [7] J. LI ET. AL. (2007). A random-effects Markov transition model for Poisson-distributed repeated measures with non-ignorable missing values. *Stat. Med.* **26**, 2519–2532.
- [8] BOATWRIGHT, P., MCCULLOCH, R., AND ROSSI, P. (1999). Account-level modeling for trade promotion: an application of a constrained parameter hierarchical model. *J. Am. Stat. Assoc.* **94**, 1063–1073.
- [9] MICHALOPOULOU, Z. AND PICARELLI, M., (2005) Gibbs sampling for time-delay-and amplitude estimation in underwater acoustics. *J. Acoust. Soc. Am.* **117**, 799–8087.
- [10] BOATWRIGHT, P., MCCULLOCH, R. AND ROSSI, P. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *J. Roy. Stat. Soc. C-App.* **50**, 187–200.
- [11] RAY, B. AND TSA, R. (2002). Bayesian methods for change-point detection in long-range dependent processes. *J. Time. Ser. Anal.* **23**, 687–705.
- [12] ARDIA, D., HOOGERHEIDE, L. AND DIJK, H. (2009). Adaptive mixture of Student-t distributions as a flexible candidate distribution for efficient simulation: The R Package AdMit. *J. Stat. Softw.* **29**, 1–32.
- [13] CHEN, C. AND WEN, Y. (2001). On goodness of fit for time series regression models *J. Stat. Comput. Sim.* **69**, 239–256.
- [14] AUSNA, M. AND GALEANO, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model *Comput. Stat. Data. An.* **51**, 2636–2652.
- [15] RAY, B., MCCULLOCH, R. AND TSA, R. (1997). Bayesian methods for change-point detection in long-range dependent processes. *Stat. Sinica.* **7**, 451–472.
- [16] BARNARD, J., MCCULLOCH, R. AND MENG, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Stat. Sinica* **10**, 1281–1311.

- [17] BAUWENS, L. AND ROMBOUTS, J. (2007). Bayesian inference for the mixed conditional heteroskedasticity model. *Economet. J.* **10**, 408–425.
- [18] BAUWENS, L. AND ROMBOUTS, J. (2007). Bayesian clustering of many GARCH models. *Economet. Rev.* **26**, 365–386.
- [19] WALKERA, D., PREZ-BARBERAB F. AND MARION, G (2006). Stochastic modelling of ecological processes using hybrid Gibbs samplers. *Ecol. Model.* **198**, 40–52.
- [20] RITTER, C. (1994). Statistical analysis of spectra from electron spectroscopy for chemical analysis. *J. Roy. Stat. Soc. D-Stat.* **43**, 111–127.
- [21] DONG, W., TANG, X., YU, Y., NILSEN, R., KIM, R., GRIFFITH, J., ARNOLD, J, H. BERND SCHUTTLE (2008). Systems Biology of the Clock in *Neurospora crassa*. PLoS ONE 3: e3105.
- [22] DINH, V., RUNDELL, A.E. AND BUZZARD, G.T (2014). Experimental Design for Dynamics Identification of Cellular Processes. To appear on *Bull. Math. Biol.*
- [23] LIPNIACKI T., HAT B., FAEDER J.R., HLAVACEK W.S, (2008). Stochastic effects and bistability in T cell receptor signaling. *Journal of Theoretical Biology* **254** 110–122.
- [24] DONAHUE, M.M., BUZZARD, G.T. AND RUNDELL, A.E. (2010). Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids. *IET System Biology* **4**, 249–262.
- [25] LILLACCI, G. AND KHAMMASH, G. (2012). A distribution-matching method for parameter estimation and model selection in computational biology. *Int. J. Robust. Nonlinear Control* **22**, 1065–1081.
- [26] ROBERTS, G.O. AND ROSENTHAL, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys* **1**, 20–71.
- [27] BUZZARD, G.T. (2012). Global sensitivity analysis using sparse grid interpolation and polynomial chaos. *Reliability Engineering and System Safety* **17**, 82–89.
- [28] LASSER, R. (1996.) Introduction to Fourier Analysis. CRC Press.
- [29] ROBERTS, G.O., ROSENTHAL, J.S., AND SCHWARTZ, P.O. (1998). Convergence properties of perturbed Markov chains. *Journal of applied probability*, 1–11
- [30] MITROPHANOV, A.Y. (2005). Sensitivity and convergence of uniformly ergodic Markov chains. *Journal of Applied Probability*, 1003–1014.
- [31] ANDRIEU, C., AND ROBERTS, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Stats.* **37.2**, 697–725.

CHAPTER 6. CONCLUSIONS AND FUTURE WORK

This dissertation is concluded here with a comprehensive summary of the work performed, presentation of future work, and discussion of the conclusions drawn.

6.1 Summary of work

The work presented in this dissertation comprises the development of a procedure to enable uncertainty quantification and experimental design of non-linear mathematical models within the practical and theoretical limitations of the biological contexts. This work partially overcomes these limitations through several approaches and successfully addresses multiple problems arising from the field. Those include

- (i) The problem of experimental design for dynamics identification:

We proposed a novel estimator and formalized a process to quantify the uncertainty in prediction the dynamics of interest, as well as provided theoretical foundation for the use of the Maximal Informative Next Experiment approach and improved its performance in several settings.

- (ii) The problem of behavior discrimination in nonlinear models:

We considered the problem of choosing effective data sampling schemes for behavior discrimination of nonlinear systems in two different settings: the low-discrepancy sampling scheme and the uncertainty-based sequential sampling scheme. In both cases, we successfully derived theoretical results about the convergence of the expected boundary to the true boundary of interest. Both methods have also proven to be effective in studies of expensive high-dimensional biological systems in various contexts. The analysis proposed in the work is

novel and may be applicable to other settings, while the performance of the algorithm is state-of-the-art to the best of our knowledge.

- (iii) The problem of data-free identifiability analysis and data-free uncertainty quantification.

We introduced and explored the novel concept of data-free identifiability, which further extends the concept of structural identifiability, taking into account any constraints on the experimental setting. We also proposed a Bayesian approach to address system identifiability when data are not yet available. This approach is global, strongly theoretically supported, amenable to high-dimensional cases, can be used to study various types of identifiability and is compatible with a large class of experimental settings. The framework is also built not only to assess parameter identifiability but also to quantify the uncertainty in prediction of any quantity of interest.

This work also draws a direct connection between studies of identifiability and the concept of uncertainty quantification in predictive sciences. With this method, we attempt to lay a unifying framework for the problems of structural/practical identifiability analysis, dynamics identifiability analysis and data-free uncertainty quantification.

- (iv) The convergence of perturbed Monte Carlo Markov Chains:

We investigated the performance of the Griddy Gibbs sampling in different biological examples and provided a theoretical foundation for the use of Griddy Gibbs sampling and other Monte Carlo Markov Chain methods. The distinguishing feature between our approach and previous work on convergence of perturbed Markov Chains is that by considering the invariant measures as fixed points of linear operators on function spaces, we don't need to impose any further conditions on the rate of convergence of the Markov Chain. For example, the results we derived in this paper can address the case when the considered

Monte Carlo Markov Chains are not uniformly ergodic, which had not been addressed in the perturbed Markov Chain literature.

6.2 Future work: Localized analysis/uncertainty quantification and unsupervised behavior discrimination of biological systems

As emphasized earlier, one of the objectives of this research is to lay a theoretical framework for the problem of localized analyses and uncertainty quantification of high-dimensional biological systems in the face of system unidentifiability and discontinuous/sharp responses. With the successful use of behavior discrimination in mapping the parameter space by qualitative behaviors, we are equipped with a powerful tool to tackle the problem of localized analysis/uncertainty quantification. The only obstacle remaining is the approximation of smooth functions on arbitrarily shaped domains with unstructured samples, which has recently emerged as an active research direction in uncertainty quantification [2].

In Figure 6.1, a preliminary example is presented, in which we contrast the performances of global versus localized methods of uncertainty quantification. This example is extracted from my collaborative work [3] in the theory of robust explicit model predictive control, that extends and adapts our proposed framework to fit the new application context. This outstanding performance is also expected in analyses of biological systems, such as sensitivity analysis or model order reduction.

Another revenue for future work comes from a limitation of our behavior discrimination framework, namely, the necessity of strict definitions of the contrasting behavior of interest. In our setting, the number of contrasting behaviors as well as their definitions has to be specified in advance. The lack of information about what behaviors are relevant to a given biological systems motivates us to build an unsupervised behavior discrimination framework, in which the algorithm would propose and define the behaviors itself. This question is directly related to the task of unsu-

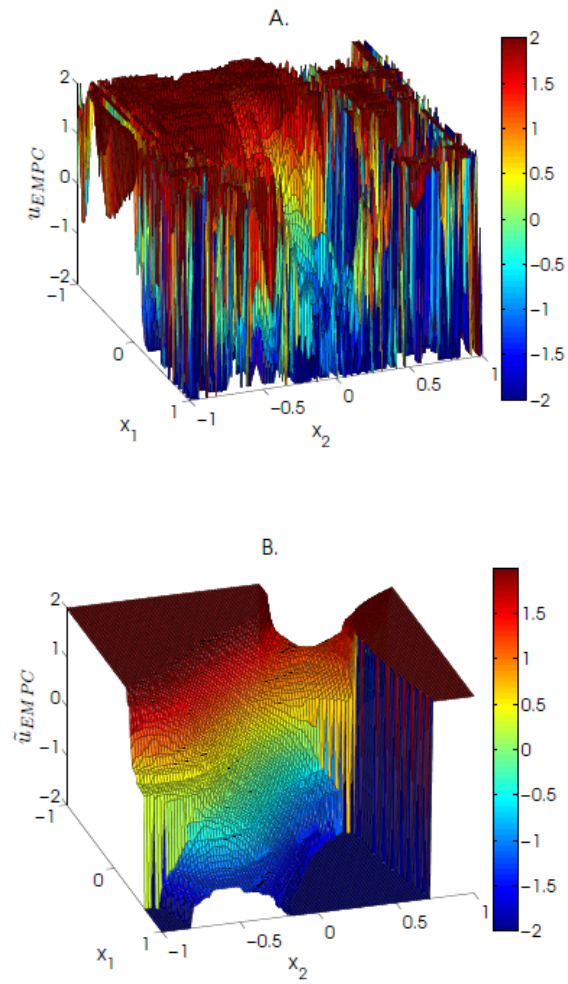


Figure 6.1. Comparison of (Top) non-smooth EMPC control surface and (Bottom) Localized EMPC controller

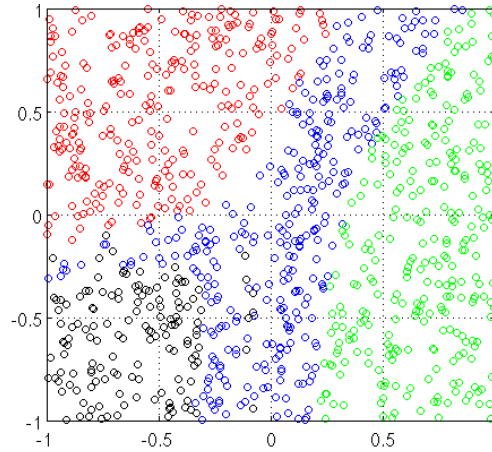


Figure 6.2. The 2-dimensional parameter space of the Fitz Hugh-Nagumo model is partitioned according to dynamics behavior using the spectral clustering method based on the Pearson correlation distance: Red and green regions correspond to the oscillatory and transient behavior of the membrane potential. Blue and black regions both correspond to cases when the membrane potential saturates at a high level; the distinguishing feature is that in the black region, the membrane potential decreases at the beginning before being activated and saturating.

pervised learning that has been of huge interest in the field of machine learning for many years [4].

In Figure 6.2, we provide a preliminary example in which an unsupervised behavior discrimination is performed on the Fitz-Hugh Nagumo model [5], a simplified version of the Hodgkin-Huxley model, which models in a detailed manner activation and deactivation dynamics of a spiking neuron. In this example, the 2-dimensional parameter space of the Fitz Hugh-Nagumo model is partitioned according to dynamics behavior using the spectral clustering method [6] based on the Pearson correlation distance [7]. Despite being simple, the Fitz-Hugh Nagumo model produces several different types of dynamics, and the fact that the algorithm can distinguish between major behaviors without a labeling process or a strict definition of behavior promise a lot of potential for further study of the approach.

6.3 Conclusion

Through the newly developed procedures, this dissertation has addressed many challenges in studying biological systems and successfully created a general probabilistic framework for uncertainty quantification and experimental design in the face of unidentifiability, sharp model responses with limited number of model simulations, constraints on experimental setting, and even in the absence of data. The proposed methods have strong theoretical foundations and have also proven to be effective in studies of expensive high-dimensional biological systems in various contexts.

These procedures are particularly suited to enable immediate gains in biological studies. By taking advantage of these proposed procedure for UQ and experimental design, discovery sciences can more efficiently evaluate hypotheses and allocate resources for experimentation. Thereby, the framework and procedure developed herein are poised to benefit many applications of computational biology, systems biology as well as further uses of mathematical modeling in biology.

However, it is worth noting that this work is structured to address those problems in analyzing non-linear systems in a general mathematical setting. For that reason, the strategies developed herein are not limited by either the type of model or application contexts and are applicable beyond the scope of biological studies to improve the efficiency in analyzing mathematical models in other fields of predictive science.

6.4 References

- [1] Vu Dinh, Ann E. Rundell and Gregory T. Buzzard, (2014), Experimental Design for Dynamics Identification of Cellular Processes. *Bulletin of Mathematical Biology*, 76.3: 597-626.
- [2] Akil Narayan, and Dongbin Xiu, (2012), Stochastic collocation methods on unstructured grids in high dimensions via interpolation. *SIAM Journal on Scientific Computing* 34.3: A1729-A1752.
- [3] Ankush Chakrabarty, Vu Dinh, Gregory T. Buzzard, Stanislaw H. Zak, and Ann E. Rundell, (2014), Robust explicit nonlinear model predictive control with integral sliding mode. *American Control Conference (ACC)*, pp. 2851-2856, IEEE.
- [4] Sotiris B. Kotsiantis,, I. D. Zaharakis, and P. E. Pintelas, (2007), Supervised machine learning: A review of classification techniques.
- [5] Eugene M. Izhikevich and Jeff Moehlis, (2008), Dynamical Systems in Neuroscience: The geometry of excitability and bursting. *SIAM Review* 50.2: 397.
- [6] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss, (2002), On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems* 2 (2002): 849-856.
- [7] J. L. Rodgers and W. A. Nicewander, (1988), Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):5966, February 1988.

VITA

VITA

Vu Cao Duy Thien Dinh was born in Di Linh, Lam Dong Province, Vietnam. He received his Bachelor of Science in Mathematics and Computer Science at the University of Science, National University of Vietnam at Hochiminh City (Vietnam) in 2008. From 2008 through 2009, he pursued a Master of Science in Analysis and Applied Mathematics at the University of Orleans (Orleans, France). In 2009, he came to the Department of Mathematics at Purdue University to pursue his doctorate under the instruction of Professor Gregory T. Buzzard. He has published three journal articles and presented his work at several conferences as both oral and poster presentations.